Combinatorial Approach to Statistical Design of Experiment

Petya Valcheva

Sofia University "St. Kliment Ohridski"
(e-mail: pvalcheva@unwe.bg)

Abstract. The paper describes the various types of combinatorial designs applied in Statistical Design of Experiment. We give more detailed information about Balanced Incomplete Block Designs (*BIBDs*) and Orthogonal Arrays (*OAs*). The "difference matrix"-method for achievement of *OAs* is also described. Based on it two new constructions of orthogonal arrays $(6,15)$ and $(6,20)$ are obtained. They are non-isomorphic to the already known constructions of such arrays. The known $(7,12)-OA$ is also discussed.

This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract $BG051PO001-3.3.06-0052(2012/2014)$.

Keywords: difference matrix, balanced incomplete block design, quasi-difference matrix, orthogonal array.

1 Introduction

Orthogonal arrays(*OAs*), first introduced by Rao in 1947, are essential combinatorial structures. Their mathematical theory is inspiring, beautiful and closely related to combinatorics, geometry, finite fields and error-correcting codes. They are used in various scientific fields such as computer science, cryptography and statistics. The Statistical Design of Experiment is that branch of the statistics, where these structures are widely applied and that is the reason to be immensely important in areas, where a lot of research interests are concentrated, like manufacturing, medicine, agriculture and many others. More detailed information can be found in [1].
The statistical theory of the Experimental design was mainly initiated by R.A.Fisher [2] in the 1935*s* at the Rothamsted Experimental Station as a performance of agricultural experiments, but later it has been applied successfully in the military and industry. For example, Besse Day, working at the U.S. Naval Experimentation Laboratory, used the methods to solve problems such as finding the cause of bad welds at a naval shipyard during World War II [3].
Many experiments involves the study of the effects of two or more factors. The most efficient way to see the relationship between the independent variables(factors) is so called *factorial design*. By this technique, we mean that in each complete trial or replication of the experiment all possible combinations of the levels of the factors are investigated. But in many cases the number of treatments to be tested is large, which need more materials and respectively will increase the cost of experimentation in terms of labor, time and moreover money. In certain situations, we may not

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

be able to run all the treatment combinations in each block. In incomplete block designs, the block size is smaller than the total number of treatments to be compared. But when all comparisons are equally important, the combinations used in each block should be selected in so called *balanced manner*, that is, we must construct such block that any pair of treatments occur together in the same number of times as any other pair. These are the balanced incomplete block designs (*BIBD*), in which any two treatments occur together an equal number of times. Such designs are introduced by Yates in 1936 [6] and are similar to the construction of the orthogonal arrays, as we will see below.
Consider two sets T and B, whose elements are t treatments and b blocks respectively in a *BIBD*, that satisfy the following conditions:
(i) each block contains exactly k members
(ii) each treatment occurs in exactly r blocks(or is replicated r times)
(iii) each pair of treatments occurs in the same block exactly λ times.
A balanced incomplete block design has five parameters t, b, r, k, λ and can be defined as a collection of b subsets of size k from a set of t treatments, such that (i),(ii) and (iii) are satisfied. The parameter λ must be an integer. If we denote by n the total number of observations, we have the following relations between the parameters:
(i) $r.t = b.k = n$
(ii) $\lambda(t-1) = r(k-1)$
If $t = b$, the *BIBD* is said to be a symmetric *BIBD*.
The statistical model for the *BIBD* is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

where y_{ij} is the $i-th$ observation in the $j-th$ block, μ is the overall mean, τ_i is the effect of the $i-th$ treatment, β_j is the effect of the $j-th$ block and ε_{ij} is the $NID(0, \sigma^2)$ random error component.
The second part of the article gives detailed information about the technique for finding difference matrix and quasi-difference matrix, which will help us to construct orthogonal array. As we mentioned, they are used extensively in factorial designs because of their similar structure to the block designs.
The last section demonstrates some results, derived by computer realization of the methods, described in the previous part.

2 On the Construction of Difference Matrices

The purpose of this section is to describe some different techniques for finding combinatorial designs. It will be discussed various approaches with so called *difference matrix(DM)* and *quasi-difference matrix(QDM)*. Such matrices are constructed via algebraic arguments. More detailed information is given in [4,5].
One of the relevant and interesting questions in the theory of Latin squares is how

to determine the maximum possible number, denoted by $N(\nu)$, of *mutually orthogonal Latin squares of order* ν *(MOLS*(ν)*)*. It is well known that for each $\nu \geq 2$, $N(\nu) \leq \nu - 1$. Summarized results for $N(\nu)$ can be found in [7].
Let S be a non-empty set of order ν. A *Latin square of order* ν is an $\nu \times \nu$ matrix L, in which ν distinct symbols are arranged so that each element of S appears once in each row and column. Let L_1 and L_2 be Latin squares of same order, say $\nu \geq 2$. We say that L_1 and L_2 are orthogonal if, when superimposed, each of the possible ν^2 ordered pairs occur exactly once. A set $\{L_1, ..., L_t\}$ of $t \geq 2$ Latin squares of order ν is orthogonal if any two distinct Latin squares are orthogonal. We call this a set of mutually orthogonal Latin squares(*MOLS*).
The existence of $k-2$ $MOLS((\nu))$ is equivalent to the existence of an orthogonal array $OA(k,\nu)$, which is defined as a $k \times \nu^2$ matrix

$$A = \{(a_{ij}), i = 1, ..., k, j = 1, ..., \nu^2\}$$

over a ν-set S, such that any two rows contain all the ordered pairs of elements from S exactly once. In this case, it is customary to say that the orthogonal array has index unity $(\lambda = 1)$ and strength k. It is known that $k \leq \nu + 1$. A $k \times \nu$ submatrix of an $OA(k,\nu)$ is called a *parallel class*, if every row in it is a permutation of the elements of S. An $OA(k,\nu)$ is called *resolvable*, if its columns can be split into ν disjoint parallel classes. The existence of an $OA(k,\nu)$ possessing a parallel class is equivalent to the existence of $k-2$ *idempotent* $MOLS((\nu))$. The existence of a resolvable $OA(k,\nu)$ is equivalent to the existence of $k-1$ $MOLS((\nu))$ [4,5]. Orthogonal arrays are combinatorial designs, that can be obtained from difference and quasi-difference matrices, which will be defined below.
Let $\Gamma = \{1, g_2, ..., g_\nu\}$ be a group of order ν. A $k \times \lambda\nu$ matrix

$$D = \{(d_{ij}), i = 1, ..., k, j = 1, ..., \lambda\nu\}$$

with entries from Γ is called a $(\nu, k; \lambda)$-difference matrix over Γ and is denoted by $DM(\nu, k, ; \lambda)$ if it satisfies the *difference property*: for each $1 \leq i < j \leq k$, the multi-set

$$\{d_{il} d_{jl}^{-1}; 1 \leq l \leq \lambda\nu\}$$

(the difference list) contains every element of Γ λ times. When Γ is *abelian*, i.e. $\Gamma = \{0, g_2, ..., g_\nu\}$, typically additive notation is used, so that differences $d_{il} - d_{jl}$ are employed. Removing any row from a $(\nu, k; \lambda)$-difference matrix gives a $(\nu, k-1; \lambda)$-difference matrix, i.e. the difference property still holds. A $DM(\nu, k; \lambda)$ does not exist if $k > \lambda\nu$. A (ν, k)-difference matrix over Γ gives rise to a resolvable $OA(k,\nu)$ and hence to an $OA(k+1,\nu)$.
In the case when $\lambda = 1$ a $(\nu, k; 1)$-difference matrix, denoted by $DM(\nu, k)$, gives rise to $OA(k,\nu)$ by developing it through the group Γ in the following way:

$$OA(k,\nu) = (DM|DM.g_2|\ldots|DM.g_\nu)$$

In many cases *MOLS* are obtained from quasi-difference matrices, that is a matrix which contains an additional to the group elements point often denoted by ∞. Below

we discuss the construction of these matrices.
Extend the group Γ by an additional element $\{\infty\}$(point infinity) and denote it by $\Gamma_\infty = \Gamma \cup \{\infty\} = \{1, g_2, ..., g_n, \infty\}$. Consider a $k \times \lambda(v+2)$ matrix over the group Γ_∞, so that:
(i) each row contains the point $\{\infty\}$ exactly λ times and each column contains it at most once
(ii) for any two distinct rows, the difference property stay in force, i.e. for every $i,t \in \{1,2,...,k\}, i \neq t$ the list of differences $\{d_{ij}d_{tj}^{-1}\}$ contains each element of Γ exactly λ times (differences with the additional element ∞ are undefined).
Such matrix is called $(v,k;\lambda)$ quasi-difference matrix over (Γ_∞, G) and is denoted by $QDM(v,k;\lambda)$. In case when $\lambda = 1$, denote it by $QDM(v,k)$.An orthogonal array $OA(k, v+1)$ can be obtained developing $QDM(v,k;\lambda)$ over the subgroup G and adding an additional column consisting of points infinity - $(\infty,\infty,...,\infty)^T$ in the following way:

$$OA(k,v+1) = \left(\begin{array}{c|c|c|c|c} & & & & \infty \\ & & & & \infty \\ QDM & QDM.u_2 & \dots & QDM.u_n & \vdots \\ & & & & \infty \end{array} \right)$$

Another way of deriving orthogonal arrays and $MOLS$ from difference matrices with $\lambda > 1$ was introduced in [8]. Let Γ be a group of order $v = \lambda.n$ and let $G = \{1, u_2, ..., u_n\}$ be an subgroup of $\Gamma (G \leq \Gamma)$ of order n, where $n = \frac{v}{\lambda}$. A $(v,k;\lambda)$-difference matrix $DM = (d_{ij})$ is said to be a difference matrix over (Γ, G), and denoted by $DM(v,k;\lambda)$, if the difference property stay true, but is of the form: $d_{ij}d_{tj}^{-1} = d_{is}d_{ts}^{-1}$ yields $d_{ij}^{-1}d_{is} \notin G$, i.e. d_{ij}, d_{is} are from different left cosets of G. In this case $OA(k,v)$ is obtained from $DM(v,k;\lambda)$, developing it over the subgroup G in the following way:

$$OA(k,v) = (DM|DM.u_2|\dots|DM.u_v)$$

3 Some Results on $MOLS$ of order 12,15 and 20

3.1. The case $v = 12$
A construction of a resolvable $DM(12,6)$ over the group $C_6 \times C_2$ was obtained in [8], which leads to the existence of 5 $MOLS(12)$. We try to improve this result by investigation of $DM(12,k;2)$. There exist 5 groups of order 12, but from the Proposition 1 in [10] it is sufficient to consider Z_{12} and $S_3 \times C_2$ only. The exhaustive computer search for $(12,6;2)$-difference matrix over these groups with respect to their order 6 subgroups shows that there exists no $DM(12,k;2), k > 6$. Moreover all the resolvable $DM(12,6;2)$ found this way give no new $OA(7,12)$, i.e. under the above restrictions the construction of 5 $MOLS(12)$ obtained in [8] is unique.

3.2. The case $v = 15$

Four $MOLS$ of order 15 were obtained in [9] by a construction of $DM(15,6)$ over the group Z_{15}. We performed an exhaustive computer search for $QDM(14,k;2)$ over $Z_{14} \cup \{\infty\}$. All of the obtained $QDM(14,4;2)$ were developed to the orthogonal array $OA(4,15)$ and put to extension(without preserving the "difference" property). None of these arrays can be extended to $OA(7,15)$ for 5 $MOLS(15)$. Below is given one example of $QDM(14,4;2)$ and its extension to 4 $MOLS(15)$. Using *Todorov's* program package *OADM* for *OAs* and *DMs* processing it was shown that this construction is not isomorphic to the one in [9].

$QDM(14,4;2)$

$$\left\|\begin{array}{cccccccccccccc|cccccccccccccc}
\infty & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \infty & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \infty & 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \infty \\
0 & \infty & 0 & 3 & 11 & 1 & 10 & 8 & 2 & 12 & 5 & 13 & 9 & 4 & 6 & 7 & 1 & 7 & \infty & 1 & 4 & 12 & 2 & 11 & 9 & 3 & 13 & 6 & 0 & 10 & 5 & 8 \\
0 & 10 & 6 & \infty & 2 & 8 & 0 & 4 & 13 & 11 & 3 & 7 & 12 & 5 & 1 & 9 & 1 & 2 & 11 & 7 & \infty & 3 & 9 & 1 & 5 & 0 & 12 & 4 & 8 & 13 & 6 & 10
\end{array}\right\|$$

Extension to 4 $MOLS(15)$

$$\begin{array}{cccccccccccccc}
\infty & 0 & 3 & 11 & 1 & 10 & 8 & 2 & 12 & 5 & 13 & 9 & 4 & 6 & 7 \\
7 & \infty & 1 & 4 & 12 & 2 & 11 & 9 & 3 & 13 & 6 & 0 & 10 & 5 & 8 \\
6 & 8 & \infty & 2 & 5 & 13 & 3 & 12 & 10 & 4 & 0 & 7 & 1 & 11 & 9 \\
12 & 7 & 9 & \infty & 3 & 6 & 0 & 4 & 13 & 11 & 5 & 1 & 8 & 2 & 10 \\
3 & 13 & 8 & 10 & \infty & 4 & 7 & 1 & 5 & 0 & 12 & 6 & 2 & 9 & 11 \\
10 & 4 & 0 & 9 & 11 & \infty & 5 & 8 & 2 & 6 & 1 & 13 & 7 & 3 & 12 \\
4 & 11 & 5 & 1 & 10 & 12 & \infty & 6 & 9 & 3 & 7 & 2 & 0 & 8 & 13 \\
9 & 5 & 12 & 6 & 2 & 11 & 13 & \infty & 7 & 10 & 4 & 8 & 3 & 1 & 0 \\
2 & 10 & 6 & 13 & 7 & 3 & 12 & 0 & \infty & 8 & 11 & 5 & 9 & 4 & 1 \\
5 & 3 & 11 & 7 & 0 & 8 & 4 & 13 & 1 & \infty & 9 & 12 & 6 & 10 & 2 \\
11 & 6 & 4 & 12 & 8 & 1 & 9 & 5 & 0 & 2 & \infty & 10 & 13 & 7 & 3 \\
8 & 12 & 7 & 5 & 13 & 9 & 2 & 10 & 6 & 1 & 3 & \infty & 11 & 0 & 4 \\
1 & 9 & 13 & 8 & 6 & 0 & 10 & 3 & 11 & 7 & 2 & 4 & \infty & 12 & 5 \\
13 & 2 & 10 & 0 & 9 & 7 & 1 & 11 & 4 & 12 & 8 & 3 & 5 & \infty & 6 \\
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \infty
\end{array}$$

$$\begin{array}{cccccccccccccc}
10 & 6 & \infty & 2 & 8 & 0 & 4 & 13 & 11 & 3 & 7 & 12 & 5 & 1 & 9 \\
2 & 11 & 7 & \infty & 3 & 9 & 1 & 5 & 0 & 12 & 4 & 8 & 13 & 6 & 10 \\
7 & 3 & 12 & 8 & \infty & 4 & 10 & 2 & 6 & 1 & 13 & 5 & 9 & 0 & 11 \\
1 & 8 & 4 & 13 & 9 & \infty & 5 & 11 & 3 & 7 & 2 & 0 & 6 & 10 & 12 \\
11 & 2 & 9 & 5 & 0 & 10 & \infty & 6 & 12 & 4 & 8 & 3 & 1 & 7 & 13 \\
8 & 12 & 3 & 10 & 6 & 1 & 11 & \infty & 7 & 13 & 5 & 9 & 4 & 2 & 0 \\
3 & 9 & 13 & 4 & 11 & 7 & 2 & 12 & \infty & 8 & 0 & 6 & 10 & 5 & 1 \\
6 & 4 & 10 & 0 & 5 & 12 & 8 & 3 & 13 & \infty & 9 & 1 & 7 & 11 & 2 \\
12 & 7 & 5 & 11 & 1 & 6 & 13 & 9 & 4 & 0 & \infty & 10 & 2 & 8 & 3 \\
9 & 13 & 8 & 6 & 12 & 2 & 7 & 0 & 10 & 5 & 1 & \infty & 11 & 3 & 4 \\
4 & 10 & 0 & 9 & 7 & 13 & 3 & 8 & 1 & 11 & 6 & 2 & \infty & 12 & 5 \\
13 & 5 & 11 & 1 & 10 & 8 & 0 & 4 & 9 & 2 & 12 & 7 & 3 & \infty & 6 \\
\infty & 0 & 6 & 12 & 2 & 11 & 9 & 1 & 5 & 10 & 3 & 13 & 8 & 4 & 7 \\
5 & \infty & 1 & 7 & 13 & 3 & 12 & 10 & 2 & 6 & 11 & 4 & 0 & 9 & 8 \\
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \infty
\end{array}$$

$$\begin{array}{cccccccccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \infty \\
12 & 11 & \infty & 7 & 6 & 0 & 5 & 4 & 10 & 1 & 3 & 8 & 13 & 2 & 9 \\
7 & 2 & 13 & 9 & 10 & 4 & 11 & 5 & 0 & 8 & \infty & 6 & 1 & 12 & 3 \\
2 & 10 & 7 & 12 & 3 & 8 & 0 & 13 & 5 & 11 & 1 & 9 & 4 & \infty & 6 \\
9 & 7 & 8 & 10 & 2 & 6 & 1 & 11 & 12 & 5 & 13 & \infty & 3 & 0 & 4 \\
11 & 3 & 10 & 1 & 8 & 7 & 4 & \infty & 13 & 2 & 5 & 12 & 9 & 6 & 0 \\
4 & 13 & 6 & 8 & \infty & 1 & 10 & 0 & 9 & 12 & 7 & 5 & 2 & 3 & 11 \\
6 & 0 & 12 & 4 & 1 & 9 & \infty & 8 & 11 & 3 & 2 & 10 & 5 & 7 & 13 \\
10 & 4 & 11 & 2 & 0 & \infty & 3 & 9 & 1 & 13 & 6 & 7 & 8 & 5 & 12 \\
5 & 8 & 0 & 13 & 7 & 11 & 9 & 6 & 3 & \infty & 12 & 4 & 10 & 1 & 2 \\
\infty & 5 & 1 & 11 & 12 & 10 & 13 & 3 & 4 & 6 & 9 & 2 & 0 & 8 & 7 \\
1 & 9 & 5 & \infty & 13 & 2 & 8 & 12 & 6 & 0 & 4 & 3 & 7 & 11 & 10 \\
13 & \infty & 3 & 5 & 9 & 12 & 7 & 1 & 2 & 4 & 11 & 0 & 6 & 10 & 8 \\
8 & 12 & 9 & 6 & 5 & 3 & 2 & 10 & \infty & 7 & 0 & 13 & 11 & 4 & 1 \\
3 & 6 & 4 & 0 & 11 & 13 & 12 & 2 & 7 & 10 & 8 & 1 & \infty & 9 & 5
\end{array}$$

$$\begin{array}{cccccccccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \infty \\
13 & 10 & 9 & 4 & 7 & 1 & 0 & 3 & \infty & 6 & 5 & 12 & 2 & 8 & 11 \\
6 & 13 & 12 & 5 & 1 & \infty & 9 & 10 & 7 & 11 & 3 & 0 & 8 & 4 & 2 \\
1 & 3 & 13 & 8 & 0 & 9 & 11 & 5 & 12 & \infty & 2 & 7 & 10 & 6 & 4 \\
3 & 9 & 7 & 13 & 6 & 10 & 5 & 2 & 0 & 8 & 11 & 4 & \infty & 12 & 1 \\
8 & 7 & 5 & \infty & 13 & 3 & 12 & 0 & 4 & 10 & 6 & 2 & 1 & 11 & 9 \\
2 & 6 & \infty & 0 & 11 & 13 & 7 & 8 & 10 & 1 & 12 & 3 & 4 & 9 & 5 \\
5 & 4 & 3 & 11 & 10 & 2 & 13 & \infty & 6 & 12 & 9 & 8 & 7 & 1 & 0 \\
9 & 0 & 1 & 7 & 2 & 12 & 4 & 13 & 11 & 3 & 8 & 5 & 6 & \infty & 10 \\
11 & 5 & 10 & 9 & \infty & 4 & 8 & 1 & 13 & 2 & 7 & 6 & 0 & 3 & 12 \\
7 & 2 & 0 & 12 & 5 & 11 & 1 & 6 & 9 & 13 & 4 & \infty & 3 & 10 & 8 \\
12 & \infty & 4 & 10 & 8 & 0 & 2 & 9 & 3 & 5 & 13 & 1 & 11 & 7 & 6 \\
\infty & 8 & 11 & 1 & 12 & 6 & 10 & 4 & 5 & 7 & 0 & 13 & 9 & 2 & 3 \\
4 & 11 & 6 & 2 & 9 & 8 & 3 & 12 & 1 & 0 & \infty & 10 & 13 & 5 & 7 \\
10 & 12 & 8 & 6 & 3 & 7 & \infty & 11 & 2 & 4 & 1 & 9 & 5 & 0 & 13
\end{array}$$

3.3. The case $v = 20$

Here we give our version of four $MOLS(20)$, which are constructed in [11] and [12] in a $QDM(19,6;1)$ form. We investigate $DM(20,k;2)$ over Z_{20} with respect to its order 10 subgroup. It was supposed that corresponding orthogonal array admits an order 3 automorphism. Am example of $DM(20,6;2)$ is given below, but no $DM(20,7;2)$ was obtained. By the construction, it is obviously that the corre-

sponding $OA(6,20)$ possesses an order 5 automorphism and since the constructions in [11] and [12] admit no such automorphism, our example is non-isomorphic to them.

$DM(20,6;2)$

$$\left\|\begin{array}{cccccccccccccccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 \\
12 & 2 & 17 & 7 & 13 & 10 & 5 & 4 & 8 & 16 & 15 & 1 & 0 & 9 & 11 & 18 & 3 & 6 & 19 & 14 \\
18 & 19 & 6 & 17 & 15 & 1 & 5 & 11 & 13 & 12 & 3 & 2 & 9 & 0 & 14 & 8 & 10 & 7 & 4 & 16 \\
9 & 4 & 10 & 12 & 16 & 19 & 2 & 0 & 6 & 18 & 11 & 3 & 5 & 18 & 1 & 13 & 7 & 14 & 17 & 15 \\
5 & 6 & 15 & 7 & 12 & 13 & 1 & 14 & 17 & 11 & 9 & 3 & 18 & 10 & 8 & 4 & 16 & 0 & 2 & 19
\end{array}\right\|$$

$$\left\|\begin{array}{cccccccccccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 \\
11 & 1 & 0 & 15 & 6 & 18 & 16 & 9 & 12 & 3 & 4 & 10 & 5 & 19 & 17 & 13 & 2 & 8 & 14 & 7 \\
3 & 16 & 4 & 11 & 19 & 10 & 15 & 8 & 0 & 9 & 18 & 13 & 2 & 12 & 1 & 7 & 17 & 6 & 14 & 5 \\
2 & 16 & 6 & 14 & 10 & 5 & 9 & 17 & 3 & 15 & 4 & 13 & 12 & 0 & 11 & 19 & 1 & 18 & 8 & 7 \\
3 & 16 & 19 & 2 & 14 & 15 & 17 & 0 & 4 & 7 & 11 & 12 & 10 & 9 & 6 & 1 & 18 & 8 & 5 & 13
\end{array}\right\|$$

References

1. Hedayat, A.S., Sloane, N, J.A., Stufken, John, *Orthogonal arrays: theory and application.*, Springer, 1999.
2. John, Peter William Meredith, *Incomplete block designs.*, Marcel Dekker, 1980.
3. Davim, J.Paulo (Editor), *Statistical and Computational Techniques in Manufacturing.*, Springer, 2012.
4. C. Colbourn and J. Dinitz (Editors), *CRC Handbook of Combinatorial Designs.*, CRC Press, Boca Raton, 2007.
5. T. Beth, D. Jungnickel, and H. Lenz, *Design Theory.*, Cambridge University Press, Cambridge, England, 1999.
6. Montgomery, Douglas.C, *Design and analysis of experiments - 5th edition.*, John Wiley and Sons, 2001.
7. Brouwer, A.E., *The number of mutually orthogonal Latin squares - a table up to order 10000.*, Report ZW 123, Math. Centr., Amsterdam 1978
8. Johnson, D.M., Dulmage, A.I., Mendelson, N.S., *Orthomorphisms of group and orthogonal Latin squares.*, Canadian J. Math., 13, 356-372, 1961.
9. Schellenberg, P.J., van Rees, G.M., Vanstone, S.A., *Four pairwise orthogonal Latin squares of order 15.*, Ars Combinatoria, 6, 141-150, 1978.
10. Todorov, D.T., *Four mutually orthogonal Latin squares of order 14.*, Journal of Combinatorial Designs, 20, 1-5, 2012.
11. Todorov, D.T., *Four mutually orthogonal Latin squares of order 20.*, Ars Combinatoria, 27-C, 63-65, 1989.
12. Abel, R.J.R., Todorov, D.T., *Four mutually orthogonal Latin squares of orders 20, 30, 38 and 44.*, J. Comb. Theory(A), 64, 144-148, 1993.

ON 21 ST CENTURY'S MISUSING OF THE CLASSICAL PEARSON'S SUM AND POWER DIVERGENCE STATISTICS

Vassilly Voinov, Rashid Makarov

KIMEP University, 2 Abai Ave., and IMMM of MES RK, 125 Pushkin Str.,
Almaty 050010, Republic of Kazakhstan, E-mail: voinovv@mail.ru, rashidm@kimep.kz

Abstract. The extant researchers continue misusing the Cardoso De Oliveira & Ferreira's test and power divergence statistics by assuming erroneous chi-squared limiting distribution. Neither these tests nor classical Pearson's sum reject the null hypothesis as often as it should be. Moreover, even in case of correct implementation of the tests, their power with respect to alternatives close to multivariate normal distribution is much lower than that for other multivariate normality tests known by the date. An extensive simulation study and two classical data sets are used to illustrate the theoretical arguments and derived exact limiting distribution.

1. Introduction

Karl Pearson has invented the famous chi-squared test in 1900. Nowadays everybody knows that this statistic is distribution free and will possesses in the limit the chi-squared distribution with M - 1 degrees of freedom, where M stands for the number of grouping classes, if and only if, a null hypothesis is simple, and, hence, parameters of the null hypothesis are known. Until 1934, Karl Pearson believed that the limiting distribution of his test will not change if unknown parameters are replaced by their sample estimates. In 1924 Ronald Fisher, a founder of modern statistics, showed that if parameters are estimated by grouped sampled data, then the Pearson's sum is distributed in the limit as chi-squared with M - s - 1 degrees of freedom, where s is the number of parameters under estimation. Since then during many years researchers thought that that result is true if parameters are estimated by raw (non-grouped) data. In 1954 Chernoff and Lehmann proved that, if parameters (in the univariate case) are estimated by maximum likelihood method (MLE) based on raw data, then the limit distribution of Pearson's sum does not follow the chi-squared distribution with M - s - 1 degrees of freedom, and actually will be distributed as a weighted sum of independent chi-squared random variables. They (Chernoff and Lehmann (1954), p.586) noted that "the error is not serious in the case of fitting a Poisson distribution", but "in the normal case the use of maximum

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

likelihood estimates (based on raw data) may lead to a more serious underestimate of the probability of type *I* error". In spite of this important result, there are instances of repeated misusing of the classical Pearson's test with MLEs based on raw data in textbooks of statistics (see, e.g., Weiers (1991), p. 602, Clark (1997), p. 273).

Moore and Stubblebine (1981) considered a possibility to use the classical Pearson's test for testing the multivariate normality (MVN). Using the fundamental result of Chernoff and Lehmann (1954) they show that under the null hypothesis of MVN the Pearson's statistic with parameters estimated by MLEs based on raw data for equiprobable fixed grouping cells will follow the weighted sum of two independent chi-squared random variables with *M - 2* and *1* degree of freedom, where *M* stands for the number of equiprobable fixed grouping cells. This result has not been taken into consideration by Cardoso De Oliveira and Ferreira (2010), and Batsidis, Martin, Pardo, Zografos (2013), who consider tests for MVN based on unbiased estimates that are asymptotically equivalent to MLEs. The authors of those papers erroneously decided that their tests approximately follow the chi-squared distribution with *M - 1* degrees of freedom in asymptotic.

In Section 2 we derive the exact limit distribution of the Cardoso De Oliveira and Ferreira (2010) statistic (CarFer statistic hereafter). To verify the validity of the theory we developed in Section 2 a simulation study of CarFer test and power divergence statistics introduced in Batsidis *et al.* (2013) is done in Section 3. The overall conclusion is presented in Section 4. Throughout the article boldfaced non-italicized letters are used to denote vectors and matrices.

2. The limit null distribution of the CarFer statistic

We intend to test the composite hypothesis that a set $\boldsymbol{X}_1,...,\boldsymbol{X}_n$ of n independent identically distributed p - dimensional random vectors follow a multivariate normal distribution with the joint probability density function

$$f(\boldsymbol{x}\,|\,\boldsymbol{\theta}) = (2\pi)^{-p/2}\,|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] \qquad (1)$$

where $\boldsymbol{\mu}$ is the p - vector of means and $\boldsymbol{\Sigma}$ is a positive definite $p\times p$ covariance matrix. Let a given vector of unknown parameters be $\boldsymbol{\theta} = (\mu_1,...,\mu_p,\sigma_{11},\sigma_{12},\sigma_{22},...,\sigma_{1j},\sigma_{2j},...,\sigma_{jj},...,\sigma_{pp})^T = (\theta_1,...,\theta_m)^T$, where the elements of the matrix $\boldsymbol{\Sigma}$ are arranged column-wise by taking the elements of the upper-triangular submatrix of $\boldsymbol{\Sigma}$, and $m = p + p(p+1)/2$. The unbiased estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is the vector $(\bar{\boldsymbol{X}}, \boldsymbol{S})^T$, where $\bar{\mathbf{X}} = \sum_{j=1}^n \mathbf{X}_j / n$, and $\mathbf{S} = \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T / (n-1)$ are unbiased

estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively. Cardoso De Oliveira and Ferreira (2010) considered the squared radii

$r_j^2 = (\mathbf{X}_j - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}),\ j = 1, ..., n.$

It is known that if $\mathbf{X}_j$ follows (1) , then statistics

$$b(\mathbf{X}_j) = \frac{n}{(n-1)^2} r_j^2 = \frac{n}{(n-1)^2} (\mathbf{X}_j - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}),\ j = 1, ..., n, \qquad (2)$$

follow the beta-distribution with parameters $p/2$ and $(n-p-1)/2$ (Gnanadesikan and Kettering (1972)) . Define grouping cells

$$E_i(\boldsymbol{\theta}) = \{\mathbf{X} \in R^p : c_{i-1} \le \frac{n}{(n-1)^2} (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) < c_i\},\ i = 1, ..., M,$$

and expected cell probability $p_i(\boldsymbol{\theta}) = \int_{E_i(\hat{\boldsymbol{\theta}}_n)} f(\mathbf{x} \mid \boldsymbol{\theta}) d\mathbf{x},\ i = 1, ..., M.$

Let ends of equiprobable grouping cells be $0 = c_0 < c_1 < \cdots < c_{M-1} < c_M = 1,$ where $c_i,\ i = 1, ..., M-1,$ is the i/M point of the beta-distribution with parameters $p/2$ and $(n-p-1)/2$. If N_i denotes the number of $b(\mathbf{X}_j),\ j = 1, ..., n,$ falling in $E_i(\hat{\boldsymbol{\theta}}_n)$, and, since the expected frequency $np_i(\hat{\boldsymbol{\theta}}_n) = n/M$, then the CarFer's statistic Z , that actually is the classical Pearson's sum, will be

$$Z = \sum_{i=1}^{M} M (N_i - n/M)^2 / n. \qquad (3)$$

Note that $N_i,\ i = 1, ..., M,$ depend on unbiased estimators $\bar{\mathbf{X}}$ and $\mathbf{S}$ of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and that those estimates are based on the non-grouped (raw) data. Following Moore and Stubblebine (1981), p. 719 for a specified $\boldsymbol{\theta}_0$, namely $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, define

$$p_i(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \int_{E_i(\boldsymbol{\theta}_0)} f(\mathbf{x} \mid \boldsymbol{\theta}) d\mathbf{x}, \qquad (4)$$

and the $M \times m$ matrix $B(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ with entries

$$\frac{1}{\sqrt{p_i(\boldsymbol{\theta}, \boldsymbol{\theta}_0)}} \frac{\partial p_i(\boldsymbol{\theta}, \boldsymbol{\theta}_0)}{\partial \theta_j},\ i = 1, ..., M,\ j = 1, ..., m. \qquad (5)$$

The $m \times m$ Fisher information matrix for one observation from (1) is

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{bmatrix},$$

where **Q** is the $p(p+1)/2 \times p(p+1)/2$ covariance matrix of **w,** a vector of the entries of $\sqrt{n}\mathbf{S}$ arranged column-wise by taking the upper triangular elements

$\mathbf{w} = (s_{11}, s_{12}, s_{22}, s_{13}, s_{23}, s_{33}, ..., s_{pp})^T$.

Under the regularity conditions of Moore and Spruill (1975), p. 602 which hold in our case, the M-vector $\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n)$ of standardized frequencies with components $V_i(\hat{\boldsymbol{\theta}}_n) = (N_i - n/M)/\sqrt{n/M},\ i = 1, ..., M,$ follow asymptotically the M -dimensional multivariate normal distribution $N_M(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{V})$ with **0**-vector of means and the covariance matrix

$$\boldsymbol{\Sigma}_\mathbf{V} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T, \qquad (6)$$

where $\mathbf{J} = \mathbf{J}(\boldsymbol{\theta}_0)$, $\mathbf{B} = \mathbf{B}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$, and $\mathbf{q}$ is the M-vector with components $p_i(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)^{1/2} = 1/\sqrt{M},\ i = 1, ..., M.$ Moore and Stubblebine (1981), p. 720 justified this result for maximum likelihood estimators (MLEs) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Since unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are asymptotically equivalent to their MLEs, the formula in (6) is valid in our case as well. Thus the asymptotic null distribution of Z is determined by the eigen-values of the matrix $\boldsymbol{\Sigma}_\mathbf{V} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$. From the formula (A.8) it follows that all columns of the matrix $\mathbf{B}$ are scalar multipliers of the vector $(d_1, ..., d_M)^T$, and, hence, rank of $\mathbf{B}$ and also rank of $\mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ is 1. From the above it follows that eigen-values of $\boldsymbol{\Sigma}_\mathbf{V} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ are M - 2 1's, one 0 and $0 < \Lambda < 1$ which is nonzero eigen-value of $\mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$. The explicit value of Λ is given in (A.10).

Theorem.
The limit null probability distribution function of the CarFer's statistic Z is

$$F(z) = \frac{1}{\sqrt{\pi}\,\Gamma\left((M-2)/2\right)}\,\gamma\left(\frac{1}{2}, \frac{z}{2\Lambda}\right)\gamma\left(\frac{M-2}{2}, \frac{z}{2}\right),\ M > 2, \qquad (7)$$

where $\gamma(s, x) = \int_0^x t^{s-1}e^{-t}dt,\ s > 0,$ is the incomplete gamma-function.

Proof.
Since eigen-values of $\boldsymbol{\Sigma}_\mathbf{V} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ are M - 2 1's, one 0 and $0 < \Lambda < 1$, the limit null distribution of the CarFer's quadratic form will be the same as that of $X + \Lambda Y$, where the probability density function of X is

$f(x) = \chi^2_{M-2}(x) = x^{\frac{M-4}{2}} e^{-x/2} / \left[2^{\frac{M-2}{2}} \Gamma\left((M-2)/2\right) \right],\ x > 0,$ and the probability density function of Y is $f(y) = \chi^2_1(y) = e^{-y/2} / \sqrt{2\pi y},\ y > 0.$ Due to the independence of X and Y the limit null probability distribution function of Z will be

$$F(z) = P(Z \le z) = \iint\limits_{x+\Lambda y \le z} f(x) f(y) dx dy = \int\limits_0^{z/\Lambda} f(y) dy \int\limits_0^z f(x) dx =$$

$$= \frac{1}{\sqrt{\pi}\Gamma\left(\frac{M-2}{2}\right)} \gamma\left(\frac{1}{2}, \frac{z}{2\Lambda}\right) \gamma\left(\frac{M-2}{2}, \frac{z}{2}\right),\ M > 2.$$

The corresponding probability density function is

$$f(z) = \frac{\chi^2_1(z/\Lambda)}{\Lambda\Gamma\left(\frac{M-2}{2}\right)} \gamma\left(\frac{M-2}{2}, \frac{z}{2}\right) + \frac{\chi^2_{M-2}(z)}{\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{z}{2\Lambda}\right),\ M > 2. \quad (8)$$

3. A simulation study

3.1 A Simulation study of the CarFer statistic

The Pearson's sum, and the CarFer statistic are discrete random variables and so are their probability distributions. Those distributions can be approximated by continuous functions only for very large samples of size *n*. It has to be noted that the exact limit null distribution function of the CarFer statistic in (7) does not exist if $M = 2$, because the gamma function $\Gamma(x)$ is not defined for $x = 0$.

To simulate samples from the null multivariate normal distribution with a random positive definite matrix Σ we use functions genPositiveDefMat(⊔) and rmvnorm(⊔) from R-packages {clusterGeneration} and {mvtnorm} respectively. Using those samples we simulated statistics in equation (3). We simulate the CarFer statistic when $M = 2$ to see the discreteness and to assess the simulated critical values. The discreteness of the simulated CarFer statistic if $M = 2$, $n = 250$, $p = 2$ is clearly seen in Figure 1.

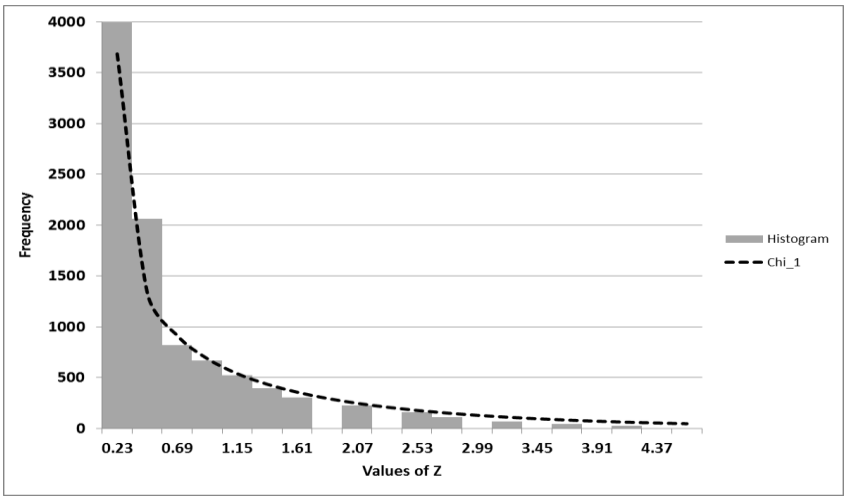


Figure 1. The histogram of N = 50,000 simulated CarFer statistics (shaded area), and the chi-squared distribution with M-1=1 degree of freedom (dashed line).

