WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Text Entities' Metrics

Vol. 3
No. 1
Spring
2008

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

Contents

Page

Editors' Note – JAQM 2007 Awards 1

Text Entities' Metrics

Adina Florenta GIUCLEA, Ciprian Costin POPESCU, Marius GIUCLEA
Metrics on Entities Spaces 3

Cosmin TOMOZEI
Hypertext Entities' Semantic Web-Oriented Reengineering 9

Ion IVAN, Daniel MILODIN, Sorin-Nicolae DUMITRU, Dragos PALAGHITA
C# Entities Quality Analysis 20

Quantitative Methods Inquires

Alexandru ISAIC-MANIU, Viorel Gh. VODĂ
Some Notes on the Logistic Distribution 32

Ali ÜNLÜ, Michael D. KICKMEIER-RUST
Implementation and Applications of a Three-Round User Strategy for Improved Principal Axis Minimization 44

Laura OBREJA BRASOVEANU, Victor DRAGOTA, Delia CATARAMA, Andreea SEMENESCU
Correlations Between Capital Market Development and Economic Growth: the Case of Romania 64

Siu Loon HOE
Issues and Procedures in Adopting Structural Equation Modeling Technique 76

Comments on Software

Dumitru TODOROI
Adaptable Software Shells Versus Microsoft Software Shells 84

Iulia SURUGIU, Manole VELICANU
Academic Research Integration System 97

Anniversary Moment

Gheorghe NOSCA
Professor Marcel Dragos Stoica at 75th Anniversary 106

Contents

Page

Book Reviews

Daniel Traian PELE
Book Review on Econometrics ("Econometrie") by Tudorel ANDREI and Régis BOURBONNAIS, Editura Economica, Bucharest, 2008 — 108

Claudiu HERTELIU
Critical Observations upon the Methodological Aspects in the "China Study" Volume — 110

Gheorghe NOSCA
Book Review on Subtle Sets and their Applications ("Multimi Subtile si Aplicatiile lor ") by Marcel STOICA, Marin ANDREICA, Dan NICOLAE and Romulus ANDREICA, Cibernetica MC Publishing House, Bucharest, 2008 — 114

Editors' Note – JAQM 2007 Awards Announcement

Ion IVAN
PhD, University Professor, Department of Economic Informatics
Faculty of Cybernetics, Statistics and Economic Informatics
University of Economics, Bucharest, Romania

E-mail: ionivan@ase.ro ; **Web Page:** http://www.ionivan.ro

Claudiu HERTELIU
PhD, Assistant Professor, Department of Statistics and Econometrics,
University of Economics, Bucharest, Romania

E-mail: claudiu.herteliu@gmail.com, **Web page:** http://www.hertz.ase.ro

Gheorghe NOSCA
PhD, Association for Development through Science and Education, Bucharest, Romania

E-mail: r_g-nosca@yahoo.com

*We are happy to announce the winners of JAQM 2007 Awards. We are grateful to **Prof. dr. Alexandru ISAIC-MANIU** who had coordinated the nomination committee.*

After deliberations, the winners are:

1st Category

For his very important contribution in promoting and developing Quantitative Methods in Academia, as a recognized scholar, we grant to:

Emilian DOBRESCU

from **Romanian Academy of Science, Romania**

the ***2007 JAQM Medal of Honor***

2nd Category

For the most valuable Quantitative Method related, single author paper published in JAQM, we grant to:

Mehmet MENDEŞ

from **Çanakkale Onsekiz Mart University, Turkey**

the **2007 JAQM Special Award**

3rd Category

For the most promising young researcher in Quantitative Methods area, we grant to:

Anatolie BARACTARI

from **Academy of Economic Studies, Kishinev, Moldova**

the ***2007 JAQM Distinction***

JOURNAL OF APPLIED QUANTITATIVE METHODS

METRICS ON ENTITIES SPACES

Adina Florenta GIUCLEA
"Dimitrie Leonida" Technical College, Bucharest, Romania

E-mail: agiuclea@yahoo.com

Ciprian Costin POPESCU
PhD, University Assistant
Department of Mathematics, University of Economics, Bucharest, Romania
Calea Dorobantilor 15-17 Street, Sector 1, Bucharest, Romania

E-mail: cippx@yahoo.com

Marius GIUCLEA
PhD, Assistant Professor
Department of Mathematics, University of Economics, Bucharest, Romania
Calea Dorobantilor 15-17 Street, Sector 1, Bucharest, Romania

E-mail: mgiuclea@yahoo.com

Abstract: *In this paper, we deal with concept of "metric". At first, we briefly discuss some issues regarding this very shaded notion in human knowledge. Secondly, we emphasize its usefulness in Mathematics, particularly in the relatively recent field of fuzzy models.*

Key words: *algebraic structure; vectorial space; metric; norm; fuzzy space; fuzzy number; fuzzy random variable*

1. Introduction

"Metric" is one of the words with great spreading and equally many senses, depending on the scope we think about. For instance, in computer networking, it characterizes a way (or a route), while in general relativity theory it describes the spacetime complex, in software it's one important tool for pointing out certain characteristics. In connection with this last assertion, an interesting investigation regarding text entities and adjacent evaluation algorithms (among the text characteristics, "orthogonality" has received great attention) are presented in [9][1]. However, in this paper, we focus our attention on some categories of metrics and their utility in the field of mathematics. In fact, it's the domain in which this concept proves its true value. It is sufficient to enumerate geometry, algebraic theory and fuzzy systems.

2. Metric spaces

Definition 2.1. [11]

Consider a set $X \neq \phi$.

A function $d : X \times X \to R$ is called metric or distance if:

i) $d(x, y) = d(y, x), \forall x, y \in X$;

ii) $d(x, y) \geq 0, \forall x, y \in X$ and

$d(x, y) = 0 \Leftrightarrow x = y$;

iii) $d(x, y) \leq d(x, z) + d(z, y), \forall x, y, z \in X$.

Thus the ordered pair (X, d) is a metric space.

Definition 2.2. [7]

A group is an algebraic structure $(G, \circ), G \neq \phi$, which fulfill the conditions:

i) closure: $a \circ b \in G, \forall a, b \in G$;

ii) "$\circ$" is associative: $a \circ (b \circ c) = (a \circ b) \circ c, \forall a, b, c \in G$;

iii) identity element: $\exists e \in G$ such that $a \circ e = e \circ a = a, \forall a \in G$;

iv) inverse element: $\forall a \in G, \exists a' \in G$ such that $a \circ a' = a' \circ a = e$.

If the commutative rule ($a \circ b = b \circ a, \forall a, b \in A$) is also satisfied, then $(G, \circ)$ is called abelian (or commutative) group.

Definition 2.3. [7]

We define a field as a set $K \neq \phi$ with two binary algebraic (symbolically writed as addition and multiplication) operations

$$+ : K \times K \to K; (a, b) \mapsto a + b$$

$$\cdot : K \times K \to K; (a, b) \mapsto a \cdot b$$

which satisfies the following requirements:

i) $(K, +)$ is abelian group with identity element called zero: $e_{(+)} = 0$;

ii) $(K, \cdot)$ is group with identity element called unity: $e_{(\cdot)} = 1$;

iii) $1 \neq 0$.

If $(K, \cdot)$ is commutative group, then $(K, +, \cdot)$ is a commutative field.

Definition 2.4. [1]

Let $X \neq \phi$ and $(K, +, \cdot)$ commutative field.

We define the following arithmetical operations:

1) $\oplus : X \times X \to X; (x, y) \mapsto x \oplus y$ (the sum of two vectors);

2) $\otimes : K \times X \to X; (a, x) \mapsto a \otimes x$ (scalar multiplication).

We say that (X, K) is a vectorial space if and only if:

i) $(K, \oplus)$ is a commutative group.

ii) $(a + b) \otimes x = (a \otimes x) \oplus (b \otimes x), \forall a, b \in K, \forall x \in X$;

iii) $a \otimes (x \oplus y) = (a \otimes x) \oplus (a \otimes y), \forall a \in K, \forall x, y \in X$;

iv) $a \otimes (b \otimes x) = (ab) \otimes x, \forall a, b \in K, \forall x \in X$;

v) $1 \otimes x = x, \forall x \in X$, where $k \in K$ and $1k = k1 = k, \forall k \in K$.

Example 2.1. [1]

A classical example of vectorial space is (R^n, R) where:

i) $R^n = \{x = (x_1, ..., x_n) \mid x_j \text{ are real numbers for all } j = \overline{1,n}\}$

ii) If $a \in R, x, y \in R^n$ then

$$x + y = (x_1 + y_1, ..., x_n + y_n)$$

and

$$ax = (ax_1, ..., ax_n).$$

Definition 2.5. [11]

Consider (X, K) a vectorial space with dimension equal to n and $(K = R) \vee (K = C)$.

A function $\|\cdot\| : X \to K$ is called norm if the following relations hold:

i) $\|x\| \geq 0, \forall x \in X$ and $\|x\| = 0 \Leftrightarrow x = 0_X$;

ii) $\|ax\| = |a| \cdot \|x\|, \forall a \in K, \forall x \in X$;

iii) $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in X$.

Example 2.2. [11]

For instance, if $x, y \in R^n$ it results that

$$\|x\| = \sqrt{x_1^2 + ... + x_n^2}$$

and

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2} .$$

3. Metrics on fuzzy spaces

Definition 3.1. [8]

A fuzzy subset of a given set X is described by a function $u(x)$ (membership degree of x in X) with $u : X \to [0,1]$. The set $F(x) = \{u \mid u : X \to [0,1]\}$ contain all fuzzy subset previously defined.

Definition 3.2. [8]

The $\alpha - \text{level set}$ of $u(x)$ is defined as

$$L_\alpha(u) = \{x \in X \mid u(x) \geq \alpha\}.$$

Definition 3.3. [8]

If X, Y are two subsets of R^n, the Hausdorff metric between X and Y is given by the formula:

$$d(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y}\|x-y\|, \sup_{y\in Y}\inf_{x\in X}\|x-y\|\right\} =$$

$$= \max\left\{\sup_{x\in X} d(x,Y), \sup_{y\in Y} d(y,X)\right\}.$$

Remark that Hausdorff metric satisfies the requirements given in Definition 2.1, namely symmetry, nonnegativity and triangle inequality. Moreover it is important to be mentioned that the space of subsets of R^n with this distance is a complete space.

It is necessary to remark that this metric plays an important role in some calculation on fuzzy random variable (at short, FRV).

Definition 3.4. [8]

It's possible to define a metric on $F(R^n)$ as follows:

$$d_\infty(u_1,u_2) = \sup_{\alpha\in[0,1]} d(L_\alpha(u_1), L_\alpha(u_2)).$$

One can prove that $(F(R^n), d_\infty)$ is a complete space.

Another way to define a such type of metric is:

$$d_p(u_1,u_2) = \left(\int_0^1 d^p(L_\alpha(u_1), L_\alpha(u_2))d\alpha\right)^{\frac{1}{p}}.$$

Remark 3.1.

An interesting development of a distance between two fuzzy numbers is presented in [5]. First, a fuzzy number x is described as a pair $(\underline{x}(r), \overline{x}(r))$ of two special functions defined on closed interval $[0,1]$. Next, the distance between x and y is measured through a metric (which admit the quality of complectness on the respective fuzzy space) given by:

$$d(x,y) = \left(\int_0^1 (\underline{x}(r) - \underline{y}(r))^2 dr + \int_0^1 (\overline{x}(r) - \overline{y}(r))^2 dr\right)^{\frac{1}{2}}.$$

This metric was useful in the process of managing a regression model with fuzzy data and real parameters.

Remark 3.2.

A special type of fuzzy numbers is LR (left-right); in such a case, we have [5]:

$$u(t) = \begin{cases} L\left(\dfrac{m-t}{a}\right), & t \le m, a > 0 \\ R\left(\dfrac{t-m}{b}\right), & t \ge m, b > 0 \end{cases}$$

where L, R are nonincreasing and nonnegative functions with $L(0)=R(0)=1$. The triangular form is a particular case of LR form, with many applications in fuzzy statistics.

In [6] two LR *f*-numbers are compared by Hukuhara difference as in the following lines.

Consider $x, y \in F_{LR}$, $x=(u_x, l_x, r_x)_{LR}$, $y=(u_y, l_y, r_y)_{LR}$, (where l and r mean the left spread and the right spread, respectively). Under assumption that $l_x \geq l_y$ and $r_x \geq r_y$, it is possible to define the Hukuhara difference between these two numbers as it follows:

$$x\Theta_H y=(u_x-u_y, l_x-l_y, r_x-r_y)_{LR}.$$

The Hukuhara difference appears in some theoretical results regarding covariance between two fuzzy random variables [6].

Remark 3.3.

As regression models, fuzzy clustering is one field in which a proper choice of a metric is very important, too. A classical model is fuzzy *c*-means [3]. Generally, it is based on searching the minimum of the following function:

$$J(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}u_{ik}^{m}\|x_k-v_i\|^2,$$

where v_i represents the "center" for the *i*'th cluster, x_k the *k*'th data, u_{ik} the membership degree of x_k in the *i*'th group of data, and m is a fuzzification exponent. In dedicated literature, many improvements of this method were published and among them are eFFCM and geFFCM, suitable for managing a large number of data [3].

4. Conclusions

At the end, beyond all doubt we can say that the "metric" is one essential tool in all kinds of (roughly speaking) measurements. It is useful for point out "distances" between somehow abstract things such as mathematical objects and is a vital question in the process of building theoretical pattern which reproduce with more rigour the complex phenomena of nature.

References

1. Cenusa, G., Raischi, C., Baz, D., Toma, M., Burlacu, V., Sacuiu, I. and Mircea, I. **Matematici pentru economisti**, Ed. Cison, 2000
2. Coppi, R., Gil, M. A. and Kiers, H. A. L. **The fuzzy approach to statistical analysis,** Computational Statistics & Data Analysis, Vol. 51, 2006
3. Hathaway, R. J. and Bezdek, J. C. **Extending fuzzy and probabilistic clustering to very large data sets,** Computational Statistics & Data Analysis, Vol. 51, 2006
4. Lopez-Diaz, M. and Ralescu, D. A. **Tools for fuzzy random variables: Embeddings and measurabilities,** Computational Statistics & Data Analysis, Vol. 51, 2006
5. Ming, M., Friedman, M. and Kandel, A. **General fuzzy least squares,** Fuzzy Sets and Systems 88, 1997
6. Nather, W. **Regression with fuzzy random data,** Computational Statistics & Data Analysis, Vol. 51, 2006

7. Nastasescu, C., Nita, C. and Vraciu, C. **Bazele algebrei,** Ed. Academiei RSR, 1986
8. Negoita, C. V. and Ralescu, D. **Simulation, Knowledge-Based Computing and Fuzzy Statistics,** Van Nostrand Reinhold, New York, 1987
9. Popa, M. **Evaluation methods of the text entities,** Journal of Applied Quantitative Methods, Vol. 1, 2006
10. Popescu, C. and Giuclea, M. **A model of multiple linear regression,** Proceedings of the Romanian Academy Series A, Vol. 8, 2007
11. Serban, R. **Algebra liniara,** Ed. Dacia Europa Nova Lugoj, 2000

[1] Codification of references:

[1]	Cenusa, G., Raischi, C., Baz, D., Toma, M., Burlacu, V., Sacuiu, I. and Mircea, I. **Matematici pentru economisti**, Ed. Cison, 2000
[2]	Coppi, R., Gil, M. A. and Kiers, H. A. L. **The fuzzy approach to statistical analysis,** Computational Statistics & Data Analysis, Vol. 51, 2006
[3]	Hathaway, R. J. and Bezdek, J. C. **Extending fuzzy and probabilistic clustering to very large data sets,** Computational Statistics & Data Analysis, Vol. 51, 2006
[4]	Lopez-Diaz, M. and Ralescu, D. A. **Tools for fuzzy random variables: Embeddings and measurabilities,** Computational Statistics & Data Analysis, Vol. 51, 2006
[5]	Ming, M., Friedman, M. and Kandel, A. **General fuzzy least squares,** Fuzzy Sets and Systems 88, 1997
[6]	Nather, W. **Regression with fuzzy random data,** Computational Statistics & Data Analysis, Vol. 51, 2006
[7]	Nastasescu, C., Nita, C. and Vraciu, C. **Bazele algebrei,** Ed. Academiei RSR, 1986
[8]	Negoita, C. V. and Ralescu, D. **Simulation, Knowledge-Based Computing and Fuzzy Statistics,** Van Nostrand Reinhold, New York, 1987
[9]	Popa, M. **Evaluation methods of the text entities,** Journal of Applied Quantitative Methods, Vol. 1, 2006
[10]	Popescu, C. and Giuclea, M. **A model of multiple linear regression,** Proceedings of the Romanian Academy Series A, Vol. 8, 2007
[11]	Serban, R. **Algebra liniara,** Ed. Dacia Europa Nova Lugoj, 2000

HYPERTEXT ENTITIES' SEMANTIC WEB-ORIENTED REENGINEERING

Cosmin TOMOZEI[1]
PhD Candidate, University Assistant
Mathematics and Computer Science Department
University of Bacau, Romania

E-mail: cosmin.tomozei@ub.ro

Abstract: *This paper's aim is to define the concept of Hypertext Semantic Web-Oriented Reengineering (HSR) as a process of distributed applications development which takes into consideration the semantic aspect in information retrieval and communication. It is virtually possible to apply the reengineering on web applications concerned being about the efficiency of the ideas of data structures and implementation than to mainly being troubled with the language or syntactic point of view. This research also brings some examples of distributed applications types, some small segments of them being mainly explained as well, in order to make our theory strongly connected with the practical work from software companies.*
It is very important that semantic approaches to be implemented while developing software applications, mostly when reengineering is integrated in the development process, as a step for the evolution to the next generation of web.

Key words: *distributed systems; distributed applications; web semantic; reengineering; web service; interoperability; distributed databases; dependable applications*

1. Semantic Reengineering

Reengineering appears as a result of the necessity to improve the quality of the results of one initial entity, such as software or text in order to realize new objectives which do not differ significantly from the old ones. Reengineering also may be formalized by a function, that may be simply called reengineering function which consists of the amount of methodologies, methods, procedures and techniques that applied to a software entity makes possible the transformation and integration of the new semantic principles.

This function depends of the modules that appear as variables or arguments initially, after that reengineering appearing as a product of the initial function, which is called development function in the initial moment (1). The development function from the stage 1 includes the function of reengineering.

$$Dev(\bigcup_{i=0}^{n-1} Mo_i) = \sum_{i=1}^{t} Obj_i = Obj \qquad (1)$$

The development process in this case appears as a function which depends of the modules of the software application. Semantic aspects are not taken into consideration that means the modules are only considered to be semantically insignificant. This point of view is very efficient and easy but reductionist, because each module of a software application is enhanced with semantic.

If we consider the semantic aspects of computer science and their implication in distributed software applications, semantic web will appear as a new paradigm that implies changes in our thinking about web development.