The simulated critical value of level $\alpha = 0.05$ for the data used in Fig. 1 equals to 1.936. Note that the critical value of the central chi-squared probability distribution with M - 1 = 1 degree of freedom, assumed by Cardoso De Oliveira & Ferreira (2010), is 3.841 that is almost two times more than simulated one. From Fig. 1 it can be seen that the limit null distribution of the CarFer test does not follow the chi-squared probability distribution with M - 1 = 1 degree of freedom (even approximately). The same is observed if one will compare the simulated critical values of level $\alpha = 0.05$ for CarFer's test if n = 250, p = 2, $M > 2$ with theoretical critical values for $f(z)$ (we defined them solving the equation $F(z) = 0.95$ with the help of the R-function nleqslv(□)), and those for the central chi-squared probability distribution with M - 1 degrees of freedom (see Table 1).

Table 1. Simulated critical values of level $\alpha = 0.05$ for CarFer's test if n = 250, p = 2, theoretical critical values for $f(Z)$, and those for the chi-squared probability distribution with M - 1 df. We did 5 simulations by N = 10,000 replications each, and give the assessed simulated standard deviation for the mean of those 5 runs.

M	CarFer	$f(z)$	Chi_M-1
2	**1.94 ± 0**	**N/A**	**3.841**
3	**4.27 ± 0.01**	**3.870**	**5.991**
5	**8.09 ± 0.04**	**7.815**	**9.488**
10	**15.58 ± 0.05**	**15.507**	**16.919**
20	**28.82 ± 0.08**	**28.869**	**30.144**
30	**41.41 ± 0.09**	**41.337**	**42.557**
40	**53.36 ± 0.10**	**53.384**	**54.572**

From this table we see that the simulated critical values for $M > 3$ almost coincide with the theoretical ones, but are noticeably less than critical values for the chi-squared probability distribution with M - 1 degrees of freedom. This evidently contradicts to the incorrect conclusion of Cardoso De Oliveira and Ferreira (2010), p. 516. The same conclusion holds for $p = 5$ (see Table 2).

Table 2. Simulated critical values of level $\alpha = 0.05$ for CarFer's test if $n = 250$, $p = 5$, theoretical critical values for $f(z)$, and those for the chi-squared probability distribution with M - 1 df. We give $\pm$ one simulated standard deviation assessed as in Table 1.

M	CarFer	$f(z)$	Chi_M-1
2	**1.60 ± 0**	**N/A**	**3.841**
3	**4.18 ± 0.06**	**3.846**	**5.991**
5	**7.94 ± 0.05**	**7.815**	**9.488**
10	**15.49 ± 0.08**	**15.507**	**16.919**
20	**28.94 ± 0.11**	**28.869**	**30.144**
30	**41.31 ± 0.14**	**41.337**	**42.557**
40	**53.30 ± 0.19**	**53.384**	**54.572**

Consider the simulated histogram of the CarFer's test if $M = 5$, $n = 250$, $p = 5$ (see Fig. 2).

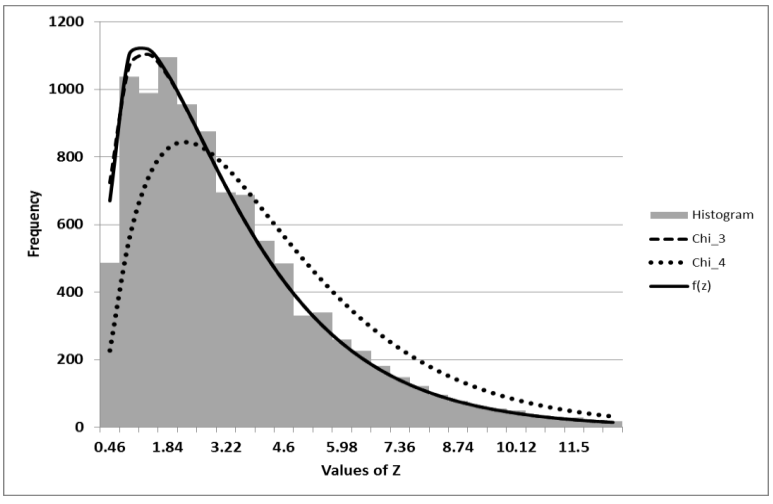


Figure 2. The histogram of $N = 50{,}000$ simulated CarFer statistics (shaded area), the chi-squared distribution with M - 1 = 4 df (dotted line), M - 2 =3 df (dashed line), and the theoretical limit probability density function $f(z)$ defined in (8) (solid line).

From Fig. 2 we see that even for not very large samples of size n = 250 the simulated values of the CarFer's test almost perfectly follow the theoretical limit probability density function $f(z)$ in (8) that can be very well approximated by the chi-squared probability distribution with M - 2 = 3 degrees of freedom. Again one sees that the simulated values of the CarFer's statistics do not follow the chi-squared probability distribution with M - 1 = 4 degrees of freedom.

Being a function of the Mahalanobis distances, the CarFer test in (3) is invariant with respect to all affine transformations of a sample space. Moreover, using arguments of Henze (2002), p. 484, we see that the test is consistent. Assume now that the CarFer's statistic will be used correctly with, say, simulated critical values or exact theoretical ones that can be defined from (7). The very important feature of any goodness-of-fit test is its power with respect to alternatives close to the hypothetical null distribution. Power is an ability of a goodness of fit test to discriminate between the hypothetical null and alternative distributions. Let us compare the power of the CarFer's statistic with respect to the multivariate Student t probability distribution with 10 degrees of freedom that is very close to the multivariate normal probability distribution. Let the sample size be n = 250 and p = 2. In Figure 3 we compared the simulated power of the CarFer statistic (line 4 on Fig. 3) with those of: Dzhaparidze & Nikulin (1974) (line 1 on Fig. 3), McCull test proposed in Voinov, Pya, Makarov, Voinov (2013) (line 2 on Fig. 3) , ChLeh test considered by Moore and Stubblebine (1981) (line 3 on Fig. 3), Doornik & Hansen (2008) test (line 5 on Fig. 3) , Royston's (1992) test (line 6 on Fig. 3), Anderson-Darling test as per Henze (2002), p. 483 (line 7 on Fig. 3), and the Energy test of Székely & Rizzo (2005) (line 8 on Fig. 3).

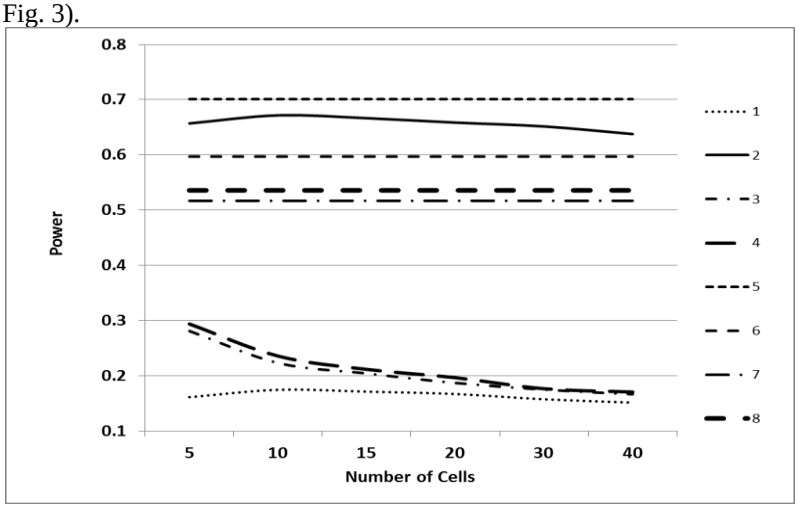


Figure 3. Power of different MVN tests for n = 250, p = 2 against the multivariate Student t probability distribution with 10 df as a function of the number of equiprobable grouping cells. The number of replications was N = 10,000.

From Fig. 3 we see first that the Dzhaparidze & Nikulin (1974) test like in the univariate case (see Voinov, Pya, Alloyarova (2009), p. 359) possesses very low power. The same can be said about the ChLeh test (line 3 on Fig. 3) considered by Moore and Stubblebine (1981) (see also Voinov *et al*. (2013), p. 431). We see also that power of the CarFer's statistic is very low as compared with the well known tests for multivariate normality introduced by Royston (1992), Anderson-Darling (see Henze (2002), p. 483), Székely & Rizzo (2005), Doornik & Hansen (2008), and Voinov *et al*. (2013).

3.2 A simulation study of the Power divergence tests

For testing the multivariate normality Batsidis *et al*. (2013) suggest to use the family of power divergence statistics

$$Z(\lambda)=\begin{cases}\dfrac{2}{\lambda(\lambda+1)}\sum\limits_{i=1}^{M}O_i\left(\left(\dfrac{O_i}{E_i}\right)^{\lambda}-1\right), & -\infty<\lambda<\infty,\ \lambda\neq-1,0,\\ 2\sum\limits_{i=1}^{M}E_i\log\dfrac{E_i}{O_i}, & \lambda=-1,\\ 2\sum\limits_{i=1}^{M}O_i\log\dfrac{O_i}{E_i}, & \lambda=0,\end{cases} \tag{9}$$

where O_i and E_i are the observed and expected frequencies for the i^{th} equiprobable interval, $i=1,...,M$. Particular values of a parameter λ correspond to: CarFer chi-squared test ($\lambda=1$) considered in detail in Sections 2 and 3.1, likelihood ratio test ($\lambda=0$), Freeman-Tukey statistic ($\lambda=-0.5$), modified chi-squared test ($\lambda=-2$), and Cressie-Read statistic ($\lambda=2/3$). The authors of Batsidis *et al*. (2013), p. 2256 announced that under the same procedure as in Section 2 the limiting null distribution of $Z(\lambda)$ will follow the chi-squared distribution with $M-1$ degrees of freedom. We have already shown that this is not true if $\lambda=1$. The same conclusion can be made for other values of λ. To avoid theoretical work for deriving the limit null distributions of $Z(\lambda)$ for different λ, we investigated simulated critical values of $Z(\lambda)$ having compared them with those for the chi-squared distribution with $M-2$ and $M-1$ degrees of freedom (see Table 3).

Table 3. Simulated critical values of level $\alpha = 0.05$ for $Z(\lambda)$ if n = 250, p = 2. We give $\pm$ one simulated standard deviation assessed as in Table 1.

M	Chi_M-2	$\lambda = -1$	$\lambda = 0$	$\lambda = 2/3$	Chi_M-1
2	**N/A**	**1.94 ± 0**	**1.94 ± 0**	**1.94 ± 0**	**3.841**
3	**3.841**	**4.39 ± 0.05**	**4.38 ± 0.03**	**4.33 ± 0.03**	**5.991**
5	**7.815**	**8.28 ± 0.06**	**8.02 ± 0.03**	**8.08 ± 0.06**	**9.488**
10	**15.507**	**16.47 ± 0.06**	**15.83 ± 0.06**	**15.53 ± 0.05**	**16.919**
20	**28.869**	**32.34 ± 0.04**	**29.43 ± 0.07**	**28.88 ± 0.08**	**30.144**
30	**41.337**	**49.89 ± 0.07**	**42.40 ± 0.08**	**41.19 ± 0.07**	**42.557**
40	**53.384**	**∞**	**55.53 ± 0.10**	**53.22 ± 0.04**	**54.572**

From Table 3 we see that for $\lambda = 0$ and $\lambda = 2/3$, except possibly $M = 30$, and $M = 40$ the critical values of $Z(\lambda)$ are between those of the critical values for the chi-squared distributions with $M - 2$ and $M - 1$ degrees of freedom as it was observed by Moore and Stubblebine (1981) for the classical Pearson's sum based on raw data. The case of $\lambda = -1$ is the most interesting. The critical value of $Z(-1)$ for M = 40 is infinitely large. This is quite expected because some equiprobable cells can be empty and, hence, $Z(-1)$ in (9) will degenerate. In the above simulation there were more than 5% infinities. Moreover for n = 250 and M = 5 the histogram of simulated values of $Z(-1)$ is well approximated by the chi-squared distribution with $M - 2$ (not $M - 1$ as it was announced in Batsidis *et al*. (2013), p. 2256) degrees of freedom (see Fig. 4).

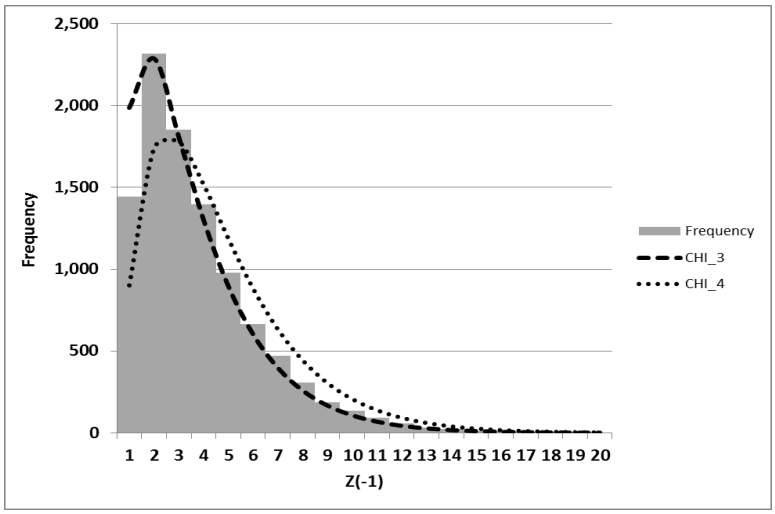


Figure 4. The histogram of N = 50,000 simulated statistics $Z(-1)$ (shaded area), the chi-squared distributions with M - 1 = 4 (dotted line), and M - 2 = 3 (dashed line) df.

3.3 An application for two well-known data sets

Let us consider the iris setosa data set which is available, e.g., in R. For this four-dimensional data set Batsidis *et al*. (2013), p. 2270 under the same procedure as described in Section 2 applied power divergence statistics with $\lambda = -0.5, 1, 2$ (see Table 4). In the third and fourth columns the values of $Z(\lambda)$ and p-values corresponding to the chi-squared distributions with M - 1 = 6 degrees of freedom are reproduced. Simulated critical values and simulated p-values are given in the second and fifth columns. In the last column we give the p-value obtained with the use of (7). One sees that even for a small sample of size 50 it is rather close to the simulated value.

Table 4. Critical values and p-values of $Z(\lambda)$ for Iris Setosa data set. n = 50, p = 4, M = 7, N = 50,000.

λ	Simulated critical values	$Z(\lambda)$ of Batsidis *et al.*	p-values of Batsidis *et al.*	Simulated p-values	"Exact" p-value
-0.5	**12.56 ± 0.07**	**10.1080**	**0.1202**	**0.1028**	-
1	**11.10 ± 0.10**	**7.9600**	**0.2410**	**0.1650**	**0.1584**
2	**11.74 ± 0.03**	**7.5328**	**0.2744**	**0.1997**	-

From Table 4 we see that the simulated p-values are less than those corresponding to the chi-squared distribution with $M-1=6$ degrees of freedom, because simulated histograms are better fitted by the chi-squared distributions with $M-2=5$ degrees of freedom. From these results it follows that at a level $\alpha = 0.05$ the null hypothesis about multivariate normality is not rejected. As we noted in Section 3.1 the power of $Z(1) = Z$ with respect to the multivariate Student t probability distribution with 10 degrees of freedom is very low (see Fig. 3). The value of the McCull statistic introduced in Voinov *et al*. (2013) equals 0.0014 with the p-value of 0.97. We see that this rather powerful test also does not reject the null hypothesis but the probability value of McCull test is much higher than that of power divergence statistics.

Consider the six-dimensional data set of Royston (1983). Like in previous case we applied power divergence statistics with $\lambda = -0.5,\ 1,\ 2$ used by Batsidis *et al*. (2013) (see Table 5).

Table 5. Critical values and p-values of $Z(\lambda)$ for the Royston's data. n = 103, p = 6, M = 10, N = 50,000.

λ	Simulated critical values	$Z(\lambda)$ of Batsidis *et al*.	p-values of Batsidis *et al*.	Simulated p-values	"Exact" p-value
-0.5	**16.60** $\pm$ **0.02**	**13.5993**	**0.1373**	**0.1204**	-
1	**15.54** $\pm$ **0.09**	**15.5437**	**0.0770**	**0.0518**	**0.0494**
2	**16.28** $\pm$ **0.08**	**18.0593**	**0.0345**	**0.0324**	-

From Table 5 we see that as for iris setosa data set the simulated p-values (fifth column of the table) are also less than those corresponding to the chi-squared distribution with $M-1=9$ degrees of freedom (fourth column). From sixth column of the table we see that "exact" p-value is very close to simulated one.

The McCull test statistic for the Royston's data equals 11.76 with the probability value of 0.0006. So, the multivariate normality of Royston's data is rejected not only by $Z(2)$ but by the much more powerful McCull test as well.

4. Conclusion

Chernoff and Lehmann (1954) and Moore and Stubblebine (1981) in their classical papers have shown that if one use raw (ungrouped) data, then the Pearson's sum does not follow in the limit the chi-squared distribution and, actually, will be distributed as a weighted sum of independent chi-squared random variables. The main goal of this article is to prevent incorrect usage of Pearson's sum in extant research when unknown parameters are estimated by the raw data, e.g., by maximum likelihood or unbiased estimators as in the case considered. The power divergence statistics when used to test for multivariate normality possess low power against close alternatives, such as, e.g., multivariate Student *t* distribution with 10 degrees of freedom that is very close to a multivariate normal distribution.

Appendix

Derivation of the Elements of matrix B

$$p_i(\theta,\theta_0)=(2\pi)^{-p/2}\,|\Sigma|^{-1/2}\int\limits_{\mathbf{x}\in\Delta} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}d\mathbf{x},\ \mathbf{x}\in R^p, \qquad \text{(A.1)}$$

where

$\Delta=\dfrac{(n-1)^2}{n}c_{i-1}\le(\mathbf{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\mathbf{x}-\boldsymbol{\mu}_0)\le\dfrac{(n-1)^2}{n}c_i$,

$0=c_0<c_1<\cdots<c_M=1$ are $i/M,\ i=1,...,M-1,$ points of the beta

distribution with parameters $p/2$ and $(n-p-1)/2$, $\mathbf{x}=(x_1,...,x_p)^T$, $\mathbf{z}=(z_1,...,z_p)^T$, $\boldsymbol{\mu}$, $\boldsymbol{\mu}_0$ are *p*-vectors, and $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_0$ are positive-definite $p\times p$ covariance matrices. Let us use the following transformation

$$\mathbf{z}=\boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_0)\Rightarrow(\mathbf{x}-\boldsymbol{\mu}_0)=\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}\Rightarrow\mathbf{x}=\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}+\boldsymbol{\mu}_0.$$

Under this transformation the determinant of the Jacobian will be

$|\mathbf{Jac}|=\left|\frac{\partial(x_1,...,x_p)}{\partial(z_1,...,z_p)}\right|=|\boldsymbol{\Sigma}_0|^{1/2}$, and, hence, $d\mathbf{x}=|\boldsymbol{\Sigma}_0|^{1/2}\,d\mathbf{z}$. Formula (A.1) becomes

$$p_i(\theta,\theta_0)=(2\pi)^{-p/2}\frac{|\boldsymbol{\Sigma}_0|^{1/2}}{|\boldsymbol{\Sigma}|^{1/2}}\int\limits_{\mathbf{z}\in\Delta_p}e^{-\frac{1}{2}\left[(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})\right]}d\mathbf{z},\quad\text{(A2)}$$

where $\Delta_p=\frac{(n-1)^2}{n}c_{i-1}\le\mathbf{z}^T\mathbf{z}\le\frac{(n-1)^2}{n}c_i$.

Differentiating under the sign of integral we have from (A.2)

$$\frac{\partial p_i(\theta,\theta_0)}{\partial\boldsymbol{\mu}}=(2\pi)^{-p/2}\frac{|\boldsymbol{\Sigma}_0|^{1/2}}{|\boldsymbol{\Sigma}|^{1/2}}\int\limits_{\mathbf{z}\in\Delta_p}\frac{\partial}{\partial\boldsymbol{\mu}}e^{-\frac{1}{2}\left[(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})\right]}d\mathbf{z}.$$

Due to the formula (11.1) of Dwyer (1967), p.617

$$\frac{\partial}{\partial\boldsymbol{\mu}}e^{-\frac{1}{2}\left[(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})\right]}=$$

$$=e^{-\frac{1}{2}\left[(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})\right]}\Sigma^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu}).$$

Taking into account that $\theta=\theta_0$ means $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$, it follows that

$$\frac{\partial}{\partial\boldsymbol{\mu}}e^{-\frac{1}{2}\left[(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}-\boldsymbol{\mu}_0-\boldsymbol{\mu})\right]}\Bigg|_{\theta=\theta_0}=$$

$$=e^{-\frac{1}{2}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z})^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{z})}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}=\boldsymbol{\Sigma}_0^{-1/2}\mathbf{z}e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}},$$

and, finally, due to symmetry of Δ_p

$$\frac{\partial p_i(\theta,\theta_0)}{\partial\boldsymbol{\mu}}\Bigg|_{\theta=\theta_0}=(2\pi)^{-p/2}\boldsymbol{\Sigma}_0^{-1/2}\int\limits_{\mathbf{z}\in\Delta_p}\mathbf{z}e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}}d\mathbf{z}=\mathbf{0},\ i=1,...,M,\quad\text{(A.3)}$$

where $\mathbf{0}$ is a zero *p*-vector. From (A.3) it follows that matrix **B** has *p* zero columns.

Changing $\mathbf{u}=\mathbf{x}-\boldsymbol{\mu}$ in (A.1), $\mathbf{x}=\mathbf{u}+\boldsymbol{\mu}$, $d\mathbf{x}=d\mathbf{u}$, gives

$$p_i(\theta,\theta_0) = (2\pi)^{-p/2} \mid \Sigma \mid^{-1/2} \int\limits_{\mathbf{u}\in\Delta_{p1}} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} d\mathbf{u},\ \mathbf{u}\in \mathrm{R}^p, \tag{A.4}$$

where $\Delta_{p1} = \dfrac{(n-1)^2}{n} c_{i-1} \le (\mathbf{u}+\boldsymbol{\mu}-\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{u}+\boldsymbol{\mu}-\boldsymbol{\mu}_0) \le \dfrac{(n-1)^2}{n} c_i$. Due to the product rule and Dwyer (1967), formulas (11.3), (11.8)

$$\frac{\partial}{\partial\boldsymbol{\Sigma}}\left\{\mid\boldsymbol{\Sigma}\mid^{-1/2} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}}\right\} = \frac{\partial \mid\boldsymbol{\Sigma}\mid^{-1/2}}{\partial\boldsymbol{\Sigma}} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} + \mid\boldsymbol{\Sigma}\mid^{-1/2} \frac{\partial}{\partial\boldsymbol{\Sigma}} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} =$$

$$= -\frac{1}{2}\mid\boldsymbol{\Sigma}\mid^{-1/2} \boldsymbol{\Sigma}^{-1} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} + \frac{1}{2}\mid\boldsymbol{\Sigma}\mid^{-1/2} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} \boldsymbol{\Sigma}^{-1}\mathbf{u}\mathbf{u}^T\boldsymbol{\Sigma}^{-1} = \tag{A.5}$$

$$= \frac{1}{2}\mid\boldsymbol{\Sigma}\mid^{-1/2} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} \left\{\boldsymbol{\Sigma}^{-1}\mathbf{u}\mathbf{u}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\right\}.$$

Differentiating under the sign of integral in (A.4) we get

$$\frac{\partial p_i(\theta,\theta_0)}{\partial\boldsymbol{\Sigma}} = (2\pi)^{-p/2} \int\limits_{\mathbf{u}\in\Delta_{p1}} \frac{\partial}{\partial\boldsymbol{\Sigma}}\left\{\mid\boldsymbol{\Sigma}\mid^{-1/2} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}}\right\} d\mathbf{u}.$$

From this and (A.5) it follows that

$$\left.\frac{\partial p_i(\theta,\theta_0)}{\partial\boldsymbol{\Sigma}}\right|_{\theta=\theta_0} = (2\pi)^{-p/2} \int\limits_{\mathbf{u}\in\Delta_{p1}} \frac{1}{2}\mid\boldsymbol{\Sigma}_0\mid^{-1/2} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} \left\{\boldsymbol{\Sigma}_0^{-1}\mathbf{u}\mathbf{u}^T\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}\right\} d\mathbf{u} = \tag{A.6}$$

$$= \frac{1}{2}\boldsymbol{\Sigma}_0^{-1}\left\{(2\pi)^{-p/2}\mid\boldsymbol{\Sigma}_0\mid^{-1/2} \int\limits_{\mathbf{u}\in\Delta_{p1}} \mathbf{u}\mathbf{u}^T e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}} d\mathbf{u}\right\}\boldsymbol{\Sigma}_0^{-1} - \frac{1}{2} p_i(\theta_0,\theta_0)\boldsymbol{\Sigma}_0^{-1}.$$

Changing $\mathbf{u} = \boldsymbol{\Sigma}_0^{1/2}\mathbf{z}$ and taking into account that $d\mathbf{u} = \mid\boldsymbol{\Sigma}_0\mid^{1/2} d\mathbf{z},\ \mathbf{u}^T\boldsymbol{\Sigma}_0^{-1}\mathbf{u} = \mathbf{z}^T\mathbf{z},\ \mathbf{u}\mathbf{u}^T = \boldsymbol{\Sigma}_0^{1/2}\mathbf{z}\mathbf{z}^T\boldsymbol{\Sigma}_0^{1/2}$ we obtain from (A.6)

$$\left.\frac{\partial p_i(\theta,\theta_0)}{\partial\boldsymbol{\Sigma}}\right|_{\theta=\theta_0} = \frac{1}{2}\boldsymbol{\Sigma}_0^{-1/2}\left\{(2\pi)^{-p/2} \int\limits_{\mathbf{z}\in\Delta_{p2}} \mathbf{z}\mathbf{z}^T e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}} d\mathbf{z}\right\} - \frac{1}{2} p_i(\theta_0,\theta_0)\boldsymbol{\Sigma}_0^{-1/2}, \tag{A.7}$$

where $\Delta_{p2} = \dfrac{(n-1)^2}{n} c_{i-1} \le \mathbf{z}^T\mathbf{z} \le \dfrac{(n-1)^2}{n} c_i$. Using arguments of Moore and Stubblebine (1981), pp. 734-735 the formula (A.7) reduces to

$$\left.\frac{\partial p_i(\theta,\theta_0)}{\partial\boldsymbol{\Sigma}}\right|_{\theta=\theta_0} = d_i\boldsymbol{\Sigma}_0^{-1},\ i = 1,\ldots,M, \tag{A.8}$$

where

$$d_i = \left(c_{i-1}^{p/2} e^{-\frac{(n-1)^2}{2n}c_{i-1}} - c_i^{p/2} e^{-\frac{(n-1)^2}{2n}c_i}\right)\frac{(n-1)^p b_p}{2n^{p/2}}, \tag{A.9}$$

$b_p = [p(p-2)\cdots(4)(2)]^{-1}$ if p is even,

and $b_p = (2/\pi)^{1/2}[p(p-2)\cdots(5)(3)]^{-1}$ for odd p.

From (A.3) and (A.8) it follows that all columns of the matrix **B** are scalar multipliers of the vector $(d_1,...,d_M)^T$, and, hence, rank of **B** is 1.

Using the rest arguments of Moore and Stubblebine (1981), pp. 735-736 it can be shown that for equiprobable grouping cells

$$\Lambda = 1 - 2Mp\sum_{i=1}^{M} d_i^2. \qquad \text{(A.10)}$$

References

1. Batsidis, A., Martin, N., Pardo, L. & Zografos, K. (2013). A necessary power divergence type family tests of multivariate normality. *Communications in Statistics - Simulation and Computation*. 42, 2253-2271.
2. Cardoso De Oliveira, I.R. & Ferreira, D.F. (2010). Multivariate extension of chi-squared univariate normality test. *Journal of Statistical Computation and Simulation*. 80, 513-526.
3. Chernoff, H. & Lehmann, E.L. (1954). The use of maximum likelihood estimates in tests for goodness of fit. *The Annals of Mathematical Statistics*. 25, 579-589.
4. Clark, J. (1997). *Business Statistics*. Chicago: Barron's Educational Series, Inc.
5. Cressie, N. & Read, T.R.C. (1984). Multinomial goodness of fit test. *Journal of the Royal Statistical Society, Series B*. 46, 440-464.
6. Dzhaparidze, K.O. & Nikulin, M.S. (1974). On a modification of the standard statistic of Pearson. *Theory of Probability and its Applications*. 19, 851–853.
7. Doornik, J.A. and Hansen, H. (2008). Practitioners' corner: An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, supplement . 70, 927-939.
8. Farrell, P.J., Saliban-Barrera, M. & Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation* . 77, 1065-1080.
9. Gnanadesikan, R. & Kettenring, J.R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*. 28, 81-124.
10. Greenwood , P.E. & Nikulin , M.S. (1996). A Guide to Chi-squared Testing . New York: John Wiley.
11. Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical Papers*. 43, 467-506.
12. Moore, D.S. & Spruill, M.C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *The Annals of Statistics*. 3, 599-616.
13. Moore, D.S. & Stubblebine, J.B. (1981). Chi-square tests for multivariate normality with application to common stock prices. *Communications in Statistics - Theory and Methods*. A10(8), 713-738.
14. Read, T.R.C. & Cressie, N. (1988). *Goodness of Fit Statistics for Discrete Multivariate Data*. New York : Springer-Verlag.
15. Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk. *Applied Statistics*. 32, 121-133.
16. Royston, J. P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing* . 2, 117-119.
17. Székely, G.J. & Rizzo, M.L. (2005). A new test for multivariate normality. *J. of Multivariate Analysis*. 93, 58-80.
18. Voinov, V., Pya, N., Makarov, R. & Voinov, Y. (2013). A new invariant and consistent chi-squared type goodness-of-fit test for multivariate normality. In *JSM*

Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association. 425-439.
19. Voinov, V., Pya, N. & Alloyarova, R. (2009). A comparative study of some modified chi-squared tests. *Communications in Statistics - Simulation and Computation*. 38, 355-367.
20. Weiers, R.M. (1991). *Introduction to Business Statistics*. Chicago, Illinois : The Dryden Press.

Control charts for arbitrage-based trading

Angeliki Vyniou[1],Stelios Psarakis[1] and Kostas Triantafyllopoulos[2]
[1]Department of Statistics, Athens University of Economics & Business, Athens, Greece
[2]Department of Probability and Statistics, University of Sheffield, Sheffield, UK

Abstract: This paper concerns the role and contribution of SPC methods to pairs trading, a relative-value statistical arbitrage trading technique. Pairs trading and its numerous extensions have gained increased attention over the recent years and is a popular trading strategy among hedge funds and investments boutiques. The core idea of pairs trading is "buy low" and "short-sell high" and it is based on the assumption that the low-valued asset will gain value and the high-valued asset will lose value, so that the two assets are co-evolving or mean-reverting. Several authors have recently suggested that this mean-reversion may be appropriate only at some periods of time (in which profits may be realized) while in other periods of time mean-reversion is lost (resulting to significant losses, e.g. one could buy an asset which loses its value). In this paper we propose the use of appropriate SPC methods in order to detect mean-reversion, hence to identify tradable periods. The literature on pairs trading will be reviewed and examples will illustrate the need for a monitoring procedure. Finally, control charts for autocorrelated processes are proposed for the detection of mean-reversion.
Keywords: pairs-trading, control charts, SPC, autocorrelated processes, financial data, mean-reversion

1. Introduction

Pairs trading is a very common trading strategy among hedge funds and institutional investors because of the consistent, though usually modest, profitability. The strategy requires the choice of two assets whose prices are influenced by the same economic forces. The consequence of this is that the prices of these two assets actually co-move. The second step is to detect the time point that these prices diverge from their long term equilibrium so as to take action and short sell the overvalued share and buy the undervalued one. Profit is generated when the prices return to their long term equilibrium state. Several researchers propose methods and techniques for selecting the appropriate pairs of assets, identifying the spread magnitude that should trigger a trade, predict the next step of the process so as to proceed or not to opening a position.

More specifically, in order to apply a pairs trading strategy traders choose a pair of assets and decide on trading their spread. The spread of these assets could be defined as the difference of the prices of these assets or another linear combination of the two prices. The rationale behind profit generation is quite simple. When the price of one asset is low the trader buys this asset while simultaneously short-sell the other whose price is at that time high. Then it is said that the trader opens a position. Short-selling means that the trader lends an amount of shares and sells them. These shares though must be returned sometime in the future. To close the position the trader must sell the number of shares bought and buy the shares of the "short-sold" asset, in order to return them. Profits are generated when the price of the undervalued asset increases whereas the price of the overvalued asset decreases. For book length versions on pairs trading one can refer to Ehrman [1], Vidyamurthy [2] and Whistler [6].

There is a main assumption made for financial markets, which affects trading strategies, that is efficient market hypothesis (Fama [4]). This assumption is based on that information is considered to be instantly and equally distributed among all market participants. Therefore, it is argued that all relevant information is reflected and incorporated in current asset prices. A direct result of this assumption is that no excessive returns can be gained. Applying a pairs trading strategy, though, gives investors the opportunity to exploit temporary market inefficiencies, disturbances of the levels of asset prices that do not last long, and therefore, benefit from the excessive returns

which then can be generated.

There are three main methods on pairs trading considered in the literature. The distance method proposed by Gatev *et al.* [5], according to which the pairs chosen are the ones that have the smallest sum of squared deviations. Although this method is cost efficient, it is based on the assumption that the price differences follow a standardized normal distribution while it is known that share prices usually follow a log-normal distribution.

The second is the correlation method discussed in Ehrman [2]. In his book he proposes to consider the correlation of a pair of shares as a factor affecting the selection of appropriate pairs to be traded. It is suggested that in cases when the correlation coefficient is greater than or equal to 0.7 then a pair of assets could be considered tradable. Moreover, it is recommended that the correlation coefficients be measured and monitored at several time intervals so as to detect any changes in the assets relation.

The third is cointegration method introduced in Engel and Granger [6] and Engel and Yoo [7]. According to this method two non-stationary price time series are said to be co-integrated when there exists at least one linear combination of them that is a stationary process see Figure 1. Moreover, the co-integrated pair is assumed to have a long-term equilibrium and a self adjustment mechanism that is triggered when deviations from the equilibrium state occur.

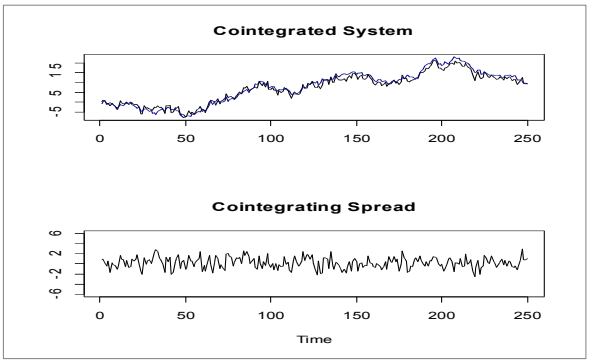


Figure 1. Cointegrated assets and their spread

Other approaches that can be found in the literature are those of stochastic modeling, copulas and artificial neural networks. Scholars also model the spread process either in the context of continuous (Ornstein-Ulenbeck processes) or discrete time. Several empirical studies have also been conducted to investigate the effectiveness of this trading strategy in different markets under various economic conditions.

As previously mentioned there are many studies published concerning pairs-trading. In this paper a review of some recently proposed methodologies will be presented as well as the potential of using Statistical Process Control methods in the pairs trading context. The studies that investigate the statistical properties of this trading strategy are to be presented in the next section. In Section 3 the potential of the use of Statistical Process Control techniques in the context of pairs-trading is to be investigated. In the last section the main conclusions of the study as well as some issues for further research are to be summarized.

2. Literature review – Pairs trading

Several researchers propose methods and techniques for selecting the appropriate pairs of assets, identify the spread magnitude that should trigger a trade, predict the next step of the process so as to proceed or not to opening or closing a position. As mentioned above, there are three main methods to approach a pairs trading strategy that are considered in the literature; the distance method Gatev [5], the correlation method Ehrman [1] and the co-

integration mehtod Engel and Granger [6] and Engel and Yoo [7] see Figure 1. Other approaches that can be found in the literature are those of stochastic modeling, copulas and artificial neural networks. Scholars also model the spread process in the context of continuous time – Ornstein-Ulenbeck processes (Ulenbeck-Ornstein [8])– or discrete time. Several empirical studies have also been conducted to investigate the effectiveness of this trading strategy in different markets under various economical conditions. In this talk a review of the several proposed methodologies will, first, be presented .

Reviewing pairs trading literature it is obvious that most scholars focus on improving the modeling of either spread or prices time series attempting to develop models that better reflect reality. Various models have been developed so far. Some can be found in the studies of Elliot *et al.* [9] who propose the use of a mean reverting Gaussian Markov chain model for the spread of pairs trading strategies. In another study, Dattasharma [10] introduce a general framework that can be used to predict the dependence between two stocks based on any user-defined criterion by applying the concepts of events and episodes. Triantafyllopoulos and Montana [11] propose a Bayesian state-space model for spread processes with time varying parameters. In this study the researchers also developed an on-line estimation algorithm that could be used to monitor data for mean reversion. Gatarek *et al.* [12] suggest a combination of Dirichlet process prior techniques with Bayesian estimation to estimate co-integrated models with non-normal disturbances. Triantafyllopoulos and Han [13] propose a methodology for detecting mean-reverted segments of data streams in algorithmic pairs trading using a state-space model for the spread and propose two new recursive least squares (RLS) algorithms with adaptive forgetting for predicting mean-reversion in real time. Tourin and Yan [14] in their study suggest the use of an optimal stochastic control model to address the problem of analyzing dynamic pairs trading strategies. In Fasen [15] the asymptotic properties of the least squares estimator for the model parameter of a multivariate Ornstein-Uhlenbeck model are investigated. Alrasheedi and Al-Ghamedi [16] apply a Vector Auto-Regressive model (VAR) for the simulation of the time series of two stocks and examine the influence of some of the model parameters on the total profits earned. Improving and developing models is indeed very important since more accurate predictions of assets future prices can then be obtained. Unlike other processes, financial processes are difficult to be predicted because of the nature of financial data and it is generally argued that the best prediction of a tomorrow's asset price is the price of the asset today. This explains the vast amount of studies published on modeling financial time series considering continuous or discrete time. There are also non-parametric approaches proposed for handling financial data, as well. An example is the study of Bogomolov [17] in which a novel non-parametric approach for pairs trading is proposed in which the only assumption to be made is that the statistical properties of the volatility of the spread process remain reasonably constant.