The product of reengineering and development function brings the possibility of having realized the new objectives. Reengineering functions have a variable number of arguments, each argument becoming a type of reengineering. If the semantic aspect has the main proportion in this function, then semantic reengineering is created or the main share of reengineering is the semantic one.

$$\mathrm{Re}\,eng(\arg_1 ...\arg_n) * Dev(\bigcup_{i=0}^{n-1} Mo_i) = \sum_i Obj_i{}^1 = Obj^1 \quad (2)$$

The above equation (2) describes synthetically the process of reengineering due to the changing of the application's objective. Each software product has one main objective that appears as a sum or reunion of smallest objectives. Changing or modifying the objective will affect the entire thinking about the existent entity.

Objective's modification may appear as a quality driven process that will imply reengineering in order to grow the quality of the software. It will be a lot easier to measure the quality of each module if decided to hold the objectives and predictive results in a table of association. We consider, for the straightforwardness' sake that each module only has one objective as a segment of the main objective. The realization is measured by a numeric scale from 1 to 100, but other ways to measure it numerically are also accepted.

Table 1. The correspondence between objectives and modules

Module / ***Objective***	Objective $_1$	Objective $_2$	………….	Objective $_n$
Module $_1$	80	0		2
Module $_2$	15	76		0
…………				
Module $_n$	5	14		55
Realization	**100**	**90**		**57**

Semantic reengineering is not a standalone process. It does not exclude the other arguments and needs them to complete the tasks. Language semantics is very important in information exchange as well as data semantics and code semantics. The main idea is to develop structural and architectural transformations with minimum efforts with the result of restructuring the entities' architecture, design and source code that will bring evolution in the meaning, in contrast with the syntactic point of view. We would like to show that syntax must subordinate itself to semantics, because *the meaning* is the most important. Syntactic modification is only a way for bringing evolution in semantics.

Operational semantics consists of rigorousness of functions, procedures and programs that implement mathematic algorithms with mathematical meaning.

Reengineering in operational semantics comes as a transformation or metamorphosis of mathematic formulae in source code.

Operational semantics is stalwartly related with the transition of a software system. In [Plotkin81][2] system transition is described as being formed by a set of configurations and a binary relation which brings the system from an initial configuration (γ) to a final one (γ^1). We consider each configuration related with a meaning which is not significantly different from the previous one. Transformation from an initial to a final phase that brings a qualitative growth in the meaning and a better realization of the objectives is defined as *semantic reengineering*.

The process of semantic reengineering is finite, deterministic, value orientated and progressive. The following elements will be more explanative [TOVA08] and [IVPOTO05]:

- *finiteness* which means that reengineering is bounded limited and will be finalized after realizing all the transformations.
- *determinism*, due to the necessity of doing a number of iterations until the process is finalized and the meaning of the software faced a new qualitative upward;
- *continuity* during the development cycle;
- *flexibility* regarding the development platforms;
- *appropriate management* which guides the development team in the process.

Denotational semantics is the approach in which the programming languages meanings are used by constructing mathematical objects that describes them. Denotational semantics is very important as well, being related with the states, in which commands are partial functions of the domains of states, denotation of data types and datta structures, such as graphs, trees or vectors.

Denotational semantics of concurrency introduced new models for concurrent computation such as the actor model or Petri nets. It is as well studied from the denotational point of view the sequentiality of programs and source to source translation. For example if we consider the following web application functions it may be possible to use semantic reengineering and denotational semantics to translate it from one language to another.

```
Protected Sub Button1_Click(ByVal sender As Object, ByVal e As  System.EventArgs) Handles Button1.Click
      Dim strcon As String = "Data Source=ASE-EY5776NZ4X6\SQLEXPRESS;Initial Catalog=seminarii;Integrated Security=True"
      Dim scoma As String = "select * from firme"
      Dim sconn As SqlConnection = New SqlConnection(strcon)
      Dim ds As DataSet = New DataSet()
      sconn.ConnectionString = strcon
      Dim sq1 As SqlCommand = New SqlCommand()
      sq1.CommandText = scoma
      Dim sdadpt As SqlDataAdapter = New SqlDataAdapter(scoma, sconn)
      sdadpt.Fill(ds, "sursa")
      For Each dr As DataRow In ds.Tables("sursa").Rows
         If Convert.ToInt32(dr("id")) = 1 Then
            TextBox1.Text = Convert.ToString(dr("id_firme"))
            TextBox1.Text + Convert.ToString(dr("den_firme"))
              End If
      Next
      MsgBox("Salut")
   End Sub
```

An experienced programmer may as well transform it for reducing the number of code lines and the keeping the same functionalities. Mathematical semantic translation may be used to keep the functions and the objectives of the web application correct and good.

```
protected void Button3_Click(object sender, EventArgs e)
  {
      string strconn =
System.Configuration.ConfigurationManager.ConnectionStrings["string_config_seminar"].ConnectionString;
      SqlConnection scon = new SqlConnection(strconn);
      string str="select * from firme";
      SqlCommand scm = new SqlCommand(str, scon);
      SqlDataAdapter sadapt = new SqlDataAdapter();
      sadapt.SelectCommand = scm;
      DataSet ods = new DataSet();
      sadapt.Fill(ods,"tabel1");
      foreach (DataRow dr in ods.Tables["tabel1"].Rows)
      {
         TextBox2.Text += dr["id_firme"];
         TextBox2.Text += dr["den_firme"];

      }
  }
```

Axiomatic semantics is an approach that provides rigorousness and correctness to computer programs and is based on mathematical logic. Operators from programming languages are deeply connected with mathematical logic. If we consider as an example XOR, Logical AND, Logical OR we will see how mathematical logic affect the way of representing reality by computer programs. Axiomatic semantics has strong associations with Hoare Logic [www1], having as central element a formalism which is called Hoare Triple {P} C {Q}, where P and Q are assertions and Q is the command. It describes how a software code is changing the state of computation.

Web orientation of software applications manages to make transformations in software code and specifications from both directions, syntactic and semantic. Web services, which allow communication to take on between software programs by SOAP [www2] use web methods as traditional software applications use normal methods. Reengineering will adapt the piece of code offering it the possibility to more general and interoperable. Transformation by reengineering produces the following piece of code as web method with the result of an XML Dataset, understood by software applications which communicate over the Internet. The clients will retrieve information from the relational database just by referring and invoking web methods.

```
 [WebMethod (Description ="Citeste tot si afiseaza xml,Cosmin Tomozei")]
public System.Data.DataSet citeste_tot()
{
        SqlConnection scon = new
        SqlConnection(ConfigurationManager.ConnectionStrings["conn1"].ConnectionString);
        scon.Open();
        string  st =  @"SELECT intalnire.dataintalnire, intalnire.ora,intalnire.durata,intalnire.linki,
        perscon.nume + ' ' + perscon.prenume AS Invitat, Institutie.deninstitutie, utilizator.prenume + ' ' +
        utilizator.nume AS Participant, tara.codtara,Localitate.denloc,Locuri.denl FROM intalnire INNER JOIN
        perscon ON intalnire.persconid = perscon.persconid INNER JOIN Institutie ON intalnire.codinstitutie =
        Institutie.codinstitutie INNER JOIN utilizator ON intalnire.utid = utilizator.utid;
  SqlDataAdapter adpt1 = new SqlDataAdapter();
   SqlCommand scmd1 = new SqlCommand(st, scon);
```

```
        System.Data.DataSet ods1 = new System.Data.DataSet();
        adpt1.SelectCommand = scmd1;
        adpt1.Fill(ods1, "tabela_noua1");
        return ods1;
    }
```

2. Semantics in Distributed Web Applications

Resource description framework (RDF) is a language which allows information to be represented and exchanged on the web. Semantic web appears as a new challenge for web developers to create applications that share data across any barriers, in a more efficient way.

RDF is based on the graph data model [www3] that uses triples formed by *subject, predicate* and *model* and represents a kind of relation between things. Efficiency appears in consequence when metadata is widely used by web applications during communication.

The syntax of this new language uses the XML syntax because is widely used in machine to machine communication over the Web. Another important aspect of the XML language is that is also human readable and understandable with very little effort. XML provides to hypertext entities more power and capacity of being machine readable.


```
HTTP/1.1 200 OK
Content-Type: application/soap+xml; charset=utf-8
Content-Length: length
<?xml version="1.0" encoding="utf-8"?>
<soap12:Envelope xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:soap12="http://www.w3.org/2003/05/soap-envelope">
  <soap12:Body>
    <citeste_tot1Response xmlns="urn:ub.ro/seminarii">
      <citeste_tot1Result>
        <xsd:schema>schema</xsd:schema>xml</citeste_tot1Result>
    </citeste_tot1Response>
  </soap12:Body>
</soap12:Envelope>
```


The result after invoking the web method consists of an XML machine to machine communication file. The file serializes the dataset returned by the web method, which contains data extracted from the distributed databases, or other information that is necessary in our development process. Below is just a piece of information regarding costumers and meetings from an SQL Server Database.


```
<NewDataSet>
<tabela_noua1 diffgr:id="tabela_noua11" msdata:rowOrder="0">
<dataintalnire>25/11/2008</dataintalnire>
<ora>12:20</ora>
        <Invitat>Cosmin Tomozei</Invitat>
        <deninstitutie>CCIT</deninstitutie>
        <Participant>Popescu Pop</Participant>
        <codtara>196</codtara>
        <denloc>Johannesbourg</denloc>
        <denl>str k 1</denl>
</tabela_noua1>
<tabela_noua1 diffgr:id="tabela_noua12" msdata:rowOrder="1">
        <dataintalnire>30/11/2008</dataintalnire>
```

<ora>12:20</ora>
<Invitat>Simona Varlan</Invitat>
<deninstitutie>CCIT</deninstitutie>
<Participant>Popescu Pop</Participant>
<codtara>196</codtara>
<denloc>Johannesbourg</denloc>
<denl>str k 1</denl>
</tabela_noua1>
</NewDataSet>

RDF has the possibility of cooperating with the relational model in distributed web applications that have relational SQL or Oracle Databases, many specialists claiming that a relational view of the Semantic Web [NEWM07] can be developed due to the efficiency, openness and robustness of the relational model. We deeply agree with that idea and consider that the relational model as well as the object oriented model can be used efficiently in Semantic Web.

SPARQL is the query language for RDF, which allows retrieval of information from the graph model and has the same relational algebra query operations as in the relational model, such as *projection*, *selection* or *join* and also RDF graphs structure may be easily transformed in relational tables as vice versa. SPARQL [NEWM07] can be seen as an extension of the relational model, as we also believe.

Administration of distributed applications [EBER06] has evolved, implying semantic management. *Ontology* has to be implemented as an important component of semantic web distributed applications. Ontology represents a set of attributes or characteristics that are in a specific domain, including the relations between them. Instances (objects), classes, assertions rules and also events are components of ontology. They can be merged, generalized and also inheritance can take place in order to create an ontology which is more specific for a particular objective of a semantic web application. Ontologies formalize concepts and relations between concepts similarly to the class diagram from the UML [EBER06]. In this case they may be seen as a source for interoperability, which is an important metric for distributed applications, and make them easier to integrate in distributed systems. Due to the formalization based on computational logics they are clear and very easy to understand, not being ambiguous and unformalized.

Taxonomy contributes as a science of classifications to create ontology hierarchies and ordinate them due to our interests and their semantics. OWL is the language. The following example points out reengineering results of transformation from LISP to OWL, both languages being semantic web related.

```
(make-class student
(agent (computer_science_student))
(is-a (value computer_science_faculty)))
(instrument  (software))
(location (Computer_Lab_Corp_C_Universitatea_Bacau)))
```

The entity *student* now can be modeled thanks to reengineering in OWL [BAKU05], showing how to transform the description of an entity from the Common LISP dialect to OWL.

```
<owl:Class rdf:about="#student">
            <rdfs:subClassOf rdf:resource="#Faculty"/>
            <owl:onProperty rdf:resource="#Computer Science"/>
<owl:Class rdf:about="#Corp_C_Universitatea_Bacau"/>
            </owl:Class>
            <owl:ObjectProperty rdf:ID="Descriere_student">
              <rdfs:domain rdf:resource="#Student" />
              <rdfs:range  rdf:resource="#Media" />
            </owl:ObjectProperty>
        <owl:ObjectProperty rdf:ID="Bursier">
              <rdfs:subPropertyOf rdf:resource="#Are_Bursa" />
              <rdfs:range rdf:resource="#Bursa" />
            </owl:ObjectProperty>
</owl:Class>
```

The reason *web services* are different in comparison with other kind of web applications is because while HTML pages or web forms from ASP.Net web applications are dedicated to human final users, web services are to be understood by remote machines which refer them over the distributed system. Companies that are dealing with advertising over the Internet or E-commerce provide their own web services to other companies or to final users or just developers to access their information. The main idea is that web services have methods that access information from different data sources and serialize it in XML. The web methods are to be invoked remotely in the distributed system, network or Internet. Web services are the perfect solution for a web developer to build easily distributed applications, having in consideration the semantic prevalence, access data in short time and low costs, productiveness being in this case the suited word for describing this activity. As semantic web is, web services are platform independent, the client having the possibility to get and process XML data independently of his type, such as hardware platform, operating system, etc.

Grid computing may involve web services as well. In [WILEY04] grid computing is defined as connecting supercomputers into metacomputers that are remotely controlled. The greed reunites remotely placed computers for serving the scientific community in a more cohesive way. However, grid computing offers the possibility to have access to diverse resources which are not available from a single machine or a single kind of platform. Grid computing and *cluster computing* use distributed software reengineering in order to realize new objectives, to add new workstations or servers into the distributed system, to make possible cooperation and collaborative work within distributed systems so as to contain heterogeneous software platforms.

Languages such as Common LISP, OWL, HTML and XML as well as object oriented languages such as VisualC#.NET, Visual C++.NET or Java being integrated in distributed web applications, frameworks and developer kits in order to resolve problems of heterogeneity and distributiveness.

This paper's aim is not only to describe technologies, concepts already well known in the distributed application development area but also to reflect semantic evolution, metrics of semantic reengineering and the importance of software reengineering in the semantic approach. The next generation of web is for sure semantically oriented. The syntax is not the final objective, but just a way in achieving semantic efficiency and clarity of the ideas. Transformation and redefinition of data structures play an important role in this scenario as well as key factors of distributed applications reengineering.

3. Distributed Web Applications Semantic Reengineering Metrics

Reengineering becomes a must in distributed applications development, due to the efficiency it brings in the development cycle. Software companies or just research laboratories do not have the physical resources or time to start projects from scratch over and over again. Even if they had, it would have been inefficient and absolutely chaotic to abandon good work which has been already done with money and time spending and not to reuse it in the following projects. Experience and expertise have already been achieved by the specific situations in which the development team worked in the past in order to realize the objectives from the past. This expertise has to be implemented in the future projects, experience being important for them as well.

In [TOVA08] we described code transformation as being an important component of software reengineering while having in consideration aspects related to security, concurrency, openness, scalability, dependability or transparency. For distributed systems, as well as for distributed applications it is essential to follow this metrics as important approaches of software quality. This metrics must convey in which proportion, the software quality characteristics are realized by software developers. First of all, as a premise for reengineering the entity that would be subjected to it must have passed quality certification. Having passed quality certification time before guarantees a good premise for a successful reengineering process.

Concurrency is one of the most important distributed application's metrics. Concurrency is defined as a property of software processes to be executed in the same time, with the possibility that each process (thread) communicates with the others. In distributed applications, users may interact simultaneously with parts or domains of the application. When talking about distributed databases, data may be accessed from many physical locations. Software applications have to consent maximum concurrency with maximum safety. The increasing of concurrency should not affect the functionalities of the application. The database resources should be accessible to a big number of users that can be considered to be infinite in formalization.

$\lim NrUsr(APP_i) = \infty$,where $NrUsr = \sum_i Usr_i$ (3)

Shall we consider the rights each user has when accessing the software application, distinctions have to be made, depending of the category in which the user belongs. In this case, the formula has to be transformed in order to reflect the user's role about the application.

$$NrUsr = \sum_{i=1}^{n}\sum_{j=1}^{m} Usr_{ij}, \quad (4)$$

where the index *j* represents the category and *i* the index of the user from the category. *NrUsr* in this case have to be maximal and considered infinite as well.

While doing reengineering to the distributed application, each category of users should considered, and attention should be focused on the rights each category of users have. Optimization of user accounts is very important in distributed applications development. Categories and rights have a semantic background for the software developers. The associations between users and rights have semantic intelligence.

The ***openness*** characteristic is very common for distributed systems and it is also compulsory for a good and maintainable distributed application. Apart from autonomous sequential software programs that cannot be modified nor may be configuration updatable during their compilation and execution, distributed systems and applications be updated during the execution nods or procedures being added, modified or deleted. A distributed application may change it's configuration during the lifecycle. The indicator ***Icop*** (openness indicator) reflects the number of actualization procedures during a period of time. It is advisable to compare these indicators from each two periods of time t_i, t_{i+1} and to build time series with them.

$$Icop = \sum_{i=1}^{n} Ins_i + \sum_{j=1}^{t} Upd_j + \sum_{k=1}^{r} Del_k \qquad (5)$$

If it is desired to make comparisons between the numbers of actualizations from two periods of time it is very simple and efficient just forming a ratio index.

$$Incop = \frac{Icop_1}{Icop_0} \qquad (6)$$

The openness of the application is in direct proportion with the ***Icop*** indicator. Every chain of transformations and modifications creates the reengineering of the distributed application. It is very important to have an open distributed application in order to make reengineering, but what is the most important factor in building reliable distributed applications is that the system works while transformations are being made.

Openness is important for the semantic transformation. An open application can be subjected to software reengineering easily by adding new components with new meanings and objectives.

Transparency and ***fault tolerance*** means that the user must not be restricted to use the application and the main functions of the system must not be affected if there are some components that do not work. Transparency means that from the user's point of view, the software application appears as a single non divided unit which helps him, without showing it's distributiveness. It is not important as well as not recommended to show the users the fragments or clusters of the software system. It will not be accessible to see the fragment or the cluster from which the information brought by the query resides, whether is ***replicated*** or subjected to ***concurrency*** while using it. Committing or finalizing transactions in distributed application is also done independently and transparently. In this case, reengineering may be more difficult in comparison with the traditional, non distributed applications because the development team must see the image behind the transparent image, as the back-side one.

The formalism that represents a grid from a distributed software application may be referred as :

$$Grid_i = \bigcup_{j=0}^{n} Frag_{ij} \qquad (7)$$

as a reunion between the j indexed fragments of software or database fragments from the network. If we have replication in our distributed system, we must put one condition that will optimize the access to one replica and that the fragments should be distinct. We must evaluate each fragment or replica and see if there are any updates and modifications.

The whole image of the application will appear transparently and can be described by the following formalism as ***TrImage*** or transparent image.

$$Trimage = \bigcup_{i=0}^{m} \bigcup_{j=0}^{n} Frag_{ij} = \bigcup_{i=0}^{m} Grid_i \qquad (8)$$

Transparency makes possible semantic constancy not depending of the deficient functionality of some modules of the system. Architectural reengineering may be also realized without notifying the users voluntarily or involuntarily by malfunctions or stops in the system's operation.

4. Conclusions

This paper described software reengineering in distributed applications on the subject of the hypertext entities semantic perspective. Semantic reengineering is very important and may be as well considered as the main component of software reengineering. Particular metrics have been shown regarding semantic reengineering and distributed applications. In addition to the text entities reengineering, development of distributed applications presumes technological implementations and platform dependence. However, the openness of distributed systems involves autonomy and context independence.

Semantic reengineering appears as an argument of the reengineering function, this concept being implemented and developed by the author.

As the other types of reengineering, semantic reengineering is also related with the renewing of the application's objectives, each objective being quantitatively measured as well as about the quality of the results.

References

1. Barzdins, G., Gruzitis, N. and Kudins, R. **Re-engineering OntoSem Ontology Towards OWL DL Compliance, Knowledge-Based Software Engineering,** in "Proceedings of the Seventh Joint Conference on Knowledge-Based Software Engineering"
2. Eberhart, A. **Semantic Management of Distributed Web Applications,** IEEE DISTRIBUTED SYSTEMS ONLINE, 1541-4922, 2006, IEEE Computer Society Vol. 7, No. 5; May 2006
3. Ivan, I., Popa, M. and Tomozei, C. **Reingineria Entitatilor Text,** Revista Romana de Informatica si Automatica, vol. 15, no. 2, 2005
4. Newman, A. **A Relational View of the Semantic Web,** XML.COM, March 14, 2007
5. Plotikin, G. D. **A Structural Approach to Operational Semantics,** Computer Science Dept., Aarhus University, 1981
6. Tomozei, C. and Varlan, S. **Distributed Applications Reengineering Metrics,** Studii si Cercetari Stiintifice, no. 17, Seria Matematica, Universitatea din Bacau
7. Reilly, E. D. **Concise Encyclopedia of Computer Science,** Wiley, 2004
8. http://www.w3.org/TR/owl-features/ 07.03.2008
9. http://www.w3.org/2001/sw/ 07.03.2008
10. http://dsonline.computer.org/portal/site/dsonline/index.jsp

---

[1] Cosmin Tomozei is Universtiy Assistant at Mathematics and Computer Science Department from Faculty of Sciences of the University of Bacau. He PhD candidate from October 2007 at Economic Informatics Department from University of Economics, Bucharest. He holds an Master in Science - Databases- Business Support from University of Economics, Bucharest. He graduated in Economic Informatics at Faculty of Economic Cybernetics, Statistics and Informatics in 2006. His main research areas are: - object oriented programming; - functional programming; - software reengineering; - distributed applications.

[2] Codification of references:

[BAKU05]	Barzdins, G., Gruzitis, N. and Kudins, R. **Re-engineering OntoSem Ontology Towards OWL DL Compliance, Knowledge-Based Software Engineering,** in "Proceedings of the Seventh Joint Conference on Knowledge-Based Software Engineering"
[EBER06]	Eberhart, A. **Semantic Management of Distributed Web Applications,** IEEE DISTRIBUTED SYSTEMS ONLINE, 1541-4922, 2006, IEEE Computer Society Vol. 7, No. 5; May 2006

[IVPOTO05]	Ivan, I., Popa, M. and Tomozei, C. **Reingineria Entitatilor Text,** Revista Romana de Informatica si Automatica, vol. 15, no. 2, 2005
[NEWM07]	Newman, A. **A Relational View of the Semantic Web,** XML.COM, March 14, 2007
[PLOTKIN81]	Plotikin, G. D. **A Structural Approach to Operational Semantics,** Computer Science Dept., Aarhus University, 1981
[TOVA08]	Tomozei, C. and Varlan, S. **Distributed Applications Reengineering Metrics,** Studii si Cercetari Stiintifice, no. 17, Seria Matematica, Universitatea din Bacau
[WILEY04]	Reilly, E. D. **Concise Encyclopedia of Computer Science,** Wiley, 2004
[www1]	http://www.w3.org/TR/owl-features/ 07.03.2008
[www2]	http://www.w3.org/2001/sw/ 07.03.2008
[www3]	http://dsonline.computer.org/portal/site/dsonline/index.jsp

C# ENTITIES QUALITY ANALYSIS

Ion IVAN[1]

PhD, University Professor, Department of Economic Informatics
University of Economics, Bucharest, Romania
Author of more than 25 books and over 75 journal articles in the field of software quality management, software metrics and informatics audit. His work focuses on the analysis of quality of software applications.
E-mail: ionivan@ase.ro , **Web page:** http://www.ionivan.ro

Daniel MILODIN[2]

University of Economics, Bucharest, Romania
He graduated "The Informatized Project Management" Master Program

E-mail: daniel.milodin@ase.ro

Sorin-Nicolae DUMITRU[3]

University of Economics, Bucharest, Romania

E-mail: sorin.dumitru@yahoo.com

Dragos PALAGHITA[4]

4th year student, University of Economics, Bucharest, Romania

E-mail: dpalaghita@gmail.com

Abstract: *In this paper we will present the requirements for C# batch program development. Also there is constituted a quality analysis system of indicators for C# programs.*
The software structure used to analyze the C# program batch and computation of regarded indicators is presented. C# program classes are constructed and the belonginess criterion of a C# program to a particular class is established through an aggregation process.

Key words: *C#; quality indicators; C# entities; quality analysis*

1. C# program batch definition

In order to analyze C# programs it is necessary to build a representative program batch.

The size of the batch, expressed as number of programs, is based on probability, which guarantees that the results given by the programs in the batch are correct and represent the type of programs of which they are a part of, based on the result dispersion and on the error margin.

For a 95% probability, a variance of 0.25 and a maximum error of 3% the batch must contain 1067 programs. The result is based on the formula used to determine sample volume, in the case of extraction with return [ISAIC99][5]:

$$n = \frac{z^2\sigma^2}{\Delta_x^2}$$

where:

z^2 – the Gauss-Laplace repartition quantum for the requested probability level (for P-95%, z=1,96);

σ^2 - the population variance;

Δ_x^2 - the error margin.

The programs must be characterized by:

- homogeneity, the program length must not differ significantly;
- must be developed using C#;
- must be syntactically correct;
- the degree of difficulty must be similar;
- must not contain calls to extern libraries or modules.

To obtain the batch the following procedure was applied:

A collectivity of 50 programmers is considered.

They have significantly similar levels of experience and qualification.

The proposed issues to be solved have restrictions regarding:

- the type of the procedures;
- the nature of obtained results;
- the data structures used;
- the complexity of the software product;
- type of the results returned by the procedures;
- the parameter number for each procedure;
- the types of parameters used by each procedure;
- the complexity of each procedure;
- the number and type of variables that may be used inside the procedures.

The programmers have tested the programs on 100 data sets.

The associated tree like structure was elaborated, and the tests were meant to cover the largest possible scenario number.

The programmers made a series of recordings regarding:

- the time spent coding the program;
- number of errors;
- number of runs;
- running time;
- number of runs with errors;
- number of runs which produced correct results;
- number of runs which produced flawed results,

which they freely declared.

2. Program quality indicator system

There are numerous indicators. The ones for which data is automatically collected and are automatically computed, having input data a number of files which contain in fact programs are called operational, like in figure 1:

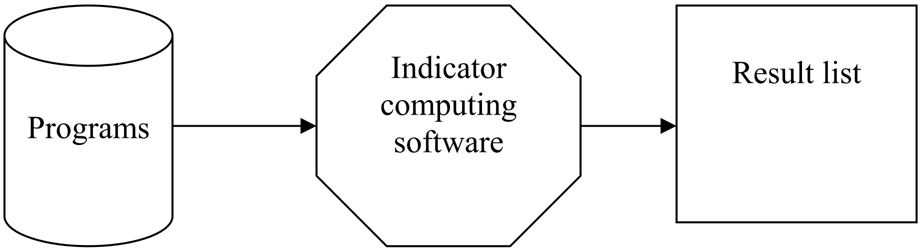


Figure1. Automatic indicator computing.

For program P_1 in which over 80% of the components represent routine activity, having 10000 lines of code, the following activities have been unrolled:

- specification elaboration: 2 days;
- product analysis: 2 days;
- product design: 2 days;
- product implementation: 8 days;
- product testing: 3 days;
- product deployment: 3 days,

resulting in a total of 20 days.

This indicator is influenced by:

- personnel experience, because an experienced personnel will foresee implementation errors, and will decrease the needed time period to implement the software product;
- the complexity, which leads to increasing the time allocated to create the software product; medium complexity programs are preferred, as they do not solicit personnel and computing resources;
- personnel qualification influences the quality and the time needed to complete the development of the software product;
- reused components greatly reduce the time allocated to developing the software product by obtaining them using program libraries;
- the routine of unrolled activities affect personnel attention to details; the reallocation of personnel on different types of programs is preferred such that the state of indifference is avoided.

Source code lines and the length of the file are indicators that apply to programs that have a large number of source code lines.

Complexity is a quality indicator for software products; it establishes the program level of difficulty, expressed as number of source code lines, number of instructions used, and number of repetitive cycles.

The operands n_1 and n_2 are considered. The complexity is determined using the formula:

$$C = n_1 \lg n_1 + n_2 \lg n_2$$

There is software that automatically receives programs sources and identifies the operators, +, -, *, /, %, <, >, !, &&, ||, <<, >>, ==, it counts them and computes their complexity.

Time of completion represents the number of days or hours needed to go through the stages of the software development cycle.

Program length is given as a number of source code lines.

The programmers are instructed to realize the instruction alignment in a convenient way.

The coding of the programs is made depending on the personal way of aligning instructions, specific to each programmer. The C++ language does not enforce any restrictions regarding the arrangement of instructions in the page.

The following instructions are considered:

```
int a,b,c;
a = b + c;
```

There are several ways of writing these instructions:

- all instructions on one line: *int a,b,c; a = b + c;*
- each operand and operator in a new line:

```
int
a,
b,
c;
a
=
b
+
c;
```

- several combinations that imply the methods stated above.

In a structural way, aligning instructions on source code lines must respect several rules:

- a procedure is called with its parameters on one line of code;
- the repetitive cycles are written on several lines of code to clearly identify the running conditions, the operations inside the cycle and the exit conditions from the respective cycle;
- simple instructions are written on one source line;
- the simple conditional instructions are written on several lines to determine the condition to respect, the body that is executed if the condition is true, and the block that is executed if the condition is false;

- the alignment must be done in such a way that the program will be well structured.

To obtain a better view of the program or program block content, each new instruction must be tabbed right compared to the previous instruction if it depends on it.

The procedure to determine the maximum of three integer numbers is considered. It is written in several different readings considering instruction alignment.

Reading R_1

```
procedure maxim(int a, int b, int c)
{
int nr;
if (a>b) nr:= a;
else nr:= b;
if (nr<c) nr:=c;
printf("%d", nr);
}
```

Reading R_2

```
procedure maxim
(int a, int b, int c)
{
int nr;
if (a>b)
 nr:= a;
else
nr:= b;
if (nr<c)
nr:=c;
printf("%d", nr);
}
```

Reading R_3

```
procedure
maxim
(int a, int b, int c)
{
int nr;
if
(a>b)
 nr:= a;
else
 nr:= b;
if
 (nr<c)
nr:=c;
printf("%d", nr);
}
```

In table 1, centralization is made for the number of specific source lines of each reading of the procedure:

Table 1. Procedure sizing

Reading	Size
R_1	8
R_2	12
R_3	15

The designated procedure to compute the maximum has a number of 8 source code lines. The length of the file is influenced by the arrangement of lines in blocks method.

Other traits of the software that is processed to give information regarding the indicators are:

- integrability expresses the ability of the software product to be called and integrated as a component module in a complex software product; Integrability assumes the uniformity of naming used in each module, for naming the variables as well as naming created functions, function parameterization on the basis of new frame formats conveyed upon during design, returning results that can serve as input in another module and processing results obtained from other modules;
- interoperability represents the capacity of the software product to couple with other products; this attribute allows the reuse of some programs to build complex applications; the coupling is assured directly or using interfaces;
- orthogonality is a quality characteristic which establishes the degree of resemblance between two or more software products; orthogonality represents the base of defining the software reuse indicator and determines a program's level of orthogonality.

There are quality characteristics that are highlighted by the behavior of the software product in exploitation.

Viability is perceived as a capacity metric of the software product to function correctly in all given conditions from the beginning. There are quantitative subscriptions of viability, expressed through the probability that a software product to fulfill its functions with certain performances and errors in a time interval and given exploitation conditions, as well as there are qualitative subscriptions which look at viability as a capacity of the software product. The ISO/IEC 9126 standard defines viability as a set of attributes that are based on the capability of the software product to maintain its level of performance in an established time period and conditions. The viability limits of a software product are caused by errors in the requirements definition, design and implementation stages. The problems caused by these errors depend on the conditions in which the software product is used. Corresponding to the software product life cycle, there is a projected or provisioned viability, on experimental viability and an operational viability (at the beneficiary). Viability defines the capacity of a software product to fulfill the functional parameters covering the whole use interval.

Maintenance is a specific process of software products meant to function over a large time interval, meaning longer then three years. In time, because of the technological processes evolution, law changes, structural collectivity modifications, the software products must answer the real requirements of users in order to be chosen.

A software product is maintainable if it allows a quick and easy actualization such that it can be used in the best conditions. A product that respects the maintenance criterions has a sufficiently long use period to amortize the production costs. Exceeding the cost recovery period and ensuring that the software product is viable contributes to the increase of the product's efficiency.

Portability is defined as a set of attributes based on the facility that software products should have, to transfer from one environment to another. The environment is represented by the hardware and/or software context in the organizational framework. Programs are said to be portable, not only if they are implemented on several computers directly, without any modifications, but also if the execution on other types of computer systems needs little modification and a reduced programming effort compared to redeveloping the program. Another aspect of this characteristic is tied to the portability of compilers, to different video and audio facilities offered by the computing systems and used by the application programs. A portable product is used easily in the organization, in organization branches, in departments; this relives the organization of the effort to buy new software programs.

Correctness gives the degree in which a software product satisfies or not the specifications of the problem that needs to be solved. A program is correct if for the input data that satisfies the specification of the problem, the obtained results are correct. The correctness of a software product represents the product's capacity to unroll a group of operations necessary for supplying results used in the analysis and prognosis process through respecting the set of norms at implementation time. The correctness of the program does not refer only to its capacity to respect implemented rules to obtain results, but also to the implementation of correct norms that supply correct results.

There are estimative and effective quality levels. The estimative quality levels are the ones computed and set at application design time. They are based on the formulas that are implemented, repetitive cycles used and the quantity of memory needed to run the software product. The effective levels are the ones obtained after running the software product and making modifications to the source code in order to reach the objectives that the application was developed to fulfill. Between the estimative and effective levels there are some differences caused by the necessity of adapting the software product to the requirements, the improvement of algorithm efficiency, and the appearance of new objectives. The quality indicators that characterize the software product are validated if the differences between the estimated and effective values are acceptable.

3. Software structure

The application developed to analyze the source file quality is available on the internet, being freeware.

The user inputs C# source files, which are tested with the help of the application.

The values returned by the application correspond to applying the quality indicators on the source files inputted by the users.

The software product includes:

- the human – computer interface;
- the computing modules;
- the modules used to validate the data inputted by the user;
- access authentication;
- the problem definition module;
- program storage base for each user.

To ensure user created product testing, a C# program batch is created, which allows the following operations:

- visualization of stored programs
- the selection of a program product from the batch;
- indicator computation for the batch, and individually for each program product ;
- creation of back-up copies and ensures their processing, to avoid the loss of basic information.

The program batch is meant to centralize the existing tendencies in the tested programs, to identify programming and logic errors, and to create code sequences used frequently by programmers.

4. Program classes

The program batch allows the computation of a multitude of indicators.

In [MACES85] the importance coefficients for characteristics are presented. The importance coefficients are aggregated in a characteristic model based on the correspondence between the weight of a characteristic and the value of the indicator of that specific characteristic, as it results from table 2.

Table 2. The correspondence between quality characteristics and their importance

Characteristic / Program	C_1	C_2	…	C_j	…	C_m
P_1						
P_2						
…						
P_i				X_{ij}		
…						
P_n						
Weights	p_1	p_2	…	p_j	…	p_m

Where p_i is the quality characteristic's importance weight C_i and $\sum_{i=1}^{m} p_i = 1$.

The quality characteristic importance coefficients are used to determine the program quality aggregated indicator using the formula:

$$IA(P_i) = \sum_{i=1}^{m} p_j * x_{ij} = IA_i$$

The determination of the aggregated program quality indicator underlines the basis of grouping the programs into quality classes.

The grouping into classes is done by establishing the number of classes and their size. Choosing the number of classes requires the knowledge of the program quality values variation, the elaboration of several class patterns until the optimum solution of the phenomena is reached.

To determine the quality class existence intervals the minim and maximum value of the quality values needs to be established. The difference between the two values is the amplitude, A.

To establish the quality classes, their number must be determined. It is determined using the Sturges rule:

$$N = 1+3{,}322 * \log_{10} n$$

where:
n – the number of programs for which the quality classes are determined;
N – the number of quality classes.

Based on the number of classes the size of each class, *d*, is determined according to the formula:

$$d = A / N$$

After the number of quality classes is computed, it is delimited by lower and upper bounds.

The bounds are determined in the following manner:

- the upper bound of each interval, will take the value of the lower bound of the following interval, this way repetitive bound intervals are obtained;
- the upper and lower bounds of the intervals are differenced by one unit;

The stages that must be covered to establish the program quality classes are the following:

S1: individual quality characteristic importance weight allocation
S2: determination of the program quality aggregated indicator;
S3: building the program's quality classes;

The covering of the above stages ensures the delimitation of program quality on value intervals, quality classes contributing to the grouping of programs considering the aggregated quality level.

5. Establishing the belonging of a program to a quality class

Program quality estimation imposes the creation of value delimited quality classes of quality indicators. The belonging of a program to a certain quality class supposes the computation of quality indicators applied to the respective program and the identification of the bounds between which the program must fit.

The analyzed programs must respect a series of frame conditions which certify their belonging to one of the quality classes:

- to be written in C#;
- to allow data input;
- to allow data visualization;
- not to contain syntax errors, or interpretation errors;
- to validate processed data.

The quality level for the *m* characteristics must not differ significantly compared to the average level.

The frame conditions have the role of limiting programs that are the subject of analysis. Program analysis is realized based in quality characteristics computed for every program.

Program quality is composed of a set of quality characteristics which are divided in:

- technical and usage characteristics;
- economical characteristics;
- social characteristics.

Quality characteristics are imposed not only by the client, but by the need that the software product runs efficiently, as well. Based on the quality characteristics the quality cost in determined, respectively the implied quality characteristics implementation cost and the cost of the quality characteristics requested by the client.

The quality cost is determined based on the formula:

$$CC(P_i)=\sum_{i=1}^{m} p_j * x_{ij} * c_j = CC_i$$

where:

m – number of the quality characteristics;

c_j – the quality cost of characteristic *j*;

x_{ij} – quality value of characteristic C_j for program $P_{i.}$

The belonging of a program to a quality class is established after the building the quality classes, considering the quality aggregated value of each program.