Gatev *et al.* [5] proposed the GGR model for applying a pairs trading strategy. The study leads to the conclusion that excessive returns are likely to be generated for market participants that have relatively low transaction costs and the ability to short sell securities. It is also observed that there is a latent risk factor that affects the profitability of pairs trading over time. Papadakis and Wysocki [18] examine whether accounting information events, such as earnings announcements and analysts' earnings forecasts, have an effect on the profitability of the pairs trading strategy proposed by Gatev *et al.* [5]. Broussard and Vaihekoski [19] extended the work of Gatev *et al.[5]* through an empirical study showing that the aforementioned investment strategy is profitable even in markets with reduced liquidity. In the study of Wang and Mai [20] a comparison of GGR, Herlemont and FTBD pairs trading openning position strategies is conducted. The main conclusion obtained from this study is that after deducting the trading cost, the absolute income of the three strategies considered is significantly bigger than zero.

Portfolio optimization, i.e. the choice of which assets and the number of stocks from each asset are to be traded in order to attain maximum profits, is another issue considered in pairs trading literature. Perlin [21] suggests a

multivariate version of pairs trading which can be used to create an artificial pair for a specific stock using the information associated to m assets. Mudchanatongsuk et al. [22] propose a stochastic control approach to address pairs trading portfolio optimization. Chiu and Wong [23] investigate the continuous-time mean-variance portfolio selection problem considering co-integrated assets. Alsayedand and McGroarty [24] introduce a solution to the portfolio optimization problem when risky arbitrage trading is considered through the introduction of a nonlinear generalization of Ornstein-Uhlenbeck model which takes into consideration important risk factors.

Moreover, the trading costs are also taken into consideration in some studies. Transaction costs are those associated to opening or closing a position. In the study of Lin *et al.* [25] researchers propose the integration of loss limitation within the statistical modeling of pairs trading strategies. In several empirical studies transaction costs are also considered in order to assess the performance of the different methodologies. Trading costs can be significant and if not taken into consideration the returns of applying a pairs trading strategy could be minimized.

Various trading rules have also been proposed in order to perform successful pairs trading. In the study of Song and Zhang [26] pairs trading is investigated and a pairs trading rule is proposed which takes into account profit maximization or losses minimization. The approach of the researchers to address the problem considered is dynamic programming.

Some recent studies also consider the process's microstructure using intra-day data. Microstructure theory focuses on how specific trading mechanisms affect the price formation process. Zebedee and Kasch-Haroutounian [27] in their study examine the microstructure of the co-movement among the returns of stocks on an intra-day basis applying a combination of a traditional lead-lag model with a pseudo-error correction mechanism. Marshall *et al.* [28] investigate the microstructure of pairs trading on an intra-day basis. In other words, they examine the intra-day market characteristics that can be observed when arbitrage opportunities appear. Since pairs trading could be applied on an a daily basis and traders exploit daily market disturbances the examination of process' microstructure is indeed very interesting.

A basic step in pairs trading is to be able to identify suitable pairs so that the pairs trading to be profitable. To this end, several researchers try to develop an optimal methodology for choosing the most suitable pairs. Gatev *et al.* [5] proposes choosing the pairs having the smallest sum of squared deviations for trading. Ehrman [1] suggests pairs to be chosen using the correlation coefficient. When this coefficient is greater than or equal to 0.7, the pair is tradable. Engle [6] introduce co-integration approach and proposed choosing pairs whose prices are co-integrated. In the study of Huck [29] the sensitivity of pairs trading strategies' returns on the length of the pairs formation period is investigated. Through an example it is shown that the choice of the formation period affects the returns of the strategy employed and after taking into consideration the data snooping bias this result does not change.

Various empirical studies have also been published. In Matteson *et al.* [30] researchers introduce a new methodology in identifying local stationarity of non-stationary processes. Through an empirical approach robust estimates of time varying windows of stationarity are “produced”. Moreover, it is proven that using the adaptive window leads to higher returns and, in some cases, holding the positions open for a shorter period of time. Mai and Wang [31] published a limited study on the impact of the structure of the market on the returns of a pure statistical pairs trading. The researchers suggest that the annual rate of return of pairs trading can be improved by choosing the markets the traders are operating in.

Some different approaches have also been recently developed. Huck [29] proposes a methodology that can be used for pairs selection in a highly non-linear environment. The researcher combines forecasting techniques (Neural Networks) and multi-criteria decision making methods to select and trade pairs under pairs trading strategies. Artificial Neural Network models

(ANN) are presented by Gomide and Milidiu [32] that are used to predict spread time series. Through obtaining spread predictions, times of the day when to perform a particular Pair Trading can be recommended. The use of copulas in development of pairs trading strategies is investigated in Liew and Wu [33] It is suggested that copulas approach is a good alternative to the traditional ones – distance approach and co-integration approach – since it is not necessary to assume the existence of correlation among the values of the assets to be traded and thus, argued to be realistic and robust.

3. Potential of the use of SPC tools

The main questions that arise considering pairs trading are summarized in the following: When is the optimal time to open and close a position? How many trades are optimal in a given time period? What is the optimal time for a position to be kept open? What is the optimal way to handle non-stationary spread processes? What is the optimal way to select pairs? All these questions should be answered considering the ultimate goal of investors, i.e. maximum profitability, and the distinctive characteristics of financial time series.

Some of these questions could be addressed using Statistical Process Control techniques. Control charts signal when the process monitored is observed to be out of control. Thus, as long as we can model a process and we are able to estimate its long term parameters control charts can be constructed and used to aid traders identify tradable periods, in which they will have to take action, i.e. to open or close a position.

Moreover, control charts could be used to identify mean-reverting periods of a process that is locally mean-reverting (Figure 2). Mean reverting processes are those that are characterized by a long term equilibrium, they present a constant mean and variance in the long run. In case these processes are disturbed and deviate from their equilibrium they are expected to revert within a short time interval.

While this can be observed in some cases mean-reversion could be, as stated before, local, meaning that a series could be mean-reverting during some periods while in others could be non-stationary, non mean-reverting. In this case control charts could be constructed so as to signal when mean reversion is observed. In this case it is of great importance for practitioners to be able to identify mean reverting segments in non-mean reverting processes. This would actually mean that the possible pairs of stocks to be traded are not only the "obvious" ones. Practically a practitioner could arbitrary choose pairs of stocks and monitor their spread to detect periods where it is mean reverting. That would lead to more trading strategies since the pairs to be traded could be "uniquely" chosen by each trader without using any specific rule. That would actually mean higher returns for the traders. Unit root tests, Dickey – Fuller (Dickey and Fuller [34]), Phillips – Perron (Phillips and Perron [35]) KPSS (Kwiatkowski *et al.* [36]), can be employed to check whether a process is integrated of order one or stationary (mean-reverting).

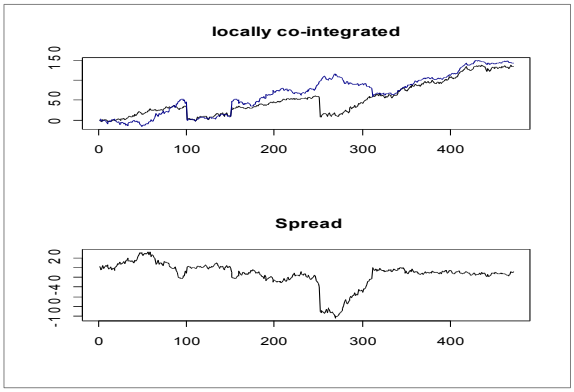


Figure 2. Locally co-integrated assets and their spread.

Several surveillance methodologies for detecting I(0) and I(1) segments of time series are developed in the studies of Steland [39], [40], [39] and [40] where he respectively proposes a control chart (stopping time) based on sequential Dickey-Fuller unit root test statistic to detect stationary segments of a time series, a control chart based on weighted Dickey-Fuller unit root test statistic to detect stationarity, a control chart for monitoring sequentially a time series with stopping times based on a sequential version of a kernel-weighted variance-ratio statistic and a sequential monitoring procedure that relies on KPSS unit root test statistic to detect whether the error terms in a polynomial regression model behave as a random walk or as a stationary process. In the study of Steland and Weidauer [41] the researchers investigated a monitoring procedure based on KPSS unit root test to address the problem of detecting sequentially stationary error terms in a multiple regression model, examining the case of co-integration as a special case, as well. The proposed monitoring procedures can be used for developing modifications that would be suitable for being used in the context of pairs trading. This issue is going to be examined further in an other study.

For all the above reasons, this study proposes the use of Statistical Process Control tools, i.e. control charts, as a means to enhance timely detection of potential market inefficiencies. Since mean reverting models are commonly used to describe the spread between the prices of two assets when a pairs trading investment strategy is applied, SPC techniques could be employed to identify tradable periods. The concept of co-integration is largely used in pairs trading literature to describe the relation between two financial variables. Co-integration refers to the co-movement of two time series, which actually means that the two processes considered have common stochastic trends. It has been proven that if the prices are "tied" together there will be a linear combination of them that will be stationary. This linear combination can be the spread of the two prices of interest. In order to estimate the long term mean and variance of the spread we have to estimate the parameters of this process. The unconditional expected value of the dependent variable, spread, is the estimated long-term mean and the unconditional variance is the long-term process variance.

In the context of SPC the long-run mean provides the center line of the control chart to be constructed and the variance assists in the construction of the respective control limits. Autocorrelated processes have been thoroughly investigated in the SPC literature. Several control charts have been proposed for monitoring stationary processes. A residuals control chart for the spread of two co-integrated processes is presented in Figure 3. The combined Shewhart – EWMA control chart of Lu and Reynolds [42] could also be employed to detect shifts of the disequilibrium factor. The rationale is quite simple considering that in pairs trading an appropriate control chart should timely detect step changes (shocks) as well as slow drifts of the process mean. This way short term disturbances would be detected and traders would act accordingly in order to make profitable investments. When the chart signals that would mean that the process is out-of-control which would lead the trader to decide upon opening or closing a position. While opening a position can be achieved by using a control chart like the one mentioned before, the right time to close the position so as to maximize profits must then be determined. A lower one-sided CUSUM control chart for autocorrelated data could be used to determine this time point. This issue is usually tackled by using an impulse response function which gives detailed information about the effects of a possible shock over time (Alexander [43]). The combined Shewhart-EWMA and the lower one-sided CUSUM chart in the context of pairs trading are left to be investigated in a future study.

There are several tools used in the context of technical analysis, most of which are usually empirical. Traders use oscillators and indices to make decisions upon their trades. Bollinger Bands is such a technical analysis tool invented by John Bollinger. The construction of a chart using Bollinger Bands can be used to determine the time of opening and closing a position. Bollinger Bands are actually envelopes that surround the price bars plotted two standard deviations away from a simple moving average, which may or may not be displayed and they are defined as $\left(MA \pm K\sigma\right)$. The window for this

moving average can vary but it is usually 20 and $K=2\sigma$. In the case of the spread process a simple moving average of the spread process is used and the band surrounds this moving average. The time of opening a position is determined when the spread crosses the upper control limit while the position is closed when the process returns in the in-control area.

An example of the use of a simple residuals control chart and the respective Bollinger Bands are shown below in Figue 3. and Figure 4. The following charts were

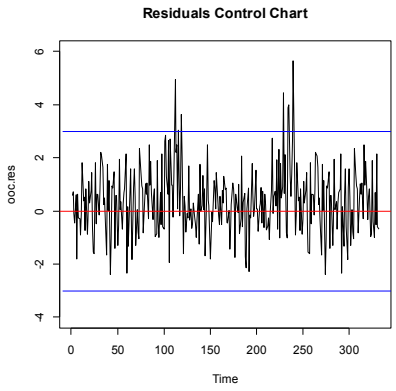


Figure 3. Residuals control chart for the spread of two co-integrated assets.

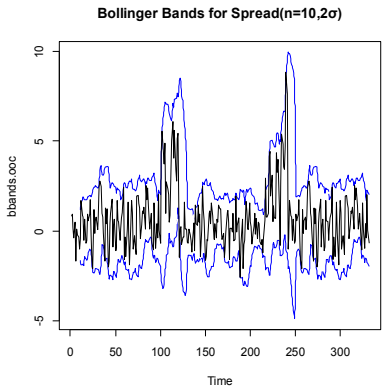


Figure 4. Bollinger Bands for the spread of the same two co-integrated assets.

constructed using a set of simulated data. The data were generated so as the co-integrating system of the two assets considered to present short term deviation from its long term equilibrium state. The two out-of-control segments can be easily observed. In the out-of-control segments both the mean and the variance of the spread process were different from the long term, true parameters. The residuals of the generated spread process are presented in Figure 3 in the residuals control chart. The Bollinger Bands for the respective spread process is presented in Figure 4.

Observing the two figures one can realize that the traditional residuals chart is able to detect the out of control segments while the technical analysis tool seems to be far less appropriate for this purpose. It can be seen that the upper limit of the Bollinger Bands chart is approached only in the second out-of-control period, whereas the first the out of control situation is not detected. In the trading context this actually can be interpreted as a missed opportunity of making profit.

4. Conclusions and Further Research

Since a symmetry of assumptions in Pairs-trading and Statistical Process Control literature is evident, one can straightforward conclude that SPC techniques can be effectively used in the pairs trading context. In the case of co-integrating assets, examined in this study, the use of the traditional residuals chart was proven to be more effective in identifying tradable periods than that of Bollinger Bands, the technical analysis tool which is usually used to trigger potential trades. Although this study is quite limited, it introduces a new area of research for the SPC scholars. The adjustment of SPC tools and theory to account for financial data and more specifically the use of these tools in the decision making process for traders employing relative-value statistical arbitrage trading techniques such as pairs-trading is a promising area of investigation.

The conclusions drawn from this study are to be substantiated further in a new study using real data, closing share prices. Furthermore, the use of all the control charts mentioned in the study (combined Shewhart-EWMA and lower CUSUM) are to be used in order to investigate the effectiveness of each under various circumstances. This investigation will also enable a comparison

of the proposed charts, when used in the Pairs-trading context, that will eventually lead to the selection of the most appropriate charts to be used as trading decision making tools. Another issue to be investigated in the future is the use of control charts for detecting mean reverting segments of non-stationary processes.

References

[1] D. S. Ehrman, *The handbook of pairs trading*. Wiley, New York, 2006.

[2] G. Vidyamurthy, *Pairs Trading: quantitative methods and analysis*, vol. 217. John Wiley & Sons, 2004.

[3] M. Whistler, *Trading Pairs: capturing profits and hedging risk with statistical arbitrage strategies*, vol. 216. John Wiley & Sons, 2004.

[4] E. F. Fama, "Efficient capital markets: A review of theory and empirical work*," *J. Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[5] E. Gatev, W. Goetzmann, and K. Rouwenhorst, "Pairs Trading: Performance of a Relative-Value Arbitrage Rule," *Rev. Financ. Stud.*, vol. 19, no. 3, pp. 797–827, Feb. 2006.

[6] R. F. Engle and C. W. J. Granger, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, Mar. 1987.

[7] R. F. Engle and B. S. Yoo, "Forecasting and testing in co-integrated systems," *J. Econom.*, vol. 35, no. 1, pp. 143 – 159, 1987.

[8] G. E. Uhlenbeck and L. S. Ornstein, "On the Theory of the Brownian Motion," *Phys Rev*, vol. 36, no. 5, pp. 823–841, Sep. 1930.

[9] R. J. Elliott, J. Van Der Hoek *, and W. P. Malcolm, "Pairs trading," *Quant. Finance*, vol. 5, no. 3, pp. 271–276, Jun. 2005.

[10] A. Dattasharma, P. K. Tripathi, and S. Gangadharpalli, "Identifying stock similarity based on episode distances," in *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, 2008, pp. 28–35.

[11] K. Triantafyllopoulos and G. Montana, "Dynamic modeling of mean-reverting spreads for statistical arbitrage," *Comput. Manag. Sci.*, vol. 8, no. 1–2, pp. 23–49, Apr. 2011.

[12] L. T. Gatarek, L. Hoogerheide, H. K. Van Dijk, and M. Verbeek, "A simulation-based Bayes' procedure for robust prediction of pairs trading strategies," *Tinbergen Inst. Discuss. Pap.*, pp. 09–061, 2011.

[13] K. Triantafyllopoulos and S. Han, "Detecting Mean-Reverted Patterns in Algorithmic Pairs Trading," in *Mathematical Methodologies in Pattern Recognition and Machine Learning*, Springer, 2013, pp. 127–147.

[14] A. Tourin and R. Yan, "Dynamic pairs trading using the stochastic control approach," *J. Econ. Dyn. Control*, vol. 37, no. 10, pp. 1972–1981, Oct. 2013.

[15] V. Fasen, "Statistical estimation of multivariate Ornstein–Uhlenbeck processes and applications to co-integration," *J. Econom.*, vol. 172, no. 2, pp. 325–337, Feb. 2013.

[16] M. A. Alrasheedi and A. A. Al-Ghamedi, "Some Quantitative Issues in Pairs Trading," *Res. J. Appl. Sci. Eng. Technol.*, vol. 5, no. 6, pp. 2264–2269, Feb. 2013.

[17] T. Bogomolov, "Pairs trading based on statistical variability of the spread process," *Quant. Finance*, vol. 13, no. 9, pp. 1411–1430, Sep. 2013.

[18] G. Papadakis and P. Wysocki, "Pairs trading and accounting information," *Boston Univ. MIT Work. Pap.*, 2007.

[19] J. P. Broussard and M. Vaihekoski, "Profitability of pairs trading strategy in an illiquid market with multiple share classes," *J. Int. Financ. Mark. Inst. Money*, vol. 22, no. 5, pp. 1188–1201, Dec. 2012.

[20] S. Wang and Y. Mai, "The GGR and Two Improved Pairs Trading Open Position Strategies," *Int. J. Adv. Inf. Sci. Serv. Sci.*, vol. 4, no. 7, pp. 326–334, Apr. 2012.

[21] M. Perlin, "M of a kind: A Multivariate Approach at Pairs Trading," 2007.

[22] S. Mudchanatongsuk, J. A. Primbs, and W. Wong, "Optimal pairs trading: A stochastic control approach," in *American Control Conference, 2008*, 2008, pp. 1035–1039.

[23] M. C. Chiu and H. Y. Wong, “Mean–variance portfolio selection of cointegrated assets,” *J. Econ. Dyn. Control*, vol. 35, no. 8, pp. 1369–1385, Aug. 2011.
[24] H. Alsayed and F. McGroarty, “Optimal portfolio selection in nonlinear arbitrage spreads,” *Eur. J. Finance*, vol. 19, no. 3, pp. 206–227, Mar. 2013.
[25] Y.-X. Lin, M. McCrae, and C. Gulati, “Loss protection in pairs trading through minimum profit bounds: A cointegration approach,” *J. Appl. Math. Decis. Sci.*, vol. 2006, pp. 1–14, 2006.
[26] Q. Song and Q. Zhang, “An optimal pairs-trading rule,” *Automatica*, vol. 49, no. 10, pp. 3007–3014, Oct. 2013.
[27] A. A. Zebedee and M. Kasch-Haroutounian, “A closer look at co-movements among stock returns,” *J. Econ. Bus.*, vol. 61, no. 4, pp. 279–294, Jul. 2009.
[28] B. R. Marshall, N. H. Nguyen, and N. Visaltanachoti, “ETF arbitrage: Intraday evidence,” *J. Bank. Finance*, vol. 37, no. 9, pp. 3486–3498, Sep. 2013.
[29] N. Huck, “Pairs trading and outranking: The multi-step-ahead forecasting case,” *Eur. J. Oper. Res.*, vol. 207, no. 3, pp. 1702–1716, Dec. 2010.
[30] D. S. Matteson, N. A. James, W. B. Nicholson, and L. C. Segalini, “Locally stationary vector processes and adaptive multivariate modeling,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 8722–8726.
[31] Y. Mai and S. Wang, “Whether Stock Market Structure Will Influence the Outcome of Pure Statistical Pairs Trading?,” 2011, pp. 291–294.
[32] P. Gomide and R. L. Milidiu, “Assessing Stock Market Time Series Predictors Quality through a Pairs Trading System,” 2010, pp. 133–139.
[33] R. Q. Liew and Y. Wu, “Pairs trading: A copula approach,” *J. Deriv. Hedge Funds*, vol. 19, no. 1, pp. 12–30, 2013.
[34] D. A. Dickey and W. A. Fuller, “Distribution of the Estimators for Autoregressive Time Series with a Unit Root,” *J. Am. Stat. Assoc.*, vol. 74, no. 366a, pp. 427–431, 1979.
[35] P. C. B. Phillips and P. Perron, “Testing for a unit root in time series regression,” *Biometrika*, vol. 75, no. 2, pp. 335–346, Jun. 1988.
[36] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?,” *J. Econom.*, vol. 54, no. 1, pp. 159–178, 1992.
[37] A. Steland, “On detection of unit roots generalizing the classic Dickey-Fuller approach,” Technical Report/Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen, 2005.
[38] A. Steland, “Weighted Dickey–Fuller processes for detecting stationarity,” *J. Stat. Plan. Inference*, vol. 137, no. 12, pp. 4011 – 4030, 2007.
[39] A. Steland, “MONITORING PROCEDURES TO DETECT UNIT ROOTS AND STATIONARITY,” *Econom. Theory*, vol. null, no. 06, pp. 1108–1135, 2007.
[40] A. Steland, “Sequentially Updated Residuals and Detection of Stationary Errors in Polynomial Regression Models,” *Seq. Anal.*, vol. 27, no. 3, pp. 304–329, Aug. 2008.
[41] A. Steland and S. Weidauer, “Detection of Stationary Errors in Multiple Regressions with Integrated Regressors and Cointegration,” *Seq. Anal.*, vol. 32, no. 3, pp. 319–349, 2013.
[42] C. . Lu and R. M. Reynolds, “Control charts for monitoring the mean and variance of autocorrelated processes,” *J. Qual. Technol.*, vol. 31, no. 3, pp. 259–274, 1999.
[43] C. Alexander, *Market risk analysis. quantitative methods in finance 2 2*. Chichester: Wiley, 2008.

Weighted Multivariate Fuzzy Trend Model for Seasonal Time Series

Erika Watanabe[1] and Norio Watanabe[2]

[1] Graduate School of Chuo University
Tokyo, Japan
(e-mail: a09.3ng6@g.chuo-u.ac.jp)
[2] Industrial and Systems Engineering, Chuo University
Tokyo, Japan
(e-mail: watanabe@indsys.chuo-u.ac.jp)

Abstract. Analyzing trends in multivariate time series is an important issue. A fuzzy trend model has been proposed for estimating trends in multivariate time series. This fuzzy trend model can decompose trends into common and individual trends. However, seasonality is not considered in this model. In this paper we propose a model including seasonality. Another problem of the former model is that common trend might differ from each series when there are large differences among series. Therefore we propose a scaling model which can decompose trends effectively by introducing weights. Usability of proposed models is demonstrated by a numerical example.
Keywords: Decomposition of trend, Common trend, Seasonal component, Fuzzy system.

1 Introduction

It is important to analyze trend included in multivariate time series. Most of models and methods are proposed for stationary time series or time series whose trends are removed previously. Models or methods for analyzing trend have not been developed sufficiently.

The moving average method and polynomial regression are typical methods for estimating trend. However, it is not easy to determine the length of interval for moving average. Polynomial regression can estimate trend easily but cannot follow irregular movement. On the other hand the fuzzy trend model ([1]) occupies an intermediate position between moving average method and polynomial regression and can analyze trend objectively and flexibly. Fuzzy trend models are also available for multivariate time series and can decompose trends into common and individual trends ([2], [3]). Whether time series is multivariate or scalar, most time series have seasonality or periodic components. However, seasonality is not considered in these models. In this paper we propose a multivariate fuzzy trend model including seasonal components.

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

The former model has another problem that the common trend might differ from each series when there are large differences among series. In this case standardization is difficult, since time series with trends are nonstationary, and mean values and variances depend on time points generally. To resolve this problem we propose a weighted model which can decompose trends effectively by introducing weights for scaling. We also consider a non-weighted fuzzy trend model with seasonal components and weighted fuzzy trend model without seasonal components. We provide an identification method for our models. Applicability of proposed models is demonstrated by a numerical example.

2 Fuzzy trend model

Let $\{y_{ln}|n = 1, ..., N,\ l = 1, 2, ..L\}$ denote the observed time series, where L is a number of time series and N is a length of time series. The new model proposed in this paper is defined as follows:

$$y_{ln} = \frac{1}{r_l}(\mu_{ln} + u^S_{ln} + x_{ln}) \tag{1}$$

$$\mu_{ln} = \sum_{k=1}^{K} \nu_{lk}(n)\mu_k \tag{2}$$

$$R_{lk}\text{: If } n \text{ is } A_{lk}, \text{ then } \mu_k = \alpha_{lk}(n - a_k) + \beta_{lk}, \tag{3}$$

where R_{lk} is a fuzzy if-then rule, $\nu_{lk}(n)$ is a membership function of a fuzzy set A_{lk}, $u_{lk} = \{(\alpha_{lk}\ \beta_{lk})'\}$ is an unobserved bivariate series (u' means the transpose of u), and x_{ln} is a zero-mean stationary process with variance σ^2.

When u_{lk} is a stochastic process, we assume that x_{ln} and u_{lk} are independent. The weight r_l is for standardization and we assume that $\sum_{l=1}^{L} r_l = 1$ and $r_l > 0$. The parameter a_k satisfies the equation:

$$a_k = a_{k-1} + d, \tag{4}$$

where d is a positive integer and $a_1 = 1$. We use the following membership function:

$$\nu_{lk}(n) = \frac{1}{2}\{\cos(\pi(n - a_k)/d) + 1\}. \tag{5}$$

Fig. 1 shows the membership functions of $A_{l1}, ..., A_{lK}$. The model given by (2)-(3) is a kind of Takagi-Sugeno's fuzzy system ([4]).

The term u^S_{ln} is a deterministic seasonal component, where $u^S_{l,n+p} = u^S_{ln}$ and p is the period. As a constraint we set $\sum_{n=1}^{p} u^S_{ln} = 0$ for all l. In this paper we assume that p is known.

Each trend is lead by the latent process $u_l = (u'_{l1}, ..., u'_{lK})'$ in this model. We assume that u_{lk} can be decomposed as follows:

$$u_l = u^C + u^I_l, \tag{6}$$

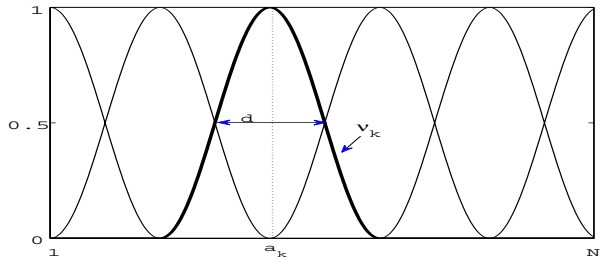


Fig. 1. Membership functions

where the latent processes u^C and u_l^I imply the common and individual trends respectively. Such decomposition requires a constraint condition. We will describe this condition later (cf. Eq. (26)).

As special cases of the model (1), the non-weighted model with seasonal components is given by

$$y_{ln} = \mu_{ln} + u_{ln}^S + x_{ln}, \tag{7}$$

and the weighted model without seasonal components is given by

$$y_{ln} = \frac{1}{r_l}(\mu_{ln} + x_{ln}). \tag{8}$$

When $u_{ln}^S = 0$ in (7), this model is identical with the former one.

3 Identification

In this section we consider the weighted model (1) only. Identification methods for the models (7) and (8) are derived easily.

In fuzzy trend models estimation of latent processes plays an important role. We rewrite the model as follows:

$$z^R = G_1 u^C + G_2 u^I + x, \tag{9}$$

where

$$z^R = \begin{pmatrix} z_1 \\ \vdots \\ \vdots \\ z_L \end{pmatrix} = \begin{pmatrix} r_1 y_1 - B^S u_1^S \\ \vdots \\ \vdots \\ r_L y_L - B^S u_L^S \end{pmatrix} \tag{10}$$

$$y_l = (y_{l1}, ..., y_{lN})' \tag{11}$$

$$u^I = (u_1^{I\prime}, ..., u_L^{I\prime})' \tag{12}$$

$$u_l^S = (u_{l1}^S, ..., u_{lp}^S)' \tag{13}$$

$$G_1 = (B_1', ..., B_L')' \tag{14}$$

$$G_2 = \begin{pmatrix} B_1 & \cdots & \cdots & 0 \\ \vdots & B_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_L \end{pmatrix} \tag{15}$$

$$B_l = \begin{pmatrix} \nu_{l1}(1) & \nu_{l1}(1)(1-a_1) & \cdots & \nu_{lK}(1) & \nu_{lK}(1)(1-a_K) \\ \vdots & \vdots & & \vdots & \vdots \\ \nu_{l1}(N) & \nu_{l1}(N)(N-a_1) & \cdots & \nu_{lK}(N) & \nu_{lK}(N)(N-a_K) \end{pmatrix}. \tag{16}$$

The matrix B^S is the upper $N \times p$ submatrix of the matrix consisting of sufficient numbers of p-dimensional identity matrices. The vector x is defined similarly.

Now we consider an estimation procedure. First we estimate a seasonal component for each series by the least squares method for given d and r_l. From (1) – (16) we can rewrite

$$y_l = \frac{1}{r_l} B_l u_l + \frac{1}{r_l} B^S u_l^S + \frac{1}{r_l} x_l \tag{17}$$

$$= B_l \tilde{u}_l + \tilde{B}^S \tilde{u}_l^S + \tilde{x}_l \tag{18}$$

$$= [B_l \quad \tilde{B}^S] \begin{pmatrix} \tilde{u}_l \\ \tilde{u}_l^S \end{pmatrix} + \tilde{x}_l, \tag{19}$$

where

$$\tilde{u}_l^S = \frac{1}{r_l} (u_{l1}^S, ..., u_{l,p-1}^S)', \tag{20}$$

$$\tilde{B}^S = B^S \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \end{pmatrix}, \tag{21}$$

and so on. The size of the last matrix in (21) is $p \times (p-1)$. Note that u_l^S follows the constraint condition. From (19) the ordinary least squares method can be applied for estimation of $\tilde{u}_l^S$. We represent the estimated seasonal component by $\hat{u}_l^S$. The latent processes u_l's are re-estimated by using all series in the following way.

We define the new series by removing the seasonal component from the original series as follows:

$$\hat{z}^R = (\hat{z}_1', ..., \hat{z}_L')' \tag{22}$$

$$\hat{z}_l = r_l y_l - B^S \hat{u}_l^S. \tag{23}$$

From (9) the latent process can be estimated by the least squares method also. However, we cannot apply the ordinary least squares method to our model directly. Then we propose a two stage estimation procedure. The least squares estimators $\hat{u}^C$ and $\hat{u}^I$ are given by

$$\hat{u}^C \equiv (G_1'G_1)^{-1}G_1'\hat{z}^R \tag{24}$$

and

$$\hat{u}^I \equiv (G_2'G_2)^{-1}G_2'(\hat{z}^R - G_1\hat{u}^C) \tag{25}$$

respectively. We can show that the estimator $\hat{u}^I$ satisfies the equation:

$$G_1'G_2\hat{u}^I = 0. \tag{26}$$

Thus we adopt the condition (26) as a restriction. That is, the constraint of our model is $G_1'G_2u^I = 0$.

Next we propose a recursive procedure for estimating r_l.

Step 1. Set $r_l = 1/L$ $(l = 1, ..., L)$.
Step 2. Estimate $\hat{u}^C$ by (24).
Step 3. Calculate r_l as follows:
$r_l^{(0)} = \hat{z}_l'B_l\hat{u}^C/\hat{z}_l'\hat{z}_l$
$\lambda = (\sum_{l=1}^{L} r_l^{(0)} - 1)/(\sum_{l=1}^{L} 1/\hat{z}_l'\hat{z}_l)$
$r_l = r_l^{(0)} - \lambda/\hat{z}_l'\hat{z}_l$.
Step 4. Go to Step 2, if r_l's are not converged.
Step 5. Estimate $\hat{u}^I$ by (25).

The step 3 in the above procedure is lead by the method of Lagrange multiplier. Finally we have to determine d from data, since d is unknown generally. We apply the quasi Bayesian Information Criterion given by the equation:

$$BIC = NL\log(\hat{\sigma}^2) - N\sum_{l=1}^{L}\log r_l^2 + (2KL + L + p - 1)\log NL, \tag{27}$$

where

$$\hat{\sigma}^2 = (\hat{z}^R - G_1\hat{u}^C - G_2\hat{u}^I)'(\hat{z}_R - G_1\hat{u}^C - G_2\hat{u}^I)/NL. \tag{28}$$

The width parameter d is selected by minimizing BIC. The length of the latent process K is determined from d.

4 Numerical example

We apply the proposed models to artificial time series shown by Fig. 2 for demonstration. The length of series N is 116 and number of series L is three. (These series are real data of carbon dioxide concentration at three places in Japan. However, this multivariate time series is not real, since the original time points are not aligned.)

We fit the non-weighted model (7) and weighted model (1), since seasonality appears clearly. Both selected d's of two models are 68. The result by the non-weighted model (7) is shown by Fig. 3. The bold line is the common trend, solid line is the original series and dotted line is the sum of common and original trends and seasonal component. The figure shows that seasonality is estimated well. However, the common trend differs from each series, since there are large differences among series. This means that the common movement in each series is not estimated appropriately.

The result by the weighted model (1) is shown by Fig. 4. The solid line is the weighted series $r_l y_l$. Comparing Figs. 3 and 4 it is found that the weighted model provides more natural results. The estimated weight is (0.3430, 0.3292, 0.3278). Fig. 5 shows the estimated seasonal components and individual trends. Similarity or dissimilarity of three series can be considered from Fig. 5.

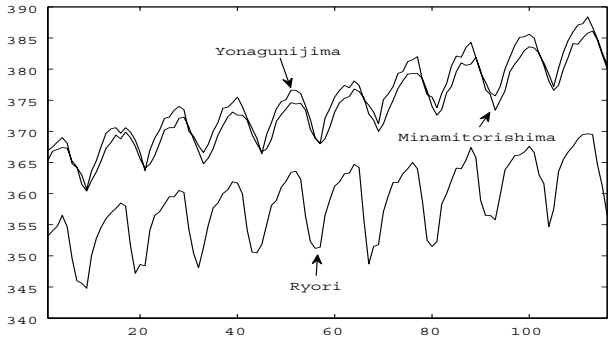


Fig. 2. Original series

5 Conclusion

In this paper we proposed the non-weighted and weighted fuzzy trend models for multivariate time series with or without seasonality.

The non-weighted model cannot estimate common trend well, when there are large differences among series. The weighted model resolves this problem. Moreover the proposed models can estimates seasonal components directly. The numerical example shows that the proposed identification method works well.

It is expected to apply our models to multivariate time series widely. However, simulation studies and practical examples are required for further evaluation of the identification procedure.

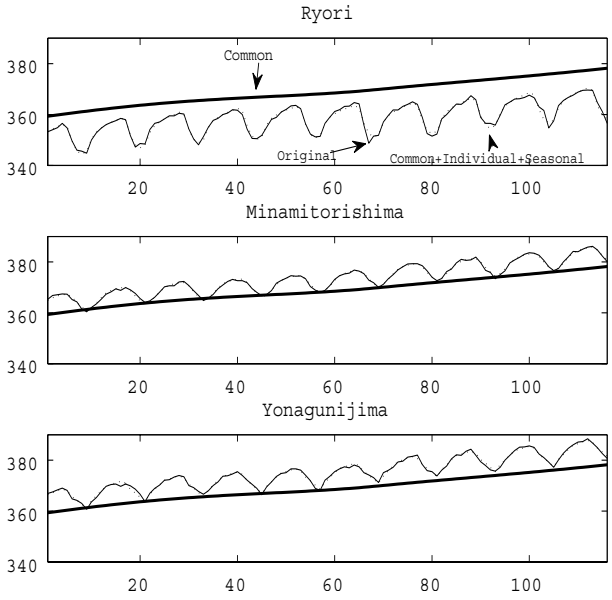


Fig. 3. Result by non-weighted model

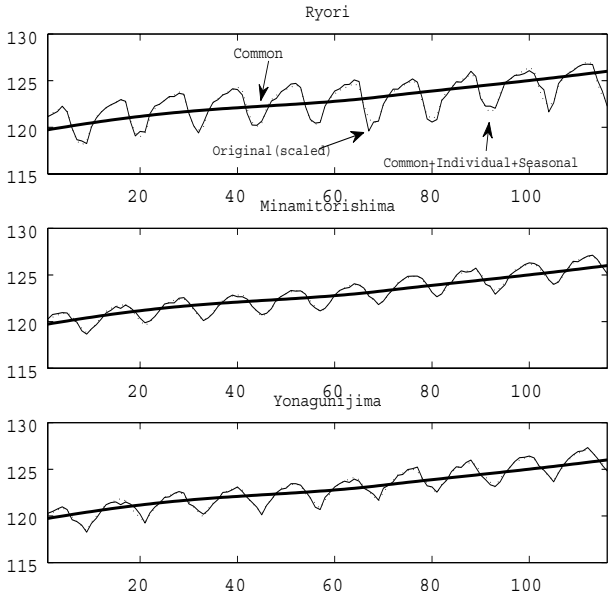


Fig. 4. Result by weighted model

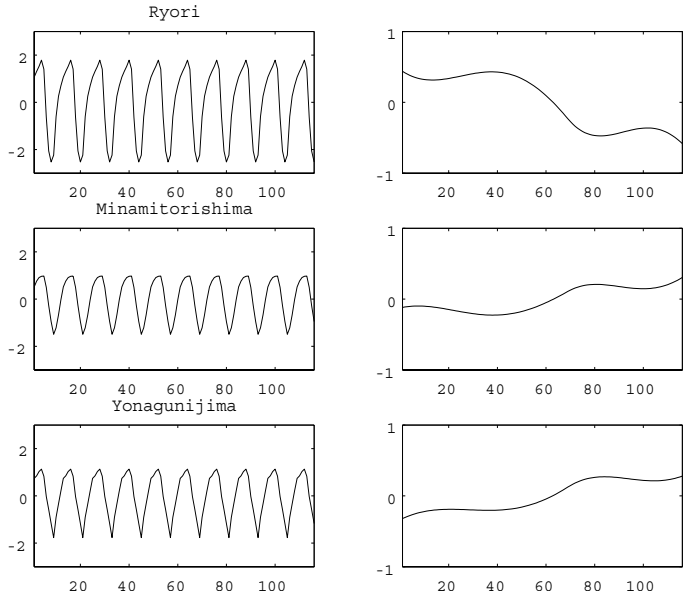


Fig. 5. Seasonal components and individual trends

References

1. Kuwabara, M. and Watanabe, N., "Correlation analysis of long-term financial time series by a fuzzy trend model", *The Porceedings of 11th IASTED International Conference on Artificial Intelligence and Soft Computing* 63–67 (2007).
2. Kuwabara, M. and Watanabe, N., "Financial time series analysis based on a fuzzy trend model" (in Japanese), *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 20, 2, 244–254 (2008).
3. Kuwabara, M. and Watanabe, N., "A multivariate fuzzy trend model for analysis of common and individual trends", *The Proceedings of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems*, 110–114 (2008).
4. Takagi, T. and Sugeno, M., "Fuzzy identification of systems and its application to modeling and control", *IEEE Trans. on Systems and Cybernetics*, 15(1), 116–132 (1985)

Ranking charity applications

Cong Xu and James M Freeman

Manchester Business School, The University of Manchester, Booth Street East, Manchester, M15 6PB, UK

Abstract. How can a charity ensure that only its most worthy causes are supported? One approach that has proven effective in the past is AHP (Analytic Hierarchy Process). Now a new method – based on a hybrid AHP / Evidential Reasoning (ER) adaptation - has become available, claiming distinct advantages over straight AHP. By contrasting the two procedures for a real-life dataset – we demonstrate AHP/ER's superiority in both theoretical and practical respects.