After the quality classes and their bounds are agreed upon, the programs and their belonging to quality classes are identified.

The quality characteristics dependency structure is presented in [IEEE94]:

- maintainability defines the degree of effort to repair, maintain or improve a product – C_1;
- correctness, represents the degree of effort necessary to correct errors and meet the user formulated requirements – C_2;
- modifiability, describes the effort to improve or modify the functions of the software product – C_3;
- testability, measures the effort needed to test the software product – C_4.

Table 3 presents the results of applying the above quality characteristics to a sample of 40 programs and the weight of the quality characteristics in the quality system aggregation:

Table 3. The level of quality characteristics

Characteristic / Program	C_1	C_2	C_3	C_4	IA	Characteristic / Program	C_1	C_2	C_3	C_4	IA
P_1	0.8	0.5	0.25	0.55	0.5675	P_{21}	0.6	0.68	0.55	0.84	0.6845
P_2	0.5	0.6	0.69	0.78	0.6375	P_{22}	0.7	0.77	0.77	0.82	0.764
P_3	0.6	0.1	0.65	0.76	0.5305	P_{23}	0.9	0.87	0.78	0.87	0.8655
P_4	0.3	0.98	0.89	0.92	0.7445	P_{24}	0.7	0.85	0.91	0.92	0.835
P_5	0.8	0.32	0.82	0.9	0.713	P_{25}	0.97	0.81	0.82	0.9	0.8865
P_6	0.9	0.65	0.34	0.74	0.7055	P_{26}	0.69	0.82	0.86	0.74	0.763
P_7	0.12	0.45	0.78	0.58	0.4395	P_{27}	0.18	0.88	0.8	0.84	0.646
P_8	0.54	0.89	0.79	0.64	0.695	P_{28}	0.76	0.89	0.79	0.82	0.815
P_9	0.69	0.32	0.75	0.15	0.4445	P_{29}	0.39	0.81	0.75	0.88	0.696
P_{10}	0.57	0.91	0.76	0.77	0.7435	P_{30}	0.57	0.91	0.76	0.77	0.7435
P_{11}	0.22	0.52	0.58	0.88	0.547	P_{31}	0.77	0.75	0.82	0.88	0.8055
P_{12}	0.68	0.23	0.32	0.89	0.5765	P_{32}	0.65	0.76	0.88	0.89	0.784
P_{13}	0.21	0.17	0.89	0.77	0.47	P_{33}	0.74	0.73	0.89	0.77	0.769
P_{14}	0.88	0.75	0.77	0.61	0.75	P_{34}	0.88	0.75	0.77	0.91	0.84
P_{15}	0.37	0.85	0.68	0.58	0.5995	P_{35}	0.85	0.85	0.83	0.92	0.868
P_{16}	0.89	0.83	0.69	0.46	0.716	P_{36}	0.89	0.83	0.84	0.93	0.8795
P_{17}	0.49	0.27	0.54	0.91	0.5685	P_{37}	0.85	0.77	0.87	0.91	0.851
P_{18}	0.61	0.84	0.23	0.29	0.5145	P_{38}	0.81	0.84	0.85	0.9	0.8505
P_{19}	0.55	0.55	0.61	0.39	0.511	P_{39}	0.82	0.83	0.77	0.83	0.818
P_{20}	0.92	0.64	0.88	0.94	0.85	P_{40}	0.92	0.88	0.88	0.94	0.91
Characteristic weight	**0.3**	**0.25**	**0.15**	**0.3**	**1**	**Characteristic weight**	**0.3**	**0.25**	**0.15**	**0.3**	**1**

The next stage constitutes the determination of quality classes.

The minimum value is 0.4395, and the maximum value is 0.91, the amplitude is 0.4705. The number of quality classes is 6, and the class size is 0.0784.

The quality classes are:

Class 1 $\in$ [0.4395;0.5179);
Class 2 $\in$ [0.5179;0.596);
Class 3 $\in$ [0.596;0.67);
Class 4 $\in$ [0.67;0.75);
Class 5 $\in$ [0.75;0.83);
Class 6 $\in$ [0.83;0.91].

Determining the quality classes and their lower and upper bounds makes up the starting point for the next stage, software product quality centralization.

After the centralization of data regarding the quality level of each program, the distribution of program in quality classes is obtained, given in table 4:

Table 4. Program distribution in quality classes

Class	Program number
1	5
2	5
3	3
4	9
5	8
6	10
Total	40

Corresponding to this algorithm of computing the aggregated level of software product quality, the quality characteristics costs specific to every program are added to the algorithm. Based on these the software product quality cost is determined.

By building dedicated software the processing of the 1067 programs of the sample is automated, the extension of the number of processed programs is a quantitative adaptation.

6. Conclusions

Program classes are related to costs. Thus, the level of expenses that has to be supported by customers in order to benefit from superior quality software products is identified.

The method of developing classes mixes the quality level of software products with the importance given by clients to each software quality characteristic.

For each client, the characteristics bare a level of importance dependent on the utility that they bring to the bought software product. The quality cost calculation includes the weight given to each quality characteristic, considering the level of homogeneity desired for the software product. The customers that buy the software product strictly for one of its quality characteristics will not pay extra for improvements as long as the clients who buy products with a large range of applicability need a high level of quality and homogeneity.

By implementing quality classes the software products are differentiated and catalogued, thus offering to clients to clients to find the needed software depending on its cost and desired quality.

Bibliography

1. Anghelache, C. and Niculescu, E. **Statistica: Indicatori, formule de calcul si sinteze,** Editura Economica, Bucharest, 2001
2. IEEE Standards Collection **Software Engineering, Std. 1045-1992 IEEE standard for software productivity metrics,** Published by The Institute of Electrical and Electronics Engineers, New York, 1994
3. Isaic–Maniu, Al., Mitrut, C. and Voineagu, V. **Statistica pentru managementul afacerilor,** Editura Economica, Bucharest, 1999
4. Ivan, I. and Boja, C. **Metode statistice in analiza software,** Editura ASE, Bucharest, 2004
5. Ivan, I. and Popa, M. **Entitati text – dezvoltare, evaluare, analiza,** Editura ASE, Bucharest, 2005
6. Ivan, I. and Popescu, M. **Metrici software,** BYTE România, vol.2, no.5, May 1996, pp.73-82
7. Ivan, I. and Teodorescu, L. **Managementul calitatii software,** Editura INFOREC, Bucharest, 2001
8. Ivan, I., Milodin, D. and Dumitru, S. N. **Ortogonalitatea programelor C++ software mining,** Revista Romana de Informatica si Automatica, vol. 17, no. 2, 2007, pp. 39 – 54
9. Macesatu, M., Arhire, R. and Ivan, I. **Clase de complexitate pentru produse program,** Buletinul Roman de Informatica, nr. 1, Bucharest, 1985, pp. 63 – 68

[1] Ion IVAN has graduated the Faculty of Economic Computation and Economic Cybernetics in 1970, he holds a PhD diploma in Economics from 1978 and he had gone through all didactic positions since 1970 when he joined the staff of the Bucharest University of Economics, teaching assistant in 1970, senior lecturer in 1978, assistant professor in 1991 and full professor in 1993. Currently he is full Professor of Economic Informatics within the Department of Economic Informatics at Faculty of Cybernetics, Statistics and Economic Informatics from the University of Economics. He is the author of more than 25 books and over 75 journal articles in the field of software quality management, software metrics and informatics audit. His work focuses on the analysis of quality of software

applications. He is currently studying software quality management and audit, project management of IT&C projects. He received numerous diplomas for his research activity achievements. For his entire activity, the National University Research Council granted him in 2005 with the national diploma, Opera Omnia.
He has received multiple grants for research, documentation and exchange of experience at numerous universities from Greece, Ireland, Germany, France, Italy, Sweden, Norway, United States, Holland and Japan.
He is distinguished member of the scientific board for the magazines and journals like:
- Economic Informatics; - Economic Computation and Economic Cybernetics Studies and Research; - Romanian Journal of Statistics
He has participated in the scientific committee of more than 20 Conferences on Informatics and he has coordinated the appearance of 3 proceedings volumes for International Conferences.
From 1994 he is PhD coordinator in the field of Economic Informatics.
He has coordinated as a director more than 15 research projects that have been financed from national and international research programs. He was member in a TEMPUS project as local coordinator and also as contractor in an EPROM project.

[2] He graduated "The Informatized Project Management" Master Program. He also graduated the Economy Informatics section from Cybernetics, Statistics and Economic Faculty, in 2005.
He published a series of articles in specialized magazines, articles regarding the study of the orthogonality for the alphabets, Arabian digits, Latin alphabet, using experimental methods and techniques, and articles regarding the collectivity components classification, using different criteria.
The disquisition paper in the master program follows the same research line, "The projects orthogonality, conditions for entering in the evaluation program", proposing to define the concept of project's orthogonality, methods for determine the degree of similarity between two projects, and to develop software products used to identify the project's similarity.

[3] Graduated the Economy Informatics section from Cybernetics, Statistics and Economic Faculty, in 2006. In the same year, he follows the "Informatics Security Master" program.
He published a series of articles in specialized magazines, regarding the study of software's source codes, based on text entities theory and software orthogonality using experimental methods and techniques.
The disquisition paper in the master program follows the "Informatics Security" line, "The security of online quality analizing applications of C++ programs" , proposing to define the concept of software quality, methods for determine the degree of similarity between two source codes, and to develop web-based software products used to identify the software's similarity and determine the quality indicators.

[4] Dragos Palaghita is a 4th year student in the University of Economics, Bucharest, Cybernetics Statistics and Economic Informatics faculty, Economic Informatics section. He is programming in C++ and C# and his main areas of interest are Informatics Security and Software Quality Management.

[5] Codification of references:

[ANGH01]	Anghelache, C. and Niculescu, E. **Statistica: Indicatori, formule de calcul si sinteze,** Editura Economica, Bucharest, 2001
[IEEE94]	IEEE Standards Collection **Software Engineering, Std. 1045-1992 IEEE standard for software productivity metrics,** Published by The Institute of Electrical and Electronics Engineers, New York, 1994
[VOIN99]	Isaic–Maniu, Al., Mitrut, C. and Voineagu, V. **Statistica pentru managementul afacerilor,** Editura Economica, Bucharest, 1999
[IVAN04]	Ivan, I. and Boja, C. **Metode statistice in analiza software,** Editura ASE, Bucharest, 2004
[IVPMR05]	Ivan, I. and Popa, M. **Entitati text – dezvoltare, evaluare, analiza,** Editura ASE, Bucharest, 2005
[IVAN96]	Ivan, I. and Popescu, M. **Metrici software,** BYTE România, vol.2, no.5, May 1996, pp.73-82
[IVAN01]	Ivan, I. and Teodorescu, L. **Managementul calitatii software,** Editura INFOREC, Bucharest, 2001
[MILO07]	Ivan, I., Milodin, D. and Dumitru, S. N. **Ortogonalitatea programelor C++ software mining,** Revista Romana de Informatica si Automatica, vol. 17, no. 2, 2007, pp. 39 – 54
[MACES85]	Macesatu, M., Arhire, R. and Ivan, I. **Clase de complexitate pentru produse program,** Buletinul Roman de Informatica, nr. 1, Bucharest, 1985, pp. 63 – 68

SOME NOTES ON THE LOGISTIC DISTRIBUTION

Alexandru ISAIC-MANIU
PhD, University Professor, Department of Statistics Econometrics
University of Economic Studies, Bucharest, Romania
(**Co)Author of the books:** Proiectarea statistica a experimentelor (2006), Enciclopedia calitatii (2005), Dictionar de statistica generala (2003), Statistica pentru managementul afacerilor (2000), Metoda Weibull (1983)

E-mail: al.isaic-maniu@csie.ase.ro, **Web page:** http://www.amaniu.ase.ro

Viorel Gh. VODĂ
PhD, Senior Scientific Researcher,
"Gh. MIHOC - C. IACOB" Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Bucharest, Romania
Co-author of the books: Proiectarea statistica a experimentelor (2006), Dictionar de statistica generala (2003), Manualul calitatii (1997)

E-mail: von_voda@yahoo.com

Abstract: *In this paper we study some properties of the density function* $f(x;\theta)=\theta\cdot e^{-x}\left(1+e^{-x}\right)^{-(\theta+1)}$, $x\in R$, $\theta>0$ *which may be obtained from the distribution of the last order statistic from a reduced logistic population* $F(x)=1/\left(1+e^{-x}\right)$, $x\in R$. *Its truncated variant is also discussed. The point and interval estimations for* θ *are provided also. The so-called* (P,γ) *- type statistical tolerances are constructed and a comment on the hazard rate is done also. The last paragraph is devoted to testing procedures on the parameter involved.*

Key words: *logistic distribution; Burr-Hatke family; last order statistic; truncation; MLE;* (P, γ) *- type tolerances*

1. Introduction

As it is well-known, the Belgian scientist Pierre François VERHULST (1804 - 1849) proposed in 1838 a "demographic growth cure" which was called later as the logistic function:

$$Y=\frac{A}{B+10^{a+bX}} \quad \text{or} \quad Y=\frac{A}{B+e^{a+bX}} \tag{1}$$

where $x\geq 0,\quad a,b>0,\quad a,b\in R,$ e being the Euler's number $(e\approx 2{,}71828)$. The usual form used now in econometric studies is the following :

$$Y = \frac{A}{B + e^{-CX}}, \quad a, B, C > 0, \quad x \geq 0 \tag{2}$$

(for historical details, see Iosifescu *et al*, 1985 [8, pages 280 - 281][1]).

Verhulst's function is obtained from a differential equation of the type $y' = u(y)$, where $y = y(x)$, $y_0 = u(x_0)$, $x_0 \in R$, by taking $u(y) = y - y^2$.

Therefore, we have:

$$\frac{dy}{dx} = y(1-y) \quad \text{or} \quad \frac{dy}{y(1-y)} = \frac{dx}{1} \tag{3}$$

which provides the reduced form of the logistic function:

$$y = \frac{1}{1 + e^{-x}}, \quad x \in R \tag{4}$$

If we consider y(x) as a cumulative distribution function (cdf) of a given random variable $X : y = F(x) = \Pr ob\{X \leq x\}$, the differential equation:

$$\frac{dF}{dx} = F(1-F)g(x), \quad F_0 = F(x_0), \quad x_0 \in R \tag{5}$$

where g(x) is an arbitrary positive function $(x \in R)$, provides - for several choices of g(x) - various cdf(s).

The form (4) generates the well-known BURR-HATKE family of distributions - see Burr (1942 [2]) and Hatke, (1949 [6]). Johnson, Kotz and Balkrishan (1994 [14, page 54]) enlists twelve such cdf(s), denoted from I to XII, the second one being just the Verhulst distribution function (4) if one takes $g(x) \equiv 1$. Notice that the most used cdf from these twelve cdf(s) is XII-th one: $1 - \left(1 + x^c\right)^{-k}, \quad x \geq 0, \quad c, k > 0$ (see Rodriguez, 1977 [16] or Vodă, 1982 [18]).

In this paper, we shall investigate mainly, the density function of the last order statistics from the logistic form (4).

2. Preliminaries

If X is a continuous random variable with $F_X(x)$ as a cdf and if $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(k)} \leq X_{(n+1)} \leq \ldots \leq X_{(n)}$ is an ordered sample on X, then, the distribution of the last order statistics $X_{(n)}$ is given as:

$$F_{X_{(n)}}(x) = \Pr ob\{X_{(n)} \leq x\} = \Pr ob\{all\ X_k \leq x\} = F_X^n(x) \tag{6}$$

(see David, 1970, [3 page 38]).

In our case, if $F_X(x) = F(x) = 1/\left(1 + e^{-x}\right)$, then

$F_{X_{(n)}}(x) = 1/\left(1 + e^{-x}\right)^n \quad x \in R, \quad n \in N \setminus \{0\}$. It is more interesting to change n by θ - a positive real parameter and then we obtain the cdf:

$$X : F(x;\theta) = \frac{1}{\left(1 + e^{-x}\right)^{\theta}}, \quad x \in R, \quad \theta > 0 \tag{7}$$

In a reliability frame we prefer to have $x \in [0,+\infty)$ and hence it is worth to truncate the density of X namely:

$$X : f(x;\theta) = \theta e^{-x}\left(1+e^{-x}\right)^{-(\theta+1)}, \quad x \in R, \quad \theta > 0 \tag{8}$$

as follows:

$$f_T(x;x_T,\theta) = \frac{1}{1-F(x_T;\theta)} \cdot f(x;\theta), \quad x \geq x_T \geq 0, \quad \theta > 0 \tag{9}$$

or

$$f_T(x;x_T,\theta) = \frac{\left(1+e^{-x_T}\right)^{\theta}}{\left(1+e^{-x_T}\right)^{\theta}-1} \cdot \theta \cdot e^{-x}\left(1+e^{-x}\right)^{-(\theta+1)}, \quad x \geq x_T \geq 0, \quad \theta > 0 \tag{10}$$

If the truncation point is $x_T = 0$, then:

$$f_T(x;x_T,\theta) = \frac{1}{1-1/2^{\theta}} \cdot \theta \cdot e^{-x}\left(1+e^{-x}\right)^{-(\theta+1)}, \quad x \geq 0, \quad \theta > 0 \tag{11}$$

3. Estimation problems

First, let us notice that even for the reduced form (4) the method of moments is difficult to be applied. If we consider the form (11) consider the integral:

$$m_k(x) = \int_0^x t^k f_T(t)dt \tag{12}$$

which provides the equation (the derivative of (12)):

$$m'_k(x) = x^k f_T(x) \tag{13}$$

and the k - th moment is given as:

$$E(X^k) = \int_0^\infty x^k f_T(x)dx \tag{14}$$

It is easy to see now that $E(X) = m_1(\infty) - m_1(0)$ with $m_1(0) = 0$ and the problems lies in the approximation of $m_1(\infty)$.

The MLE - maximum likelihood estimation - method gives straightforward results in the case of (7). We have the likelihood function

$$L(x_1, x_2, \ldots x_n;\theta) = \theta^n \cdot \left[\exp\left(-\sum_1^n x_i\right)\right]\prod_1^n \left(1+e^{-x_i}\right)^{-(\theta+1)} \tag{15}$$

where $x_1, x_2, \ldots x_n$ is a random sample on X. After some simple algebra, we obtain the MLE for $1/\theta$ as:

$$\left(\hat{\frac{1}{\theta}}\right) = \frac{1}{n}\sum_1^n \ln\left(1+e^{-x_i}\right) \tag{16}$$

The case of the truncated variable (11) provides successively:

$$L(x_1, x_2, \ldots x_n;\theta) = \frac{2^{n\theta}}{\left(2^{\theta}-1\right)^n} \cdot \theta^n \cdot \left[\exp\left(-\sum_1^n x_i\right)\right] \cdot \prod_1^n \left(1+e^{-x_i}\right)^{-(\theta+1)} \tag{17}$$

$$\ln L = n\theta \ln 2 - n\ln\left(2^{\theta} - 1\right) + n\ln\theta - \sum_{1}^{n} x_i - (\theta+1)\sum_{1}^{n}\ln\left(1+e^{-x_i}\right) \tag{18}$$

$$\frac{\partial \ln L}{\partial \theta} = n\ln 2 - \frac{n\cdot 2\ln 2}{2^{\theta}-1} + \frac{n}{\theta} - \sum_{1}^{n}\ln\left(1+e^{-x_i}\right) \tag{19}$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \frac{n\ln 2}{2^{\theta}-1} - \sum_{1}^{n}\ln\left(1+e^{-x_i}\right) = 0 \tag{20}$$

The likelihood equation (20) is of the following type:

$$\varphi(u) = \frac{A}{u} - \frac{B}{2^{u}-1} - C = 0 \tag{21}$$

where A, B, C > 0 and it is clear that it has a solution science $\lim\limits_{u\to 0^+}\varphi(u) = +\infty$ and $\lim\limits_{u\to +\infty}\varphi(u) = -C\cdot$

Numerical methods are needed to approximate (u).

We shall state now the following.

Proposition. If X is a logistic random variable with cdf given by (7), then the variable $Y = \ln\left(1+e^{-X}\right)$ is exponentially distributed and the consequences are:

(i) the MLE for $1/\theta$ is unbiased and with minimum variance;

(ii) the distribution of $\left(\frac{\hat{1}}{\theta}\right)$ is a Gamma one;

(iii) the statistic $U = 2n\theta/\hat{\theta}$ has a Chi - square $\left(\chi^2\right)$ distribution with 2n degrees of freedom.

Proof.We have immediately:

$$\begin{aligned} F(z) &= \Pr ob\left\{\ln\left(1+e^{-x}\right) \le z\right\} = 1 - \Pr ob\left\{x < \ln\left(e^{z}-1\right)\right\} = \\ &= 1 - \theta \int_{-\infty}^{\ln\left(e^{z}-1\right)} e^{-x}\left(1+e^{-x}\right) dx \end{aligned} \tag{22}$$

Taking into account that

$$\left(\int_{a(z)}^{b(z)} \varphi(x)dx\right)' = b'(z)\cdot\varphi[b(z)] - a'(z)\cdot\varphi[a(z)] \tag{23}$$

we obtain from (22):

$$F'(z) = \theta \exp(-\theta z) \text{ with } z \geq 0, \quad \theta > 0 \tag{24}$$

Since

$$E\left[\left(\frac{\hat{1}}{\theta}\right)\right] = \frac{1}{\theta} \quad \text{and} \quad \operatorname{Var}\left[\left(\frac{\hat{1}}{\theta}\right)\right] = \frac{1}{n\theta^2} \tag{25}$$

the property (i) is proved.

Now, since each variable is exponentially distributed, then its sum is Gamma distributed (see Johnson et al, 1994 [14, page 337 and page 494]). The density of $\left(\frac{\hat{1}}{\theta}\right)$ is therefore

$$\varphi\left[\left(\frac{\hat{1}}{\theta}\right);\theta,n\right] = \frac{n^n\theta^n}{\Gamma(n)}\left(\frac{\hat{1}}{\theta}\right)^{n-1} \cdot \exp\left\{-n\theta\left(\frac{\hat{1}}{\theta}\right)\right\} \tag{26}$$

where $\left(\frac{\hat{1}}{\theta}\right) > 0, \quad \theta > 0, \quad n \in N \setminus \{0\} \text{ and } \Gamma(n)$ is the Gamma function; here we have $\Gamma(n+1) = n!$ (the property (ii) is also proved).

To demonstrate (iii), we may write the characteristic function of U, and we have:

$$\varphi_U(t) = E\left(e^{itu}\right) = E\left[\exp\left(i(2\theta t)\sum_{j=1}^{n}\ln\left(1+e^{-xj}\right)\right)\right] = \prod_{j=1}^{n}\left(1-\frac{1}{\theta}2\lambda t\right)^{-1} = (1-2it)^{-n} \tag{27}$$

which is just the characteristic function of a Chi-square variable with 2n degrees of freedom (see for instance Wilks, 1962 [20 page 86]).

Based on this property, one can construct confidence intervals of minimum length (L) for θ. Namely, we have to determine two limits L_{inf} and L_{sup} such that

$$\Pr ob\left\{L_{inf} < 2n\theta/\hat{\theta} < L_{sup}\right\} = 1-\alpha \tag{28}$$

where $0 < \alpha < 1, \quad \alpha$ - given, with the property (29) $L = \frac{1}{2Y_0}\left(L_{sup} - L_{inf}\right) =$ minimum, where $Y_0 = \sum_{1}^{n}\ln\left(1+e^{-x_i}\right)$.

In accordance with Tate and Klett`s results (1959, [17]) the following system has to be solved:

$$\int_{L_{inf}}^{L_{sup}} u^{n-1}e^{-u/2}du = 2^n(1-\alpha)\Gamma(n); \quad \left(\frac{L_{sup}}{L_{inf}}\right)^{n-1} = \exp\left(\frac{L_{sup}-L_{inf}}{2}\right) \tag{30}$$

The solutions (L_{inf}, L_{sup}) are founded by entering Tate - Klett`s tables in the cell corresponding to 2n degrees of freedom (see also Isaic-Maniu and Vodă, 1989 [10]).

4. (P, γ) - type tolerance limits

The literature devoted to the problem of statistical (or "natural") tolerance is very wide. In 1981, Miloš Jílek (Prague) compiled a large bibliography on this subject matter (see Jílek [12]) and in 1988 the same author (see Jílek [13]) provided a sound monograph with applications (chapter 12, pages 198 - 235 of [13]) of these tolerances.

Origin of this concept - which goes back to W. A. Shewhart (1891 - 1967), Samuel S. Wilks (1906 - 1964), Abraham Wald (1902 - 1950), Herbert Robbins (1915 - 2001) - and some other (historical aspects may be found in [7]).

The mathematical formulation of the problem is the following: if X is a continuous random variable defined on $D \subseteq R$ and having a density function $f(x;\theta)$, then we have to construct two statistics T_L and T_U (lower and upper) such that at least a proportion (P) of this population {X} will be found between T_L and T_U, and this must happen with a given probability $\gamma(0<\gamma<1)$, that is.

$$\operatorname{Prob}\left\{\int_{T_L}^{T_u} f(x;\theta)dx \geq P\right\} = \gamma \tag{31}$$

where $0<P,\gamma<1$ are previously chosen. These elements $T_L = \varphi(x_1, x_2, \ldots x_n)$ and $T_U = \Psi(x_1, x_2, \ldots x_n)$ where $\{x_i\}_{1\leq i\leq n}$ are sample values on X are called (P,γ) - type tolerance limits. In a reliability context one is interested in a lower tolerance limits since $X : D \equiv [0,+\infty)$ and we need that at most a proportion (1 - P) of the population to lie between 0 and T_L.

In our logistic case, we shall write:

$$\operatorname{Prob}\left\{\int_{T_L}^{+\infty} f(x;\theta)dx \geq P\right\} = \gamma \tag{32}$$

Two ways to deduce T_L will be presented.

a) ***The case of large sample****s* In this situation, we may state that the statistic

$$Y = \frac{\left(\widehat{\frac{1}{\theta}}\right) - \frac{1}{\theta}}{\frac{1}{\theta\sqrt{n}}} \tag{33}$$

where $\left(\widehat{\frac{1}{\theta}}\right)$ is the MLE of $1/\theta$, is approximately normally distributed with E [Y] = 0 and Var [Y] = 1.

The relationship (32) may be written as.

$$\operatorname{Prob}\left\{1-\left(1+e^{-T_L}\right)^{-\theta} \geq P\right\} = \gamma \tag{34}$$

or

$$\Pr ob\left\{\ln(1-P) \geq -\theta \ln\left(1+e^{-T_L}\right)\right\} = \gamma \tag{35}$$

$$\Pr ob\left\{\frac{1}{\theta} \leq \frac{\ln\left(1+e^{-T_L}\right)}{-\ln(1-P)}\right\} = \gamma \tag{36}$$

which may be rearranged as follows:

$$\Pr ob\left\{\frac{\left(\frac{\hat{1}}{\theta}\right) - \frac{1}{\theta}}{1/\theta\sqrt{n}} \leq \frac{\frac{\ln\left(1+e^{-T_L}\right)}{-\ln(1-P)} - \frac{1}{\theta}}{1/\theta\sqrt{n}}\right\} = \gamma \tag{37}$$

(we did replace $1/\theta$ in (36) by its MLE).

The right-hand side of the inequality in (37) is just the γ - quantile of N (0, 1) distribution - let it be u_γ - and hence we may write:

$$\frac{\ln\left(1+e^{-T_L}\right)}{-\ln(1-P)} = \left(\frac{\hat{1}}{\theta}\right)\left(1+\frac{u_\gamma}{\sqrt{n}}\right) \tag{38}$$

or

$$\ln\left(1+e^{-T_L}\right) = \left(\frac{\hat{1}}{\theta}\right)\left(1+\frac{u_\gamma}{\sqrt{n}}\right)\ln\frac{1}{1-P} \tag{39}$$

Since $0 < e^{-T_L} < 1$, then we could use the approximation of ln (1+x), 0 < x < 1 given in Abramowitz and Stegun (1964, [1]) to find a polynomial equation in T_L :

$$\ln(1+x) = a_1x + a_2x^2 + \ldots + a_5x^5 + \varepsilon(x) \tag{40}$$

where the error $\varepsilon(x)$ is $|\varepsilon(x)| < 10^{-5}$ and $a_1 \approx 0.99949556$ a.s.o. One may restrict the approximation to the roughest one, namely:

$$\ln\left(1+e^{-T_L}\right) \approx a_1 \cdot e^{-T_L} \tag{41}$$

and hence T_L may be deduced easily by taking logarithms in (39).

We did call this method to find statistical tolerances as "normalizing" one (see Isaic-Maniu and Vodă, 1981, [9], and 1993, [10].

b) The general case. When we have an arbitrary sample size, we may re-write (35) as follows:

$$\Pr ob\left\{\frac{2n\theta}{\hat{\theta}_{MLE}} \leq \frac{1\ln\frac{1}{1-P}}{\ln\left(1+e^{-T_L}\right)} \cdot \sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right)\right\} = \gamma \tag{42}$$

Since $2n\theta/\hat{\theta}_{MLE}$ is Chi-square (χ^2) distributed with 2n degrees of freedom, the righ-hand side in the inequality of (42) is just the γ - quantile of the (χ^2) distribution - let it be $(\chi^2{}_{2n;\gamma})$. Therefore, a similar equation with (39) is obtained:

$$\ln\left(1+e^{-T_L}\right)=\frac{1}{\chi^2{}_{2n;\gamma}}\cdot\ln\frac{1}{1-P}\cdot\sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right) \tag{43}$$

5. A discussion on the hazard rate

If X is a continuous random variable representing the time-to-failure of a certain device, then the hazard (or failure) rate associated to X is given as:

$$h(x)=\frac{f(x)}{1-F(x)} \tag{44}$$

where F(x) is the cdf of X and $F'(x)=f(x)$ is the corresponding density. The variable X is assumed to be positive.

In our case we have to work with the truncated variable. Suppose that our logistic is truncated at x = 0, that is $x\in[0,+\infty)$. Therefore:

$$F_T(x;\theta)=\frac{1}{1-F(0;\theta)}\int_0^x F(u;\theta)du=K\cdot[F(x;\theta)-F(0,\theta)] \tag{45}$$

where $K=2^\theta/(2^\theta-1)$ and $F(0;\theta)=1/2^\theta$.

In this situation, we have:

$$h_T(x;\theta)=\frac{f_T(x;\theta)}{1-F_T(x;\theta)}=\frac{\frac{2^\theta}{2^\theta-1}\cdot\theta e^{-x}\left(1+e^{-x}\right)^{-(\theta+1)}}{1-\left[\frac{1}{\left(1+e^{-x}\right)^\theta}-\frac{1}{2^\theta}\right]} \tag{46}$$

or

$$h_T(x;\theta)=\frac{\frac{\theta\cdot 2^\theta}{2^\theta-1}e^{-x}\cdot\frac{1}{\left(1+e^{-x}\right)^{\theta+1}}}{\frac{2^\theta-1}{2^\theta}-\frac{1}{\left(1+e^{-x}\right)^\theta}} \tag{47}$$

The study of $h_T(x;\theta)$ may be performed denoting $e^{-x}=y$ (we have 0 < y < 1) and consequently, we get.

$$h_T(y;\theta)=\frac{K\theta y\cdot\frac{1}{(1+y)^{\theta+1}}}{\frac{1}{K}-\frac{1}{(1+y)^\theta}} \tag{48}$$

or

$$h_T(y;\theta)=\frac{K^2\theta y}{\left[(1+y)^\theta-K\right](1+y)},\quad 0<y<1 \tag{49}$$

An interesting situation arises when we consider the reduced form of the logistic $F(x)=1/\left(1+e^{-x}\right),\quad x\in R$ and write formally:

$$h(x)=\frac{\frac{dF}{dx}}{1-F}=\frac{F(1-F)}{1-F}=F(x) \tag{50}$$

In this case, the hazard rate coincides with the distribution function and the behaviour of h(x) is now obvious.

6. Testing a simple statistical hypothesis

Since only one parameter is involved in $F(x;\theta)$, we may make use of the fact that

$$U=\frac{2n\theta}{\hat{\theta}_{ML}}=2\theta\cdot\sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right) \tag{51}$$

is Chi-square distributed with 2n degrees of freedom and hence to test:

$$h_0:\theta=\theta_0 \text{ versus } h_1:\theta=\theta_1\quad(\theta_0<\theta_1) \tag{52}$$

the critical value will be $\chi^2_{2n;\alpha}$ (α - quantile of the χ^2 distribution with 2n degrees of freedom).

More economical is to apply SPRT (Sequential Probability Ratio Test) of Abraham WALD (1902 - 1950) - see his well - known book [19 , Chapter 3, pages 37 - 54].

We shall write straightforwardly the logarithm of likelihood ratio (r_n):

$$\ln r_n=n\cdot\ln\frac{\theta_1}{\theta_0}+(\theta_0-\theta_1)\cdot\sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right) \tag{53}$$

where $x_i\,(i=1,2,\ldots,n,\ldots)$ is the sequential sample.

In accordance with Wald`s rules [19, page 49] we could take the following decisions:

a) If

$$\sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right)\geq\frac{\ln A}{\theta_1-\theta_0}+n\cdot\frac{\ln\frac{\theta_1}{\theta_0}}{\theta_1-\theta_0} \tag{54}$$

we accept (H_0) and automatically reject the alternative (H_1):

b) If

$$\sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right)\leq\frac{\ln B}{\theta_1-\theta_0}+n\cdot\frac{\ln\frac{\theta_1}{\theta_0}}{\theta_1-\theta_0} \tag{55}$$

then reject the null hypothesis (H_0) and accept (H_1)

c) If

$$\frac{\ln B}{\theta_1-\theta_0}+n\cdot\frac{\ln\frac{\theta_1}{\theta_0}}{\theta_1-\theta_0}<\sum_{i=1}^{n}\ln\left(1+e^{-x_i}\right)<\frac{\ln A}{\theta_1-\theta_0}+n\cdot\frac{\ln\frac{\theta_1}{\theta_0}}{\theta_1-\theta_0} \quad (56)$$

then, the procedure continues by taking the next observation.

Here $A=(1-\beta)/\alpha$ and $B=\beta/(1-\alpha)$, α,β being the classical statistical risks in the theory of hypothesis testing (see Wilks, 1962 [20]).

The OC - function (the Operative Characteristic) of the test is given by:

$L(\theta)=\left(A^h-1\right)/\left(A^h-B^h\right)$ where h is the solution of Wald`s equation $E\left(e^{th}\right)=1,\ h\neq 0,\ z=\ln\frac{f(x;\theta_1)}{f(x;\theta_0)}$.

In our case:

$$z=\ln\frac{\theta_1}{\theta_0}+(\theta_0-\theta_1)\cdot\ln\left(1+e^{-x}\right) \quad (57)$$

and consequently:

$$e^{zh}=\left(\frac{\theta_1}{\theta_0}\right)^n\cdot\left(1+e^{-x}\right)^{h(\theta_0-\theta_1)} \quad (58)$$

and

$$E\left[e^{zh}\right]=\left(\frac{\theta_1}{\theta_0}\right)^h\cdot\int_{-\infty}^{+\infty}e^{-x}\left(1+e^{-x}\right)^{h(\theta_0-\theta_1)-\theta-1}dx=1 \quad (59)$$

Imposing the condition $\theta>h(\theta_0-\theta_1)$, we get a parametric representation of the OC - function as

$$\theta=\frac{h(\theta_1-\theta_0)}{\left(\frac{\theta_1}{\theta_0}\right)^h-1} \quad (60)$$

The ASN (Average Sample Number) needed to perform SPRT is given be

$E_\theta(n)=\{L(\theta)\cdot\ln B+[1-L(\theta)]\ln A\}/E_\theta(z)$

where - in one case:

$$E_\theta(z)=\ln\frac{\theta_1}{\theta_0}+(\theta_0-\theta_1)\cdot\frac{1}{\theta}$$

Therefore, the sequential test is completely constructed. We did not re-stated the whole theory - all details are given in Wald [19], Dixon and Massey (1972, [4, pages 300 - 312]) or in more recent works such as those of Govindarajulu (2000 [5]) or Pham (2006, [15]).