Keywords: Analytic Hierarchy Process, Dummy variables, Evidential Reasoning, Intelligent Decision Software, Utility

1 Introduction

Buxton and District Lions Club (http://www.buxtonlions.com/index.html) belongs to the International Association of Lions Clubs. As such, each year, the Club runs a variety of fund-raising events (see for example Fig.1), the income from which is then used to resource good causes - primarily within the local area.

Fig. 1. Participants in the BDLC's Ladies' Ruff Stuff Challenge

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

On average, BDLC raises £5,000 - £6,000 in donations annually. From experience this is never sufficient to meet the charitable demands upon the Club - hence the need for applications to be systematically evaluated (by the body's charity committee) to determine which, if any of them, should be earmarked for BDLC support.

In an effort to make the committee's screening process more objective and scientific, AHP [2] [3] was tested by Pang [1] on the thirteen grant applications received by BDLC in 1999. Seven criteria were considered in the application - four of them, quantitative and three, qualitative (binary) – as described in Table 1.

QUANTITATIVE ATTRIBUTES
"How long will the benefit last" (**Duration**)
"Numbers of people who benefit" (**Numbers**)
"How well resourced" (**Resource**)
"Impact of funding" (**Impact**)
QUALITATIVE ATTRIBUTES
"Any possibility of alternative funding" (**Alternative funding**)
"Direct or indirect applications" (**Direct**)
"Help with daily living" (**Living**)

Table 1. Decision criteria

Because of theoretical limitations with AHP at the time, only the quantitative criteria could be used in the resultant (EXCEL-based) analysis - final rankings from which are summarised in Table 2. However, providing strong vindication for the procedure, these were found to significantly correlate with actual funding decisions made by the Club.

Alternatives	Priority	Overall Ranking
Buxton Mountain Rescue Team	**0.190**	**1**
Buxton Opportunity club	**0.177**	**2**
Heartbeat	0.144	3
Burbage Football Club	**0.119**	**4**
Bereaved lady	**0.061**	**6**
Buxton Samaritans	0.095	5
Disabled man	0.039	8
Disabled riders	**0.040**	**7**
Holidays for disabled	0.037	9
PC for disadvantage school pupil	0.032	10
Wheelchair applicant	0.027	11
Chapel band	0.022	12
Nepal travel	0.017	13

Bold entries in the table correspond with applications that were finally funded by BDLC.

Table 2. Original AHP Summary

Building on this promising start, the data have now been re-analysed using a combined AHP/ER approach with the advantage that both quantitative and qualitative data (see Table 3) can be taken into account in the computations. Relevant results are detailed in the next section of the paper. Beforehand, background is provided on ER and the Intelligent Decision System (IDS) software[4] used to operationalise the analysis.

	Any possibility of alternative funding?	direct or indirect applications?	Help with daily living?
Buxton Mountain Rescue Team	1	1	0
Buxton Opportunity club	1	0	0
Heartbeat	1	1	0
Burbage Football Club	1	1	0
Bereaved lady	0	1	1
Buxton Samaritans	1	1	0
Disabled man	0	1	1
Disabled riders	1	1	1
Holidays for disabled	1	0	0
PC for disadvantage school pupil	1	1	0
Wheelchair applicant	1	1	1
Chapel band	1	1	0
Nepal travel	1	1	0

YYes is scored ‘1’ and No is scored ‘0’ here.
Table 3. Qualitative data details

2 Evidential reasoning and IDS

ER significantly extends the application of multiple criteria decision analysis (MCDA) methods by allowing formal belief structures to be incorporated into the modelling under conditions of uncertainty.
The approach itself is very flexible enabling uncertainty to be accommodated in many different guises e.g. as single numerical values, probability distributions, subjective judgments with degrees of belief … leading to greater realism and reliability in the overall assessment.
The re-analysis of the BDLC data was performed using the IDS software. As well providing a systematic interface for the model formulation, IDS offers a range of powerful facilities – not least its ability to incorporate different risk

outlooks into the analysis as well as an exhaustive sensitivity testing provision.

Typically, four distinct stages are involved in an IDS modelling application: for the BDLC data these can be illustrated as follows:

1. **"Define the alternatives"** (See Table 2)
2. **"Define the attributes"** (see Table 1)
3. **"Assign attributes weights"** For this stage of the project the **Eigenvector (AHP)** IDS option was selected over the **Geometric mean,** and **Mixed approach** alternatives - see the values obtained in Table 4 which compare very closely with those based on the traditional method set out in the Appendix**)**
4.

	Weight
Duration	0.424
Number	0.201
Resource	0.161
Impact	0.08
Living	0.066
Alternative Funding	0.034
Direct	0.033

Table 4. Attribute weights generated by IDS for the 7-Criteria Model

5. **"Convert grades"**.

As there are two levels of attributes for the BDLC data, grades from lower level attributes (*applications*) have to be converted and aggregated into the higher-level attributes (*criteria*). However, the process for handling qualitative data and quantitative data is different. IDS provides two different ways of aggregating them. One way is by **rule based transformation** and the other – the one used for the project - is **utility based transformation**[5]

In the latter case, IDS offers two sub-options for determining managers' utility types: **Visual Scoring**, and **Direct Assignment**. Visual scoring, the choice used here, involves computer graphical manipulation whereas Direct Assignment allows managers' utilities to be represented by specific utility values.

Following on, utility scores - assuming a risk neutral attitude to risk - were

obtained from IDS as follows:

Alternatives	IDS Utility score	Ranking
Buxton Mountain Rescue Team	0.907	1
Buxton Opportunity club	0.776	2
Heartbeat	0.683	3
Burbage Football Club	0.505	4
Bereaved lady	0.281	6
Buxton Samaritans	0.463	5
Disabled man	0.189	7
Disabled riders	0.156	8
Holidays for disabled	0.072	11
PC for disadvantage school pupil	0.078	10
Wheelchair applicant	0.087	9
Chapel band	0.038	12
Nepal travel	0.016	13

Table 5. IDS Rankings (7 criteria model) Risk neutral attitude

Corresponding graphical output is shown in Figure 2.

Equivalent graphs for risk averse and risk welcoming attitudes are shown in Figures 3 and 4 respectively:

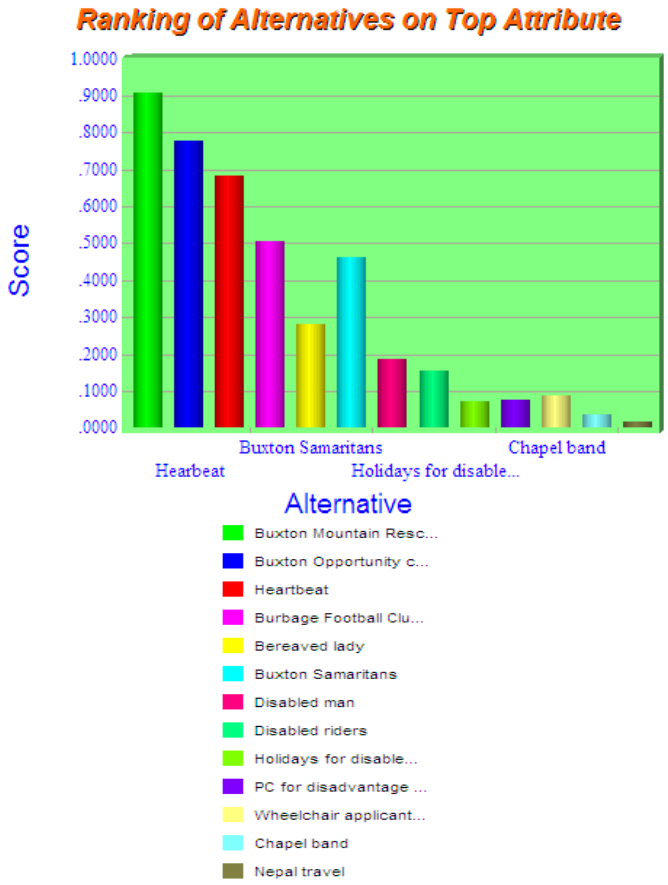


Fig. 2. Utility scores by alternative. Risk neutral attitude

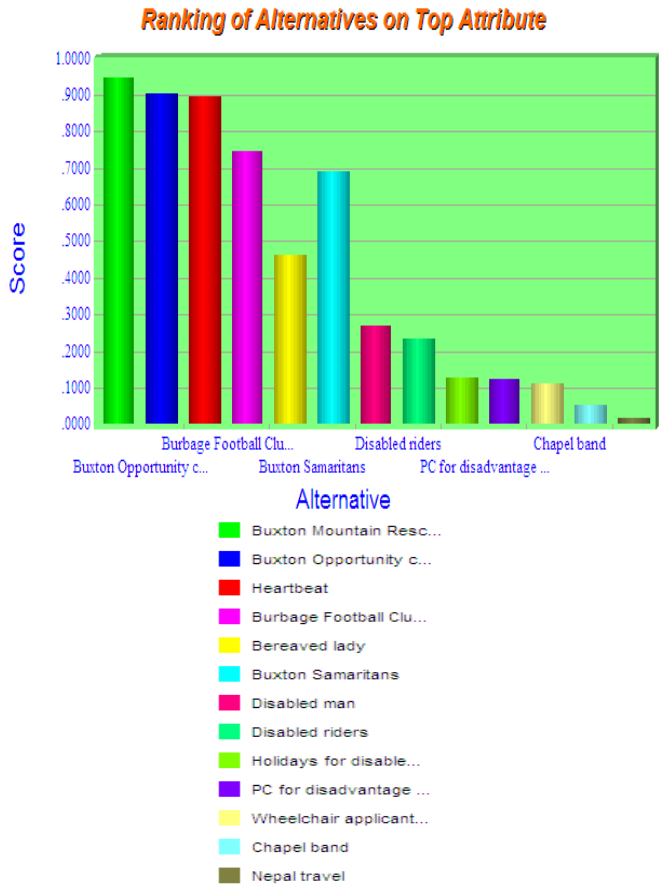


Fig. 3. Utility scores by alternative. Risk averse attitude

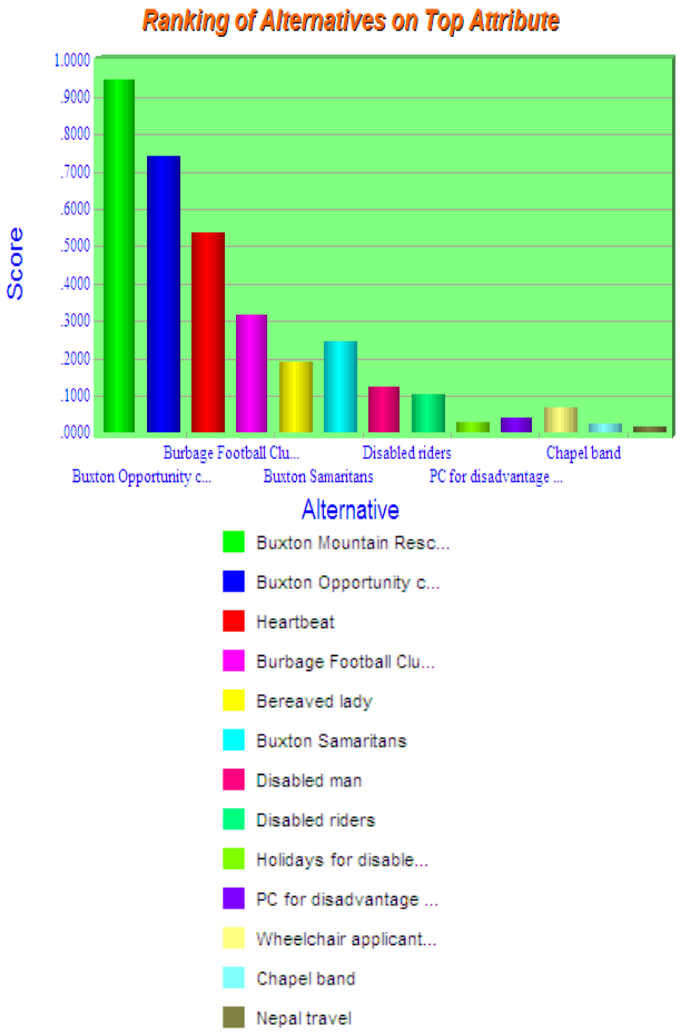
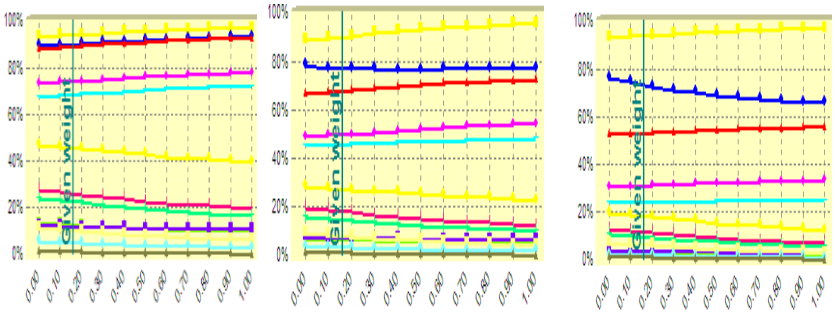


Fig. 4. Utility scores by alternative. Risk welcoming attitude

Of interest, all three rankings here can be shown to be significantly correlated. This overall consistency backed up by selected sensitivity results – see e.g. Figure 5 – suggest the ER/AHP rankings obtained for this particular dataset are remarkably robust.



(7-Criteria model Risk averse) **(7-Criteria model Risk neutral)** **(7-Criteria model Risk taking)**

Fig. 5. Sensitivity analysis re changes in weight value for "How well resourced"

4 Conclusions

Results from an AHP/ER analysis of a historical dataset on charity applications contrast markedly with those from a longstanding analysis based on set-piece AHP. More to the point, the new approach was found to outperform its predecessor in virtually every respect:

1. Whereas AHP was able to only handle quantitative criteria in the modeling, AHP/ER was able to deal with both quantitative and qualitative criteria.
2. IDS – the system used for the AHP/ER modeling here –substantially outclassed the open-ended EXCEL-based code generated for the AHP in scope and user-friendliness.
3. Utility stereotypes can be automatically taken into account in an IDS analysis enabling decision-makers' preferences to be directly incorporated into the results.
4. Similarly, IDS's sensitivity analysis capability is impressively comprehensive: not only does the system highlight the specific points where changes in data inputs cause overall rankings to change but it routinely maps out feasible regions associated with a given solution.

Irrespective of the utility type considered, the rankings for the first 8 of the BDLC alternatives remained the same: 1-Buxton Mountain Rescue Team, 2-Buxton Opportunity club, 3-Heartbeat, 4-Burbage Football Club, 5- Buxton Samaritans, 6- Bereaved lady, 7-Disabled man, 8-Disabled riders. Similarly, the rankings for the last 2 alternatives were also found to be unchanged: 12- Chapel band, 13- Nepal travel. Not surprisingly, this translated into significant agreement between all three of the seven criteria rankings obtained and indeed between them and the old AHP-based ranking.

References

1. J. Freeman and H.C. Pang. Separating the Haves from Have-nots – how the Analytic Hierarchy Process was used to Priorities applications for charitable funding. OR Insight 13, 4, 14{20, 2000.

2. T.L. Saaty. What is the analytic hierarchy process? Springer Berlin Heidelberg. 109{121, 1988.

3. T.L. Saaty. How to make a decision: the analytic hierarchy process. Interfaces, 24, 6, 19{43, 1994.

4. D. L Xu and J. B Yang. Intelligent Decision System for self-assessment. Journal of Multi-criteria Decision Analysis, 12, 1, 43{60, 2003.

5. J. B. Yang. Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties. European Journal of Operational Research. 131, 1, 31{61, 2001.

Appendix

The basis of the application comparisons that follow is the fundamental scale:

Verbal judgement or preference	Numerical rating
Extremely preferred	9
Very strongly to extremely	8
Very strongly preferred	7
Strongly to very strongly	6
Strongly preferred	5
Moderately to strongly	4
Moderately preferred	3
Equally to moderately	2
Equally preferred	1

Step 1: Calculate the sum of each column.

	Duration	Number	Resource	Impact	Living	Alternative funding	Direct
Duration	1	4	4	5	5	8	8
Number	0.25	1	2	4	3	5	5
Resource	0.25	0.5	1	4	3	4	5
Impact	0.2	0.25	0.25	1	1	4	4
Living	0.2	0.33	0.33	1	1	2	2
Alternative Funding	0.125	0.2	0.25	0.25	0.5	1	1
Direct	0.125	0.2	0.2	0.25	0.5	1	1
Total	**2.15**	**6.48**	**8.03**	**15.5**	**14**	**25**	**26**

Step 2: Normalization

	Duration	Number	Resource	Impact	Living	Alternative funding	Direct
Duration	0.465	0.617	0.498	0.323	0.357	0.320	0.308
Number	0.116	0.154	0.249	0.258	0.214	0.200	0.192
Resource	0.116	0.077	0.124	0.258	0.214	0.160	0.192
Impact	0.093	0.039	0.031	0.065	0.071	0.160	0.154
Living	0.093	0.051	0.041	0.065	0.071	0.080	0.077
Alternative Funding	0.058	0.031	0.031	0.016	0.036	0.040	0.038
Direct	0.058	0.031	0.025	0.016	0.036	0.040	0.038
Total	1	1	1	1	1	1	1

Step 3: Calculate Row average

	Duration	Number	Resource	Impact	Living	Alternative funding	Direct	**Row average**
Duration	0.465	0.617	0.498	0.323	0.357	0.320	0.308	**0.412**
Number	0.116	0.154	0.249	0.258	0.214	0.200	0.192	**0.198**
Resource	0.116	0.077	0.124	0.258	0.214	0.160	0.192	**0.163**
Impact	0.093	0.039	0.031	0.065	0.071	0.160	0.154	**0.087**
Living	0.093	0.051	0.041	0.065	0.071	0.080	0.077	**0.068**
Alternative Funding	0.058	0.031	0.031	0.016	0.036	0.040	0.038	**0.036**
Direct	0.058	0.031	0.025	0.016	0.036	0.040	0.038	**0.035**

The row averages in the last table correspond with those summarised in Table 4 using IDS. The consistency index (Saaty, 1980) for the latter can be shown to be zero signifying the weights from this analysis are perfectly consistent.

Management of technical maintenance of water pipeline networks on the basis of reliability characteristics

Boli Ya. Yarkulov
Samarkand State Institute of Building and Architecture,
Samarkand, Uzbekistan

Email: yarkulov@rambler.ru

Abstract

This article considers the new devising methodological approach on the basis of reliability characteristics (failure intensity), to be applied in the system of controlling the technical maintenance of water pipeline networks. Specifically:
- Define the types of reliability characteristics of the controlling objects, which characterizes on its actual condition.
- Define the main parameters of controlling object (spare parts, workers) technique maintenance on the basis of reliability characteristics.
This approach needs solving the problem in the following sequence:
- Define the types of failure of water pipeline networks;
- Determine algorithm for each type of failure flow of water pipeline networks;
- Define the quantity of resources that is necessary for the maintenance of workability of water pipeline networks on each type of failure;
- Estimate the reliability of water pipeline networks operation for various types of failure.

Carried out research allowed to distinguish four types of failure of elements and pipelines of water-supply networks dependent on time of their introduction, putting into exploitation, and outer influences:
- Running-in failure, which occurs in the initial period of exploitation;
- Failures in condition of system's normal operation;
- Failures in condition of system's physical depreciation (aging);
- Instantaneous (random) or gradual failures that occur as a result of outer strikes, i.e. earthquakes.

To determinate the amount of spare parts, the mathematical model in the form of optimization problem with one non-linear restriction and algorithm that builts

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

on the base of quickest and descent method, which its solution under the increasing intensity of failures of elements and pipelines of water-supply networks was formulated.

Keywords: Water pipeline networks, technique maintenance, reliability, failures, modeling, earthquakes, algorithm, workability, probability.

1. Introduction

It is known fact that engineering networks including water pipeline network in big cities and settlements is not built simultaneously because the water pipeline networks are constructed as result of cities and settlements widening. That is why there are buildings (constructions) of water pipeline networks of different remoteness year in the same city.

This situation complicates in a certain extent the process of controlling technical maintenance of water pipeline networks, as the intensity of failure flows in districts differs from one another dependent on their construction time remoteness. The traditional method causes insufficient or excess planning of amount of spare parts and employee group, which is necessary for the maintenance of workability of water pipeline networks.

The reasons are: existence of input information in a big capacity, missing information on some of networks' characteristics, insufficiency of data information on some parameters (on network state, influence of outer strikes), which require the special approach to solving this problem.

The dynamical character of development of water pipeline network system requires a constant improvement of system of controlling the technical maintenance, as lagging in development of subject of control from object of control causes huge social-economical losses.

On the other hand, the condition of water networks may change as a result of various earthquakes. Especially, during strong earthquakes, the condition of water networks will change instantly. As a result of earthquake, the water supply and distribution system will be partially or fully out of order. In hot climates, this situation leads not only to social and economic losses, but also increases the danger of expansion of epedemii. For example, as a result of earthquake in Armenia in 1989, there was the danger of expansion of the epidemic, even though it was the winter time. Because the water supply system was not restored by the deadline, and there was no beforehand developed system management.

2. Research of Changing the Failure Intensity Properties and Solving the Problem of Technical Maintenance of Water Pipeline Networks

Let the system of water pipeline networks consist of a_r, ($r = \overline{1,R}$)elements (r - signifies element's type). The system is concentrated in a sufficiently big territory, i.e. it is territorially dispersed system, object of control (water pipeline networks) is divided onto definite repair-exploitation

areas and technique control carried out for the maintenance of their workability. Controlling system consists of two levels, i.e. upper level represents city municipal government of water pipeline network, and lower level includes repair-exploitation area administration.

The failures of one or several elements do not cause the general failure in system. This depends on exactly in what part of system of water pipeline network the failure element (area) is situated, i.e. in the lines of main water pipeline (in initial network), in crosspieces and so on. We mark out four types of failure.

1. Running-in, early failure, occurring in the initial period of exploitation;
2. Failures in condition of system's normal operation;
3. Failures in condition of system's physical aging, depreciation (obsolescence);
4. Random, chance or gradual failures, that occur as a result of outer strikes, i.e. earthquakes, strikes of heavy transport means and so on.

Random, chance failure occurs at the same moment after one or series of strikes under strong earthquakes. Gradual failure occurs as a result of saving of deformation of several not strong strikes under not strong or insensible earthquakes.

It is known that failure characteristics (intensity, deepness) reflect the system state. To control the technical systems first their state should be determined. For that we introduce the parameter of system's elements condition $x_r(t)$, $r = \overline{1,R}$

$x_r(t)$ = { 0 on faultless r element in the moment t }

$x_r(t)$ = {1 on defective r element in the moment t }

If to mark through $\lambda_r(t_i)$ the intensity of failures of r- element in the moment t_i (i =1,4) then vector of systems condition is determined:

In the periods of running-in failures

$X_r(t_1)=\{X_1(t_1),\ \lambda_1(t_1);\ X_2(t_1),\lambda_2(t_1);...;\ X_R(t_1),\ \lambda_R(t_1)\ \}$,

In the periods of normal operating

$X_r(t_2)=\{X_1(t_2),\ \lambda_1(t_2);\ X_2(t_2),\ \lambda_2(t_2);...;\ X_R(t_2),\ \lambda_R(t_2)\}$,

In the periods of physical depreciation

$X_r(t_3)=\{X_1(t_3),\ \lambda_1(t_3);\ X_2(t_3),\ \lambda_2(t_3);...;\ X_R(t_3),\ \lambda_R(t_3)\}$,

In earthquakes (extraordinary, force major situations)

$X_r(t_4)=\{X_1(t_4),\ \lambda_1(t_4);\ X_2(t_4);\ \lambda_2(t_4);...;\ X_R(t_4),\ \lambda_R(t_4)\}$

Besides, arises necessity in defining the function of allotment of failure intensity on appropriate periods. Proceeding from features of changing of the failure intensity of water pipeline networks there considered four types of functions of failure intensity allotment $\lambda_r(t_i)$:

1. Running-in failure is corresponded by Weybulla-Gnedenko distribution, as on $\beta<1$ the intensity of failures decreases.

In the period of normal operation of failures intensity system is usually considered by constant value, i.e. λ=const and Weybulla-Gnedenko distribution corresponds to this, (and also exponential distribution), as if β=1, λ is constant value.

2. In case of occurring of many deterioration failures, i.e. obsolescence occurrence is essential, then it causes strong change in intensity of failures during time (Figure 1).

Besides, intensity of failures monotonously increase (period t_2,t_3) and intensity of failures change corresponded by Weybulla-Gnedenko distribution, where $\lambda(t)$ increases if β>1 (Beichelt& Franken, 1988).

3. In case of weak earthquakes, i.e. when there is outer strike influence, it is accepted that operating time of the equipment and pipelines of water-supply networks have distribution of increasing in average function intensity. Usually, in the models of impact load occur operating times from the class of increasing in average functions intensity. This means that equipments and pipelines of water-supply networks have undergone outer strikes that occur in casual moments of time and cause damage (accident) in the system. Damages accumulate in the equipments and pipelines until some of critical level won't be reached or exceeded, if this critical level is reached, then in equipments and pipelines occur failure (gradual).

Besides, in case of strong earthquakes instantaneous, chance failures usually occur, which happen in seismic active areas of the Globe. For example, such cases were observed in 1966 in Tashkent (Republic of Uzbekistan), in 1968 in Ashgabat(Turkmenistan) earthquakes and these earthquakes reached up to 9 points on the Richter scale.

It should be marked that analysis of failure intensity quality for all period of exploitation of water-supply pipeline networks shows that there exists the following determination.

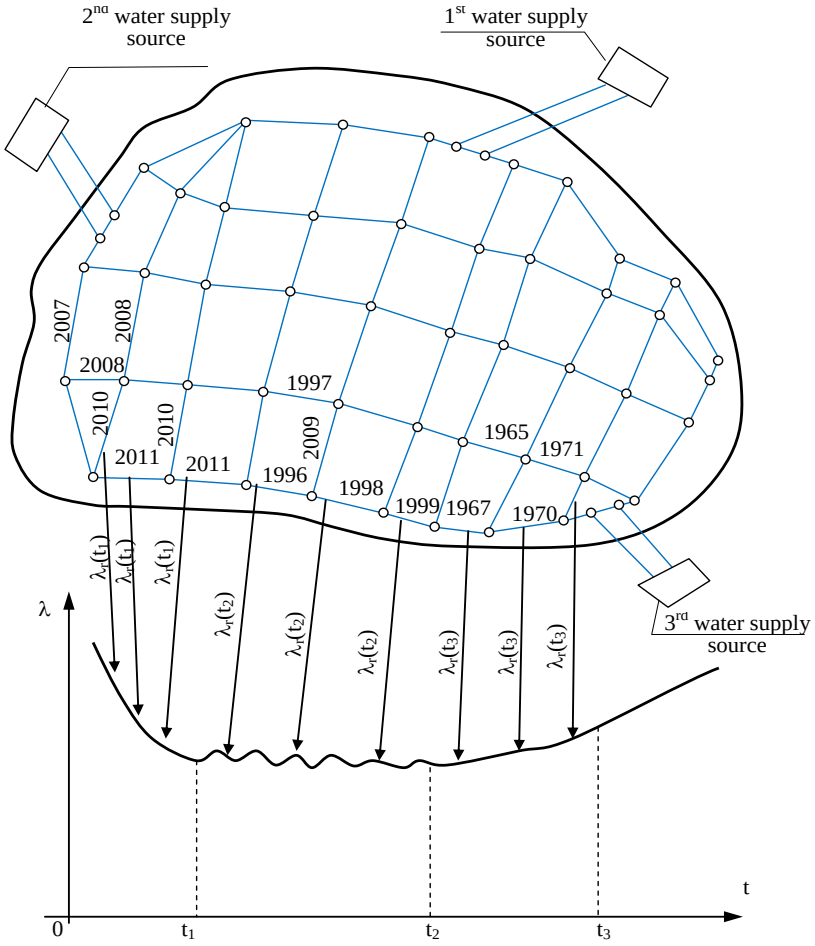
In running-in stage of exploitation of water-supply pipeline networks an interval of $(0,t_1)$ in real probability of faultless operation begins to grow after the beginning of exploitation.

Determination 1. Probability of faultless operation of technical system elements $\overline{F}_t(a_r)$, that are worked less t_1, monotonously increases on t, $0<t<t_1$.

According to the given determination, the intensity of failures in the interval of $(0,t_1)$ monotonously decreases, it is considered decreasing function of intensity.

Determination 2. Probability of faultless operation of works of elements of technical system $\overline{F}_t(a_r)$, that worked for a time t_2, monotonously decreases on t, $t_2<t<\infty$, t_2 – beginning of obsolescence stage of the elements of technical systems.

Fig. 1. Failures intensity change dependent on water networks exploitation time.



3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

According to this determination, beginning from t_2 time failure intensity monotonously arises, i.e. increasing function of intensity.
Now we will examine the process of the task solution of the quantity determination of the resources (spare parts), which are necessary for the workability efficiency support in the condition of the little-studied gradual failure, which happens as a result of weak earthquakes.
It is established that resources quantity determination task in weak earthquake conditions should be solved in three stages:

1. Determination of failure flow.
2. Determination of the probability of the faultless work.
3. Determination of resources quantity necessary for the liquidation of the earthquakes aftereffects.

Under the supervision of professor Abramov N.N. (Sabitov,1977), it is established that failure flow dependent on earthquakes intensity is expressed by the following dependency formula:

$$\lambda = a + b\,m - c\,m^2 \qquad (1)$$

where - a, b, c are the unknown parameters, m earthquakes intensity scale.
As it is seen from formula (1), failure flow depends from random factors, at the same time for determination of their meaning we used the Monte-Carlo method which is given in work (Yarkulov, 2004).
On the second stage with the known meaning of the failure flow intensity the water-supply networks' system faultless work probability is determined.

3.Algorithm for the Faultless Work Probability Determination of the Water Pipeline Networks at the External Influences.

At the imperceptible and weak earthquakes the water plumbing networks are subject to the weak influences and as a result critical failure crashes (damages) appear. It leads to damages accumulation and this process will continue till certain critical level will be achieved and exceeded, afterward the failure comes in the system.

When damages accumulate till the failure threshold, the system's faultless work probability is expressed not by the exponential law, but by the Weybulla's distribution law (close to Weybulla's distribution). It is explained by the fact that at the damages stack as a result of external influences and physical wear and tear, and the time among the failures will be decreased.

Therefore, at the gradual failure the system's faultless work probability is determined on the base of the full probability formula:

$$F(t) = \sum_{0}^{\infty} e^{-/_t} \frac{/\,k}{k!} P\left[x_1 + x_2 + \ldots + x_k < x\right] \qquad (2)$$

Where P $[x_1+x_2+\ldots+x_k<x]$ – is a probability that common damages that present k-hits sum, don't exceed acceptable limits equal to x.

$(\lambda e)^k e^{-\lambda t} / k!$ is a probability that by (0, t_1) time k hits sharp ensue.

As it is seen from formula (2) numerical values of probability P $[x_1+x_2+\ldots+x_k<x]$ are not always possible to get. At the same time it's possible to determine the limiting boundaries of failure origin probability on the basis of the past (registered) earthquakes.

It leads to the imitation of the given process, i.e. probability enactment, exceeding the possible, limiting boundaries, equal to x at k-hits.

Algorithm imitating the given process must provide the possibility of modeling interval change and change boundaries of damages origin possibility. Suppose that during a month 10-15 imperceptible earthquakes happen, it means that hits quantity equal to k=10÷15 (usually it happens in Central Asia country). In the capacity of modeling interval length t=30 days, i.e. a month period should be taken.

When considering the water supply system it's possible to divide the whole system into the R equivalent elements, i.e. every area is accepted as one equivalent element. Now we will consider the algorithm, enacting system's faultless work probability.

Algorithm's Description

1. Determination of equivalent elements quantity and elements types $a_r(r=1,R)$.
2. x=0.1 is specified.
3. Modeling interval is determined (0,t), t=30 days.
4. Hits quantity is specified, i.e. k (k=10÷15).
6. Random value is enacted ξ_i .
7. The necessary tests number is specified $n_g \geq \log(1-p)/\log(1-\varepsilon)$, where p= 0.95, ε = 0.01 .
8. $i \geq n_g$ is verified. If $i \geq n_g$, then transition to the following paragraph, otherwise - to the paragraph 6.

9. $\xi_i = \sum_i \xi_i / n_g, \quad (i = \overline{1, n_g},\ j = \overline{1,k})$ is calculated.
10. j≥ k conditions verification. If this condition is satisfied, then transition to the following paragraph, otherwise - to the paragraph 6.
11. $\xi_1 + \xi_2 + ... + \xi_k < x$ conditions verification. If this condition is satisfied, then $P[\xi_1 + \xi_2 + ... + \xi_k \le x] = 0$ and transition to the paragraph 12, otherwise $P[\xi_1 + \xi_2 + ... + \xi_k \le x] = p(\eta)$ and transition to the following paragraph (where $0 < p(\eta) \le 1$).

p(η) – probable value, which is determined by big numbers law depending on tests and factors quantity.

12. $F_r(t) = \sum_{k=0}^{L} e^{-\lambda_r a_r t} \frac{(\lambda_r a_r t)^k}{k!}$ $P[\xi_1 + \xi_2 + ... + \xi_k < x]$ is calculated

Where a_r – r type element quantity, (r=1,R), L – possible hits quantity.

13. r< R conditions verification. If this condition is satisfied, then transition to the paragraph 6, otherwise - to the following paragraph.
14. Printing $F_r(t)$ and algorithm end.

The developed algorithm allows taking into account water supply network condition change under the impact of the imperceptible and weak earthquakes gradual failures can occur.

On the third stage for spare parts quantity determination originally singularity of the given system should be taken into account. As it is known, water supply networks are connected logically serial, chained and in parallel.

We will consider logically connected elements of the water supply network.

4. **The Task of Spare Parts Quantity Determination**

If take the designations such as m_r – spare elements quantity, T_D – acceptable probability of the system's faultless work, (reliability norm α – water supply percentage), then the spare elements optimal quantity determination task in damages stacking conditions is presented in this way, i.e. find m_r, which may be solved at different criteria and limitations.

$$\min C_N = \sum_{r=1}^{R} C_r m_r \qquad (3)$$

We choose in the capacity of limitations system:

$$\prod_{r=1}^{R} F_r m_r \geq T_D \qquad (4)$$

Conditions (3)-(4) are the task of the multidimensional optimization with one limitation. Its solution practically comes to the one-dimensional tasks sequence solution on every step of optimization. Therefore, for (3)-(4) tasks solution it is possible to use the fastest descent method. On the basis of the last we will compose their solution algorithm.

The Algorithm Description

1. $m_r = 1$, $r = \overline{1,R}$ are specified.
2. $\theta = \prod_{r=1}^{R} F_r m_r$ is calculated.
3. $\prod_{r=1}^{R} F_r m_r \geq T_D$ Conditions of verification. If this condition is satisfied, then transition to the paragraph 9, otherwise - to the following paragraph.
4. $F_r(m_r) = \sum_{k=0}^{m_r} \frac{(\lambda_r a_r t)^k}{k!} e^{-\lambda_r a_r t}$ is calculated
5. $\Delta F_r(m_r) = F_r(m_r) - F_r(m_r+1)$, $(r = \overline{1,R})$ is calculated.
6. $C_r/\Delta F_r(m_r)$, $(r = \overline{1,R})$ is calculated.
7. $\min_{m_r} C_r / \Delta F_r (m_r)$, $(r = \overline{1,R})$ is found.
8. The variable (m_r) is increased on one unit to which $\min_{m_r} C_r/\Delta F_r(m_r)$ corresponds and transition to the paragraph 3.
9. Objective function value calculation $\min C_N = \sum_{r=1}^{R} C_r m_r.$
10. The end of the algorithm.

The analysis of the results obtained shows that the developed algorithm is suitable for the practical accounts of the spare elements quantity when under the influence of the external hits gradual failures arise. The main sources of the gradual refusals emergence in seismic active districts (zones) are the weak earthquakes.

The given task was solved for the consecutively serial connected elements of the system, and for the systems with parallel joined elements it is solved analogically, only with the limitations change (4).

5.Conclusions:

1. It is necessary to manage the technical maintenance (efficiency support) of the water-supply networks only by their condition, as it allows minimizing the exploitation expenses and water loss. To support the water plumbing networks efficiency at minimal expenses it is necessary to determine the failure flow intensity on the network areas. In the inhabited districts and cities, which are situated in seismic active zones, it is necessary to determine the failure flow in condition of imperceptible, weak and strong earthquakes.
2. With due regard for the data of the failure flow intensity it is necessary to determine the optimal resources quantity (spare parts, working teams) on the basis of the above mentioned models and algorithms which do not admit shortage or plenty of the resources.
3. To develop the complete system of models and algorithms, which are necessary for the water plumbing network efficiency support, and as a result, to create the information system of the network technique maintenance management.

6.Recommendation

This project is virtually internationl project because no country is safe from earthquakes in the world. It would be good if the concerned organizations (enterprises) from other countries will be able to support (finans support) for this project.

Create emergency reserve resources for individual cities in the region (e.g., state or province), which is enough to rebuild water systems within the same city. Create a separate section "control in extreme situations" in the system engineering management.

References

1. Beichelt F., Franken P.(1988) Reliability and Technical Maintenance. Mathematical approach. Translated from German, edited by I.A. Ushakov.“ Radio and communication”, Moscow.

2. Sabitov A.D.(1977) Water in Areas with Research Reliability of Systems Supply and Distribution of Increased Seismicity .Thesis of diss. Ph.D., Moscow.

3. Yarkulov B. Ya. (2004) Management of Technical Maintenance and Appraisal of Reliability of Water Line Networks. “Fan” Publishing House, Tashkent.