Bibliography

1. Abramowitz, M. and Stegun, Irene, A. **Handbook of Mathematical Functions**, NBS Applied Math. Series. Nr. 55, Washington, D. C. (Russian edition, 1979, Moscow, NAUKA), 1964
2. Burr, I. W. **Cumulative frequency functions**, Annals of Mathematical Statistics, vol. 13, 1942, pp. 215 - 232
3. David, H. A. **Order Statistics**, John Wiley and Sons Inc, New York, 1970
4. Dixon, W. J. y Massey, F. J. (Jr) **Introduccion al Analisis Estadistico** (segunda edicion, revisado por D. Sixto Rios; traducido par J. P. Vilaplana), Edicion Revolucionaria, Instituto Cubano del Libro La Habana, Cuba, 1972
5. Govindarajulu, Z. **Statistical Techuiques in Bioassay** (second, revised and enlarged edition), KARGER Publishers, Basel, Switzerland, 2000
6. Hatke, M. A. (Sister) **A certain cumulative probability function**, Annals of Mathematical Statistics, vol. 20, 1949, pp. 461 - 463
7. Iliescu, D. V. and Vodă, V. Gh. **Statistica si tolerante (Statistics and Tolerances - in Romanian)**, Ed. Tehnica, Bucharest, 1977
8. Iosifescu, M., Moineagu, C., Trebici, Vl. and Ursianu, E. **Mica Enciclopedie de Statistică (Little Statistical Enciclopedia- in Romanian)**, Editura Stiintifica si Enciclopedica, Bucharest, 1985
9. Isaic-Maniu, Al. and Vodă, V. Gh. **Tolerance limits for the Weibull distribution using MLE for the scale parameter**, Bull. Math. Soc. Sci. Math de la R.S. de Roumanie (Bucarest), Tome 24(72), no. 3, 1981, pp. 273 - 276
10. Isaic-Maniu, Al. and Vodă, V. Gh. **Estimarea duratei medii de viata a produselor prin intervale de încredere de lungime minima (Estimation of product average life via minimum length confidence intervals - in Romanian)**, Stud. Cerc.Calc.. Econ. Cib. Econ. Anul XXIII,no. 1, 1989, pp. 33 - 41
11. Isaic-Maniu, Al. and Vodă, V. Gh. **On a special case of Sedrakian class of distributions applications**, Econ. Comp. Econ. Cyb. Stud. Res. (Bucharest), Vol. XXVII, no. 1 - 4, 1993, pp. 33 - 42
12. Jilek, M. **A bibliography of statistical tolerance regions, Mathematische Operations Forschung und Statistik,** Ser. Statistics (Berlin), vol. 12, 1981, pp. 441 - 456
13. Jilek, M. **Statisticke Tolerancni Meze. (Statistical Tolerance Limits - in (czech)**, SNTL/TKI, Praha, 1988
14. Johnson, N. L., Kotz, S. and Balakrishan, N. **Continuous Univariate Distributions**, Vol. 1 (Second Edition), John Wiley and Sons Inc., New York, 1994
15. Pham, H. (editor) **Springer Handbook of Engineering Statistics**, Springer Verlag, Berlin, 2006
16. Rodriguez, R. N. **A guide to the Burr Type XII distributions**, Biometrika, vol. 64, no. 1, 1977, pp. 129 - 134
17. Tate, R. F. and Klett, G. W. **Optimum confidence intervals for the variance of a normal population**, J. Amer. Statist. Assoc., vol. 54, 1959, pp. 674 - 682
18. Vodă, V. Gh. (1982): **Burr distribution revisited,** Rev. Roum. Math. Pures et Appl**.**, Tome XXVII, no. 8,1982, pp. 885 - 893
19. Wald, A. **Sequential Analysis. Dover Publications Inc. New York** (Student edition: unabridged republication of the original (1947) - John Wiley and Sons Inc.), 1973
20. Wilks, S. S. **Mathematical Statistics** (Second Edition), John Wiley and Sons Inc., New York, 1962

[1]Codification of references:

[1]	Abramowitz, M. and Stegun, Irene, A. **Handbook of Mathematical Functions**, NBS Applied Math. Series. Nr. 55, Washington, D. C. (Russian edition, 1979, Moscow, NAUKA), 1964
[2]	Burr, I. W. **Cumulative frequency functions**, Annals of Mathematical Statistics, vol. 13, 1942, pp. 215 - 232
[3]	David, H. A. **Order Statistics**, John Wiley and Sons Inc, New York, 1970

[4]	Dixon, W. J. y Massey, F. J. (Jr) **Introduccion al Analisis Estadistico** (segunda edicion, revisado por D. Sixto Rios; traducido par J. P. Vilaplana), Edicion Revolucionaria, Instituto Cubano del Libro La Habana, Cuba, 1972
[5]	Govindarajulu, Z. **Statistical Techuiques in Bioassay** (second, revised and enlarged edition), KARGER Publishers, Basel, Switzerland, 2000
[6]	Hatke, M. A. (Sister) **A certain cumulative probability function**, Annals of Mathematical Statistics, vol. 20, 1949, pp. 461 - 463
[7]	Iliescu, D. V. and Vodă, V. Gh. **Statistica si tolerante (Statistics and Tolerances - in Romanian)**, Ed. Tehnica, Bucharest, 1977
[8]	Iosifescu, M., Moineagu, C., Trebici, Vl. and Ursianu, E. **Mica Enciclopedie de Statistică (Little Statistical Enciclopedia- in Romanian)**, Editura Stiintifica si Enciclopedica, Bucharest, 1985
[9]	Isaic-Maniu, Al. and Vodă, V. Gh. **Tolerance limits for the Weibull distribution using MLE for the scale parameter**, Bull. Math. Soc. Sci. Math de la R.S. de Roumanie (Bucarest), Tome 24(72), no. 3, 1981, pp. 273 - 276
[10]	Isaic-Maniu, Al. and Vodă, V. Gh. **Estimarea duratei medii de viata a produselor prin intervale de încredere de lungime minima (Estimation of product average life via minimum length confidence intervals - in Romanian)**, Stud. Cerc.Calc.. Econ. Cib. Econ. Anul XXIII,no. 1, 1989, pp. 33 - 41
[11]	Isaic-Maniu, Al. and Vodă, V. Gh. **On a special case of Sedrakian class of distributions applications**, Econ. Comp. Econ. Cyb. Stud. Res. (Bucharest), Vol. XXVII, no. 1 - 4, 1993, pp. 33 - 42
[12]	Jilek, M. **A bibliography of statistical tolerance regions, Mathematische Operations Forschung und Statistik,** Ser. Statistics (Berlin), vol. 12, 1981, pp. 441 - 456
[13]	Jilek, M. **Statisticke Tolerancni Meze. (Statistical Tolerance Limits - in (czech)**, SNTL/TKI, Praha, 1988
[14]	Johnson, N. L., Kotz, S. and Balakrishan, N. **Continuous Univariate Distributions**, Vol. 1 (Second Edition), John Wiley and Sons Inc., New York, 1994
[15]	Pham, H. (editor) **Springer Handbook of Engineering Statistics**, Springer Verlag, Berlin, 2006
[16]	Rodriguez, R. N. **A guide to the Burr Type XII distributions**, Biometrika, vol. 64, no. 1, 1977, pp. 129 - 134
[17]	Tate, R. F. and Klett, G. W. **Optimum confidence intervals for the variance of a normal population**, J. Amer. Statist. Assoc., vol. 54, 1959, pp. 674 - 682
[18]	Vodă, V. Gh. (1982): **Burr distribution revisited,** Rev. Roum. Math. Pures et Appl., Tome XXVII, no. 8,1982, pp. 885 - 893
[19]	Wald, A. **Sequential Analysis. Dover Publications Inc. New York** (Student edition: unabridged republication of the original (1947) - John Wiley and Sons Inc.), 1973
[20]	Wilks, S. S. **Mathematical Statistics** (Second Edition), John Wiley and Sons Inc., New York, 1962

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

IMPLEMENTATION AND APPLICATIONS OF A THREE-ROUND USER STRATEGY FOR IMPROVED PRINCIPAL AXIS MINIMIZATION[1, 2]

Ali ÜNLÜ[3]

PhD, Assistant Professor, Department of Computer Oriented Statistics and Data Analysis,
Institute of Mathematics, University of Augsburg, Germany

E-mail: ali.uenlue@math.uni-augsburg.de

Michael D. KICKMEIER-RUST[4]

MSc, Senior Project Coordinator,
Department of Psychology, University of Graz, Austria

E-mail: michael.kickmeier@uni-graz.at

Abstract: *This paper presents a three-round user strategy (EPM), extending the C implementation of Brent's PRAXIS algorithm by Gegenfurtner. In a first round, EPM applies a multistart procedure for global optimization, randomly generating and evaluating multiple sets of start values drawn from weighted primary and secondary intervals. Using the parameter estimates of the smallest first round minimum, in a second and third round, EPM performs iterative minimization runs and applies an additional break-off criterion to improve and stabilize the approximated minimum and parameter estimates. Moreover, EPM increases the precision of the original PRAXIS implementation by a conversion from the double to the long double data type. This conversion is not trivial and even seen to be essential for minimizing a complex empirical function from psychometrics. Important special cases of EPM are discussed and promising strategies for the handling of EPM are proposed. EPM's advantages over PRAXIS are illustrated using two different functions: a 'well-behaved' Rosenbrock function and an 'ill-behaved' psychometric likelihood function.*

Key words: Evaluation and improvement of optimization software; Extended principal axis minimization; Numerical optimization; PRAXIS; Psychometrics

Introduction

Remarks on the PRincipal AXIS (PRAXIS) minimization

Minimizing functions of several variables is a common and important problem in psychology and in natural sciences in general. A popular method is the *PRincipal AXIS* (PRAXIS) minimization by Brent (1973), an algorithm for numerical minimization of multivariable functions without the use of derivatives. PRAXIS is a modification of Powell's (1964) direction-set method; for a review of PRAXIS, see Gegenfurtner (1992). Gegenfurtner (1992) also provides an implementation of Brent's algorithm in the programming language C which is freely available on the Internet (see Section Availability).

The PRAXIS algorithm has been applied to a variety of problems in psychology; for instance, see Hartinger (1999) and Regenwetter, Falmagne, and Grofman (1999) in decision making, D'Zmura, Rinner, and Gegenfurtner (2000) and Heller (2001) in visual perception, or Doignon and Falmagne (1985, 1999), Fries (1997) and Ünlü (2006) in the psychometric theory of knowledge spaces. Further references on PRAXIS applications in psychological sciences are Erdfelder and Buchner (1998) in process-dissociation modeling, Heller (1997) in psychophysics, and McClelland and Chappell (1998) in memory research. PRAXIS type algorithms have been also applied in fields other than psychology; for instance, see Carcione, Mould, Pereyra, Powell, and Wojcik (2001) in computational acoustics, Hwang and Tien (1996) in physics, Leroy, Mozer, Payan, and Troccaz (2004) in medical image computing, Ren, Chen, Wu, and Yang (2002) and Ren, Wu, Yang, and Chen (2002) in nuclear medicine, Tubic, Zaccarin, Beaulieu, and Pouliot (2001) in medical physics, or Woelk (2000) in magnetic resonance.

Basic motivations for an Extended Principal axis Minimization (EPM)

In a realistic context, optimizing an objective function, in general, is more than to use a computer and a software application (e.g., PRAXIS), implicitly and unjustifiably assuming that (a) the actual optimization exercise utilizing the software is trivial and (b) the obtained results are accurate (cf. McCullough & Vinod, 1999). Quite the contrary, (a) optimization problems in practice often concern functions with many local extrema rather than being globally convex, concave, or otherwise 'good-natured.' Moreover, (b) numerical optimization algorithms and their software implementations differ in quality (e.g., accuracy of results), and one software application is not as good as any other, in general. This, however, is mostly hard to assess, especially for software users merely interested in application.

These points, in particular, apply to Brent's algorithm and its C implementation by Gegenfurtner. (a) PRAXIS generally converges to local minima; hence, minimization results strongly depend on the selection of suitable start values (cf. Table 1). Thus, in order to be also able to handle complex optimization problems, procedures for global optimization are required. Global optimization strategies for the PRAXIS algorithm have not been investigated, implemented and freely supplied so far, although there is a large body of strategies that could be considered. Methods are, for example, controlled random search (Price, 1983), evolutionary and genetic algorithms (Back & Schwefel, 1993), or multistart methods (Törn & Zilinskas, 1989). For a review of these approaches, see Pintér (1995). In this paper, we extend the original PRAXIS algorithm to *Extended Principal axis Minimization* (EPM) to offer a natural, flexible multistart procedure for global optimization. This provides a facilitation of usage resulting in improved and effective applications of PRAXIS (PRAXIS, currently, being merely available as a 'bare' C function).

Moreover, (b) the accuracy of the PRAXIS algorithm can be improved by three natural techniques: iterative minimization runs with reset program settings, an additional break-off criterion besides the PRAXIS algorithm's default criterion, and a conversion from the double to the long double data type. The EPM strategy proposed in this paper implements these techniques.

In the iterative part, EPM successively uses resulting parameter estimates as vectors of start values for new runs of the complete PRAXIS routine with reset internal program settings. The use of a second break-off criterion based on a minimal change in approximated minima, while PRAXIS applies a default criterion based on a minimal change in parameter estimates, supplements the iterative part to offer the possibility of improving and stabilizing the approximated minimum and parameter estimates. Finally, to gain more computational precision for complex optimization problems, EPM uses the long double data type instead of the double data type of the original PRAXIS implementation.[5]

A conversion from the double to the long double data type, however, is not trivial. It cannot be accomplished by simply altering the declaration/definition of variables in the original C source files (e.g., 'long double tol' instead of 'double tol'). A conversion requires the use of long double functions instead of double functions (e.g., 'long double sqrtl(long double)' instead of 'double sqrt(double)'), and it should also consider changing the data type of relevant constants (e.g., '0.0L' instead of '0.0'), throughout the entire source files. On the other hand, the conversion from the double to the long double data type is even seen to be essential for minimizing the empirical psychometric function in Subsection LCMRE function. Optimization of the *Latent Class Model with Random Effects* (LCMRE) likelihood function, which is based on real psychological test data, was not possible using the double (data type) versions of PRAXIS and EPM. Apparently, this was due to rounding errors that resulted in undefined operations (e.g., division by zero), and consequently, in undefined values for a minimum or parameter estimate. Use of the long double data type simply accommodated this problem.

In the end, the different components of EPM can be combined with each other to give important special cases of the EPM strategy in practice. Additionally enhanced by a multiplicity of intra-component strategies for the handling of individual components, this offers a great flexibility of usage in the actual minimization exercise utilizing the EPM extension (cf. Fig. 3).

Additional notes

Similar to optimization heuristics such as *simulated annealing*, *genetic algorithms*, or *ant colonies* (for details and further references, see, e.g., Winker, 2001, and Winker & Gilli, 2004), EPM allows 'uphill moves', that is, larger minima in consecutive iterations than in previous ones. Throughout its three rounds, however, EPM stores the smallest minimum found and the corresponding parameter estimates globally, and outputs them as the final solution (cf. Fig. 2).

EPM is implemented in C programming language (ANSI C 99). The function is held very general, so it can be easily used with other algorithms available in C/C++ without extensive modifications. The source code for EPM is freely available from the authors.

It shall be also noted that the long double (data type) versions of PRAXIS and EPM do not run on Windows 32 systems (95, 98, NT, Me, 2000, XP) properly. On these systems, the long double data type is directly mapped to the double data type. To overcome this limitation on Windows systems, extensive workarounds are necessary.

Architecture of EPM

The EPM strategy consists of three rounds which are described next. In all rounds, the user can additionally specify the PRAXIS settings.[6] An overview of the EPM strategy is schematically shown in Fig. 2. In Section Examples, EPM and the original PRAXIS routine are implemented using both the double and the long double data type.

Round 1

In Round 1, EPM applies a variant of a multistart procedure to cover global optimization. Such procedures generally produce a set of random start value vectors and evaluate an objective function at these vectors. Then a number of start value vectors with the lowest function values are selected from this initial set of vectors and local search methods are applied. EPM uses a variant of such a multistart procedure. A number of start value vectors are randomly generated based on a specific 'three-interval-uniform-sampling' design. An objective function is evaluated at these vectors using the PRAXIS algorithm. The parameter estimates of the best candidate vector of start values resulting in the smallest minimum are then locally investigated in subsequent rounds of EPM.

More precisely, in Round 1, a number $N_1 \in \mathbb{N} := \{1, 2, \ldots\}$ of start value vectors are randomly generated based on user specified, weighted primary and secondary intervals (see Fig. 1). The primary interval is determined by a center point C and a number *d* > 0 as [C – *d*, C + *d*]. Two secondary intervals surrounding the primary interval are specified by a number e > *d* as [C – e, C – *d*) and (C + *d*, C + e]. The probability for sampling a value from the primary interval is specified by a weight $0 \le w \le 1$, and consequently, the probability for sampling a value from any of the two secondary intervals is set to (1 – *w*) / 2 each. For all intervals, start values are randomly drawn using uniform distributions. That is, the density function for the primary interval is $f_p(t) := \begin{cases} 1/2d & \text{for } C - d \le t \le C + d \\ 0 & \text{else} \end{cases}$, and for the secondary intervals they are given by $f_{s1}(t) := \begin{cases} 1/(e-d) & \text{for } C - e \le t < C - d \\ 0 & \text{else} \end{cases}$ and $f_{s2}(t) := \begin{cases} 1/(e-d) & \text{for } C + d < t \le C + e \\ 0 & \text{else} \end{cases}$, respectively (*t*, a real number). The constants C, *d*, e, and *w* can be chosen from the set $\mathbb{R}$ of real numbers, within the limitations of the respective data type used in the implementation.[5] The objective function is evaluated at each of the N_1 randomly generated start value vectors by performing a minimization run with PRAXIS. The parameter estimates of the best candidate vector of start values resulting in the smallest minimum are then locally investigated in subsequent Rounds 2 and 3 of the EPM strategy.

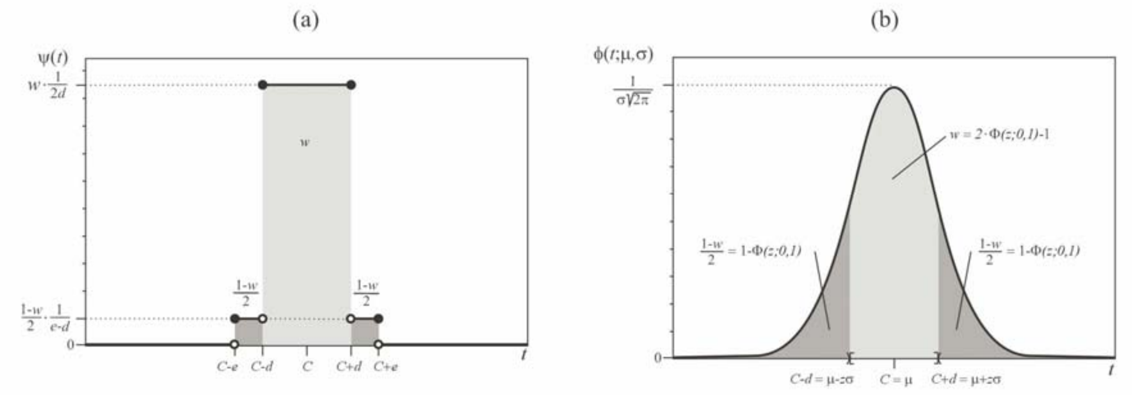


Figure 1. Random sampling of start values based on (a) the three-interval-uniform-sampling design and (b) the normal distribution

(a) *Three-interval-uniform-sampling design*: For all intervals, start values are randomly drawn using uniform distributions. Given f_p the density function for the primary interval, and f_{s1} and f_{s2} the density functions for the secondary intervals, let $\psi := w \cdot f_p + \frac{1-w}{2} \cdot \left(f_{s1} + f_{s2} \right)$.

(b) *Normal distribution*: Start values are randomly drawn using the Gaussian density $\phi(.;\mu,\sigma)$ with mean μ and standard deviation σ. To sample a value from the primary interval with a given probability (weight) *w* (strictly between zero and one), we have to solve the equation $w = 2 \cdot \Phi(z;0,1) - 1$ for z, where $\Phi(.;\mu,\sigma)$ is the Gaussian cumulative distribution function to $\phi(.;\mu,\sigma)$. Consequently, the probability for sampling a value from any of the two secondary intervals is (1 – w) / 2 each. Given this weight distribution over the primary and secondary intervals, every required spread *d* of the primary interval can be achieved by the choice of σ; simply solve the equation $d = z\sigma$ for σ (for given *d* and z).