4. Bakhvalov N.S.(1973) Numerical methods.Moskow.

4. Barlow ,R.E.,Proshan, F.(1965) Mathematical theory of reliability.New York: J. Wiley & Sons

5. Barlow, R.E., Proshan, F.(1976) Theory of maintained systems: distribution of time to first system failure. Math.oper.Res.1,32-42

On the Accuracy of the Risk Estimators

Georgios C. Zachos*

Department of Mathematics, track in Statistics and Actuarial – Financial Mathematics, University of the Aegean, 83200 Karlovassi, Samos, Greece
(E-mail: zachosg@aegean.gr)

Abstract. The purpose of this paper is to evaluate the accuracy of the beta estimations that are suggested to be free of intevalling effect bias. To this end I examine the accuracy of the asymptotic estimators of betas by comparing them to OLS assessments as well as to beta estimations adjusted according to their tendency to regress towards one. Furthermore I employ above measurement for different intervals among data observations and I re-examine my findings by taking into account the Corhay effect. In addition I utilize models that take into account Heteroskedasticity in residuals and I perform the same comparisons
Keywords: Adjusted Risk Coefficients, Intervalling Effect Bias, Asymptotic Beta
JEL classification: C22, G12, G14

**Georgios C. Zachos is indebted to D. Konstantinides for his invaluable advices, E. Lamprou for providing datasets and D. Kokonakis and E. Papanagiotou for their valuable help.*

1 Introduction

Inarguably measuring non-diversifiable risk in an investment like a fund or a security is the most common practice of financial analysts. Thanks to pioneers of financial science like Sharpe [29] systematic risk of a security could be quantified simply by regressing returns of a security to the market returns and consequently the slope, or the beta of this procedure will eventually produce a figure that will expose how volatile, thus risky, is the security compare to the market. Apart from its practical use, measuring systematic risk is prolific in academic terms. There is a vast academic effort in trying to find the best possible model that is able to measure systematic risk as accurate as possible. As an example many variations of the model can be stated. In terms of single dimension it should be mentioned the Sharpe and Lintner's [21] Capital Asset Pricing Model (CAPM) and the Market Model (MM) while in multidimensional versions the Three Factor Model of Fama and French [12] and Carhart's [4] four factor model are among the most popular.
Purpose of this Paper is to evaluate the accuracy of the estimations of betas that are suggested to be free of intervalling effect bias. A number of studies have evaluated the validity of this method, for instance McInish and Wood [23] so the matter appears to interest both academics and practitioners. Furthermore there is no other study attempting to evaluate the accuracy of intervalling effect free-betas by

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

straightforwardly comparing them to betas corrected according to their tendency to regress towards the grand beta mean.

To that end I need to employ two methodologies. The first methodology is suggested from Cohen *et al.* [6]. Taking into consideration price adjustment delays that are present into microstructure of markets and leads to interavalling-effect bias, they suggested a procedure to reduce this phenomenon. It is a two-stage procedure: In the first stage Market Model and Ordinary Least Squares (OLS) time series regression betas are calculated for a number of intervals, typically from one day to one month. Second stage is to calculate an asymptotic estimator of the systematic risk coefficient by cross-sectional regress OLS betas towards a monotonically decreased equation. Consequently the asymptotic estimator reduces the intervalling-effect bias. The methodology seems to be significant academically as researchers appear to include it in their research agenda, for instance Fung *et al.* [16] in earlier years. More recent studies that use this methodology are Milionis and Patsouri [25], Diakogiannis and Makri [9] and Milonas and Rompotis [26].
Furthermore Blume's technique [2] attempts to calculate future risk coefficients by taking into consideration the fact that betas aren't constant over time, yet they have the tendency to regress towards the market mean. By dividing a time dataset to sub-periods, for instance three or five year periods, he calculates betas for each time period. Afterwards he regress the betas of the earlier period towards the betas of the later period. With the use of the regression equation and first period's beta a researcher is able to recalculate betas of the second period, thus the adjusted assessment according to his phraseology. Final step of this methodology is an examination of the accuracy of the assessments of the latter period that are based to historical data compare to the risk factors that take into account the tendency of the betas to regress toward the grand mean. Apart for Blume's methodology, there are other techniques that attempt to correct the beta estimations according to aforementioned principal, with most important the Bayesian technique. Initially inducted by Vasicek [31], Bayesian method is widely used by prestigious companies like Merrill Lynch and as Elton *et al.* [10] suggests it could be more favorable compare Blume's technique in certain occasions. Concerning beta adjustment techniques we should mention Mantripragada [22] who applied a plethora of those methodologies to Canadian stock datasets and more recently the Sarker [27] who used both Vasicek and Blume's technique to data from Dhaka Stock Exchange and concluded that there are no significant differences in their results.
Taking into account all the above I:

1) Investigate the accuracy of the asymptotic estimators of betas by comparing them to OLS naïve assessments and to beta estimations adjusted according to Blume [2] method.
2) I employ above comparison for different intervals among data observations (daily and monthly intervals).

3) I also re-examine my results by taking into account the Corhay [8] effect in results[1].
4) Finally I extract asymptotic estimators of betas by using models that take into account Heteroskedasticity in residuals (GARCH and Exponential-GARCH) and I proceed with the same accuracy inspection.
In terms of data daily closing time security prices of the entire universe of Athens Stock Exchange are selected for calculating returns, while market returns are presented from Athens Stock General Index. Also data are collected for ten consecutive years. Moreover since Blume [2] and Cohen *et al*. [5] use the Market Model, same method should apply here hence risk free rate is not necessary. All calculations occur in Matlab statistical software package. The rest of the paper is organized as follows. Section 2 includes literature review. Section 3 discuses data, Section 4 analyzes methodology approach while results are included in section 5. Section 6 concludes.

2 Literature Review

2.1 The Model

In line with methodology used by Blume [2] and Cohen *et al*. [6], for my analysis I will use the Market Model. Formula is

$$E(ri)=\alpha_i+\beta_i E(r_m) \qquad (1)$$

Where $E(ri)$ is the expected return on the capital asset, α_i is the residual return of asset I. β_i (the beta coefficient) represents sensitivity of the asset returns compare to market returns or $\beta_i=COV(Ri,Rm)/Var(Rm)$. Total risk of the portfolio can be viewed as beta. Rephrasing, the model uses time series regression to calculate beta, so $E(r_m)$ stands for expected return of the market.
Compare to the single index CAPM model it differs in two ways: First there is absence of risk free rate as in realistic terms it makes no countable difference and second, Elton *et al*. [11] points out that MM lacks the assumption that all covariances among securities occur because of a common covariance with the market. Also multifactor models, like Fama and French three factors model [12], [14] or Carhart's [4] four factor model are not selected. As Elton et al [10] points out, maybe historical prices are better interpreted by a multi-dimensional model but in terms of predictive ability a single-index model should be preferred. Moreover

[1] As Corhay [8] points out if the interval between data observations is more than a day we might get a different result every time we choose a different starting day. To this end I perform tests by selecting every possible starting day within the interval and the final beta estimation is the average of those results.

multi-dimensional model might include more noise than information in their factors in certain occasions.

2.2 Adjusted Risk Coefficient

Blume [2] argued with common practice of investors to act as if beta coefficients are constant over time. After examining correlations of unadjusted risk coefficients he suggested a methodology where he extracted future adjusted betas according to past prices. Most important on this methodology is the assumption that betas tend to regress toward market mean and consequently his methodology measures this phenomenon. Furthermore he examined accuracy of adjusted assessment through mean square errors and found that they are significantly more accurate compare to assessments based simply on historical prices. In terms of methodology he separated his data into sub-periods and calculated via Ordinary Least Squares (OLS) time series regression the betas of those periods. In addition he performed cross section regression where betas of one period are the explanatory variables and betas of the next period are the dependent. Finally with the use of the regression equation and the data of the first period, second period's beta can be extracted and according to Blume's [2] findings, they are more accurate compare to historical prices results.

Definitely promising, yet Blume's technique has been further investigated from academic community and proved to be less than flawless. For instance Klemkosky *et al.* [18] indicated bias occurring and recommended three procedures to reduce those effects. On the contrary Blume [3] addressed the issue of order bias which leads to non-stationarity in estimated beta coefficients. He argued that it is not of major importance by suggesting that extreme betas of investments tend to become less extreme both for new or existing investments. Summing up two types that effect Blume's technique should be mentioned. One is the fact that it fails to forecast a trend in beta and assumes that any trends occurring are random. Second it fails to spot other factors except correlation with the market that effect beta changes, for instance industry effects.

In line with Blume's concept, other techniques have been suggested with most important the Bayesian technique, initially suggested by Vasicek [31]. It is widely used by prestigious firms like Merrill Lynch and as Elton *et al.* [11] explains, it assumes that beta of investments tend to be closer to average beta than historical prices suggest so adjusts each historical beta towards the average. Suffering from its own bias, for instance when a beta is greater than one, it is corrected by a bigger percentage compare to a less than the market beta, yet it is suggested to be a slightly better technique compare to Blume's by many authors like Elton *et al.* [10] and Klemkosky and Martin [18].

2.3 Asymptotic Beta

The second methodology employed in this paper is the asymptotic estimation of betas proposed by Cohen *et al.* [6]. As majority of empirical researchers in

financial economics presume no friction, Hawanini *et al.* [17], based on microstructure theory, argued by confirming the importance of friction in trading process and by indicating the complex and persistent impact friction has on generating returns practice. Also Hawanini *et al.* [17] points out in the same study that when differencing interval is increasing, price adjustment delay impact will reduce. Concluding, in their paper it is suggested that if differencing intervals are greater than aforementioned delays, then the latter will lessen. Giving a simple example of what intervalling-effect bias is, it can be stated that if an investor estimates beta for a security with daily data with OLS regression procedure, he will get a figure that will differ compare to a procedure with weekly data, and again he will get different beta for monthly intervals among observations.

Also Cohen *et al.* [6], in line with previous findings points out that when working with short differencing interval data the variation between true and observed beta is considerable. True are the beta that should be obtained in case of a frictionless environment and observed beta are the beta that can be calculated and actually observed by investors. Also Cohen *et al.* [7] denoted that price adjustment delays are associated with market value of the shares included in sample investigated. In the same work it is suggested that if intervals are increasing gradually then bias will reduce and eventually diminish. In formula terms Fung *et al.* [16] suggested the following:

$$\beta_i^* = \lim_{l \to \infty} \beta_i^0(l) \qquad (2)$$

Where β_i^* represents an inconsistent estimator of β_i, while $\beta_i^0(l)$ is the beta estimator for interval l.

The most important in Cohen *et al.* [6] work is their suggestion of a methodology where the true beta can be estimated, thus the asymptotic estimator of beta. It is a two stages procedure. First step is to calculate systematic risk coefficient, thus the slope or the beta in the Market Model with regression method for intervals from 1,2,…, l days. Regression formula is

$$r_{jLT} = a_{jL}{}^1 + b_{jL}{}^1 r_{MLT} + e_{jLT}{}^1 \qquad (3)$$

Prescript 1 denotes the first stage. Second stage is occurring in order to estimate the intervalling effect on risk coefficient. For that procedure all the estimated betas for all intervals and for each security are cross-sectional regressed with the interval effect which reduces as intervals are increasing and is expressed from the monotonically decreased equation $f_j(l) = L^{-k}$ where it is assumed that:

$$\lim_{l \to \infty} f_j(l) = 0 \, \forall j \qquad (4)$$

Formula of the second stage is:

$$b_{jL}{}^{1}=a_j{}^{2}+b_j{}^{2}L^{-k}+e_{jL}{}^{2} \qquad (5)$$

Where 2 denotes second stage of the procedure and $a_j{}^{2}$ stands for the asymptotic estimator of beta. Clearly as L increases without bound the intervalling effect reduces. Consequently L^{-k} diminishes and thus true beta will be approaching the figures of observed beta. Concerning k we follow the same methodology as in Cohen *et al.* [6] and Fung *et al.* [16] [2]. Finally Cohen *et al.* [7] highlights importance of $b_j{}^{2}$ as quantitative proxy to measure intervalling effect. If it is statistically significant, a negative price of $b_j{}^{2}$ will suggest that as differencing interval lengthens beta coefficient will rise and vice versa. If it is statistically insignificant there is no intervalling effect at all.

3 Data

Sample data were collected from Bloomberg's terminal database. Daily closing time observations are selected due to homogeneity reasons. In case an observation misses due to unforeseen constraints average of the previous and next day is calculated and serves as the missing observation. Moreover all securities are valued in euro currency. In addition the sample set is consisted from the whole universe of securities traded in Athens Stock Exchange for ten consecutive years, from 02/01/2002 until 30/12/2011. In case a security was excluded from trading during the time sample was chosen, it will be excluded from the sample as well. For Market returns ASE General Index (capitalization weighted) is used in calculations. ASE appears to be an interesting selection for a number of reasons. First it does not presents features like big capitalization of more mature markets like New York Stock Exchange or Frankfurt Stock Exchange which habitually provide data for research. Moreover it was excluded from emerging markets (and harmonized with standards of mature markets) in 2001, yet it again downgraded in 2013 by index provider MSCI so it is expected to produce very interesting results which can be compared to results provided from a mature market. In addition it is expected to be more volatile compare to a mature market as it lacks in capitalization, which is another thing that makes choice of ASE appealing. Finally Athens Stock Exchange was selected because there is evidence that friction in trading processes appear to be present. As it is suggested by Alexakis and Alexakis [1] there is evidence that Hellenic Market follows patterns of global markets with delay.

From initial observations continuously compounded rate of return on each security is calculated according to formula presented below[3].

---

[2] Cohen *et al.* [6] k estimation is approximately 0.8 and as exposed in table 1 we get a similar result only on second period (07-11) and only when we take into account Corhay's effect (OLS 0.61, GARCH 0.75 and EGARCH 0.69 respectively)

[3] Continuously compounded rate of return was also used for market returns

$$r_{pt} = \ln\left(\frac{Ppt}{Ppt-1}\right) \qquad (6)$$

Ppt stands for price observation of security P, at day t and $Ppt - 1$ represents observation of the same security one day before. ln is the natural logarithm. Working with differences and therefore interpreting coefficients as elasticities, is a common practice when analyzing such data. As Koop [19] highlights a data set of financial data will behave well in terms of stationarity[4].

3.1 Visual Inspection of Market

Appendix presents graphs for ASE General Index and the continuously compounded rate of return of the same index. ASE index appears to exhibit from 2003 onwards an excessively positive uptrend which appears to finish at the end of 2008. Reason for that inarguably is the global crisis emerging from last semester of 2007 onwards. Mortgage subprime crisis in U.S. market seem to be initial reason as Krugman [20] explains, nevertheless crisis spread worldwide afterwards. As Friedman and Schwartz [15] denoted an economic collapse could be a cumulative type process. Concerning returns of ASE index, while in uptrend almost in no occasion a 5% change is observed, on the other hand when market index is falling graphs become volatile exhibiting percentage changes even more than 10% and up to 15%. Such conclusion is to be taken into consideration, as volatility in market indicates risk and uncertainty. Concluding visual inspection, two downturns and one peak at the end of 2008 are observed in period selected for sample.

4 Methodology Approach

Purpose is to examine the accuracy of the asymptotic estimated betas compare to accuracy of betas obtained both from adjusted and naïve assessments. Initially from the whole universe of Athens Stock Exchange shares in the sample, the ones that are not traded throughout the whole ten year period are excluded. Afterwards in accordance with Blume [2] the ten year period of data are separated into two five year sub periods. Continuously compounded rates of returns of shares are calculated. Next Market Model OLS style regressions will be performed for each share return and for each five year period. The same pattern applies when I work OLS in conjunction with GARCH and EGARCH methodology. Athens Stock Exchange General Index's returns are also calculated and stand as Market variable in equation. In accordance with Cohen *et al.* [6] regressions take place for intervals from 1 to 30 days. Moreover asymptotic estimations of betas are obtained with the use of second stage Cohen *et al.* [6] recommend for same intervals. It is important to consider Corhay's [8] suggestion that estimated betas differ in case differencing interval starts on a different day so when differencing interval is bigger than a day,

[4] Further information can be provided by the author on request

then betas will be calculated by taking as a starting date every observation included in the interval. Then the average of betas obtained is the final estimated beta. Finally we examine a set of results that takes into account Corhay effect and a set that does not do so. All regressions take place in Matlab software package.
Furthermore in accordance with Blume's technique betas of the 07-11 period are cross-sectional regressed towards the 02-06 period[5]. Outcome of this procedure is a regression formula with the later period's betas as the dependent variable and earlier betas as explanatory[6]. Having in mind Blume's technique once again we are able to retrospectively make estimations of later period's betas with use of this formula and first period's beta.
Result of aforementioned procedures are assessments of betas for the period 2007-2011 that are based on historical data, results that are adjusted with Blume's correction and asymptotic estimated beta results. Final step is a valuation of accuracy of those results with the use of mean square errors as Blume suggested in his methodology. Consequently we are able to examine if the asymptotic estimation of betas provide more accurate beta estimations compare to naïve ones and also compare to beta adjustments proposed by Blume [2].

5 Results

Results are presented in table 3 and table 4 in appendix. As I am working with mean square errors the smallest price denotes the more accurate estimation. Furthermore the key finding (bolded in tables) occurs when I use as benchmark (1st period 02-06) OLS and daily results and as comparison periods OLS and daily Blume adjustment (0.2862), daily asymptotic estimator (0.2708) and daily naïve estimation (0.3882). Evidence suggests that asymptotic estimators are more accurate compare to both naïve and adjusted assessments. Furthermore when I take into account the Corhay effect in asymptotic estimator (0.2537) the results are the same and further more evidence suggest that asymptotic estimations of betas are even more accurate[7].
In addition every other test I perform evidence suggests that asymptotic estimators are more accurate compare to naïve assessments yet less accurate compare to assessments compare to blume's technique. More specifically:
1) When I use as benchmark (02-06) OLS and monthly results (No Corhay) I have the following MSE figures (Blume 0.2573, asymptotic 0.3679, naïve 0.3791).
2) When I use as benchmark (02-06) OLS and monthly results (Corhay) I have the following MSE figures (Blume 0.375, asymptotic OLS 0.4672, naïve 0.4869).

[5] We don't follow the exact pattern of Blume [2] only in terms we don't put beta prices in ascending order and also we don't categorize stocks into portfolios according to their beta prices
[6] All regression formulas are presented in table in appendix
[7] Concerning naïve assessments and adjusted ones results are the same when Corhay effect is not taken into account since these are results taken from daily data thus the interval among observations is one

3) When I use as benchmark (02-06) GARCH[8] and daily results (No Corhay) I have the following MSE figures (Blume 0.2641, asymptotic GARCH 0.2736, naïve 0.373).
4) When I use as benchmark (02-06) GARCH and daily results (Corhay) I have the following MSE figures (Blume 0.2603, asymptotic GARCH 0.2706, naïve 0.3728).
5) When I use as benchmark (02-06) GARCH and monthly results (No Corhay) I have the following MSE figures (Blume 0.3197, asymptotic GARCH 0.4202, naïve 0.4796).
6) When I use as benchmark (02-06) GARCH and monthly results (Corhay) I have the following MSE figures (Blume 0.3678, asymptotic GARCH 0.4457, naïve 0.5003).
7) When I use again as benchmark (02-06) GARCH and daily results (No Corhay) and I examine asymptotic EGARCH, I have the following MSE figures (Blume 0.2641, asymptotic EGARCH 0.3007, naïve 0.373).
8) When I use again as benchmark (02-06) GARCH and daily results (Corhay) and we examine asymptotic EGARCH, I have the following MSE figures (Blume 0.2603, asymptotic EGARCH 0.2956, naïve 0.3728).
9) When I use again as benchmark (02-06) GARCH and monthly results (No Corhay) and I examine asymptotic EGARCH, I have the following MSE figures (Blume 0.3197, asymptotic EGARCH 0.4534, naïve 0.4796).
10) When I use again as benchmark (02-06) GARCH and monthly results (Corhay) and I examine asymptotic EGARCH, I have the following MSE figures (Blume 0.3678, asymptotic EGARCH 0.4936, naïve 0.5003).
Some caveats that should be discussed seem to be present because of the special features of the ASE index composition. Specifically the capitalization of the ASE is included in only 60 shares (almost 100% of Cap), yet I have been working with 224 stocks. In other words almost ¾ of the stocks seem to contribute nothing to the index weight and as a consequence the index does not seem to be correlated with the majority of the sample. When we regress relatively uncorrelated time series we are not expected to get good R*2 values and the same applies here[9]. In an intuitive sense the capitalization's issue seems to have an effect on Blume's regressions formulas as the larger slope factor we notice gets a value of approximately 0.30 as observed in table 2 while Blume observes values that reach up to 0.75.

6 Conclusions

Inarguably the main finding in the paper is the fact that Asymptotic estimators of beta seem to provide accurate estimations of risk. In all cases examined (OLS, GARCH and E-GARCH) aforementioned technique provided more accurate beta compare to naïve assessments. When I test for daily and monthly interval between

[8] In terms of GARCH and EGARCH results I select according to AIC

[9] For instance the R*2 mean for OLS regressions (07-11 period, daily and not Corhay effect) is only 0.15

observations evidence favours the previous result. Conclusions drawn under are the same if I also take into account Corhay [8] effect. Furthermore there are two occasions where asymptotic estimations of beta give more accurate risk factors even compare to the ones Blume's [2] adjustment provides: when I am working with OLS daily data and I don't consider Corhay effect and when I am working with OLS daily data but take into account Corhay effect.

Promising as they might be, yet those findings signify the need for more research in order to provide robust evidence. More specifically I suggest:

a) A study with the same data set and methodology but only with stocks that have a considerable weight in the ASE general index. Evidence suggests that some drawbacks will be avoided if this pattern is followed.

b) In line with previous suggestion, a selection of stocks should occur according to how good the regression fits, for instance according to R*2 of regressions.

c) Apart from daily and monthly also other intervals should be examined.

d) Moreover the same methodology should apply to another market with other feature compare to the ones ASE markets exposes, preferably a mature market. The comparison between the results of an emerging and mature market will contribute to solid conclusions.

e) Apart from Blume's [2] also Bayesian techniques could be applied to the analysis. As they appear to perform slightly better (Elton *et al.* [10]) and they are used extensively by practitioners in order to correct estimations of risk they should be used as an alternative method of adjusting betas and therefore as an extra comparison measurement.

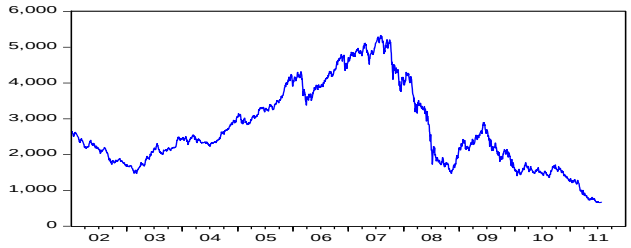
References

1. Alexakis, P., Alexakis, A. C., *Issues Concerning Efficiency in Interconnections of Big Financial Markets and Greek Financial Market*, Studies Concerning Greek Financial System, Economic university of Athens Publications, Athens (Text in Greek), 2010.
2. Blume, E. M., On the Assessment of Risk, *The Journal of Finance*, 26, 1, 1{10, 1971.
3. Blume, E. M., Betas and Their Regression Tendencies, *The Journal of Finance*, 30, 3, 785{795, 1975.
4. Carhart, M. M., On Persistence in Mutual Fund Performance, *The Journal of Finance*, 52, 1, 57{81, 1997.
5. Cohen, K., Hawanini, G, Maier, S., Schwartz, R. and Whitecomb, D., Implications of Microstructure Theory for Empirical Research on Stock Price Behavior, *The Journal of Finance*, 35, 2, 249{257, 1980.
6. Cohen, K., Hawanini, G, Maier, S., Schwartz, R. and Whitecomb, D., Estimating and Adjusting for the Intervalling Effect bias in Beta, *Management Science*, 29,1,135{148, 1983a.
7. Cohen, K., Hawanini, G, Maier, S., Schwartz, R. and Whitecomb, D., Friction in the Trading Process and the Estimation of Systematic Risk, *Journal of Financial Economics*, 12, 12, 263{278, 1983b.
8. Corhay, A. (1992), The intervalling effect bias in beta: A note, *Journal of Banking and Finance*, 16, 1, 61{73, 1992.
9. Diacogiannis, G., and Makri, P., Estimating Betas in Thinner Markets: The Case of The Athens Stock Exchange, *International Research Journal of Finance and Economis*, 1, 13, 108{123, 2008.

10. Elton, J. E., Gruber, J. M., Urich, J. T., Are Betas Best?, *The Journal of Finance*, 33, 5,1375{1384, 1978.
11. Elton, J. E., Gruber, J. M., Brown, J. S., Goetzmann, N. W., *Modern Portfolio Theory and Investment Analysis*, John Wiley & Sons (Asia) Pte Ltd, Eighth Edition, 2011.
12. Fama E., French K., Common Risk Factors in the Returns on Stocks and Bonds, *The Journal of Financial Economics*, 33, 1,3{56, 1993.
13. Fama, E. and French, K., Size and Book-to-Market Factors in Earnings and Returns, *Journal of Finance*, 50, 1, 131{156, 1995.
14. Fama E., French K., Multifactor Explanations of Asset Pricing Anomalies, *The Journal of Finance*, 51, 1, 55{84, 1996.
15. Friedman, M. and Schwartz J. A. , *A Monetary History of the United States*, Princeton University Press, 1963.
16. Fung, W., Schwartz, R. and Whitecomb, D., Adjusting for the Intervalling Effect bias in Beta: A Test using Paris Bourse Data, *Journal of Banking and Finance*, 9, 3,443-460, 1985.
17. Hawawini, G. A., Intertemporal Cross Dependence in Securities' Daily Returns and the Short-Run Intervalling Effect on Systematic Risk, *Financial Quantitave Anal.*, 15, 1, 139-150, 1980.
18. Klemkosky, C. R., Martin, D. J., The Adjustment of Beta Forecasts, *The Journal of Finance*, 30, 4, 1123{1128, 1975.
19. Koop, G., *Analysis of Economic Data*, 2nd edition, John Wiley & Sons Ltd, 2002.
20. Krugman, P., A (Subprime) Catastrophe Foretold *http://www.spiegel.de/international/0,1518,513748,00.html*, *Spiegel on Line*, 2007.
21. Lintner, J., The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *The Review of Economics and Statistics*, 47, 1, 13{37, 1965.
22. Mantripragada, K. , Beta Adjustment Methods, *Journal of Business Research*, 8, 3,329{339, 1980.
23. McInish, H. T. and Wood, A. R., Adjusting for Beta Bias: An Assessment of Alternate Techniques: A Note, *the Journal of Financial*, 41, 1 277{286, 1986.
24. Milionis E. A., A conditional CAPM; implications for systematic risk estimation, *The Journal of Risk Finance*, 12 4, 306{314, 2011.
25. Milionis, E. A. and Patsouri, K. D., A conditional CAPM; implications for the estimation of systematic risk, *Bank of Greece*, Working Paper 131**,** 2011.
26. Milonas, T. N. and Rompotis, G. G., Does Intervalling Effect Affect ETF's?, *Managerial Finance*, 39, 9, 863{882, 2013.
27. Sarker, R. M., Forecast Ability of the Blume's and Vasicek's Technique: Evidence from Bangladesh, *Journal of Business and Management*, 9, 6, 22{27, 2013.
28. Semushin, A., Parshakov, P., 'The Impact of Data Frequency on Performance Measures' *9th International Conference on Applied Financial Economics*, INEAG, 495-502, 2012.
29. Sharpe, F. W., Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, *Journal of Finance*, 19, 3, 425{442, 1964.
30. Stoukas, T., 'Greece Downgraded to Emerging Market at MSCI in World First' *Bloomberg Business week*, 11th June 2013, 2013.
31. Vasicek, A. O., A note on Using Cross-Sectional information in Bayesian Estimation of Security Betas, *The Journal of Finance*, 28, 5, 1233{1239, 1973.

Appendix

Graph 1 ASE Index



Graph 2 Continuously Compounded Rate of Return of ASE Index

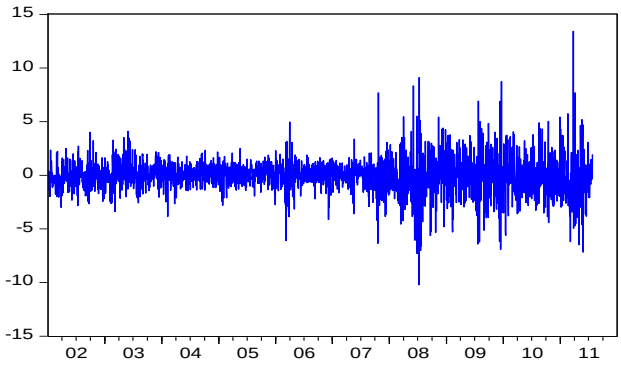


Table 1 Exponent k Prices

PRICES FOR EXPONENT K					
PERIOD 02-06					
NO CORHAY			CORHAY		
OLS	GARCH	EGARCH	OLS	GARCH	EGARCH
1.46366071	1.2540625	1.38691964	0.44526786	0.40174107	0.41767857
PERIOD 07-11					
1.13620536	1.25848214	1.17584821	0.60973214	0.74915179	0.69383929

Table 2 Blume's Regression Formulas

BLUME'S REGRESSION FORMULAS	
REGRESSION METHODOLOGY	FORMULA
OLS DAILY AND NO CORHAY	y=0.264683+0.298818*x
OLS MONTHLY AND NO CORHAY	y=0.469163+0.234048*x
GARCH DAILY AND NO CORHAY	y=0.258781+0.281737*x
GARCH MONTHLY AND NO CORHAY	y=0.4591+0.19378*x

BLUME'S REGRESSION FORMULAS	
REGRESSION METHODOLOGY	**FORMULA**
ASYM OLS AND NO CORHAY	y=0.43205+0.244549*x
ASYM GARCH AND NO CORHAY	y=0.381777+0.274477*x
ASYM EGARCH AND NO CORHAY	y=0.373206+0.248492*x
OLS MONTHLY AND CORHAY	y=0.481481+0.20592*x
GARCH DAILY AND CORHAY	y=0.247555+0.297101*x
GARCH MONTHLY AND CORHAY	y=0.424121+0.215524*x
ASYM OLS AND CORHAY	y=0.548167+0.164256*x
ASYM GARCH AND CORHAY	y=0.502987+0.175038*x
ASYM EGARCH AND CORHAY	y=0.456146+0.175427*x

Table 3 MSE Results (no Corhay correction)

MEAN SQUARE ERRORS BETWEEN ADJ OR ASYMPT AND NAÏVE BETAS (no corhay)			
OLS DAILY AS BENCHMARK (02-06)	**ADJ OLS 07-11**	**ASYMP OLS 07-11**	**NAÏVE OLS 07-11**
	0.286191552	**0.270828**	**0.388213724**
result: asymptotic OLS assesments more accurate			
OLS MONTHLY AS BENCHMARK (02-06)	**ADJ OLS 07-11**	**ASYMP OLS 07-11**	**NAÏVE OLS 07-11**
	0.257303968	0.367929	0.379163204
result: asymptotic OLS assesments less accurate but more compare to naïve			
GARCH DAILY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP GARCH 07-11**	**NAÏVE GARCH 07-11**
	0.264122376	0.273596	0.372973585
result: asymptotic GARCH assesments less accurate but more compare to naïve			
GARCH MONTHLY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP GARCH 07-11**	**NAÏVE GARCH 07-11**
	0.319749826	0.420151	0.479564789
result: asymptotic GARCH assesments less accurate			
GARCH DAILY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP EGARCH 07-11**	**NAÏVE GARCH 07-11**
	0.264122376	0.300656	0.372973585
result: asymptotic EGARCH assesments less accurate but more compare to naïve			
GARCH MONTHLY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP EGARCH 07-11**	**NAÏVE GARCH 07-11**
	0.319749826	0.453409	0.479564789
result: asymptotic EGARCH assesments less accurate but more compare to naïve			

Table 4 MSE Results (Corhay correction)

MEAN SQUARE ERRORS BETWEEN ADJ OR ASYMPT AND NAÏVE BETAS (corhay)			
OLS DAILY AS BENCHMARK (02-06)	**ADJ OLS 07-11**	**ASYMP OLS 07-11**	**NAÏVE OLS 07-11**
	0.286191552	**0.253697139**	**0.388213724**
result: asymptotic OLS assesments more accurate			
OLS MONTHLY AS BENCHMARK (02-06)	**ADJ OLS 07-11**	**ASYMP OLS 07-11**	**NAÏVE OLS 07-11**
	0.375046614	0.467201	0.486873048
result: asymptotic OLS assesments less accurate but more compare to naïve			
GARCH DAILY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP GARCH 07-11**	**NAÏVE GARCH 07-11**
	0.260343182	0.270565	0.372767352
result: asymptotic GARCH assesments less accurate but more compare to naïve			
GARCH MONTHLY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP GARCH 07-11**	**NAÏVE GARCH 07-11**
	0.367789325	0.445706	0.500254781
result: asymptotic GARCH assesments less accurate but more compare to naïve			
GARCH DAILY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP EGARCH 07-11**	**NAÏVE GARCH 07-11**
	0.260343182	0.295568	0.372767352
result: asymptotic EGARCH assesments less accurate but more compare to naïve			
GARCH MONTHLY AS BENCHMARK (02-06)	**ADJ GARCH 07-11**	**ASYMP EGARCH 07-11**	**NAÏVE GARCH 07-11**
	0.367789325	0.493591	0.500254781
result: asymptotic EGARCH assesments less accurate but more compare to naïve			

A method for calculating life tables using archive data. An example from mountainous Rhodopi.

Konstantinos N. Zafeiris[1]

[1] Laboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace. Email: kzafiris@he.duth.gr

Abstract: Using archive data, the Pomaks of Organi and Kehros (Greek Thrace) were studied. In order for the mortality transition to be evaluated a life table analysis was carried out. Results suggest a rapid mortality transition. Finally, Pomaks have converged to the total Thracian population.
Keywords: Mortality, life tables, Pomaks, Thrace.

1 Introduction

While much has been done concerning the life table analysis of mortality and health of Greece (see for example Skiadas & Skiadas [16] [15] [14]), such techniques have rarely been used in order to estimate mortality levels and transitions in the anthropological populations of the country. This paper is an attempt to apply these techniques to isolated populations, using the available archive data for the Pomaks of two former municipalities of Rhodopi, Greece (Zafeiris [21]). In a subsequent paper the Health State Theory techniques will also be applied for the calculation of several health state and demographic indicators for this population.

Map 1: The major Pomak areas (in grey).

The Pomaks (Map 1) were originally a mountainous population mainly living on either side of the Greek – Bulgarian Borders (Bacharov [2]; Georgieva [4];

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

Zafeiris [21]). They use a Slavic dialect with numerous Turkish and Greek loans and they are Muslims, in fact Islamised ex-Christian populations (Georgieva [4]; Papachristodoulou [10]).

This study concerns the mortality transition of the Pomaks of Organi and Kehros, i.e of two former administrative communities (part nowadays of the municipality of Fillyra) which are located in the most north-eastern part of the Department of Rhodopi of Greek Thrace over Rhodopi Mountain. The community of Organi consisted of 11 settlements, and covered an area of 200 Km^2. That of Kehros covered an area of 146 Km^2 and consisted of 12 settlements (National Statistical Service of Greece [8]). Pomaks, while originally geographically or even culturally isolated, lived in a harsh environment, where forests of conifers and deciduous trees were interspersed among limited cultivations and moderately extensive pasture lands. In its traditional stage, the local economy was based on family production. People were small farmers, cultivating mainly tobacco and crops and other products for home use. Others were involved in stockbreeding or forestry. Through time many of them have permanently migrated to the lowland villages and to a lesser degree to the city of Komotini, adapting to new environments and activities. Thus, the aim of this study is the analysis of mortality of a population firstly at its original mountainous place and secondly in the place in which it has gradually dispersed through time.

2 Data and Methods

We have used the civil register archives of the former municipalities of Organi and Kehros in order to reconstruct the life lines of every person of the two populations. The main book used was the Registrar General of the two municipalities, along with birth, death and marriage books. The Registrar General book is comprised of numbered family registers, each one containing registries for every member of a nuclear family which consists in its typical form of the head of the family (the husband) and his wife (or wives in the case of remarriage(s)) and the children of the family. For each one of them surname, first name, father's and mother's names are known along with demographic information: birth date, death date, marriage, transcription to another municipality, loss of citizenship, divorce. When a child from a family gets married it is transcribed to another family registry as head (husband) if it is a male or as a wife if it is a female, and it is deleted from the paternal one. However, in a special column of the book the number of the family of destination is written during this process, while the number of the family of origin is written in another column in the new family registry. It must be noted that the divorced husbands remain as heads at their family registry. Women are transcribed to a new family registry as divorced, following the same procedure as above. So formally, a person's position in the archive can be identified by a series of reference numbers consisting of its paternal family number and the number of the family registries in which it has been transcribed. In that way a

person can be followed up until their death or deletion because of their transcription following migration to another municipality.

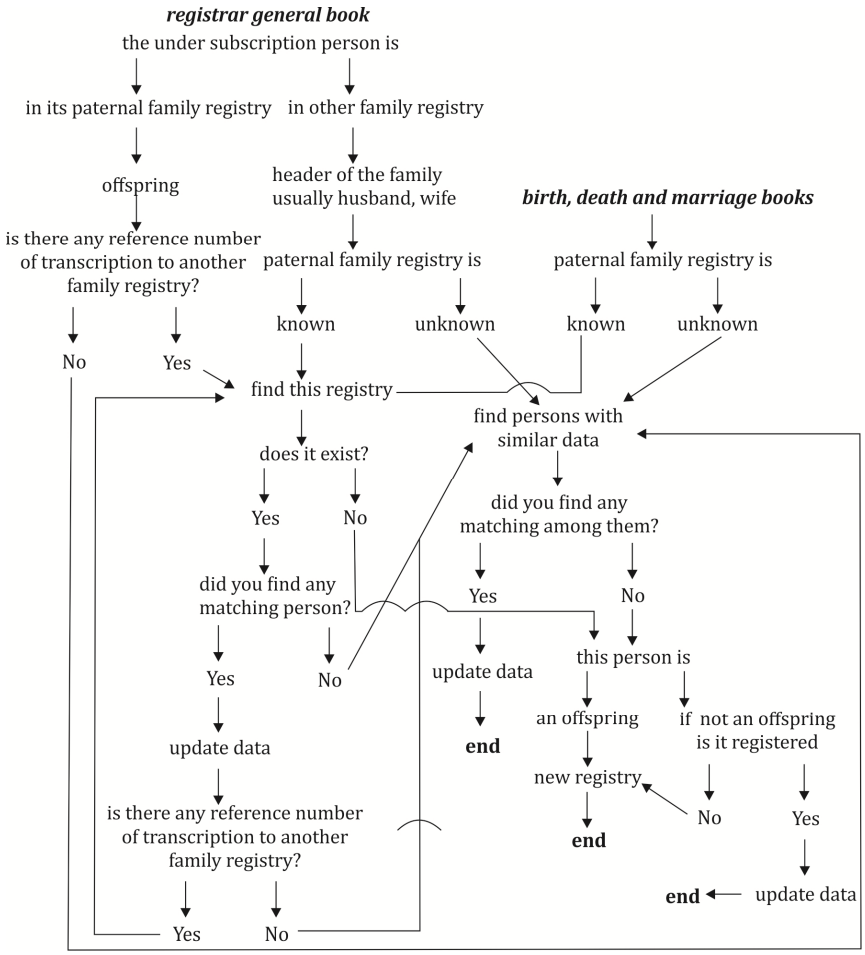


Fig. 1. An algorithm for data entry in the computer.