Some remarks are in order with respect to this sampling design.

1. In a clear and straightforward manner, this design represents a natural modeling approach to capturing subjective confidences PRAXIS users may have in promising parameter regions for global minima. This modeling approach is gradual, in the sense that stronger subjective confidence regions can be covered by primary intervals while weaker ones are covered by secondary intervals, differentiated by the choices of weights. A strategy for a concrete specification of this sampling design, of course, strongly depends on the properties of an objective function and the prior knowledge about the latter. For instance, if no prior information about a function is available, a promising strategy may be the specification of a large number of start value vectors, an extensive primary interval, large secondary intervals, and weak PRAXIS settings[2] (to reduce computational efforts). If there is, however, prior knowledge of the region for a global minimum, it may be promising to focus on that region with a narrow primary interval (*d* small), negligible secondary intervals (e close to zero), and a weight close to unity.

2. If we set the weight to unity, as an important special case in practice, we obtain uniform sampling from the primary interval only; secondary intervals are no longer considered. The width of this single sampling (primary) interval can be further gradually sharpened up by limiting the constant *d* to zero. This allows a great flexibility in narrowing down promising parameter regions. An overview of the special cases of the EPM strategy is schematically shown in Fig. 3.

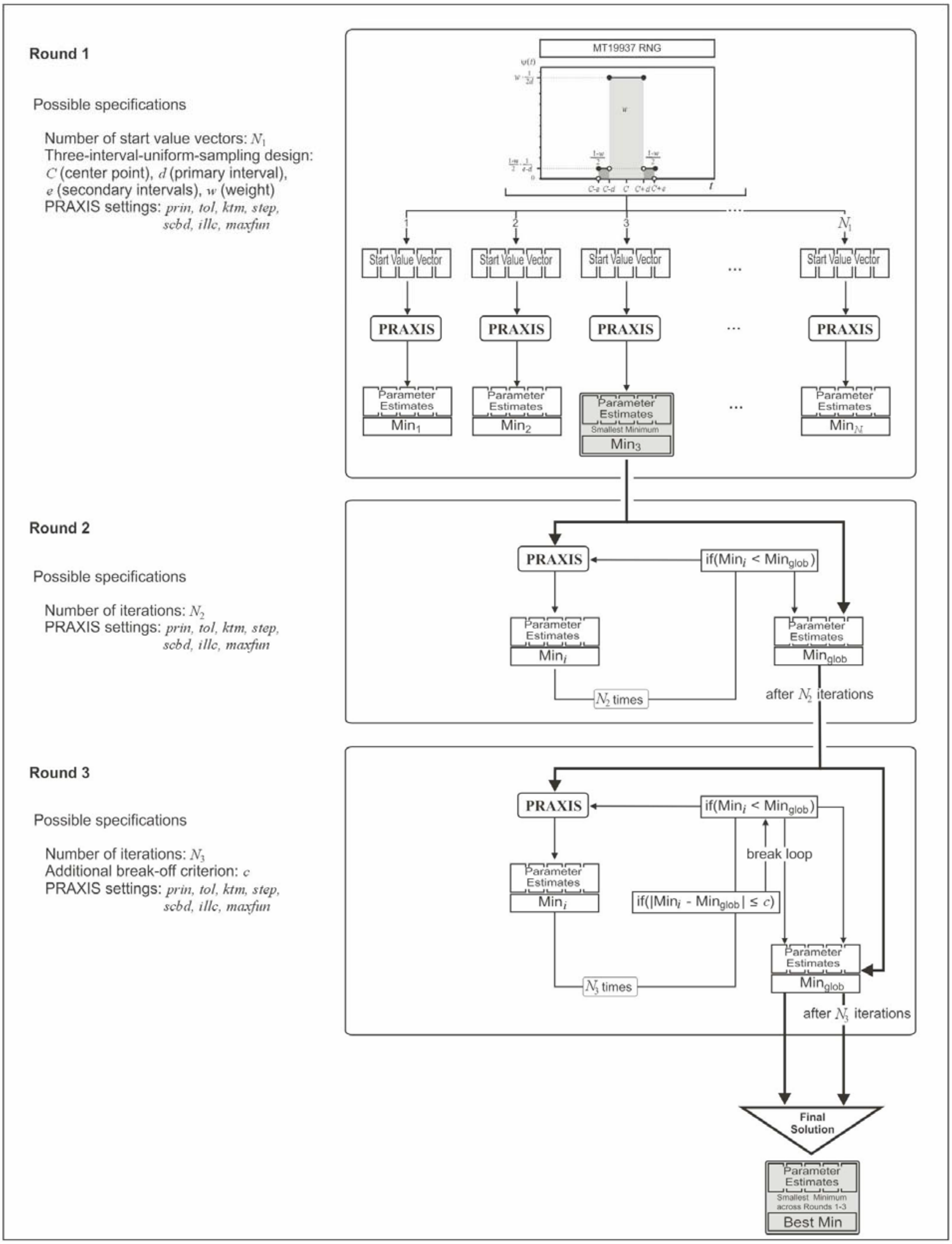


Figure 2. Overview of the EPM strategy. The EPM strategy can be implemented using both the double and the long double data type. 'MT19937 RNG' stands for the Mersenne Twister 19937 (uniform pseudo) random number generator

3. There are, of course, alternative sampling designs for the random generation of start value vectors (see Section Discussion). One alternative, for instance, is based on the normal distribution (with fixed weight distribution over the primary and secondary intervals). In this alternative, the user has to specify only two constants (mean as the center point and standard deviation as the spread) instead of having to specify the constants C, *d*, *e*, and *w* of the three-interval-uniform-sampling approach. Such alternative designs, however, generally come with a loss of modeling of subjective confidences for promising parameter regions. In general, this modeling is not that clear, straightforward, and natural anymore under such alternatives.

4. For the implementation of probabilities, EPM utilizes the *Mersenne Twister* (MT) 19937. MT19937 is a uniform pseudorandom number generator which was developed by Matsumoto and Nishimura (1998). It provides fast generation of high quality random numbers (e.g., period $2^{19937} - 1$), and rectifies many of the flaws found in older generators (e.g., Wichmann-Hill random number generator; Wichmann & Hill, 1982, 1984). MT19937 is freely available on the Internet (see Section Availability).

Round 2

The best vector of parameter estimates in Round 1 is the most promising candidate for further minimization analyses in its neighborhood. Based on these parameter estimates, in Round 2, EPM performs iterative loops successively reapplying resulting parameter estimates as start values for new runs of the PRAXIS routine. This offers the possibility of improving minimization results. The best Round 2 results are then subject to stabilization analyses in Round 3.

More precisely, in Round 2, EPM performs a number $N_2 \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ of iterative minimization runs. Starting with the best parameter estimates of Round 1, the resulting parameter estimates of each iteration are passed to the next iteration as start values for a new run of the PRAXIS routine. Though PRAXIS is internally based upon an iterative procedure, running the complete PRAXIS routine afresh with reset program settings according to this iterative paradigm generally improves the minimization results. The obtained minimum and parameter estimates can be steadily improved, allowing refined, closer approximations to the true values. The best Round 1 minimum and the corresponding vector of parameter estimates are globally stored in variables m_{glob} and v_{glob}, respectively. In each of the N_2 iterations, the current minimum m_{cur} and the corresponding vector of parameter estimates v_{cur} are globally stored in m_{glob} and v_{glob}, respectively, if $m_{cur} < m_{glob}$. This assures that the smallest minimum and the corresponding vector of parameter estimates found across all previous runs are stored, even if a slightly growing minimum is obtained for an iteration (which may occur since PRAXIS applies a minimal change in parameter estimates as the default break-off criterion; as mentioned in Subsection Additional notes, EPM allows 'uphill moves'). Round 2 finally closes with the smallest minimum found so far, and the corresponding parameter estimates. These results are then subject to stabilization analyses based on an additional break-off criterion in Round 3.

Some remarks are in order with respect to this iterative paradigm.

1. A strategy for Round 2 depends on the complexity of an objective function and the computational effort required minimizing it. The more complicated an objective function is to minimize, the more processing time is required if a larger number of iterations and

stronger PRAXIS settings are specified in Round 2. A promising strategy could be, for instance, the selection of a larger number of iterations and weak to medium PRAXIS settings.

2. Running PRAXIS afresh with reset settings according to this iterative paradigm generally improves the obtained results. In Section Examples smaller minima could be obtained for about 20 to 30 iterative runs. This was especially true for weaker PRAXIS settings.

3. The EPM strategy offers a great flexibility of usage. As a special case of EPM, the number of iterations in Round 2 can be set to zero, so the user is able to merely apply the PRAXIS routine with the multistart (global optimization) procedure of Round 1 and the stabilization stage (additional break-off criterion) in Round 3. Additionally setting the number of iterations in Round 3 to zero yields another important special case of the EPM strategy, PRAXIS enhanced for global optimization by the multistart procedure of Round 1 only. For an overview of the special cases of the EPM strategy, see Fig. 3.

Round 3

In Round 3, the iterative loops of minimization runs are supplemented by the application of an additional break-off criterion. The purpose of Round 3 is to stabilize the minimization results. The best Round 3 minimum and the corresponding parameter estimates, that is, the best minimization results across all three rounds of EPM, are output as the final solution of the EPM strategy.

More precisely, as in Round 2, in Round 3 EPM performs a maximum number $N_3 \in \mathbb{N}_0$ of iterative minimization runs. Starting with the globally stored best parameter estimates obtained in Round 2, the resulting parameter estimates of each iteration are passed to the next iteration as start values for a new minimization run. In each iteration, the current minimum (m_{cur}) and the corresponding vector of parameter estimates are globally stored if $m_{cur} < m_{glob}$. In contrast to Round 2, however, this time the iterative part is combined with the application of an additional break-off criterion. While PRAXIS internally applies a minimal change in estimated parameter vectors as the default break-off criterion (see Gegenfurtner, 1992), EPM introduces a (in general small) number $c \geq 0$ as an additional break-off criterion based on a minimal change in the approximated minimum m_{cur} of a current run and the globally stored smallest minimum m_{glob} found across all previous runs so far. If $\left| m_{cur} - m_{glob} \right| \leq c,$ EPM stops the iterative loops. As long as changes larger than the criterion c occur, upwards or downwards, EPM continues the iterative loops, until the maximum number of iterations N_3 is reached. That way the results can be stabilized in general, in the sense that the iterations stop when the obtained minimum does not vary anymore, except for minimal changes quantified by a small c. Round 3, and in particular, EPM, finally close with the smallest minimum and the corresponding parameter estimates found across all three rounds. These best results of the minimization exercise utilizing EPM are globally stored (m_{glob} and v_{glob}, respectively) and output as the final solution of the EPM strategy.

Some remarks are in order with respect to this stabilization stage.

1. A strategy for Round 3 depends on the computational effort required minimizing an objective function. The more complicated an objective function is to minimize, the more processing time is required if a larger number of iterations, a stricter break-off criterion c, and stronger PRAXIS settings are specified in Round 3. A promising strategy could be, for

instance, the selection of a smaller number of iterations, a strict break-off criterion, and strong PRAXIS settings.

2. Why do we continue the iterative procedure in Round 3, combined with the use of an additional break-off criterion not already applied in Round 2? This iterative procedure in general improves the obtained results (see Section Examples). Applying the additional break-off criterion already in Round 2 would have the disadvantage of generally stopping the iterations at an earlier stage of steady improvement; hence, yielding only suboptimal results in general. Therefore, the iterative loops are continued and the break-off criterion is first applied in Round 3, after the 'burn-in' iterations in Round 2. The break-off criterion then captures whether gained improvements stabilize in best results (except for minimal variations quantified by the criterion).

3. As a special case of EPM, the number of iterations in Round 3 can be set to zero, so the user is able to apply the PRAXIS routine with the multistart procedure of Round 1 and the iterative procedure (without the additional break-off criterion) in Round 2 only (see Fig. 3).

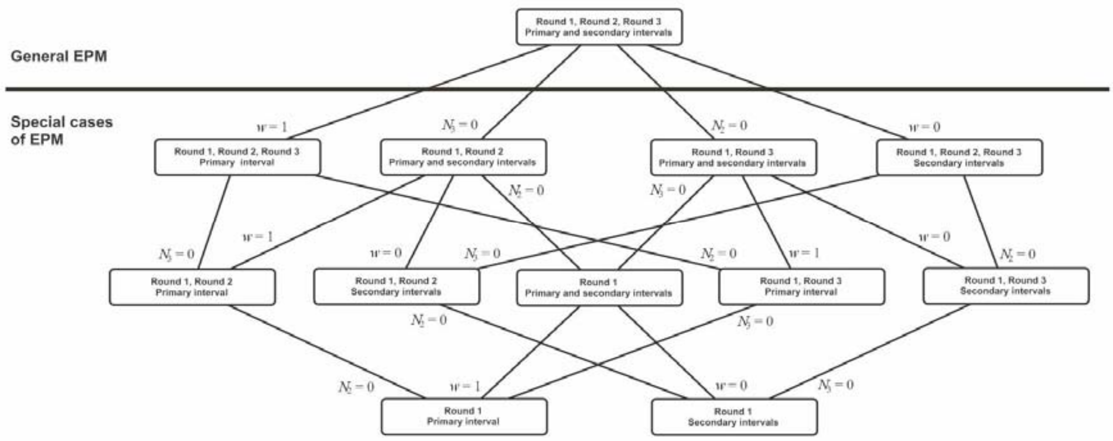


Figure 3. Diagram of the special cases of the EPM strategy

Transitivity is not explicitly depicted in the diagram; for instance, 'Round 1 / Primary interval' is a special case of 'Round 1, Round 2 / Primary and secondary intervals' (imposed restrictions: w = 1 and $N_2 = 0$). As an example of a special case of the general EPM strategy (vertex 'Round 1, Round 2, Round 3 / Primary and secondary intervals'), the number of iterations in Round 2 can be set to zero (edge $N_2 = 0$), so the user is able to merely apply the PRAXIS routine with the multistart procedure of Round 1 and the stabilization stage in Round 3 (vertex 'Round 1, Round 3 / Primary and secondary intervals'). Additionally setting the number of iterations in Round 3 to zero (edge $N_3 = 0$) yields another important special case, PRAXIS enhanced for global optimization by the multistart procedure of Round 1 only (vertex 'Round 1 / Primary and secondary intervals').

Examples

EPM improves the original PRAXIS implementation by four extensions: the automatic generation and evaluation of random start value vectors (multistart procedure for global optimization), the iterative loops to approach the true minimum and parameter values, an additional break-off criterion to stabilize minimization results, and a conversion from the double to the long double data type.

To illustrate EPM's advantages over the original PRAXIS implementation, we ran basic minimization trials using two different functions: a Rosenbrock function and a complex empirical function from psychometrics. We contrasted results obtained based on different start value vectors (for different specifications of the three-interval-uniform-sampling design), PRAXIS settings (*maxfun* = 0 and *maxfun* = 1), and double as well as long double data type versions of both the EPM and PRAXIS implementations.

Apart from the *maxfun* setting, in each case we used the following PRAXIS settings: *tol* = 1.000E-20, *ktm* = 1, *step* = 1.000, and *scbd* = 1.000. In each case the EPM strategy generated $N_1 = 20$ start value vectors (for different specifications of the sampling design), performed $N_2 = 20$ Round 2 and a maximum number $N_3 = 20$ Round 3 iterations, and applied an additional break-off criterion *c* = 1.000E-14. All computations were performed on a laptop computer with a Pentium II 366 MHz processor and 256 MB RAM running a LINUX system (SUSE LINUX 9.0).

Rosenbrock function

The Rosenbrock function in Fig. 4 is often considered as a test problem for optimization algorithms. It is a two-variable unimodal function ($x, y \in \mathbb{R}$)

$$f(x,y) := (1-x)^2 + 100 \cdot (y-x^2)^2,$$

which has a unique global minimum 0 at the point (1, 1).

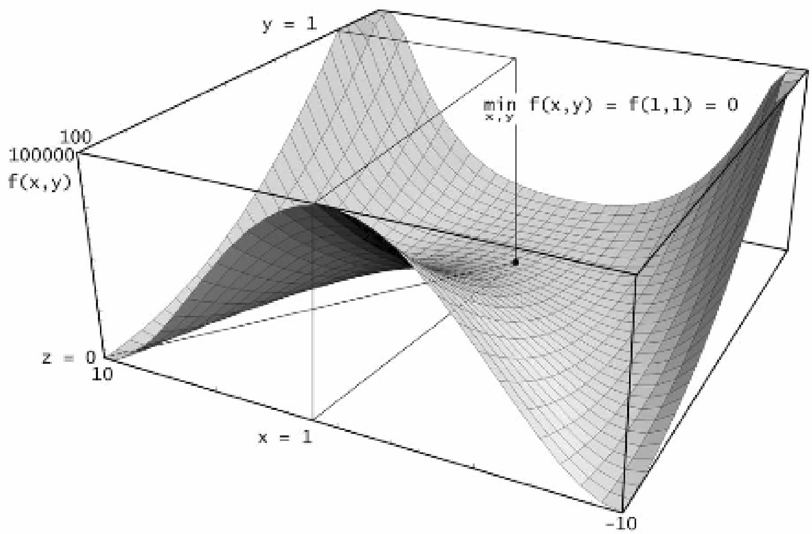


Figure 4. Plot of the Rosenbrock function

At any location other than (1, 1) this function has no local extremum. We will, however, make the following important observation (concerning the robustness of the PRAXIS routine against 'bad' choices of start values): Using vectors of start values distant to (1, 1) will cause PRAXIS to misleadingly converge and stop the minimization process at, in each case, different locations (not only far from the true location) the Rosenbrock function has no local minima at all.

For EPM's sampling design in Round 1, in each case we used the following constants: *C* = 0.000 (center point), *d* = 5.000 (primary interval), *e* = 105.000 (secondary intervals), and *w* = 0.800 (weight).

Results for the double versions: First, we consider the results obtained for the double versions of EPM and PRAXIS (see Table 1).

For the *maxfun* = 1 setting, EPM yielded a minimum of 3.420E-04. The Euclidean distance of the vector of parameter estimates to the point (1, 1) at which the global minimum is attained was 4.099E-02. For the original PRAXIS implementation we applied two different vectors of start values. First, we ran PRAXIS with the start values that resulted in the smallest minimum in EPM's Round 1; in this condition, PRAXIS yielded a minimum of 3.567E-03. The Euclidean distance of the vector of parameter estimates to the point (1, 1) was 1.367E-01. Second, we used a start value vector (8.566E+02, 3.126E+03) more distant to the location of the global minimum; in this condition, PRAXIS yielded a minimum of 5.265E+13. The Euclidean distance of the vector of parameter estimates to the coordinates of the global minimum was 3.241E+03.