However, the Registrar General and the other available books were in hand written form and the data contained within should be stored in a database manually in order for the analysis to be carried out. As obvious in the previous paragraph the majority of the people were included in the Registrar General book more than once, in some cases 3, 4 or more times and no demographic analysis can be carried out, if this is not taken into consideration. In addition, the demographic data of the different sources or book registries had to be easily updated, as is also the case with the genealogical relationships of the members of the populations. Because none of the known available commercial software

was considered suitable for this purpose a new one was developed, named in this stage of development Demstat V2 (Demographic Statistics V2).

It was built in a Visual Fox Pro environment and it was based on the SQL language syntax (see http://msdn.microsoft.com/en-us/vfoxpro/default). It retains a four-fold process. Firstly, a relational database was constructed consisting of 18 tables, connected to each other by various keys and relational expressions. A special effort was made in order to find the most parsimonious solution during this process: on one hand personal time and effort along with computer and memory requirements should be taken into consideration and on the other hand every one of the archival books used should be able to be reproduced in its original, but digital form. The last was because data entry after its finalization had to be verified digitally once again. Secondly, a friendly user interface was constructed comprised of several forms, programs and routines. The basic characteristic of this interface was that data could be entered into the computer in a dynamic way according to the strategy that the user chose every time as the most suitable for each case. One of the algorithms used for this is described in Figure 1.

Additionally, in parallel to the original data entry for the persons, a digital library of all surnames and first names existing in the population was created for spelling errors to be checked. If such errors were identified, a new corrected form of the mistakenly written entry was entered in the library. Because the population is a Muslim one a variety of sources was used (Underhill [20]; Tuncay & Karatzas [18]; Pampoukis [9]; Tzemos [19]). Afterwards, a new corrected record from the library was added for every person's mistakenly written surname or name in the database. However, the original form of every surname and name was also kept.

The third aim of the software was the construction of the genealogical trees and pedigrees of the population in any possible form: patrilineal, matrilineal, bilateral etc. The fourth aim of the software was the demographic and genealogical analysis of the population which was based on the life lines of the persons of the population. The demographic analysis was based on the two dimensional form of Lexis diagram (see Feeney [3]), which as is known for every person contains information about its cohort, the date of occurrence of a demographic event and the age of that person in that demographic event. However, because persons' life lines were known, for the analysis of any demographic phenomenon either the average population or the relevant person years lived in the population in a time period could be used. Similarly, the analysis could be based by choice either on the "squares" or the "parallelograms" of the lexis diagram.

After entering the data, a validation of the records took place following three procedures. In order for the manually entered records to be validated for their precision concerning the people linkages the Registrar General Book was

restored digitally in both original and digital state of family and people registries. Then a routine was built in SQL in order for the different reference numbers to be followed up and the records corresponding to them to be compared using several criteria like surname, name, father's and mother's name (corrected forms), birth date, death date, marriage date. If these criteria were satisfied the relevant records were considered to refer to the same person and the manual linkage of the records done before was verified. Afterwards the opposite routine was used; people were firstly identified by the criteria described above and secondly their reference numbers were compared. In the third procedure a manual check of the results of the process took place, as happened with a few problematic records like those with mistaken reference numbers or mistaken father's names etc.

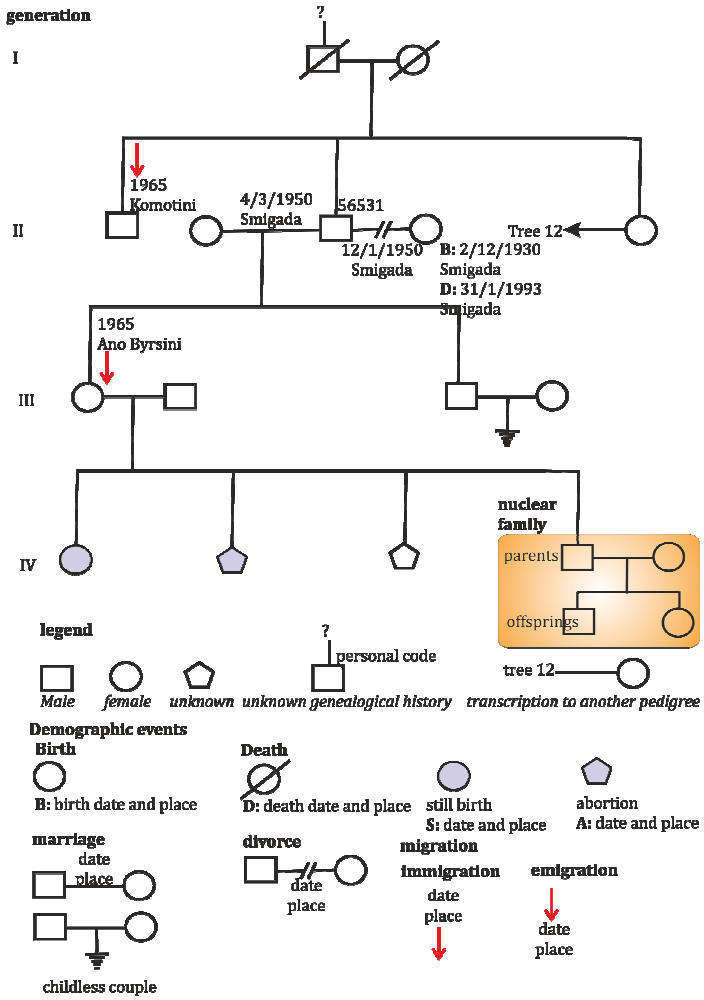


Fig. 2. A patrilineal genealogical tree or pedigree.

Then the patrilineal genealogies (Figure 2) of the population were constructed and used as a basic research tool in the field work carried out afterwards, when data were tested for their validity and completeness and the data base was once again updated if any omitted or false data were found. We have to note that inhabitants not registered in the Registrar General Book were not identified during the field work, because the population lives in a mountainous area and the only type of immigration which occurred there was the marital one which presupposes registration in the archives. However, through time many of the registered citizens of the two municipalities migrated to the Rhodopi plain, mainly to small villages of the area and to a lesser degree to the urban center of Komotini.

The life line for each of the members of the populations was constructed and used in the following analysis. By studying all these life lines at any given time period, i.e. by studying the population at risk of a specific demographic phenomenon, the person-years lived by the members of the population were calculated as well as its composition by sex and age and the demographic events which occurred (see Preston et al. [12], pp. 3-16). In this analysis, as the starting point of a person's life line was considered their birth, unless an immigrant where their date of migration is considered as the starting point, and as the terminal point of the life line were considered the death of a person, their transcription to another municipality and the end of the research. The average population for every year was calculated as the mean population between two subsequent New Year's days.

Then the age specific mortality rates per sex were calculated using standard procedures (see for example Preston et al. [12], p. 21-23) because the population is small and subjected to chance fluctuations. For each of the 5-year periods the numerator of the age specific mortality rate formula was smoothed as the average annual number of deaths per age class during that period, while as denominator the average population in the middle year of that period was used. Subsequently a life table analysis (see Preston et al. [12], pp. 38-69) was carried out for both sexes based on one year age classes up to the age of five, and five years long age classes for the older ages. Life tables were considered to be closed by the age of 85 for both sexes.

The probability of death for life table analysis was calculated using the Chiang's method (Namboodiri [6], p. 85):

$$ {}_n q_x = \frac{n^*{}_n M_x}{1 + (n - {}_n a_x)^*{}_n M_x}, $$

where ${}_nM_x$ is the age specific mortality rate between age x and [x+n), n is the length of the age interval in an abridged life table and ${}_na_x$ is the fraction of the interval between x and [x+n) birthdays lived on average by those dying in that interval. The values ${}_na_x$ were calculated directly from the data for the ages less

than 5 years. For the older age classes, because the observed number of deaths was small, deaths were considered to be equally distributed in each age class.

In order for the results of the analysis to be compared with the population of Thrace and Greece published data concerning deaths per sex and age and the relevant age distributions were used (vital events statistics and population censuses results published by the National Statistic Service of Greece, nowadays National Statistical Authority, www.statistics.gr) and life tables for the census years between 1961 and 1991 were calculated. Infant mortality rate (IMR) and though q_0 was calculated following Pressat [11] as:

$$IMR = \frac{D_{(0,t)}}{1/3 * b_{t-1} + 2/3 * b_t},$$

where D stands for infant deaths and b for births, where t is the year. The number of infant deaths was smoothed as the mean number of infant deaths in the census year and its adjusting ones. For years 1 to 5 the Reed-Merrell formula ([13], cit. Naboodiri [6]) was used:

$${}_n q_x = 1 - \exp(-n * {}_n M_x - a * n^3 * {}_n M_x{}^2)$$

where n=1, a=0.008 and ${}_nM_x$ the age specific mortality rates. For the remaining age groups the Chiang's method was used.

The calculated death probabilities were applied to an abridged life table (closing at the age of 85) prepared by Skiadas and Skiadas ([14] [15] [16]), in order for the life expectancies at birth to be estimated.

3 Results

The Pomak population of Organi and Kehros (Pomaks from now on) underwent a rapid mortality transition between 1962 and 1992 (Figure 3). During that period, female life expectancy at birth (LEB) increased by 35,4%, from 57,4 to 77,8 years. Similarly, LEB of the male population increased by 15,4 years or by 27,3%. However, if Pomaks are compared with the total, the rural and the urban population of Thrace, as well as with the analogous populations of Greece, for most of the time a three zone pattern of classification is emerging. At the upper zone, that of the highest LEB, though the lower mortality, the population of Greece is located. The middle zone is formed by the Thracian population and the third one - that of the higher mortality – by the Pomaks.

This tripartite scheme results from the economic, social and political peripheral inequalities observed in the country. It is indicative of that situation that in Thrace, even in the 1980s, the local economy was based on the primary sector, especially agricultural production, which accounted for 70% of the income of the inhabitants (Stathakis [17]). Even more, in 2001 the whole region of Eastern Macedonia and Thrace was in the first position (i.e. the worst one) of the

Human Poverty Index (HPI) ranking among the regions of Greece and in the last position according to its Human Development Index (HDI) (see Kalogirou et al. [5], pp. 50 and 57 respectively).

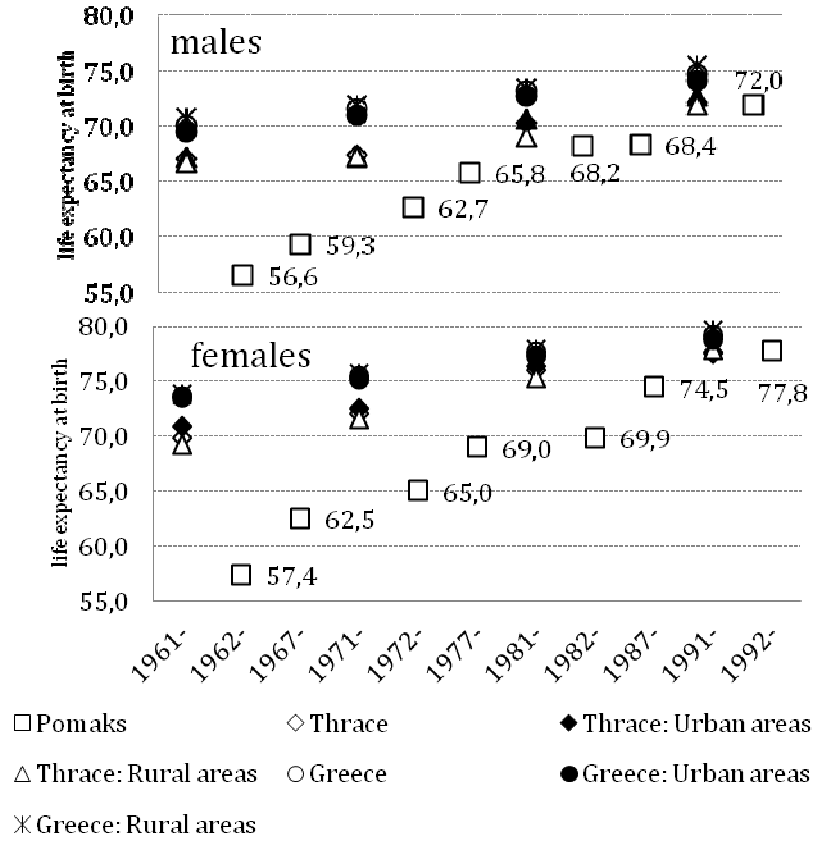


Fig. 3. Life expectancy at birth (LEB) in various populations

Table 1. Life expectancy in Thrace

	Males			Females		
	Total (1)	Urban	Rural	Total	Urban	Rural
1961-	67,1	67,1	66,7	69,9	70,9	69,3
1971-	67,4	67,2	67,3	72,1	72,5	71,6
1981-	70,3	70,6	69,1	76,1	76,4	75,4
1991-	72,2	72,7	71,9	77,7	77,5	77,7

(1) Including semi-urban population

Pomaks in their turn struggled for their survival in the past in adverse environmental conditions, especially during winter time. According to data of the Hellenic National Meteorological Service (http://www.hnms.gr/hnms/greek/index_html), the average temperature between years 1960 and 1982 in Komotini ranged between 4,62 and 6,9 ^{o}C in the winter months, while the respective lower ones ranged between -3,76 and -7,32 ^{o}C. Official data for mountainous Rhodopi is absent; however temperatures there are rather lower. According to the NGO "Arktouros" [1] in 1978-89 in Sidironero (Drama) they were less than -10 ^{o}C during winter and in Leivaditis (Xanthi) lower than -12 down to -20 ^{o}C.

More than that, the geographic isolation of the Pomak region in the past was another aggravating factor of mortality. Transportation and communication among the small settlements and the lowlands of Rhodopi was carried out mainly by mules and camels through the treacherous mountain paths. Started mainly after the 1960s, all efforts for the construction of roads were partially efficient; characteristically in the early 1990s the only paved road was the one connecting Komotini with Organi. Additionally, if not absent, electricity supply was gradually available for a few of the villages, as was the case with telephone connections and only much later did local infrastructures improve in an adequate way (Zafeiris [21]).

It is not surprising then that the local economy was rather a pre-industrial one, based on limited resources and with no real opportunities for economic development, which as a matter of fact was the problem with the mountainous Greece as a whole and especially with the Department of Rhodopi in those days. Additionally the medical care of the population was largely inadequate and even in the 1990s it was administered by few small agrarian clinics. All that, along with the very high illiteracy rate of the population, can explain the high mortality rates of this epoch. Characteristically enough, even in 1981 in the highlands of the Department of Rhodopi, only 35% of the males and 33% of the females had finished the bilingual minority elementary schools existing in Greek Thrace, while the illiteracy rate was 37,4% and 45% respectively (National Statistical Service of Greece [7]).

However, through time a general arsis of the geographic isolation took place in mountainous Rhodopi. Meanwhile a significant portion of the population had migrated to the lowland villages or the city of Komotini, and several transformation processes occurred in the whole of the Muslim minority of the area, like the modernization of agricultural production, the involvement of the population in the open market economy etc. A progressive elevation of the living standards and Health Services was observed and as a result the mortality

rates reduced (especially infant mortality) as did the "mortality gap" between the Pomaks and the other people of the area (Zafeiris [21]).

Despite the fact that Thrace is still one of the poorest areas of the Greek periphery all the populations benefited a lot from the developmental processes which occurred in the country between 1961 and 1991 (Table 1; Figure 1). During that time the LEB of the total population of Thrace increased, though at a lower rate than that of the Pomaks; 11,2% for the total female population, 9,4% for the urban and 12,4 for the rural one. The respective figures for the male population were between 7,6-8,3%. As a result by 1991 Pomaks had totally converged to the population of Thrace. In the mean time LEB in the total population of the country has increased somewhere near 6,6-8%, and both the Thracian population and the Pomaks have converged to the total population of Greece. However, important differences still existed in 1991, even though these became lesser in comparison to those which existed in the past.

In the Thracian population, females' improvements in LEB between 1961 and 1971 were more than 2 years for the total and the rural population and 1,6 for the urban one, while male gains were small. In 1981 mortality transition was accelerated and at the end, in 1991, the total, urban and rural populations per sex converged, when the observed differences among them were very small.

It is obvious that the observed LEB differences between the two sexes are positively correlated with LEB levels until 1981: the higher the LEB the greater the differences in LEB between males and females. In 1991, when LEB was at its maximum levels the differences between the two sexes narrow, though they remain big enough. Overall, females have all the time longer lives and they benefited more than males during mortality transition in the area.

In Pomaks, LEB sex differences per studied period tended to be smaller in comparison to the population of Thrace until 1987. Field work evidence suggests that on the one hand everyday life and the general living conditions in mountainous Rhodopi in the past were aggravating factors for the chances for survival. On the other hand women of reproductive ages were additionally burdened by complications during pregnancy, delivery and the postpartum period. In the past, because of their geographic isolation and the absence of qualified medical personnel, these women used to give birth to their children at home aided by the older women of the village and some uneducated midwives. Nowadays, all the Pomak women benefit from the tertiary health system of Thrace, and especially the obstetric clinics of the General Hospital of Komotini and the University Hospital of Alexandroupolis. Probably this is one of the reasons that in 1987 the differences between the two sexes were maximized.

Conclusions

Through time a general trend of decrease in mortality levels in Thrace was observed. The most rapid mortality transition was that of the Pomak population of Organi and Kehros which had the greatest improvements in life expectancy at birth. As a result by the end of the study all the Thracian populations converge to lower mortality levels.

References

1. Arktouros. *Study for Rhodopi and its natural character*. ny [In Greek].
2. M. Bacharov. The current ethnic panorama in Bulgaria. *GeoJournal*, 43,3, 215-224, 1997.
3. G. Feeney. Lexis Diagram. In Demeny, P. and McNicoll, G. (eds.) *Encyclopedia of Population, Volume 2*, Macmillan Reference USA, 586-588, 2003.
4. T. Georgieva. Pomaks: Muslim Bulgarians, *Islam and Christian - Muslim Relations,* 12, 3: 303-316, 2001.
5. S. Kalogirou, A. Tragaki, C. Tsimpos and E. Moustaki E. *Spatial inequalities of income, development and poverty in Greece. Projects 2011*. John Latsis Benefit foundation. 2011. Available at: http://www.latsis-foundation.org/en/101/projects_2011.html
6. K. Namboodiri. *Demographic Analysis.* A Stochastic Approach. Academic Press Inc, San Diego, 1991.
7. National Statistical Service of Greece. *Résultats du recensement de la population et des habitations. Effectué le 5 Avril 1981. Volume V. Fascicule 9. Epire.* Imprimerie Nationale, Athènes, 1991.
8. National Statistical Service of Greece. Distribution of the area of the country according to its principal use. Athens, 1986.
9. I. T. Pampoukis. *The Turkish Vocabulary of the Modern Greek Language*. Papazisis, Athens, 1988. [In Greek]
10. P. Papachristodoulou. The Pomaks, *Arheion Thrakikou Laografikou kai Glossologikou Thisavrou*, 23, 3-25, 1958. [In Greek]
11. R. Pressat, R. *Demographic Analysis: Methods, Results, Applications.* Aldine publishing, New York, 1980.
12. H. Preston, P. Heuviline and M. Guillot, M. *Demography. Measuring and Modeling Population Processes*. Blackwell Publishers, Oxford, 2001.
13. L. J.Reed and M. Merrel, M., A short Method for constructing an abridged life table. *American Journal of Hygiene*, 30, 33-62, 1939.
14. C. H. Skiadas, and C. Skiadas. *The Health State Function of a Population*.2nd edition. Athens, 2013a.
15. C. H Skiadas and C. Skiadas. *Supplement: The Health State Function of a Population.* ISAST, Athens, 2013b.
16. C. H. Skiadas and C. Skiadas. The First Exit Time Theory applied to Life Table Data: the Health State Function of a Population and other Characteristics, *Communications in Statistics-Theory and Methods,* 34: 1585-1600, 2014.
17. D. Stathakis. Processing in Thrace, *Thrakiki Epetirida,* Δ, 1983. [In Greek]
18. F. Tuncay and L. Karatzas. *Turkish-Greek Dictionary.* Center for Eastern Languages and Culture, Athens, 2000.

19. G. Tzemos, *The of Turkish Origins Greek Surnames*. Kessopoulos, Thessaloniki, 2003. [In Greek]
20. R. Underhill. *Turk Dili Grameri, dil, Turk dili, Turkce Grameri. Turkish Grammar.* 7th edition. The MIT Press, Cambridge, 1993.
21. K. N. Zafeiris. *Comparative analysis of the biological and demographic Structures of isolated populations in the Department of Rhodopi*. PhD Thesis. Democritus University of Thrace, Department of History and Ethnology. Komotini, Laboratory of Anthropology, 2006. [In Greek].

Demographic and Health Indicators in the Pomaks of Rhodopi, Greece

Konstantinos N. Zafeiris[1], Christos H. Skiadas[2]
[1]Laboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace, Greece (E-mail: kzafiris@he.duth.gr)
[2]ManLab, Department of Production Engineering and Management, Technical University of Crete, Chania, Greece (E-mail: skiadas@cmsim.net)

Abstract: Based on the Health State Theory several demographic and health state indicators were calculated for the Pomaks of Organi and Kehros, a Slavic speaking population from Mountainous Rhodopi (Greece). Results indicate a rapid health transition and a general improvement of the health status of the population. Gradually, by the 1990s Pomaks converge to the total population of Thrace.
Keywords: Pomaks, Thrace, Health State Indicators

1 Introduction

Life table analysis has for long been (Graunt [3], Halley [4]) used to estimate probabilities of survival and life expectancy in different ages during the course of human life. However, if applied in its classical form - as a technique aimed to describe the patterns of mortality and survival - it fails to give an overall picture of the total health status of a population and its changes with age and time, especially, if health is positively defined not simply as the absence of a disease or infirmity but in a broader way as *"a state of complete physical, mental and social well being"* (WHO [20]). Several solutions have been given to this problem.

In order to mathematically describe the state of health of a population, Chiang [1] introduced the term *"Index of health H_x",* based on the probability distribution of the number and the duration of illness and the time lost due to death (time of death) calculated from data from the Canadian Sickness Survey, 1950-1951. Sanders [8] used life table techniques to construct tables of *"effective life years"*, as a measure of the current health of the population based on mortality and morbidity rates. In that case, morbidity was measured by the functional adequacy of an individual to fulfill the role which a healthy member of his age and sex is expected to fulfill in his society. Sullivan [9] criticized these approaches on grounds of the methodology used and its effectiveness in measuring the health status of a population and later [10] he used life table techniques to calculate two related indices: the expectation of life free of

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

disability and the expectation of disability, based on published data of the current abridge life tables and surveys conducted by the National Center for Health Statistics. Torrance [19] developed a health status index model for the determination of the amount of health improvement created by a health care program, calculating a point-in-time health index (H_t), a period-of-time health index ($H_{t1,t2}$) and a health life expectancy index (E_x). The first measures the instantaneous health of all individuals in the population at a single point in time and averages these to give a group's index. The second is similar but it refers to a period of time, i.e. a year. The third is calculated by a method that uses the actual mortality and morbidity experience of the population to determine a table of age and sex specific health expectancy figures.

Jansen and Skiadas [5] introduced the general theory of dynamic models for modeling human life using life table data from France and Belgium; this was the first effort to exclusively use published death and population statistics in order for the health status of the population to be evaluated. In their model, the main assumption regarding the state of human health is that it follows a stochastic process expressed by a variable named $S_{(t)}$ and that the end of life time is reached when this stochastic variable arrives at a minimum level of health state. However an important hypothesis was set: despite the fact that the health state of an individual is unpredictable the mortality curve of human populations can be modeled and because of that the health state form derived from these data can be modeled too (see Skiadas and Skiadas [17]. Then the Health State Function (H_x) of a population was introduced, elaborated later (Skiadas and Skiadas [11] [12] [13] [14] [15] [16] [18]) and evaluated with several methods (see Skiadas and Skiadas, [17]), with particular emphasis on the calculation of the Loss of Healthy years (see Skiadas and Skiadas [17] in comparison with Mathers et al. [6] [7]). The mathematical formula for the calculation of Health State H_x is:

$$H_x = \pm\left(-2x\ln\frac{d_{(x)}\sqrt{x^2}}{k}\right)^{\frac{1}{2}}$$

where k is a parameter given by $k=\max(d_{(x)}\sqrt{x^2})$ and $d_{(x)}$ is the number of deaths per 100.000 of population provided by the classical Life Tables, or preferably they may correspond to the probability density function. According to theory developed by Jansen and Skiadas [5], the Health State Function has an improvement stage of human health during the first years after birth, followed by a decreasing one during middle and old ages.

Based on the Health State Theory (HST) many Demographic and Human Development Indicators have been proposed and calculated efficiently for numerous national populations based on data published by National Statistical Services or found in other databases (see Skiadas and Skiadas [15] [14] [13]). This paper is a first attempt to apply the Health State Theory techniques to data

already used for the evaluation of the mortality transition in the Pomaks of Organi and Kehros (Zafeiris [22] [21]), i.e. it is a first effort to apply HST on small and isolated populations.

Map 1: the geographic location of the under study population

The data used here are from two former administrative communities, namely Organi and Kehros, located in the northeastern part of the Department of Rhodopi of Greek Thrace, over the Rhodopi Mountain (Map, 1; see Zafeiris, [22] [21]). These communities were comprised of 23 settlements and extended to an area of 346 Km^2. They are inhabited by the Pomaks; in general a Muslim population, speaking a Slavic idiom with many Greek and Turkish loan words, spread on both sides of the Greek-Bulgarian borders. Pomaks of Organi and Kehros constituted a geographically and culturally isolated population, struggling for their survival in quite a harsh environment. In the past they were small farmers, stockbreeders and woodcutters. Though field work evidence suggests that a time varying trend of migration from their mountainous dwellings to the plains of Greek Thrace was observed in the past, after the arsis of geographic isolation because of the construction of roads in the Rhodopi Mountain, this trend was magnified and a significant and continuous increasing number of them settled in small lowland villages and the city of Komotini. Thus, the aim of this study is the analysis of health characteristics of a population firstly at its original mountainous place and secondly in the place to which it has gradually dispersed through time.

2 Data and Methods

The methodology used for the preparation of the abridged life tables for males and females is described in Zafeiris [22]. On these tables, the calculations based on the Health State Theory (Skiadas & Skiadas [13] [14] [15]) were made. As it is said before, the Health State Function (H_x) aims to the quantification of the health state of a population and is based on the death probability function (g_x) and a parameter called k, which is calculated from the g_x distribution (see Skiadas & Skiadas [11], p. 97). However, because the studied population is small the H_x distributions were subjected to chance fluctuations by age. In order to smooth them two order polynomial trend lines were fitted, in which as intercept the H_0 values were set, i.e. the health state level of infants (the R^2 was at all times much greater than 0,90). Because the original life tables were "closing" by the age of 85, the H_x values beyond this age were estimated with the use of the fitted line.

According to this fitted line the following health state indicators were calculated and used in this paper: H_0, H_{max} - i.e. the maximum value of the H_x distribution which is a measure of the maximum health state achieved by the population - and Total Health state (THS), which corresponds to the sum of H_x values before the zero point of the health status of the population and is then a comprehensive assessment of the health levels of the population. Afterwards, the total health state from max to zero health was estimated as a sum of the H_x values from the maximum point of the Health State Function until its zero point and will be an overall measure of health during its progressive reduction phase in the human life cycle. Additionally, the "loss of Healthy Life Years" because of severe causes (LHLY1) as well as that because of moderate and severe causes (LHLY3) were calculated (see Skiadas and Skiadas [12], p. 18). Based on that, life expectancy at birth for moderate & severe disability causes (HLEB 3) as well as that only of the severe causes (HLEB 1) were calculated simply as the difference of life expectancy at birth (LEB) and LHYL3 or LHLY1 respectively (see Skiadas & Skiadas [11], p. 101).

Rotated factor loadings		
component	1	2
LEB	,977	-,213
H_{max}	,937	-,452
THS	,938	-,479
THS from max to zero health	,873	,128
LHLY1	-,215	,958
LHLY3	-,411	,903
HLEB3	,953	-,494
HLEB1	,958	-,415
H_o	,892	-,127

Table 1. Factor loadings of PCA

The results of the analysis were compared with those of the urban, rural and total populations (including semi-urban) of Thrace and the relevant ones of the entire country, for which, in order to be comparable, exactly the same method described above was used. Then, a principal component analysis was carried out in order to visualize the geometric relationships among the populations studied, in fact in order for these to be classified according to their survival and health characteristics and the similarities they exhibited. Several rotation methods of PCA were used, giving similar results. The one cited here is the Oblimin with Kaiser Normalization which is a method for oblique (non-orthogonal) rotation (see Field [2], pp. 702-703). The variables used are seen in Table 1, however because of the scaling differences existing among them their values were transformed to their z-scores before the method appliance. The PCA was performed with 2 factors, as suggested by the scree plot (not cited here) of the eigenvalues against factors and the Kaiser criterion for keeping in a PCA all factors with eigenvalue greater than 1 (see Field [2], pp. 677-678). The KMO test (Keiser-Meyer-Olkin measure of sampling adequacy) was 0.763, well above the value of 0.5 which is considered the minimum accepted value for which the factor analysis yield distinct and reliable factors (see Field [2], pp. 684-685). The Bartlett's test of sphericity was 1929,991 ($p<0.000$) and the determinant of the correlation matrix was 2,18E-15

which indicate the validity of the PCA (See Field [2], pp. 694-695). The two factors PCA explained 91.272% of the variance observed, 72,193% by the first component and 19,079% by the second one. The scatter diagram produced by the PCA was based on the relevant regression scores of the components extracted by the analysis.

3 Results

Mortality transition in the Pomaks of Organi and Kehros (Pomaks from now on) (Zafeiris [21] [22]) was accompanied, quite predictably, by a rapid transition of the health status of the population (Figure 1).

In that course, females' total health state level (THS) was increased by 54,3% between 1962 and 1992 (calculations based on the results cited in Figure 2). Similarly, males' THS increased by 36,8% between 1967 (when its minimum value was observed) and 1992. Between 1962 and 1967 males' THS seem to have declined. Females, on the contrary, at the same time had 18,3% gains in their THS.

It must be stressed out that despite the fact that THS levels prevailed significantly in the female population, in 1962 the opposite is observed. This is because of the elevated death probabilities of women mainly of reproductive age, because of the total absence of maternity homes and in general of the basic infrastructure for ensuring the health status of the mothers and their children. This absence was also imminent later on though it not easily recognized in the analysis results. In any case, a similar but insignificant reversal is found in 1982.

If these temporal trends are compared with the urban, rural and total populations of Thrace and then with those of the entire country, a rather complicated pattern of THS levels transition emerges. While male levels seem to be more or less stable between 1961 and 1971 in the entire country, in Thrace they decreased by about 2-2,5%; a trend that continued until 1981 in the rural and urban populations while in the total one THS levels remained stable. In contrast, between 1971 and 1981 the country's populations had significant gains. Consequently, a progressive divergence of the Thrace was observed until 1981. Eventually its populations would start to converge as a result of the acceleration of the health state transition between 1981 and 1991, but this convergence was not fully completed by 1991.

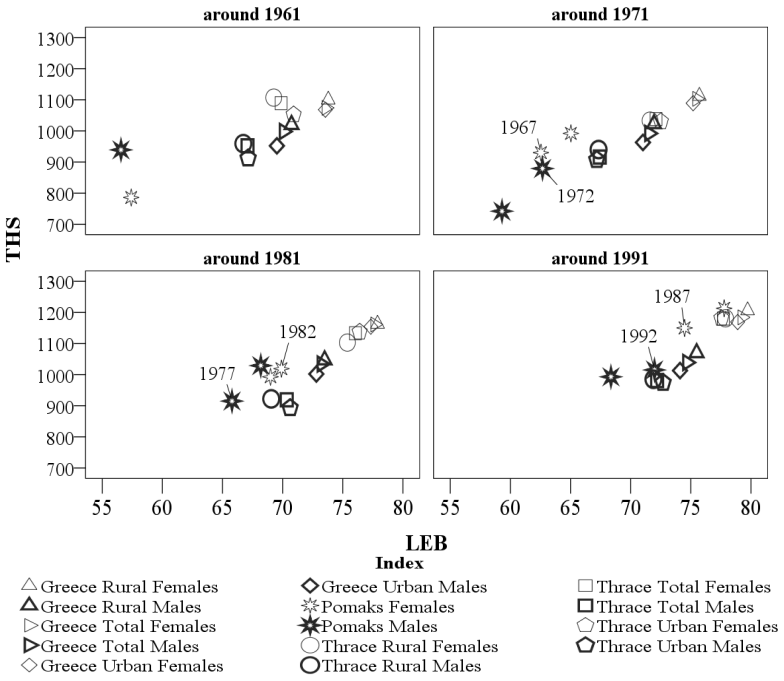


Fig. 1. Life expectance at birth (LEB) versus total health state levels (THS; fitted values).

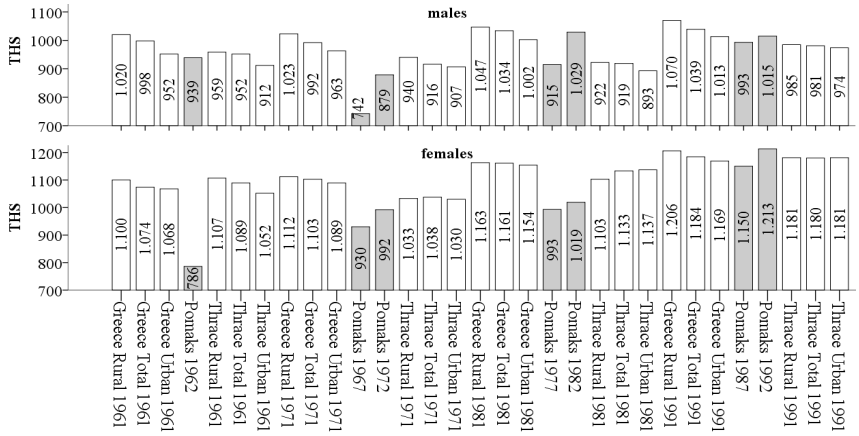


Fig. 2. Total health state levels (THS; fitted values).

Thracian females, like males, were also under held a “counter transition” of the THS levels, but in them this was more intense and finalized earlier; between 1961 and 1971 their “losses” were at the range of 2-7%, but between 1971 and

1981 their "gains" were high (~7-10%). The growth of THS levels continued later, even if it slowed significantly (~4% growth) apart from the rural population where it remained rather high (7,1%). Meanwhile, a continuous growth of THS levels was observed in the female populations of the entire country. This growth was originally minor, accelerated between 1971 and 1981 and then stalled, but still remained positive. Because of the different tempo of the THS transition, in the end the Thracian populations and those of the entire country converged significantly.

The different tempo of this transition is described by the relative growth of THS levels between 1961 and 1991. For the entire country THS gains were 4,9-6,4% for males and 9,5-10,2% for females. In Thrace they were about 2,7-6,8 and 6,7-12,3% respectively. But between 1971 and 1991 males had an increase of about 5% and females of 7,3-8,% for the entire country, while the respective figures for Thrace were 4,8-7,4% and 13,7-14,7%.

However, comparing these figures with those of the Pomaks it is obvious that THS transition was more rapid and intense in them. At the end - even if at the beginning their population was cited quite distantly - it followed the aforementioned trend of convergence with the others. Males, through a variable course, outpaced the Thracian populations in 1982, filling the "gap" observed between Thrace and the entire country. Female Pomaks followed a time variable but ultimately increasing course of their THS levels but these were consistently smaller than those of the Thracian populations until 1992. Then they prevailed in THS levels slightly and converged and even outpaced the population of the entire county.

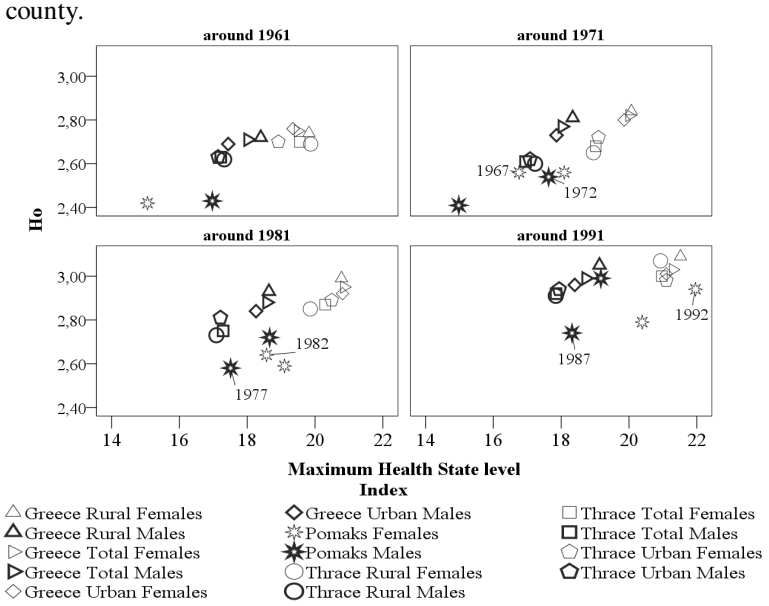


Fig. 3. H_0 versus maximum health state level.

Between 1961 and 1991 THS temporal changes were accompanied by a variable but ultimately increasing trend of the health state level of the infants (H_0) in all the populations studied (Figure 3). In males, three groups of populations were formed during that course. At the highest levels of H_0, i.e. at the best health state of the infants, the populations from the entire country were positioned, followed by the Thracian populations in the middle. Pomaks, for most of the time were located at the outskirts of this scheme, with lower H_0 levels and time-varying differences with the other populations and only in 1991 did they manage to converge with them. A similar pattern is observed for the female Pomaks, but there the differences with the other populations were greater at all times.

Meanwhile, maximum health state (H_{max}) levels of males increased significantly. While originally Pomaks were below them, in 1972 they outpaced the Thracian populations and, despite the observed discontinuities, they gradually converged with those of the entire country. Females of Pomak origin, in their turn, converged, and even outpaced, the other populations only in 1992. As a result, because of the differences observed at the starting point and the maximum level of the health state distribution by sex and age, Pomaks, either males or females, were mounted apart from the other populations studied as it is seen in Figure 3, and only after 1991 is a significant convergence.

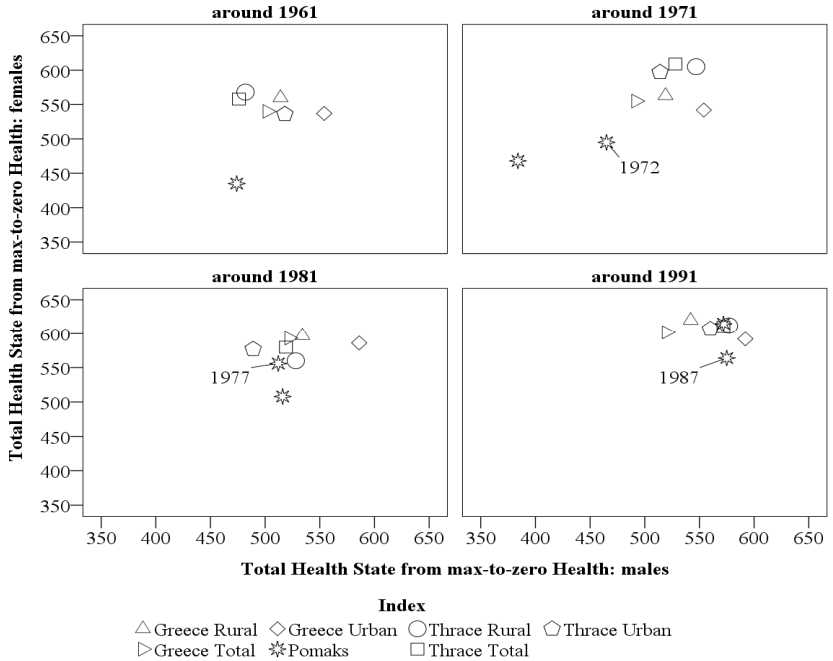


Fig. 4. Total Health State from max-to-zero Health.

The total health state from max to zero health is quite dissimilar among the populations studied (Figure 4). As an overall measure of health during its

progressive reduction phase in the human life cycle, it seems that the Pomak women, on one hand were in accordance with the general tendency of the with time improvement of their health status after middle age even if that took place in quite different tempos and followed different timetables among the populations studied. On the other hand, and for the majority of the times studied, Pomaks experienced larger burdens than those of the other women. The original convergence trend of Pomaks which was observed until 1977 was interrupted later and only in 1991 do they exhibit significant similarities with the other populations. A similar pattern of improvement of the total health status after middle age is observed in Pomak males, but their convergence was not interrupted in any way and after 1977 they exhibited major similarities with the other populations and in some cases they overtook them.

The Healthy life expectancy under severe and moderate causes (HLEB3) is in accordance with this trend of a significant lag of the health status of the Pomak population up to one time point and its convergence with the other populations later on (Figure 5). Obviously, women live longer than males - as evidenced by their LEB levels (Figure 1), and judging from the HLEB3 they are burdened by serious diseases later in their lives. However, an important exception is found for the Pomak women in 1962, as a result of their low health status in those days which was discussed previously.

Because of this, their losses of healthy years because of moderate and severe causes (LHLY3) were at their maximum values then. Afterwards, following a variable course their losses were gradually limited and at the end they were fully converged with the other populations. Pomak males followed a variable and eventually convergence course through time in both LHLY3 and HLEB3 levels. In total, Pomaks, both males and females, tended to cluster apart from the other populations until 1981, when their distances became smaller, as seen in their positions in the scatter diagrams of Figure 5. Later on, they exhibited many similarities with the other populations.

This distinction is apparent if the losses of healthy life years because of severe causes (LHLY1) and the relevant life expectancy (HLEB1) are taken into consideration (Figure 5). The temporal trends of both male and female populations are largely similar with those described above for LHLY3 and HLEB3, though more tight groupings are observed because in reality LHLY1 and HLEB1 (severe causes) is one of the components of LHLY1 and HLEB1; the other one is the component of the moderate causes.

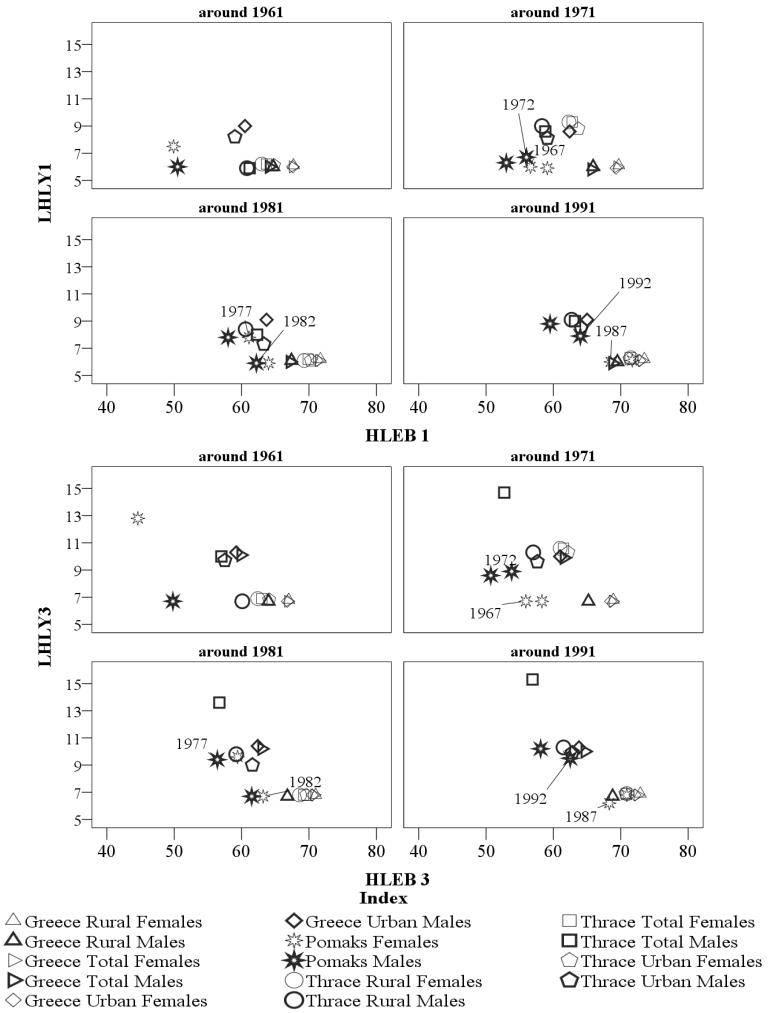


Fig. 5. LHLY1 versus HLEB1 and LHLY3 versus HLEB3.

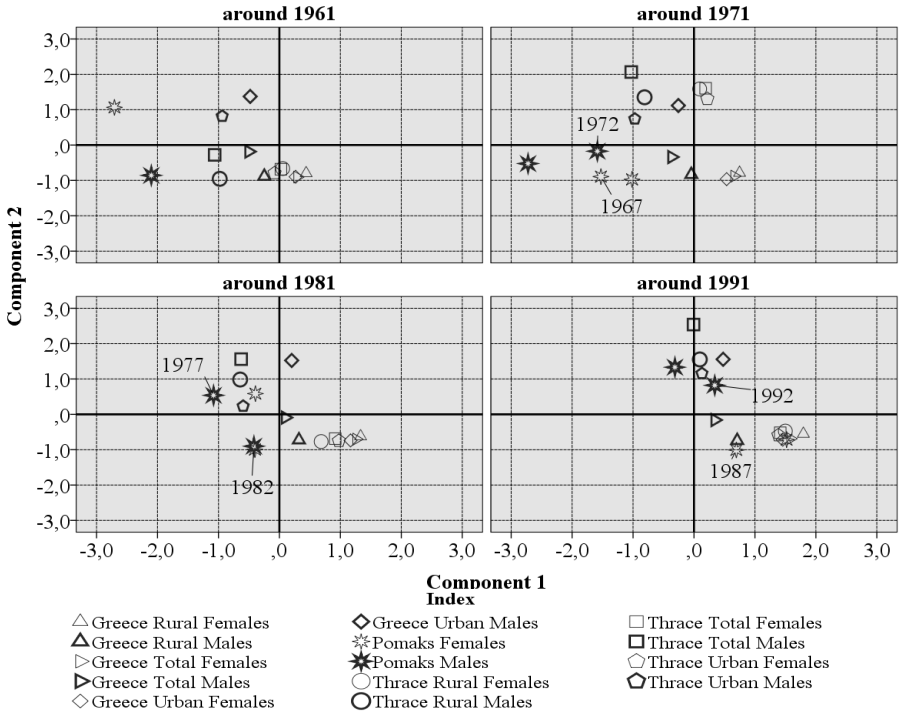


Fig. 6. PCA analysis results.

The principal components analysis (PCA) condenses the high variability observed among the populations studied (Figure 6). In general, a clear and predictable distinction of males and females is observed which is combined within the two sexes with a general tendency of convergence of all the populations studied. Around 1961, all the female populations, except the Pomaks which were placed distantly apart, clustered together according to a pattern in which those from the entire country had somewhat better health and survival characteristics. Ten years later because of the variations of component 2, which mainly summarizes the information for the years lost because of the moderate and severe causes, the Thracian women were differentiated significantly. However, their distances with the other women of the entire country because of component 1 were rather small. Pomak women in their turn reduced their distances from other populations but these remained large. From 1981 onwards, all the female populations started to exhibit more similarities and the Pomak one diminished its distances with the others even further. As a result by 1991 all the female populations had very similar health and survival characteristics.

The observed variability was higher in the male populations studied, where a tripartite pattern of differentiation in health and survival characteristics is more

visible. According to this pattern the populations from the entire country had better health and survival characteristics than the others, even if they were somewhat diverse concerning component 2 levels. They were followed by the Thracian populations and then by the Pomaks. As happened with the females, in all the populations a significant change in their health and survival characteristics was observed through time and the Pomaks, originally distantly apart, finally managed to converge with the others. However, even in 1991, a significant variability was observed for the male populations and their convergence was not as strong as that of the females.

Conclusions

Significant differences were found in the demographic and health state indicators among the populations per period of study. The mortality transition was accompanied by a health transition, which led to the gradual convergence of these populations, even if it had a different momentum and pace among the populations studied. Women benefited more during this transition.

The original differences found in the Pomak population must be attributed to the natural and manmade environment, social, economic and cultural, in which they struggled for their survival. The environmental conditions are adverse in Rhodopi, especially during wintertime, and agricultural production was limited. The population lived in conditions of underdevelopment and every day activities acted as rather aggravating factors concerning health and mortality. The geographic isolation of mountainous Rhodopi was accompanied by if not non-existent at least inadequate health infrastructure and provision of health. This situation was aggravated still further because of the low economic and educational status of the population. Originally, as field work evidence suggests, some women, giving excuses based on their perception about their religious beliefs, refused to be examined by medical doctors. This practice was eventually abandoned completely, and it is well known that the presence of medical doctors in the mountainous area has positively contributed to the decline of mortality since (and quite obviously) diseases and accidents could be addressed more easily. However, women's health and also child mortality were aggravated in the past because of the absence of qualified medical personnel, i.e. obstetricians and midwives, which caused the high divergence observed in the female population in the past. Later on, this problem diminished because Pomak women started to give birth to their children in obstetric clinics in Komotini and the nearby city of Alexandropoulis.

The convergence of the Pomak population with the Thracian ones and the populations of the entire country reflect the socio-economic transition under held by them after the arsis of the geographic isolation and gradual dispersion of part of the population in the lowlands. Gradually, either in the highlands or the lowlands, they were exposed to a different and less isolated socio-economic

environment into which they were fully integrated. Their exposure to the free economy with paid work status or because of their involvement mainly with technical occupations, accompanied by cultural diffusion processes caused several economic, social and cultural transformations in the Pomak micro society and led to a rise in living standards, health provision and consequently to the improvement of health status and a reduction in mortality (see Zafeiris [21] [22].

References

1. C. L Chiang. *An Index of Health: Mathematical Models*, U.S. Department of HEW, Public Health Service, Publication No. lCXK). Series 2, No. 5, 1965.
2. A. Field. *Discovering statistics using IBM SPSS statistics*. 4th edition. SAGE, London, 2013.
3. J. Graunt. *Natural and Political Observations Made upon the Bills of Mortality*, London, First Edition, 1662; Fifth Edition, 1676.
4. H. Halley, An Estimate of the Degrees of Mortality of Mankind, Drawn from the Curious Tables of the Births and Funerals at the City of Breslau, with an Attempt to Ascertain the Price of Annuities upon Lives, *Philosophical Transactions*, 17, 596-610, 1693.
5. J. Janssen and C. H. Skiadas. Dynamic modelling of life-table data, *Applied Stochastic Models and Data Analysis*, 11, 1, 35-49, 1995.
6. C. D. Mathers, R. Sadana, J. A. Salomon, C. JL. Murray and A. D. Lopez. Healthy life expectancy in 191 countries, 1999. *The Lancet,* 357, 1685–91, 2001.
7. C. D. Mathers, C. JL. Murray, A. D. Lopez, J. A. Salomon, R. Sadana, A. Tandon, B. L. Ustün and S. Chatterj. *Estimates of healthy life expectancy for 191 countries in the year 2000: methods and results*, Global Programme on Evidence for Health Policy Discussion Paper No. 38, World Health Organization, November 2001 (revised).
8. B. S. Sanders. Measuring Community Health Levels. *American Journal of Public Health*, 54, 1063-1070, 1964.
9. D. F. Sullivan. *Conceptual Problems in Developing an Index of Health*, U.S. Department of HEW, Public Health Service Publication No. 1000, Series 2, No. 17, 1966.
10. D. F. Sullivan. A single index of mortality and morbidity. *HSMHA Health Reports*, 86, 347-354, 1971.
11. C. H. Skiadas and C. Skiadas. Estimating the Healthy Life Expectancy from the Health State Function of a Population in Connection to the Life Expectancy at Birth. In *The Health State function of a population.* (Skiadas, C. H. and Skiadas, C., eds). 1st ed. Athens, 97-109, 2012.
12. C. H. Skiadas and Skiadas, C. *The Health State function of a Population. Supplement. Demographic and Human Development indicators. 35 countries studied.* ISAST, Athens, 2013.
13. C. H. Skiadas and C. Skiadas, *The Health State Function of a Population*.2nd edition. Athens, 2013.

14. C. H. Skiadas and C. Skiadas. *Supplement: The Health State Function of a Population.* ISAST, Athens, 2013.
15. C. H. Skiadas and C. Skiadas. *Demographic and Health indicators for 193 countries of the World Health Organization and the United Nations. Second supplement of the book The Health State Function of a Population.* Athens, 2014.
16. C. H. Skiadas and C. Skiadas. The First Exit Time Theory applied to Life Table Data: the Health State Function of a Population and other Characteristics, *Communications in Statistics-Theory and Methods,* 34, 1585-1600, 2014.
17. C. H. Skiadas and C. Skiadas. The Health State Curve and the Health State Life Table: Life Expectancy and Healthy Life Expectancy Estimates. Under review, 2014.
18. C. H. Skiadas and C. Skiadas. Exploring the State of a Stochastic System via Stochastic Simulations: An Interesting Inversion Problem and the Health State Function, *Methodology and Computing in Applied Probability.* To appear, 2014.
19. G. W. Torrance Health Status Index Models: A Unified Mathematical View. *Management Science*, 22, 9, 990-1001, 1976.
20. WHO, Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (*Official Records of the World Health Organization*, no. 2, p. 100) and entered into force on 7 April 1948.
21. K. N. Zafeiris. *Comparative analysis of the biological and demographic Structures of isolated populations in the Department of Rhodopi.* PhD Thesis. Democritus University of Thrace, Department of History and Ethnology. Komotini, Laboratory of Anthropology, 2006. [In Greek].
22. K. N. Zafeiris. A method for calculating life tables using archive data. An example from mountainous Rhodopi. Paper presented in the 3rd SMTDA Conference. 11-14 June 2014, Lisbon, Portugal. Under publication.

A Secure Communication System Based on a Modified Chaotic Chua Oscillator

Mauricio Zapateiro De la Hoz[1], Leonardo Acho[2], and Yolanda Vidal[2]

[1] Universidade Tecnológica Federal do Paraná, Av. Alberto Carazzai 1640, 86300-000 Cornélio Procópio, Paraná, Brazil
(E-mail: hoz@utfpr.edu.br)

[2] Control, Dynamics and Applications Group - CoDAlab. Departament de Matemàtica Aplicada III. Universitat Politècnica de Catalunya, Comte d'Urgell 187, 08036, Barcelona, Spain
(E-mail: leonardo.acho@upc.edu, yolanda.vidal@upc.edu)

Abstract. In this paper we propose a new scheme for secure communications using a modified Chua oscillator. A modification of the oscillator is proposed in order to facilitate the decryption. The communication system requires two channels for transmitting the message. One of the channels transmits a chaotic signal generated by the oscillator and is used for synchronization. The second channel transmits the message encrypted by a nonlinear function. This function is built in terms of one of the chaotic signals, different from that sent on the first channel. In the receiver side, a synchronizer reconstructs the chaotic oscillator signals, one of which is used for the decryption of the message. The synchronization system is designed via Lyapunov theory and chaoticity proves via Poincaré maps and Lyapunov exponents will be provided in order to demonstrate the feasibility of our system. Numerical simulations will be used to evaluate the performance of the system.

Keywords: Chaos, Secure communication, Chua oscillator.

1 Introduction

The possibility to synchronize two coupled chaotic systems has allowed the development of a variety of communication schemes based on chaotic systems. A wide variety of synchronization schemes have been developed since Pecora and Carroll [5], among others, showed it was possible to do so. In this way the use of signals generated by chaotic systems as carriers for analog and digital communications aroused great interest as a potential means for secure communications [1], [4], [9].

There are several works in the literature about chaotic secure communications. For instance, [8] addressed the problems of the chaos synchronization in a secure communication system when the observer matching condition is not satisfied. Zapateiro, Vidal and Acho [11] designed a chaotic communication system in which a binary signal is encrypted in the frequency of the sinusoidal term of a chaotic Duffing oscillator. Fallahi and Leung [2] developed a chaotic communication system based on multiplication modulation. Further examples can be found in [3], [10] and [12], to name a few.

In this paper, we present a new scheme to securely transmit a message using chaotic oscillators. It is based on a modification of the Chua oscillator

3rd SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal
C. H. Skiadas (Ed)

that allows for a simpler synchronization design and stability demonstration. A Poincaré map and the maximum Lyapunov exponent are presented as proofs of chaoticity of the modified oscillator. This scheme requires two channels for transmission. The encryption/decryption process is based on a modification of the scheme proposed by [13] in which a highly nonlinear function is used along with one of the chaotic signals. The advantage of the scheme is that neither the key signals nor the encrypted signals are transmitted over the channels.

The structure of this chapter is as follows. The problem statement is presented in Section 2. The details of the transmitter and receiver as well as the encryption/decryption blocks are given in Sections 3 - 6. Finally, conclusions are outlined in Section 7

2 Communication system scheme

The diagram of the proposed communication scheme is shown in Figure 1. It consists of the following elements:

1) *Chaotic oscillator*: It is a modified Chua oscillator that generates three signals (x_1, x_2, x_3), two of which are used for synchronization and encryption/decryption purposes.
2) *Encryption block*: It encrypts the message $m(t)$ using a nonlinear function $m_e(t) = \phi(x_2(t), m(t))$.
3) *Channels*: Two channels transmit the chaotic signal and the encrypted message. Channel noise $n_d(t)$ is added. In the receiver side, the signals are filtered with a bank of filters, producing signals $x_{1f}(t)$ and $m_{ef}(t)$.
4) *Synchronization block*: It retrieves the chaotic signals using only one signal from the chaotic oscillator ($x_{1f}(t)$).
5) *Decryption block*: It decrypts the message by using a nonlinear function $m_d(t) = \psi(y_2(t), m_{ef}(t))$. In this case, y_2 is the estimation of the chaotic signal x_2 generated by the synchronization block.
6) *Retrieving block*: In this stage, an algorithm is executed for deciding which message value was sent at an instant $t = t_k$, $k = 1, 2, 3, ...$

The details of the main blocks of the communication system are given in Sections 3 - 5

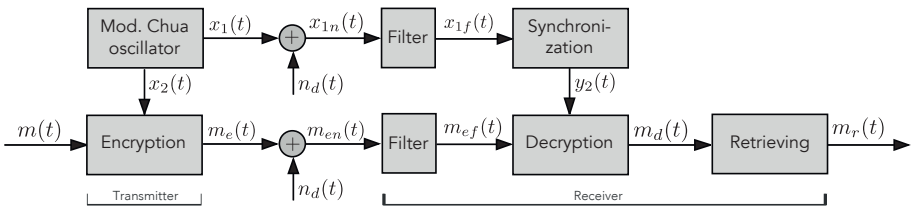


Fig. 1. Block diagram of the proposed communication system.

3 Modified Chua chaotic oscillator

The original Chua oscillator is given by the following set of equations:

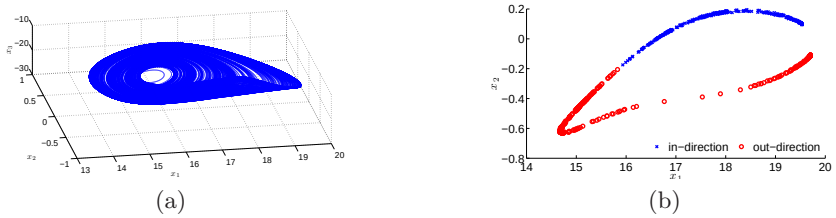
$$\dot{x}_1 = \alpha\,(x_2 - f(x_1))\,, \tag{1}$$
$$\dot{x}_2 = x_1 - x_2 + x_3, \tag{2}$$
$$\dot{x}_3 = -\beta x_2, \tag{3}$$
$$f(x_1) = m_1 x_1 + \frac{1}{2}(m_0 - m_1)(|x_1 + 1| - |x_1 - 1|). \tag{4}$$

where the overdot denotes differentiation with respect to time t; $\alpha > 0$, $\beta > 0$, m_0 and m_1 are parameters that must be chosen appropriately for obtaining chaotic behavior. In this work, we modified the original system by choosing the following characteristic function $f(x_1)$:

$$f(x_1) = -\sin x_1 \cdot e^{-0.1|x_1|}. \tag{5}$$

Note that 5 is a bounded smooth function. The system of Equations 1-3 and 5 is chaotic if $\alpha = 9.35$ and $\beta = 14.35$, as can be seen in Figure 2(a).

Figure 2(b) is the Poincaré map of the modified Chua oscillator generated when the trajectories intersect the plane $x + y + z + 1 = 0$. The map of Figure 2(b) shows the points where the trajectories intersect the plane. The two different markers show if the trajectory goes in one direction or another as it intersects the plane. The map is seen in the XY plane perspective.



(a) (b)

Fig. 2. (a) Dynamics of the modified Chua oscillator. (b) Poincaré map of the oscillator as seen in the XY plane perspective. Trajectories intersecting the plane $x + y + z + 1 = 0$.

Finally, the maximum Lyapunov exponent is calculated as another proof of chaoticity. A positive Lyapunov exponent is a strong indicator of chaos. If a system has at least one positive Lyapunov exponent, then the system is chaotic [7]. In order to determine the maximum Lyapunov exponent λ of the modified Chua oscillator, the algorithm presented in [6] was implemented in Matlab/Simulink. Figure 3 shows how λ evolves until it reaches stability. From these data, it could be found that $\lambda \approx 0.0025$ which confirms the chaoticity of the system.

4 Encryption and decryption

The encryption/decryption scheme proposed by [13] is implemented in our communication system with modified encryption/decryption functions and chaotic oscillator. In this scheme, there are two channels in order to make the synchronization process faster. The encryption/decryption process is as follows [13]:

- *Encryption*: The message $m(t)$ to be sent is encrypted by means of a nonlinear function $\phi : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ that is continuous in its first argument $x \in \mathbb{R}^3$ and satisfies the following property: for every fixed pair of $(x, m) \in \mathbb{R}^3 \times \mathbb{R}$, theres exists a unique function $\psi : \mathbb{R}^3 \to \mathbb{R}$ that is continuous in its first argument $x \in \mathbb{R}^3$ and is such that $\psi(x, \phi(x, m)) = m$. The encryption function ϕ is built in terms of the chaotic signals. The result is a signal $m_e(t)$ containing the message that is sent through one of the channels.
- *Synchronization*: A synchronization block retrieves the chaotic oscillator signals. It uses only the oscillator signal x_1 from the transmitter oscillator. This signal is sufficient to generate the signals y_1, y_2 and y_3 that are estimations of the oscillator signals x_1, x_2 and x_3, respectively. Retrieving x_2 is necessary for decrypting the message received on the second channel.
- *Decryption*: Once the oscillator signals are retrieved, the decryption function ψ can be used along with the signal $m_{ef}(t)$ in order to get the message $m(t)$.

The functions that we chose in this work to encrypt and decrypt the message are:

$$\phi : \frac{|x_2|}{x_2 + \delta} \cdot m(t) = m_e(t) \tag{6}$$

$$\psi : \frac{y_2 + \delta}{|y_2|} \cdot m_{ef}(t) = m_d(t) \tag{7}$$

where $m_d(t)$ is the decrypted message, as shown in Figure 1 and $\delta > 0$ and small compared to $|x_2|$.

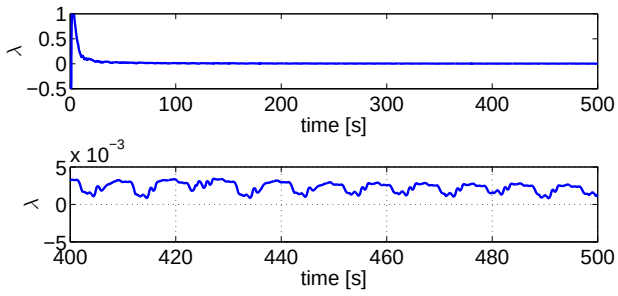


Fig. 3. Top: evolution of the maximum Lyapunov exponent. Bottom: zoom of the upper figure.

5 Synchronization

The synchronization block consists of a dynamic system that takes the signal x_1 and generates the signals y_1, y_2 and y_3 that are estimations of the oscillator signals x_1, x_2 and x_3, respectively.

Theorem 1. *Consider the modified Chua oscillator given by Equations 1 - 3 and 5 with α and β having appropriate positive values that guarantee the chaoticity of the system. Consider also a constant $\rho > 0$ such that $|x_2(t)| < \rho$. Then the system given by:*

$$\dot{y}_1 = k \cdot \mathtt{sgn}(x_1 - y_1), \tag{8}$$

$$\dot{y}_2 = y_1 - y_2 + y_3, \tag{9}$$

$$\dot{y}_3 = -\beta y_2, \tag{10}$$

where k is a design parameter such that $k > \alpha(\rho + 1)$ synchronizes with the modified Chua oscillator and thus:

i) $\lim_{t \to T_s} y_1(t) = x_1(t)$, *for a given* $T_s \in \mathbb{R}^+$.
ii) $\lim_{t \to \infty} y_2(t) = x_2(t)$.
iii) $\lim_{t \to \infty} y_3(t) = x_3(t)$.

Proof. Let the system of Equations 1 - 3 be the master and the system of Equations 8 - 10 be the slave. The function $f(x_1)$ in 5 is such that $|f(x_1)| \leq 1$, $\forall t \geq 0$. Since the system 1 - 3 is chaotic, the signal $x_2(t)$ is bounded and thus, there exists a constant $\rho > 0$ such that $|x_2(t)| \leq \rho \forall t \geq 0$. In fact, ρ depends on the initial conditions. However, assuming that $x_2(0)$ lays inside the attractor then ρ can be obtained independently of the initial conditions. The proof begins by defining the following error variable and its derivative:

$$e_1 = x_1 - y_1, \ \dot{e}_1 = \dot{x}_1 - \dot{y}_1. \tag{11}$$

Consider the terms $\dot{x}_1$ and $\dot{y}_1$ from Equations 1 and 8, respectively. Substitution of these terms into Equation 11 yields:

$$\dot{e}_1 = \alpha(x_2 - f(x_1))k - \mathtt{sgn}(x_1 - y_1). \tag{12}$$

Let $V_1 = \frac{1}{2}e_1^2$ be a Lyapunov function candidate. Then:

$$\begin{aligned}\dot{V}_1 &= e_1\dot{e}_1 = e_1\alpha x_2 - e_1\alpha f(x_1) - ke_1\mathtt{sgn}(e_1) = -k|e_1| + \alpha x_2 e_1 - \alpha f(x_1)e_1 \\ &\leq -k|e_1| + \alpha x_2 e_1 + \alpha|e_1| \leq -k|e_1| + \alpha\rho|e_1| + \alpha|e_1| \\ &= -|e_1|\left(k - \alpha(\rho + 1)\right).\end{aligned}$$

$\dot{V}_1$ will decrease and converge in finite time if and only if $k > \alpha(\rho + 1)$. Under this condition, there exists a settling time $t = T_s$ such that

$$\lim_{t \to T_s} x_1(t) = y_1(t),$$

and thus $x_1(t) = y_1(t)$, $\forall t \geq T_s$. After $t = T_s$, the synchronization system is completed with the subsystem of Equations 9 and 10. Define two new error variables e_2 and e_3 and their derivatives, as follows:

$$e_2 = x_2 - y_2,\ \dot{e}_2 = \dot{x}_2 - \dot{y}_2,$$

$$e_3 = x_3 - y_3,\ \dot{e}_3 = \dot{x}_3 - \dot{y}_3.$$

From Equations 2 and 9 we have that

$$\dot{e}_2 = x_1 - x_2 + x_3 - x_1 + y_2 - y_3 = -e_2 + e_3.$$

From Equations 3 and 10 we have that

$$\dot{e}_3 = -\beta x_2 + \beta y_2 = -\beta(x_2 - y_2) = -\beta e_2.$$

Rearrange the error variables e_2 and e_3 as a matrix system $\dot{\mathbf{e}} = \mathbf{A}\mathbf{e}$:

$$\begin{bmatrix} \dot{e}_2 \\ \dot{e}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & 1 \\ -\beta & 0 \end{bmatrix}}_{A} \begin{bmatrix} e_2 \\ e_3 \end{bmatrix}.$$

It is straightforward to show that for all $\beta > 0$, the eigenvalues of matrix $\mathbf{A}$ have negative real parts and thus:

$$\lim_{t\to\infty} y_2(t) = x_2(t),\ \text{and}\ \lim_{t\to\infty} y_3(t) = x_3(t).$$

6 Numerical results

The communication system was implemented in Matlab/Simulink. The transmitter is the implementation of Equations 1 - 3 and 5 with $\alpha = 9.35$ and $\beta = 14.35$. The receiver is the implementation of Equations 8 - 9 with $k = 1000$. The encryption and decryption functions are those of Equations 6 and 7 with $\delta = 0.01$. Noise was added to each signal and thus, a bank of filters was implemented at the input of the receiver so as to clean the signals before their processing. The message signal is assumed to be a two-valued signal that takes the values $m(t) = \{-1, +1\}$. The results to be discussed in what follows were obtained by setting the following initial conditions in the oscillator: $x_1(0) = 15$, $x_2(0) = 0$ and $x_3(0) = -15$. The initial conditions of the synchronizer were: $y_1(0) = 1$, $y_2(0) = 10$ and $y_3(0) = -1$.

Figure 4 compares the signals x_1, x_2 and x_3 of the oscillator in the transmitter side with their estimations y_1, y_2 and y_3 generated by the synchronizer. Figure 4 shows that signals x_1 and y_1 synchronize in a finite time (approximately 0.2 seconds). On the other hand, from Figure 4 we can see that the synchronization of the remaining signals takes around 5 seconds. Given that the signals $y_2(t)$ and $y_3(t)$ have an asymptotic convergence to x_2 and x_3, respectively, it could be expected that some errors might occur when retrieving the message. In order to avoid this problem, we propose sending dummy information in the beginning of the communication so as to avoid losing information.

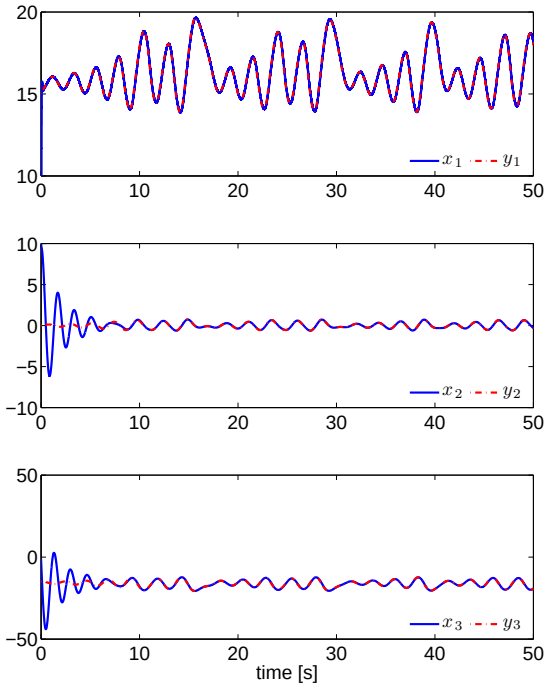


Fig. 4. Comparison of the oscillator signals and their estimations.

Figure 5 is the message that we used for the simulations. For the sake of simplicity, let us call "bit" each possible message value (+1 and −1). Thus, in this test the message $m(t)$ is sent at a rate of $T_b = 1$ bit/second. As can be seen in Figure 5, the dummy information is sent at the beginning of the transmission and afterwards, the true message is sent. Also, the message is passed through a lowpass filter in order to improve the encryption. The filter has the following transfer function:

$$H_e(s) = \frac{1}{s + 100}.$$

In this way we obtain a modified signal $M^*(s) = H_e(s)M(s)$, where $M(s) = \mathscr{L}\{m(t)\}$ and $M^*(s) = \mathscr{L}\{m^*(t)\}$. Figure 6 (top) shows the encrypted message $m_e(t)$, the signal corrupted by channel noise $m_{en}(t)$ and the filtered signal $m_{ef}(t)$. Figure 6 (bottom) shows the message sent in order to observe the differences between the original message and its encryption.

Figure 7 shows the message after the lowpass filter compared to what is obtained after the decryption, i.e. $m_d(t)$. In order to finally retrieve the message, we must determine if the bit corresponds to +1 or −1. This is done at the end of the transmission of every bit, i.e. every T_b^{-1} seconds. In this simulation, we sampled the signal $m_d(t)$ at a rate of $T_r = 0.01$ seconds. So in order to

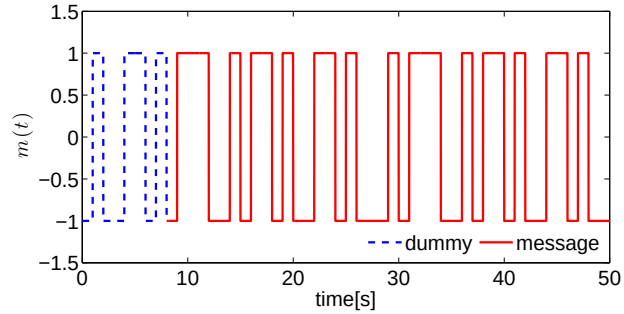


Fig. 5. Message transmitted during the test.

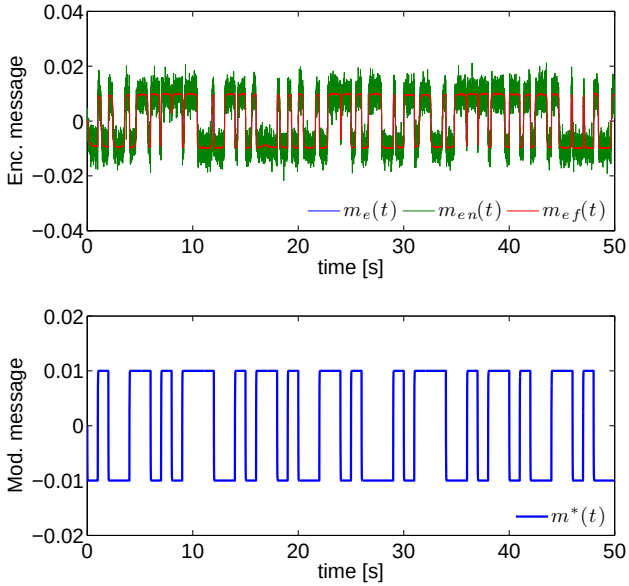


Fig. 6. Top: Encrypted message sent through the channel. Bottom: Original message (filtered).

determine the corresponding bit, we compute the sign of the sample at every instant $t = kT_b^{-1}$, $k = 1, 2, 3, ...$

Figure 8 (top) shows the result of the transmission of the message $m(t)$ which starts at $t = 8$ seconds, after a dummy message, and the retrieved message $m_r(t)$. The blue message is some dummy information sent at the beginning of the transmission in order to avoid incorrect retrieval. The true message is sent from $t = 8$ seconds. The stars in the graphic indicate the retrieved message. Figure 8 (bottom) shows the error between the original message and the retrieved message. Note that all the errors occur in the first 8 seconds of transmission of dummy information.

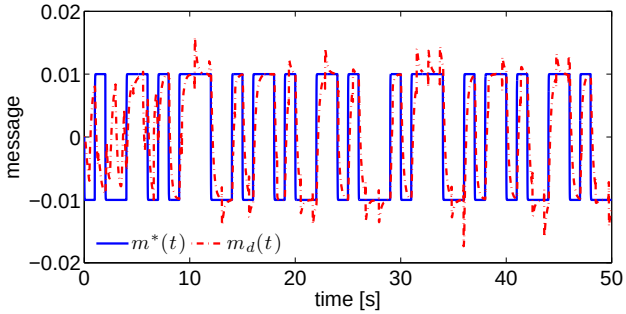


Fig. 7. Comparison between the message sent $m^*(t)$, and the decrypted message $m_d(t)$.

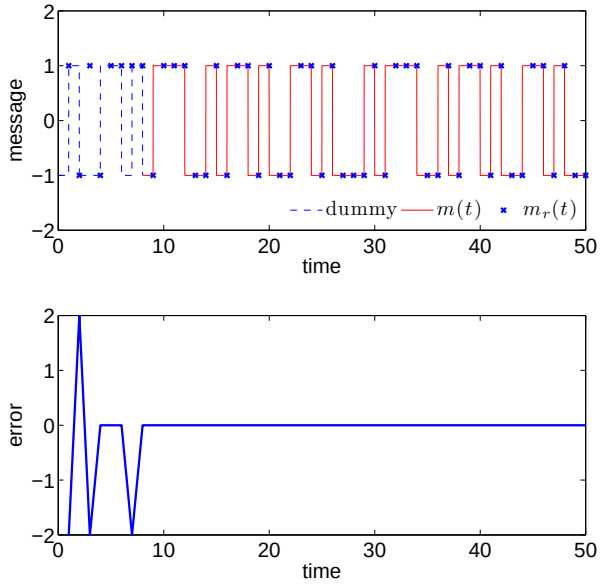


Fig. 8. Top: original and retrieved messages. Bottom: error in retrieved message.

7 Conclusion

In this paper we explored a new secure communication scheme composed of a modified Chua oscillator and an encryption/decryption scheme that makes use of nonlinear functions to encrypt the message. The oscillator characteristic function $f(x)$ was modified to make it bounded. This facilitates the synchronization because only one channel is needed and furthermore, it facilitates the demonstration of the theorem that makes possible the synchronization between the master and the slave. The encryption/decryption scheme used in this work has the advantage that the key signals and encrypted signals do not have to be transmitted over the channel and thus an increase in security is achieved.

Chaoticity proofs of the modified Chua oscillator were provided by means of a Poincaré Map and the maximum Lyapunov Exponent. The feasibility of the system was tested by numerical simulations performed in Matlab/Simulink.

Acknowledgments

Mauricio Zapateiro is supported by the fellowship from CAPES/Programa Nacional de Pos-Doutorado from Brazil. This work has been partially funded by the European Union (European Regional Development Fund) and the Spanish Ministry of Economy and Competitiveness through the research projects DPI2012-32375/FEDER and DPI2011-28033-C03-01 and by the Government of Catalonia (Spain) through 2009SGR1228.

References

1.B. Andrievsky. Adaptive synchronization methods for signal transmission on chaotic carriers. *Mathematics and Computers in Simulation* 58, **46**, 285 –293, 2002.
2.K. Fallalih, H. Leung. A chaos secure communication scheme based on multiplication modulation. *Commun. Nonlinear Sci. Numer. Simulat.* 15, pp. 368–383, 2010.
3.C. Hua, B. Yang, G.Ouyang, X. Guan. A new chaotic secure communication scheme. *Physics Letters A* 342, pp. 305–308, 2005.
4.O. Morgul, M. Feki. A chaotic masking scheme by using synchronized chaotic systems. *Physics Letters A* 251, **3**, 169 – 176, 1999.
5.L. M. Pecora, T. L. Carroll. Synchronization in chaotic systems. *Phys. Rev. Lett.* 64, 821–824, 1990.
6.J. J.Thomsen. Vibrations and Stability : Advanced Theory, Analysis, and Tools, Springer, 2003.
7.A. Wolf, J. B. Swift, H. L. Swinney, J. A. Vastano. Determining Lyapunov exponents from a time series. *Physica D* 16, pp. 285–317, 1985.
8.J. Yang, F. Zhu. Synchronization for chaotic systems and chaos-based secure communications via both reduced-order and step by step sliding mode observers. *Communications in Nonlinear Science and Numerical Simulation* 18, pp. 926–937, 2013.
9.T. Yang. A survey of chaotic secure communication systems. *Int. J. Comp. Cognition* 2, pp. 81–130, 2004.
10.T. Yang, L. O. Chua. Impulsive stabilization for control and synchronization of chaotic systems: Theory and application to secure communication. *IEEE Transaction on Circuits and Systems-I: Fundamental Theory and Applications* 44, **10**, pp. 976–988, 1997.
11.M. Zapateiro, Y. Vidal, L. Acho. A secure communication scheme based on chaotic Duffing oscillators and frequency estimation for the transmission of binary-coded messages. *Communications in Nonlinear Science and Numerical Simulation* 19, **4**, pp.991-1003, 2014.
12.X. Wang, J. Zhang. Chaotic secure communication based on nonlinear autoregressive filter with changeable parameters. *Physics Letters A* 357, pp. 323–329, 2006.
13.J. Zhon-Ping. A note on Chaotic Secure Communication Systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 49, pp. 92–96, 2002.

Strong Approximation of the Random Sums with Applications in Queuing and Risk Theories

Nadiia Zinchenko

Department of Informatics and Applied Mathematics, Nizhyn State Mukola Gogol University, Kropyv'yanskogo 2, 16600, Nizhyn, Ukraine
(e-mail: znm@univ.kiev.ua)

Abstract. We present sufficient conditions, which provide the strong approximation of the random sums, and use them to investigate certain models in the risk and queuing theories.
Keywords: Strong approximation, Invariance principle, Strong limit theorem, Random sums, Risk process, Queuing models, Law of the iterated logarithm, Risk process with stochastic premiums.

1 Introduction

Limit theorems for the random sums $D(t) = \sum_{i=1}^{N(t)} X_i$, where $\{X_i, i \geq 1\}$ are random variables (r.v.) and $N(t)$ is a counting process, became rather popular during last 20 years or so, see, for is instance, Gnedenko and Korolev[7], Whitt[15] and Silvestrov[14]. This topic is interesting not only from theoretical point of view, but also due to numerous practical applications, since mentioned random sums often appear in useful applications in queuing theory (accumulated workload input into queuing system in time interval (0,t)), in risk theory (total claim amount to insurance company up to time t), in financial mathematics (total market price change up to time t) and in certain statistical procedures. In the present work main attention is focused on the strong limit theorems for random sums. Below we consider two classes of strong limit theorem. The first class is *a strong invariance principle* (SIP), other terms are *strong approximation* or *almost sure approximation*.

We say that a random process $\{D(t), t \geq 0\}$ admits strong approximation by the random process $\{\eta(t), t \geq 0\}$ if $D(t)$ (or stochastically equivalent $D^*(t)$) can be constructed on the rich enough probability space together with $\eta(t)$ in such a way that a.s.

$$|D(t) - \eta(t)| = o(r(t)) \vee O(r(t)) \text{ as } t \to \infty, \tag{1}$$

where approximating error (error term) $r(.)$ is a non-random function.

While week invariance principle provides the convergence of distributions, the strong invariance principle describes how "small" can be the difference between trajectories of $D(t)$ and approximating process $\eta(t)$.

3^{rd} *SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal*
C. H. Skiadas (Ed)

We present some general results concerning sufficient conditions for strong approximation of random sums $D(t)$ by a Wiener or α-stable Lévy process under various conditions on the counting process N(t) and random summands $\{X_i, i \geq 1\}$. Corresponding proofs are based on the rather general theorems about the strong approximation of superposition of càd-làg processes, not obligatory connected with partial sums, Zinchenko[22]. It worth mentioning that SIP-type results itself can serve as a source of a number of limit theorems. Indeed, using (1) with appropriate error term we can easily transfer the results about the asymptotic behavior of the Wiener or α-stable Lévy process on the asymptotic behavior of random sums. Thus, the second class of limit theorems deal with the rate of growth of $D(T)$ and it's increments. As a consequence a number of limit theorems for risk processes in classical Cramer-Lundberg and renewal Sparre Andersen risk models can be obtained, particularly, strong and weak invariance principle for risk processes, diffusion and stable approximation of ruin probabilities, various modifications of the LIL and Erdös-Rényi-Csörgő-Révész-type SLLN for risk processes, which describe the rate of growth and fluctuations of mentioned processes and are useful for planning the insurance activities and reserves. The case of risk models with stochastic premiums is investigated in details.

2 SIP for superposition of the random processes

In this section we present two theorems (Zinchenko[22]), which provide strong approximation of the superposition of the random processes $X(M(t))$, when càd-làg random processes $X(t)$ and $M(t)$ themselves admit a.s. approximation by a Wiener or stable Lévy processes.

So, let $X(t)$ and $M(t)$ be independent separable real measurable càd-làg processes, $X(0) = 0$, $M(0) = 0$, $M(t)$ does not decrease with probability 1.

Theorem 1. *Suppose that there are standard Wiener processes $W_1(t)$ and $W_2(t)$, constants $m \in R^1$, $\lambda > 0$, $\tau > 0$, $\delta > 0$, for which a.s.*

$$\sup_{0 \leq t \leq T} |M(t) - \lambda t - \tau W_1(t)| = O(r(T)), \tag{2}$$

$$\sup_{0 \leq t \leq T} |X(t) - mt - \sigma W_2(t)| = O(q(T)), \tag{3}$$

where $r(t) \uparrow \infty$, $r(t)/t \downarrow 0$, $t \to \infty$, $q(t) \uparrow \infty$, $q(t)/t \downarrow 0$ as $t \to \infty$.

Let $\nu^2 = \sigma^2\lambda + m^2\tau^2$. Then $X(t)$ and $M(t)$ can be redefined on the one probability space together with a standard Wiener process $W(t)$ in such a way that a.s.

$$\sup_{0 \leq t \leq T} |X(M(t)) - m\lambda t - \nu W(t)| = O(r(T) + q(T) + \ln T). \tag{4}$$

Now let us regard a case when $X(t)$ admits a.s. approximation by α-stable process with $1 < \alpha < 2$. Condition $\alpha > 1$ is important for applications.

Theorem 2. *Suppose that $M(t)$ satisfies (2), $X(t)$ admits a.s. approximation*

$$\sup_{0\le t\le T}\big|X(t)-mt-Y_{\alpha,\beta}(t)\big|=O(q(T)), \tag{5}$$

where $Y_{\alpha,\beta}(t)$, $t\ge 0$ is α-stable process independent of $W_2(t)$, $1<\alpha<2$, $|\beta|\le 1$, $m\in R_1$. Then $M(t)$ and $X(t)$ can be redefined on the one probability space in such a way that $\forall\varepsilon>0$ a.s.

$$\Delta^*(T)=\sup_{0\le t\le T}\big|X(M(t))-(m\lambda)t-(Y_{\alpha,\beta}(t\lambda)+(m\tau)W_2(t))\big|=$$

$$=O\Big(q(T)+r(T)\Big)+o\Big(\big(r(T)+(T\ln\ln T)^{1/2}\big)^{1/(\alpha-\varepsilon)}\Big). \tag{6}$$

3 SIP for random sums

Let $\{X_i, i\ge 1\}$ be i.i.d.r.v with common distribution function (d.f.) $F(x)$, characteristic function (ch.f.) $f(u)$, $EX_1=m$, $VarX_1=\sigma^2$ if $E|X_1|^2<\infty$. Denote

$$S(t)=\sum_{i=1}^{[t]}X_i,\quad S(0)=0,\quad t>0.$$

Also suppose that $\{Z_i, i\ge 1\}$ is another sequence of i.i.d.r.v. independent of $\{X_i, i\ge 1\}$ with d.f. $F_1(x)$, ch.f. $f_1(u)$ and $EZ_1=1/\lambda>0$,

$$Z(n)=\sum_{i=1}^{n}Z_i,\quad Z(0)=0,\quad Z(x)=Z([x]),$$

and define the *renewal (counting) process* $N(t)$ associated with partial sums $Z(n)$ as

$$N(t)=\inf\{x\ge 0: Z(x)>t\}.$$

In the most interesting applications $\{Z_i\}$ are non-negative r.v. Here and in the next sections we consider *random sums (randomly stopped sums)* defined as

$$D(t)=S(N(t))=\sum_{i=1}^{N(t)}X_i,$$

where i.i.d.r.v. $\{X_i, i\ge 1\}$ and renewal process $N(t)$ are given above.

General SIP-type Theorems 1, 2 are rather convenient for investigation random sums. Really, random sum $D(t)=S(N(t))$ is a typical example of the superposition of the random processes $S(t)$ and $N(t)$, furthermore strong approximation of the partial sum processes $S(t)$ and renewal processes was rather intensively investigated since the middle of 60-th, for the wide bibliography see Csörgő and L. Horváth[5], Alex and Steinebach[1], Zinchenko[18] and more recent Bulinski and Shashkin[3], Zinchenko[22]. Concrete assumptions on summands clear up the type of approximating process and the form of error term.

When $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ have finite moments of order $p \geq 2$ both $S(t)$ and $N(t)$ admit strong approximation by a Wiener process with optimal error terms $q(t)$ and $r(t)$ presented by Csörgő and L. Horváth[5]. Denote by $\sigma^2 = VarX_1$, $\tau^2 = VarZ_1$, $\nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2$. Substituting explicit expressions for $q(t)$ and $r(t)$ in (4), we obtain following result (see also Csörgő and Horváth[5]):

Theorem 3. *(i) Let $E|X_1|^{p_1} < \infty$, $E|Z_1|^{p_2} < \infty$, $p = \min\{p_1, p_2\} > 2$, then $\{X_i\}$ and $N(t)$ can be constructed on the same probability space together with a Wiener process $\{W(t), t \geq 0\}$ in such a way that a.s.*

$$\sup_{0 \leq t \leq T} |S(N(t)) - \lambda m t - \nu W(t)| = o(T^{1/p}); \tag{7}$$

(ii) if $p = 2$ then right side of (7) is $o(T \ln \ln T)^{1/2}$; (iii) if $E \exp(uX_1) < \infty$, $E \exp(uZ_1) < \infty$ for all $u \in (0, u_o)$, then right-hand side of (7) is $O(\ln T)$.

Next suppose that $\{X_i\}$ are attracted to α-stable law with $1 < \alpha < 2$, $|\beta| \leq 1$, then approximating process for $S(t)$ is a stable process $Y_\alpha(t)$ (condition $\alpha > 1$ is needed to have a finite mean). SIP in this case was studied by Berkes *et al.*[2] in the case of symmetric stable law ($\alpha = 0$) and by Zinchenko[16] in general case with additional assumptions on ch.f. or pseudo-moments of $\{X_i, i \geq 1\}$, see also Mijnheer[13]. Below we use following

Assumption (C) : *there are $a_1 > 0, a_2 > 0$ and $l > \alpha$ such that for $|u| < a_1$*

$$|f(u) - g_{\alpha,\beta}(u)| < a_2 |u|^l, \tag{8}$$

where $f(u)$ is a ch.f. of $(X_1 - EX_1)$ if $1 < \alpha < 2$ and ch.f. of X_1 if $0 < \alpha \leq 1$, $g_{\alpha,\beta}(u)$ is a ch.f. of the stable law.

Assumption (C) not only provides normal attraction of $\{X_i, i \geq 1\}$ to the stable law $G_{\alpha,\beta}(x)$, but also leads to the rather "good" error term $q(t) = t^{1/\alpha - \varrho}, \varrho > 0$, in SIP for $S(t)$ (Zinchenko[16]). Thus, in this case random sum process $S(N(t))$ also admits a.s. approximation by α-stable process according to Theorem 2. More precise, we have

Theorem 4 (Zinchenko[21], [22]). *Let $\{X_i\}$ satisfy **(C)** with $1 < \alpha < 2$, $|\beta| \leq 1$, $EZ_1^2 < \infty$. Then $\{X_i\}$, $\{Z_i\}$, $N(t)$ can be defined together with α-stable process $Y_\alpha(t) = Y_{\alpha,\beta}(t)$, $t \geq 0$, so that a.s.*

$$\left|S(N(t)) - m\lambda t - Y_{\alpha,\beta}(\lambda t)\right| = o(t^{1/\alpha - \varrho_1}), \quad \varrho_1 \in (0, \rho_0), \tag{9}$$

for some $\varrho_0 = \varrho_0(\alpha, l) > 0$.

Corollary 1 (SIP for compound Poisson process). *Theorems 3, 4 hold if $N(t)$ is a homogeneous Poisson process with intensity $\lambda > 0$.*

4 The rate of grows of the random sums

In this section we demonstrate the possible way of application of the SIP: using SIP with appropriate error term one can easily extend the results about the asymptotic behavior of the Wiener or stable processes on the rate of growth of random sums $D(t) = S(N(t))$.

Corollary 2 (Classical LIL for random sums). *Let $\{X_i\}$ and $\{Z_i\}$ be independent sequences of i.i.d.r.v. with $EX_1 = m < \infty$, $0 < EZ_1 = 1/\lambda < \infty$, $\sigma^2 = VarX_1 < \infty$, $\tau^2 = VarZ_1 < \infty$. Then a.s.*

$$\limsup_{t\to\infty} \frac{|D(t) - m\lambda t|}{\sqrt{2t \ln\ln t}} = \nu, \quad \nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2. \tag{10}$$

Statement (10) is a straightforward consequence of the classical LIL for a Wiener process and form of error term in Theorem 3.

On the other hand, from Chung's LIL for Wiener process and Theorem 3 it easily follows

Corollary 3 (Chung's LIL for random sums). *Let $\{X_i\}$ and $\{Z_i\}$ be as in Corollary 2, then a.s.*

$$\liminf_{t\to\infty} \left(\frac{8 \ln\ln T}{\pi^2 T}\right)^{1/2} \sup_{0\le t\le T} |D(t) - m\lambda t| = \nu, \quad \nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2. \tag{11}$$

Moreover, if the stable distribution $G_{\alpha,\beta}$, $\alpha \neq 1$, is not concentrated on the half of the axe, i.e. $|\beta| \neq 1$ if $\alpha < 1$ and $|\beta| \le 1$ if $1 < \alpha < 2$, then a.s.

$$\liminf_{T\to\infty} \left(\frac{\ln\ln T}{T}\right)^{1/\alpha} \sup_{0\le t\le T} |Y_{\alpha,\beta}(t)| = C_{\alpha,\beta}, \tag{12}$$

where the constant $C_{\alpha,\beta}$ is defined with the help of so-called "I-functional of the stable process" (Donsker and Varadhan[6]). Thus, from (12) and Theorem 3 we get

Corollary 4. *Let $\{X_i, i \ge 1\}$ satisfy (**C**) with $1 < \alpha < 2$ and $\{Z_i, i \ge 1\}$ be as in Corollary 2, then a.s.*

$$\liminf_{T\to\infty} \left(\frac{\ln\ln T}{T}\right)^{1/\alpha} \sup_{0\le t\le T} |D(t) - m\lambda t| = C_{\alpha,\beta}\lambda^{1/\alpha}. \tag{13}$$

When summands $\{X_i, i \ge 1\}$ are attracted to asymmetric stable law $G_{\alpha,-1}$, we have

Corollary 5. *Let $\{X_i, i \ge 1\}$ satisfy (**C**) with $1 < \alpha < 2$, $\beta = -1$. Assume that $EZ_1^2 < \infty$. Then a.s.*

$$\limsup_{t\to\infty} \frac{D(t) - m\lambda t}{t^{1/\alpha}(B^{-1} \ln\ln t)^{1/\theta}} = \lambda^{1/\alpha}, \tag{14}$$

$$B = B(\alpha) = (\alpha - 1)\alpha^{-\theta} |\cos(\pi\alpha/2)|^{1/(\alpha-1)}, \quad \theta = \alpha/(\alpha - 1). \tag{15}$$

Proof follows from Theorem 4 and one-side LIL for the stable process $Y_{\alpha,-1}$.

Corollary 6. *Corollaries 2 – 5 hold for a compound Poisson process.*

5 How big are increments of the random sums?

When both $\{X_i\}$ and $\{Z_i\}$ have finite variance, SIP for $D(t)$ gives the possibility to extend the Erdös-Rényi-Csörgő-Révész LLN for increments of Wiener process $W(T+a_T)-W(T)$ (Csörgő and Révész[4]) on the asymptotics of $D(T+a_T)-D(T)$. Notice that additional assumptions on $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$, which determine the form of approximation term, have impact on the possible length of intervals a_T.

Theorem 5. *Let $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ be independent sequences of i.i.d.r.v., $EX_1 = m$, $varX_1 = \sigma^2$, $EZ_1 = 1/\lambda > 0$, $varZ_1 = \tau^2$,*

$$E\exp(uX_1) < \infty, \quad E\exp(uZ_1) < \infty, \tag{16}$$

as $|u| < u_0$, $u_0 > 0$, function $a_T, T \geq 0$ satisfies following conditions: (i) $0 < a_T < T$, (ii) T/a_T does not decrease in T. Also assume that

$$a_T/\ln T \to \infty \text{ as } T \to \infty. \tag{17}$$

Then a.s.

$$\limsup_{T\to\infty} \frac{|D(T+a_T) - D(T) - m\lambda a_T|}{\gamma(T)} = \nu, \tag{18}$$

where

$$\nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2, \quad \gamma(T) = \{2a_T(\ln\ln T + \ln T/a_T)\}^{1/2}.$$

Theorem 6. *Let $\{X_i, i \geq 1\}$, $\{Z_i, i \geq 1\}$ and a_T satisfy all conditions of previous Theorem 5 with following assumption used instead of (16)*

$$EX_1^{p_1} < \infty, \quad p_1 > 2, \quad EZ_1^{p_2} < \infty, \quad p_2 > 2.$$

Then (18) is true if $a_T > c_1 T^{2/p}/\ln T$ for some $c_1 > 0$, $p = \min\{p_1, p_2\}$.

When $\{X_i, i \geq 1\}$ are attracted to an asymmetric stable law, Theorem 4 and variant of Erdös-Rényi-Csörgő-Révész type law for α-stable Lévy process without positive jumps (Zinchenko[17]) yield

Theorem 7. *Suppose that $\{X_i, i \geq 1\}$ satisfy (**C**) with $1 < \alpha < 2$, $\beta = -1$, $EZ_1^2 < \infty$, $EX_1 = m$, $EZ_1 = 1/\lambda > 0$. Function a_T is non-decreasing, $0 < a_T < T$, T/a_T is also non-decreasing and provides ${d_T}^{-1} T^{1/\alpha - \varrho_2} \to 0$ for certain $\varrho_2 > 0$ determined by the error term in SIP-type Theorem 4. Then a.s.*

$$\limsup_{T\to\infty} \frac{D(T+a_T) - D(T) - m\lambda a_T}{d_T} = \lambda^{1/\alpha}, \tag{19}$$

where normalizing function $d_T = a_T^{1/\alpha}\{B^{-1}(\ln\ln T + \ln T/a_T)\}^{1/\theta}$, constants B, θ are defined in (15).

More results in this area are presented by Zinchenko and Safonova[19], Frolov[8]–[10], Martikainen and Frolov[12].

6 How small are increments of the random sums?

The answer on such question for a Wiener and partial sum processes was obtained by Csörgő and Révész[4]. For instance, they proved that for increasing $a_T > 0$ such that $(\ln(T/a_T))/(\ln\ln T) \uparrow \infty$ a.s.

$$\lim_{T\to\infty} \gamma(T, a_T) \inf_{0\le t\le T-a_T} \sup_{0\le s\le a_T} |W(t+s) - W(t)| = 1, \tag{20}$$

where

$$\gamma(T, a_T) = \left(\frac{8(\ln T/a_T + \ln\ln T)}{\pi^2 a_T}\right)^{1/2}.$$

Thus SIP for random sums with appropriate error term leads to the following statement, which holds when summands $\{X_i\}$ as well as inter-occurrence times $\{Z_i\}$ satisfy the Cramer's condition:

Corollary 7. *Assume that i.i.d.r.v.* $\{X_i, i \ge 1\}$ *and* $\{Z_i, i \ge 1\}$ *satisfy all conditions of the Theorem 5,* $\nu^2 = \lambda\sigma^2 + \lambda^3 m^2\tau^2$ *and* $a_T(\ln T)^{-3} \to \infty$ *as* $t \to \infty$*, then a.s.*

$$\lim_{T\to\infty} \gamma(T, a_T) \inf_{0\le t\le T-a_T} \sup_{0\le s\le a_T} |D(t+s) - D(t) - m\lambda a_T| = \nu. \tag{21}$$

7 Applications in queuing and risk theories

In the M/G/1 *queuing system* customers arrive according to a Poisson process $N(t)$ and ith customer requires a service time of length X_i, i.i.d.r.v. $\{X_i, i \ge 1\}$ are independent of $N(t)$. In this case the random sum process $D(t) = \sum_{i=1}^{N(t)} X_i$ is the compound Poisson process and represent the accumulated workload input in time interval (0,t]. obviously all results of the previous sections are applicable to $D(t)$ and provide SIP-type theorems (Theorems 3, 4; Corollary 1) for the accumulated workload input $D(t)$ and describe the rate of grows of $D(t)$ (Corollaries 2–6). Clearly the conditions, which provide mentioned results, are, in fact, conditions on the distributions of service times. The simplest form they have in the case of M/M/1 system. The same approach can be used for investigation the more general system G/G/1, where $N(t)$ is a renewal process. In this case conditions on inter-arrival intervals are also needed.

As the next step we consider the popular Sparre-Anderssen *collective risk model*. Within this model the risk process, which describes the evolution of reserve capital, is defined as

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i, \tag{22}$$

where: $u \ge 0$ denotes an initial capital; $c > 0$ stands for the gross premium rate; renewal (counting) process $N(t) = \inf\{n \ge 1 : \sum_{i=1}^{n} Z_i > t\}$ counts the number of claims to insurance company in time interval [0,t]; positive i.i.d.r.v. $\{Z_i, i \ge 1\}$ are time intervals between claim arrivals; positive i.i.d.r.v.$\{X_i\}$

with d.f. $F(x)$ denote claim sizes; the sequences $\{X_i, i \geq 1\}$ and $\{Z_i, i \geq 1\}$ are independent; $EX_1 = m$, $EZ_1 = 1/\lambda > 0$.

Classical Cramér-Lundberg risk model is model (22), where $N(t)$ is a homogeneous Poisson process with intensity $\lambda > 0$.

In the framework of collective risk model random sum $D(t) = \sum_{i=1}^{N(t)} X_i = S(N(t))$ can be interpreted as a total claim amount arising during time interval $[0, t]$, and increments

$$D(T + a_T) - D(T) = \sum_{i=N(T)+1}^{N(T+a_T)} X_i$$

as claim amounts during the time interval $[T, T + a_T]$.

Since process $D(t)$ is a typical example of the random sum, main results of the Sections 2 – 6 can be applied to investigation of the risk process $U(t)$. Fist of all, Theorems 3 – 4 yield the SIP-type results for $D(t)$ and $U(t)$ under various assumptions on the claim sizes $\{X_i, i \geq 1\}$ and inter-arrival times $\{Z_i, i \geq 1\}$. In the actuarial mathematics individual claim sizes are usually divided in two classes, i.e. *small* claims and *large* claims, according to the tail behavior of their distribution function $F(x)$.

Claims are called *small* if $F(x)$ is light-tailed satisfying Cramér's condition, i.e. when $M(u) = E \exp(uX_1) < \infty$ for $u \in (0, u_0)$; in opposite case, when moment generating function does not exist for any $u > 0$, the claims are called *large* ($F(x)$ is heavy-tailed). It is natural to assume that inter-arrival times Z_i have finite variance.

Thus, for small claims and $\{Z_i\}$ satisfying Cramér's condition, processes $D(t)$ and $U(t)$ admit strong approximation by a Wiener process with the error term $O(\ln t)$; for large claims with finite moments of order $p > 2$ the error term is $o(t^{1/p})$, if $p = 2$ then error term is $o((t \ln \ln t)^{1/2})$. For catastrophic events claims can be so large that their variance is infinite. In this case we assume that $\{X_i\}$ are in domain of normal attraction of asymmetric stable law $G_{\alpha,1}$ with $1 < \alpha < 2$, $\beta = 1$, and additionally satisfy condition (C). Then by Theorem 4 an approximating process for $D(t)$ is α-stable process $Y_{\alpha,1}$ with $1 < \alpha < 2$, $\beta = 1$, and risk (reserve) process $U(t)$ admits a.s. approximation by α-stable process $Y_{\alpha,-1}$, $1 < \alpha < 2$, $\beta = -1$, which has only negative jumps; the error term is presented in Theorem 4.

The form of error term in SIP is "good" enough for investigation the rate of growth of total claims and asymptotic behavior of the reserve process. Due to results of Section 4 various modifications of the LIL for $D(T)$ can be obtained almost without a proof. So, in the case of small claims or large claims (but with finite moments of order $p \geq 2$) for large t we can a.s. indicate upper/lower bounds for growth of total claim amounts $D(t)$ as $m\lambda t \pm \nu\sqrt{2t \ln \ln t}$ and for reserve capital $U(t)$ as $u + t\rho m\lambda \pm \nu\sqrt{2t \ln \ln t}$, where $\sigma^2 = VarX_1$, $\tau^2 = VarZ_1$, $\nu^2 = \lambda\sigma^2 + \lambda^3 m^2 \tau^2$, $\rho = (c - \lambda m)/\lambda m > 0$ is a safety loading.

For large claims in domain of normal attraction of asymmetric stable law $G_{\alpha,1}$ with $1 < \alpha < 2$, $\beta = 1$ (for instance, Pareto type r.v. with $1 < \alpha < 2$) Corollary 5 for large t provides a.s. upper bound for the risk process

$$U(t) \leq u + \rho m\lambda t + \lambda^{1/\alpha} t^{1/\alpha} (B^{-1} \ln \ln t)^{1/\theta}.$$

SIP-type results also help to answer on the question: how large can be fluctuations of the total claims/payments on the intervals whose length a_T increases as $T \to \infty$? Indeed, under appropriate conditions on claim size distributions and for rather "large" intervals a_T (but growing not faster then T) increments $D(T+a_T)-D(T)$ satisfy variants of Erdös-Rényi-Csörgő-Révész LLN similarly to (18) or (19).

Our general approach gives a possibility to study also more complicated risk models with stochastic premiums.

8 Strong limit theorems for the risk process with stochastic premiums

Within the **risk model with stochastic premiums** the risk process $U(t)$, $t \geq 0$, is defined as

$$U(t) = u + Q(t) = u + \Pi(t) - S(t) = u + \sum_{i=1}^{N_1(t)} y_i - \sum_{i=1}^{N(t)} x_i, \qquad (23)$$

where: $u \geq 0$ is an initial capital; point process $N(t)$ models the number of claims in the time interval $[0,t]$; positive r.v. $\{x_i : i \geq 1\}$ are claim sizes; $Ex_1 = \mu_1$; point process $N_1(t)$ is interpreted as a number of polices bought during $[0,t]$; r.v. $\{y_i : i \geq 1\}$ stand for sizes of premiums paid for corresponding polices, $Ey_1 = m_1$.

We call $U(t)$ (or $Q(t)$) the ***Cramér-Lundberg risk process with stochastic premiums*(CLSP)** if $N(t)$ and $N_1(t)$ are two independent *Poisson processes* with intensities $\lambda > 0$ and $\lambda_1 > 0$; $\{x_i\}$ and $\{y_i\}$ are two sequences of positive i.i.d.r.v. independent of the Poisson processes and of each other with d.f. $F(x)$ and $G(x)$, respectively, $\lambda_1 Ey_1 > \lambda Ex_1$.

This model, being a natural generalization of the classical Cramér -Lundberg risk model, was studied by Zinchenko and Andrusiv[20]. Korolev *et al.*[11] present an interesting example of using (23) for modeling the speculative activity of money exchange point and optimization of its profit.

Notice that process $Q(t) = \Pi(t) - S(t)$ is again a compound Poisson process with intensity $\lambda^* = \lambda + \lambda_1$ and d.f. of the jumps $G^*(x) = \frac{\lambda_1}{\lambda^*} G(x) + \frac{\lambda}{\lambda^*} F^*(x)$, where $F^*(x)$ is a d.f. of the random variable $-x_1$. In the other words

$$Q(t) = \sum_{i=1}^{N^*(t)} \xi_i, \qquad (24)$$

where $N^*(t)$ is homogeneous Poisson process with intensity $\lambda^* = \lambda + \lambda_1$ and i.i.d.r.v. ξ_i have d.f. $G^*(x)$.

Theorem 8 (SIP for CLSP, finite variance case). *(I)If in model (23) both premiums $\{y_i\}$ and claims $\{x_i\}$ have moments of order $p > 2$, then there is a standard Wiener process $\{W(t), t \geq 0\}$ such that a.s.*

$$\sup_{0 \leq t \leq T} |Q(t) - (\lambda_1 m_1 - \lambda \mu_1)t - \tilde{\sigma} W(t)| = o(T^{1/p}), \quad \tilde{\sigma}^2 = \lambda_1 m_2 + \lambda \mu_2. \quad (25)$$

(II) If premiums $\{y_i\}$ and claims $\{x_i\}$ are light-tailed with finite moment generating function in some positive neighborhood of zero, then a.s.

$$\sup_{0\le t\le T} |Q(t) - (\lambda_1 m_1 - \lambda\mu_1)t - \tilde{\sigma}W(t)| = O(\log T), \tag{26}$$

Proof immediately follows from Corollary 1 since $Q(t)$ is a compound Poisson process (see (24)) with intensity $\lambda^* = \lambda + \lambda_1$, whose jumps have mean $\frac{\tilde{a}}{\lambda^*} = \frac{\lambda_1}{\lambda^*}m_1 - \frac{\lambda}{\lambda^*}\mu_1$, and second moment $\frac{\tilde{\sigma}^2}{\lambda^*} = \frac{\lambda_1}{\lambda^*}m_2 + \frac{\lambda}{\lambda^*}\mu_2$.

Remark. In model (23) it is natural to suppose that premiums have distributions with light tails or tails which are lighter than for claim sizes. Therefore moment conditions, which determine the error term in SIP, are in fact conditions on claim sizes.

For catastrophic accidents claims can be so large that they have infinite variance, i.e. belong to the domain of attraction of a certain stable law. Thus, for Cramér-Lundberg risk process with stochastic premiums we have:

Theorem 9 (SIP for CLSP, large claims attracted to α-stable law). *Suppose that claim sizes $\{x_i\}$ satisfy **(C)** with $1 < \alpha < 2$, $\beta \in [-1, 1]$, premiums $\{y_i\}$ are i.i.d.r.v. with finite variance, then a.s.*

$$\left|Q(t) - (\lambda_1 m_1 - \lambda\mu_1)t - (\lambda + \lambda_1)^{1/\alpha} Y_{\alpha,\beta}(t)\right| = o(t^{1/\alpha - \varrho_2}), \quad \rho_2 \in (0, \rho_0), \tag{27}$$

for some $\varrho_0 = \varrho_0(\alpha, l) > 0$.

On the next step we focus on investigation the rate of growth of risk process $Q(t)$ as $t \to \infty$ and its increments $Q(t + a_t) - Q(t)$ on intervals whose length a_t grows but not faster than t.

The key moments are representation of $Q(t)$ as compound Poisson process (24) and application of the results obtained in Sections 4–6, namely, various modifications of the LIL and Erdös-Rényi-Csörgő-Révész law for random sums.

Theorem 10 (LIL for CLSP). *If in model (23) both premiums $\{y_i\}$ and claims $\{x_i\}$ have moments of order $p > 2$, then*

$$\limsup_{t\to\infty} \frac{|Q(t) - \tilde{a}t|}{\sqrt{2t \ln\ln t}} = \tilde{\sigma}, \text{ where } \tilde{a} = \lambda_1 m_1 - \lambda\mu_1, \quad \tilde{\sigma}^2 = \lambda_1 m_2 + \lambda\mu_2.$$

Notice that Theorem 10 covers not only the case of small claims, but also the case of large claims with finite moments of order $p > 2$.

Next result deals with the case of large claims with infinite variance. More precise, we shall consider the case when r.v. $\{x_i, i \ge 1\}$ in CLSP-model (23) are attracted to an asymmetric stable law $G_{\alpha,1}$, but premiums have $Ey_1^2 < \infty$.

Theorem 11. *Let $\{x_i, i \ge 1\}$ satisfy condition **(C)** with $1 < \alpha < 2$, $\beta = 1$ and $Ey_1^2 < \infty$. Then a.s.*

$$\limsup_{t\to\infty} \frac{Q(t) - (\lambda_1 m_1 - \lambda\mu_1)t}{t^{1/\alpha}(B^{-1}\ln\ln t)^{1/\theta}} = (\lambda + \lambda_1)^{1/\alpha},$$

where $B = B(\alpha) = (\alpha - 1)\alpha^{-\theta}|\cos(\pi\alpha/2)|^{1/(\alpha-1)}$, $\quad \theta = \alpha/(\alpha - 1)$.

Next theorem clarify the asymptotics of increments of the risk process with stochastic premiums and present the Erdös-Rényi-Csörgő-Révész type law for $Q(t)$.

Theorem 12 (Small claims). *Let in CLSP-model (23) claims* $\{x_i, i \geq 1\}$ *and premiums* $\{y_i, i \geq 1\}$ *be independent sequences of i.i.d.r.v. with* $Ex_1 = m$, $Varx_1 = \sigma^2$, $Ey_1 = 1/\lambda > 0$, $Vary_1 = \tau^2$, *and finite moment generating functions*

$$E\exp(ux_1) < \infty, \quad E\exp(uy_1) < \infty \text{ as } |u| < u_0, \quad u_0 > 0.$$

Assume that non-decreasing function a_T, $T \geq 0$, *satisfies following conditions: (i)* $0 < a_T < T$, *(ii)* T/a_T *does not decrease in* T. *Also let*

$$a_T/\ln T \to \infty \text{ as } T \to \infty.$$

Then a.s.

$$\limsup_{T\to\infty} \frac{|Q(T + a_T) - Q(T) - a_T(\lambda_1 m_1 - \lambda\mu_1)|}{\gamma(T)} = \tilde{\sigma},$$

where

$$\gamma(T) = \{2a_T(\ln\ln T + \ln T/a_T)\}^{1/2}, \quad \tilde{\sigma}^2 = \lambda_1 m_2 + \lambda\mu_2.$$

Remark. General Sip-type theorems give also the possibility to investigate more general cases when $\{y_i\}$ and $\{x_i\}$ are sequences of dependent r.v., for example, associated or weakly dependent, $N(t)$ and $N_1(t)$ can be renewal processes, Cox processes, ets.

References

1. M. Alex, J. Steinebach. Invariance principles for renewal processes and some applications. *Teor. Jmovirn. Mat. Stat.* 50, 22–54, 1994.
2. I. Berkes, H. Dehling, D. Dobrovski, W. Philipp. A strong approximation theorem for sums of random vectors in domain of attraction to a stable law. *Acta Math. Hung.*, 48, 161–172, 1986.
3. A.V. Bulinski, A.P. Shashkin. *Limit theorems for associated random fields and related systems*, Fizmatlit, Moscow, 2008.
4. M. Csörgő, P. Révész. *Strong Approximation in Probability and Statistics*, Acad. Press, New York, 1981.
5. M. Csörgő, L. Horváth. *Weighted Approximation in Probability and Statistics*, Wiley, New York, 1993.
6. M. D. Donsker, S. R. Varadhan. On LIL for local times.*Communs. Pure and Appl. Math.* 30, 707–753, 1977.
7. B.V. Gnedenko, V.Yu. Korolev. *Random Summation: Limit Theorems and Applications*, CRT Press, Boca - Raton, Florida, 1996.
8. A. Frolov. On Erdös-Rényi Laws for renewal processes. *Teor. Jmovirn. Mat. Stat.* 68, 164–173, 2003.
9. A. Frolov. Limit theorems for increments of sums of independent random variables. *Journal of Math. Sciences* 128, 1, 2604–2613, 2006.

10. A. Frolov. Strong limit theorems for increments of compound renewal processes. *Journal of Math. Sciences* 152, 6, 944–957, 2008.
11. V. Korolev, V. Bening, S. Shorgin. *Mathematical Foundations of Risk Theory*, Fizmatlit, Moscow, 2007.
12. A. Martikainen, A. Frolov. On the Chung law for compound renewal processes. *Journal of Math. Sciences*, 145, N 2, 4866–4870, 2007.
13. J. Mijnheer. Limit theorems for sums of independent random variables in domain of attraction of a stable law: a survey. *Teor. Jmovirn. Mat. Stat.* 53, 109–115, 1995.
14. D. Silvestrov. *Limit Theorems for Randomly Stopped Stochastic Processes*, Springer-Verlag, London, 2004.
15. W. Whitt. *Stochastic-Processes Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, Springer-Verlag, New York, 2002.
16. N. Zinchenko. The strong invariance principle for sums of random variables in the domain of attraction of a stable law. *Teor. Verojatnost. i Primenen.* 30, 131–135, 1985.
17. N. Zinchenko. Asymptotics of increments of the stable processes with the jumps of one sign. *Teor. Veroyatnost. i Primenen.* 32, 793–796, 1987.
18. N. Zinchenko. The Skorokhod representation and strong invariance principle. *Teor. Jmovirn. Mat. Stat.* 60, 51–63, 2000.
19. N. Zinchenko, M. Safonova. Erdös-Renyi type law for random sums with applications to claim amount process. *Journal of Numerical and Applied Mathematics* 1(96), 246–264, 2008.
20. N. Zinchenko, A. Andrusiv. Risk processes with stochastic premiums. *Theory of Stoch. Processes* 14, no 3–4 , 189–208, 2008.
21. N. Zinchenko. Strong invariance principle for a superposition of random processes. *Theory of Stoch. Processes* 16(32), 130–138, 2010.
22. N. Zinchenko. Almost sure approximation of the superposition of the random processes. *Methodology and Computing in Applied Probability*, DOI 10.1007/s 11009-013-9350-y, 2013.