For the *maxfun* = 0 setting, EPM yielded a minimum of 1.171E-18. The corresponding Euclidean distance was 2.420E-09. Again, for PRAXIS we applied two different start value vectors. Running PRAXIS with the start values that resulted in the smallest minimum in EPM's Round 1 yielded a minimum of 1.173E-18, with the corresponding Euclidean distance 2.423E-09. A start value vector (8.566E+02, 3.126E+03) distant to the point (1, 1) resulted in a minimum of 3.334E+03. The Euclidean distance of the vector of parameter estimates to (1, 1) was 3.450E+03.

Results for the long double versions: In a second step, we consider the results obtained for the long double versions of EPM and PRAXIS (see Table 1).

For *maxfun* = 1, EPM yielded a minimum of 7.246E-12. The corresponding Euclidean distance was 6.019E-06. Using the start values that resulted in the smallest minimum in EPM's Round 1, PRAXIS yielded a minimum of 2.827E-01. The Euclidean distance of the vector of parameter estimates to the point (1, 1) was 2.593E-01. A start value vector (-7.178E+01, 1.912E+00) more distant to the location of the global minimum resulted in a minimum of 2.235E+09, with the corresponding Euclidean distance 6.984E+01.

For *maxfun* = 0, EPM yielded a minimum of 9.284E-23. The corresponding Euclidean distance was 2.154E-11. Using the start values that resulted in the smallest minimum in EPM's Round 1, PRAXIS yielded a minimum of 3.899E-18. The corresponding Euclidean distance was 1.406E-10. Running PRAXIS with a second set of start values (-3.342E+00, 1.480E+00) yielded a minimum of 3.712E-14. The Euclidean distance of the vector of parameter estimates to the point (1, 1) was 3.902E-08.

Table 1. Minimization results for the Rosenbrock function

Method	Start Values[1]	Parameter Estimates	Minimum	Euclidean Distance[2]	t[3]
Double version using maxfun = 1					
EPM	5.990257138849760E-02	1.018301802095190E+00	3.420357271720240E-04	4.098569120465070E-02	<1
	1.860718519161620E+00	1.036672481830270E+00			
PRAXIS	5.990257138849760E-02	1.059722401501310E+00	3.566765242372130E-03	1.367427630163140E-01	<1
	1.860718519161620E+00	1.123011454735940E+00			
PRAXIS	8.566400000000000E+02	8.536400000000000E+02	5.264567074311800E+13	3.241054773017570E+03	<1
	3.125890000000000E+03	3.127890000000000E+03			
Double version using maxfun = 0					
EPM	2.997958114835880E+00	1.000000001082150E+00	1.171066722436550E-18	2.419769930378260E-09	<1
	-3.491414211106350E-01	1.000000002164310E+00			
PRAXIS	2.997958114835880E+00	1.000000001082150E+00	1.172603864548440E-18	2.423276700426840E-09	<1
	-3.491414211106350E-01	1.000000002168230E+00			
PRAXIS	8.566400000000000E+02	5.874360983932630E+01	3.334337413888250E+03	3.450306297249240E+03	<1
	3.125890000000000E+03	3.450823070877760E+03			
Long double version using maxfun = 1					
EPM	2.529086281435810E+00	1.000002691874950E+00	7.246190765643310E-12	6.019221859832670E-06	<1
	-7.754252305662780E-01	1.000005383757150E+00			
PRAXIS	2.529086281435810E+00	1.129664363507910E+00	2.827263823354180E-01	2.593196371856940E-01	<1
	-7.754252305662780E-01	1.224574769433720E+00			
PRAXIS	-7.178368032444260E+01	-6.878368032444260E+01	2.234725392149370E+09	6.984439495385250E+01	<1
	1.911608978852530E+00	3.911608978852530E+00			
Long double version using maxfun = 0					
EPM	4.217765044704490E+00	1.000000000009630E+00	9.283854970219120E-23	2.154224020857930E-11	<1
	1.595741204590470E+00	1.000000000019270E+00			
PRAXIS	4.217765044704490E+00	9.999999999699350E-01	3.899038824864630E-18	1.405531120156460E-10	<1
	1.595741204590470E+00	1.000000000137300E+00			
PRAXIS	-3.342191534662190E+00	1.000000009322850E+00	3.711604876093000E-14	3.901879148096910E-08	<1
	1.480348412059320E+00	1.000000037888660E+00			

Note: Apart from the *maxfun* setting, in each case we used the following PRAXIS settings: *tol* = 1.000E-20, *ktm* = 1, *step* = 1.000, and *scbd* = 1.000. In each case the EPM strategy generated $N_1 = 20$ Round 1 start value vectors, performed $N_2 = 20$ Round 2 and a maximum number $N_3 = 20$ Round 3 iterations, and applied an additional break-off criterion c = 1.000E-14. All computations were performed on a laptop computer with a Pentium II 366 MHz processor and 256 MB RAM running a LINUX system (SUSE LINUX 9.0). In each case we used the following constants for EPM's three-interval-uniform-sampling design in Round 1: C = 0.000 (center point), *d* = 5.000 (primary interval), e = 105.000 (secondary intervals), and *w* = 0.800 (weight).

[1] For EPM, the start values are the random start values that resulted in the smallest minimum in Round 1. In particular, we ran PRAXIS with these start values (obtained from Round 1 of EPM), and with a second collection of start values.

[2] Euclidean distance of the vector of parameter estimates to the point (1, 1) at which the unique global minimum 0 is attained.

[3] Processing time in seconds. Note that for EPM this is the processing time required across all three rounds.

As summarized in Table 1, EPM achieved (partially substantial) smaller minima and Euclidean distances than PRAXIS in all conditions. The same results were even found on the level of individual parameter estimates (i.e., not only for aggregate Euclidean distances). In any condition, EPM yielded individual parameter estimates closer to 1 than the original PRAXIS routine (cf. also the remark in Subsection LCMRE function). All this was especially true for the *maxfun* = 1 setting, under both the double and the long double versions. Finally, processing times were below 1 second across all conditions.

In particular, we made the following important observation concerning the robustness of the PRAXIS routine against 'bad' choices of start values. The results in Table 1 pointed out that PRAXIS was not robust against choices of start values more distant to the location of the global minimum. Using vectors of start values distant to (1, 1) caused PRAXIS to misleadingly converge and stop the minimization process at, in each case, different locations (not only far from the true location) the Rosenbrock function has no local minima at all. This turned out to be the case for both the double and the long double version of PRAXIS, under any of the settings *maxfun* = 1 and *maxfun* = 0. This may underline the importance of the multistart procedure implemented in EPM's Round 1 (cf. Table 1).

LCMRE function

The *Latent Class Model with Random Effects* (LCMRE) function is a negative log (kernel of) likelihood function which is derived based on a probit regression latent class modeling with random effects (for details, see Ünlü, 2006). It is a complex empirical function which has been originally proposed by Qu, Tan, and Kutner (1996) for the estimation of the accuracy (sensitivity and specificity) of a diagnostic test for screening individuals in biometrics (see also Hadgu & Qu, 1998; Hui & Zhou, 1998; Qu & Hadgu, 1998). Recently, Ünlü (2006) has adapted and applied this approach for the estimation of response error (careless error and lucky guess) probabilities when examinees respond to dichotomous test items in the psychometric theory of knowledge spaces.

The LCMRE function depends on the observed binary response data, and any information about extrema of this function is lacking. Such kind of 'ill-behaved' function is more difficult to minimize. However, this is close to optimization problems in practice.

Briefly, the LCMRE function is defined as

$$g(\theta;x) := -\sum_{R\in 2^Q}\left\{N(R)\cdot\ln\left(\sum_{u=0}^{1}\left[\tau_u\sum_{j=1}^{20}\left(w_j\prod_{l=1}^{m}\left[\Phi(a_{lu}+b_u t_j)^{s_l(R)}\cdot(1-\Phi(a_{lu}+b_u t_j))^{1-s_l(R)}\right]\right)\right]\right)\right\},$$

where

(1) $Q := \{I_l : l = 1,2,...,m\}$ is a set of $m \in \mathbb{N}$ dichotomously scored test items (a correct answer is scored 1 and an incorrect answer 0), and 2^Q is the power-set of Q;

(2) $R \in 2^Q$ denotes the set of test items solved by a subject (response pattern), and $s_l(R) \in \{0,1\}$ $(1 \le l \le m)$ are the *l*th entries of *R*'s representation as *m*-list of 0's and 1's;

(3) the data $x := (N(R))_{R \in 2^Q}$ are represented by the observed absolute counts $N(R) \in \mathbb{N}_0$ of the response patterns $R \in 2^Q$;

(4) $\theta := (a_{10}, a_{20}, ..., a_{m0}, a_{11}, a_{21}, ..., a_{m1}, b_0, b_1, \tau_1)$ is the parameter vector to be estimated which ranges over the parameter space $\Theta := \mathbb{R}^{2m+2} \times (0,1)$, and we have $\tau_0 + \tau_1 = 1$;

(5) $\Phi : \mathbb{R} \to [0,1]$ is the cumulative distribution function of a unit normal variate;

(6) $\{(t_j \in \mathbb{R}, w_j \in \mathbb{R}_{>0}) : j = 1, 2, \ldots, 20\}$ is a set of (known) constants obtained from the 20th order Gauss-Hermite quadrature.

Based on the classical unrestricted 2-classes latent class model (the psychological model assumed to underlie the responses of a subject; see Ünlü, 2006) we simulated a binary (of type 0/1) 1500×7 data matrix representing the response patterns of 1500 fictitious subjects to *m* = 7 test items. These data and the C source code for simulating these data are freely available from the first author. This data set was the basis for the subsequent analyses.

Results for the double versions: First, we consider the results obtained for the double versions of EPM and PRAXIS (see Table 2).

Computations using the double versions were not possible for the LCMRE function; we tried a great many start value vectors from a great many parameter regions, and a great many specifications of the EPM strategy. Apparently, due to rounding errors by limitations of the double data type, we received undefined operations, and consequently, undefined minima or parameter estimates. Thus, in this example, the conversion from the double to the long double data type was essential for making minimization of that psychometric function possible at all. Use of the long double data type accommodated this problem easily.

Results for the long double versions: In a second step, we consider the results obtained for the long double versions of EPM and PRAXIS (see Table 2).

Computations using the long double versions could be performed. For EPM's sampling design in Round 1, in each case we used the following constants: C = 0.000 (center point), *d* = 0.200 (primary interval), *e* = 0.500 (secondary intervals), and *w* = 0.900 (weight). Because we do not have any information about extrema of the LCMRE function, we computed the Euclidean distances between the vectors of $2 \cdot 7 + 3 = 17$ parameter estimates under EPM and PRAXIS.

For the *maxfun* = 1 setting, EPM yielded a minimum of 5.248E+03 and PRAXIS (using EPM's Round 1 best vector of start values) resulted in a minimum of 5.297E+03. The Euclidean distance of the EPM vector of parameter estimates to the PRAXIS vector of parameter estimates was 4.541E+01.

For the *maxfun* = 0 setting, EPM yielded a minimum of 5.0556071**81239010**E+**03** and PRAXIS (using EPM's Round 1 best vector of start values) resulted in a minimum of 5.0556071**99977240**E+**03**. The corresponding Euclidean distance was 2.469E+00.

Table 2. Minimization results for the LCMRE function

Method	Minimum	Euclidean Distance[1]	t[2]
Double version using maxfun = 1			
EPM	1.#QNAN	_	<1
PRAXIS	1.#QNAN		<1
Double version using maxfun = 0			
EPM	1.#QNAN	_	<1
PRAXIS	1.#QNAN		<1
Long double version using maxfun = 1			
EPM	5.247813256004320E+03	4.541479081653670E+01	624
PRAXIS	5.297099870217230E+03		17
Long double version using maxfun = 0			
EPM	5.055607181239010E+03	2.469200259958340E+00	845
PRAXIS	5.055607199977240E+03		18

Note: Apart from the *maxfun* setting, in each case we used the following PRAXIS settings: *tol* = 1.000E-20, *ktm* = 1, *step* = 1.000, and *scbd* = 1.000. In each case the EPM strategy generated $N_1 = 20$ Round 1 start value vectors, performed $N_2 = 20$ Round 2 and a maximum number $N_3 = 20$ Round 3 iterations, and applied an additional break-off criterion *c* = 1.000E-14. All computations were performed on a laptop computer with a Pentium II 366 MHz processor and 256 MB RAM running a LINUX system (SUSE LINUX 9.0). Computations using the double versions were not possible (indicated by '1.#QNAN' and '–'). Due to rounding errors by limitations of the double data type, we received undefined operations. Computations using the long double versions, however, could be performed. For the long double versions, in each case we used the following constants for EPM's three-interval-uniform-sampling design in Round 1: C = 0.000 (center point), *d* = 0.200 (primary interval), e = 0.500 (secondary intervals), and *w* = 0.900 (weight). In particular, we ran PRAXIS with the random start values that resulted in the smallest minimum in Round 1 of EPM.

[1]Euclidean distance of the EPM vector of parameter estimates to the PRAXIS vector of parameter estimates.

[2]Processing time in seconds. Note that for EPM this is the processing time required across all three rounds.

As summarized in Table 2, the conversion from the double to the long double data type turned out to be important for the psychometric LCMRE function; minimizing this function using the double versions of EPM and PRAXIS was not possible.

Computations, however, could be performed using the long double versions. EPM achieved (partially substantial) smaller minima than PRAXIS in both conditions *maxfun* = 1 and *maxfun* = 0. In particular, we found parameter estimates under EPM and PRAXIS clearly deviating from each other, even in the *maxfun* = 0 condition. Finally, processing times using EPM were 624 and 845 seconds in the conditions *maxfun* = 1 and *maxfun* = 0, respectively, contrary to PRAXIS with 17 (*maxfun* = 1) and 18 (*maxfun* = 0) seconds.

An important remark is in order with respect to deviations in the parameter estimates. Though the approximated minima in the condition *maxfun* = 0 differed only minimally (compared to the larger deviations in the condition *maxfun* = 1), the parameter estimates yielded a clearer difference. This, however, may be a crucial factor in practical applications. In general, a user is primarily interested in the optimizing parameter estimates. The LCMRE function, for instance, is used for maximum likelihood estimation of response error probabilities for dichotomous test items in psychometrics. These probabilities, however, are functions of the parameter estimates (see Ünlü, 2006). In particular, deviations in the parameter estimates may occur and be empirically important (in the current example, resulting in better maximum likelihood estimates for the response error rates of the underlying classical unrestricted 2-classes latent class model), even if the approximated minima differ only slightly. Hence, from an empirical point of view, (even little) improvements in approximated minima may be desirable, and thus longer processing times may be acceptable.

Discussion

Summary

We have introduced, implemented and applied the three-round EPM strategy to improve the original PRAXIS implementation by four extensions (cf. Fig. 2): a multistart procedure to cover global optimization, iterative loops to approach the true minimum and parameter values, an additional break-off criterion to stabilize minimization results, and a conversion from the double to the long double data type to increase computational precision for complex optimization problems. We have also seen that this strategy offers a number of important special cases in practice (cf. Fig. 3), and thus provides the user with a great flexibility in the actual minimization exercise utilizing the EPM extension.

EPM's advantages over the original PRAXIS implementation have been illustrated using two different functions: a 'well-behaved' Rosenbrock function (see Fig. 4) for which the global minimum and its corresponding coordinates are known, and an 'ill-behaved' complex empirical function from psychometrics for which any information about extrema is lacking. For both functions, across all conditions, EPM improved (partially substantial) the minimization results obtained using merely the original PRAXIS implementation (see Tables 1 and 2). Processing times, however, increased for the LCMRE function using EPM. Nevertheless, in practical applications, (even little) improvements in approximated minima may be worthwhile (e.g., yielding better maximum likelihood estimates for empirically interpreted functions of the parameter estimates), and thus longer processing times may be acceptable.

We have observed that a not necessarily trivial (see Subsection Basic motivations for an Extended Principal axis Minimization (EPM)) conversion from the double to the long double data type can significantly improve computational precision. As demonstrated with the LCMRE function, for some complex empirical minimization problems the long double data type may be essential for performing principal axis minimization based on the original PRAXIS implementation. For such problems, computational values may exceed the limitations of the double data type, resulting in undefined operations (e.g., division by zero), and consequently, yielding undefined results (see Table 2). Moreover, for the Rosenbrock function we have observed that the original PRAXIS implementation was not robust against 'bad' choices of start values. Using start value vectors more distant to the location of the global minimum caused PRAXIS to misleadingly converge and stop the minimization process at, in each case, different locations not only far from the true location, but at which the Rosenbrock function has no local minima at all. The EPM strategy easily accommodated these two observations by implementing double as well as long double versions (cf. Table 2) and a multistart procedure for global optimization (cf. Table 1), respectively.

Further extensions and modifications

As mentioned in Subsection Round 1, there are, of course, other sampling designs for the random generation of start value vectors than the three-interval-uniform-sampling approach applied in this paper. One alternative could be based on the family of two-parameter normal distributions (Gaussian densities)

$$\phi:\mathbb{R}\to[0,+\infty),\ t\mapsto\phi(t;\mu,\sigma):=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right)\quad(\mu\in\mathbb{R},\sigma\in\mathbb{R}_{>0})$$

(cf. Fig. 1). In this alternative, the user could specify two constants: the mean as the center point (i.e., $C:=\mu$) and the standard deviation as the spread (i.e., $d:=\sigma$). Then start values could be sampled based on this normal distribution. Compared to the three-interval-uniform-sampling design, the primary interval (interpreted as the stronger subjective confidence region) could be defined as $[\mu-\sigma,\mu+\sigma]$, while the secondary intervals (interpreted as the weaker subjective confidence regions) could be represented by $(-\infty,\mu-\sigma)$ and $(\mu+\sigma,+\infty)$. Under these conditions, the probability for sampling a start value from the primary interval is $2\cdot\Phi(1;0,1)-1\approx 0.682$, and consequently, the probability for sampling a start value from any of the two secondary intervals is $1-\Phi(1;0,1)\approx 0.159$ each ($\Phi(.;\mu,\sigma)$ Gaussian cumulative distribution function corresponding to $\phi(.;\mu,\sigma)$). These probabilities, however, are the same for any specification of the primary and secondary intervals, that is, for any choice of the constants (center point) $C:=\mu$ and (spread) $d:=\sigma$.

In order to imitate the weight of the three-interval-uniform-sampling design allowing a flexible distribution of unit mass over the primary and secondary intervals, one could, for instance, more generally define the primary interval as $[\mu-z\sigma,\mu+z\sigma]$ and the secondary intervals by $(-\infty,\mu-z\sigma)$ and $(\mu+z\sigma,+\infty)$; here z is any positive real number. The center point again is $C:=\mu$, the spread however is given by $d:=z\sigma$. Then the probability for sampling a start value from the primary interval is $2\cdot\Phi(z;0,1)-1$, while the probability for sampling a start value from any of the two secondary intervals is $1-\Phi(z;0,1)$ each. Since $\lim_{t\to+\infty}\Phi(t;0,1)=1$ and $\lim_{t\to0+}\Phi(t;0,1)=1/2$, we can control, by the choice of z, the probabilities for sampling start values from the primary and secondary intervals. Given any such weight distribution over the intervals, every required width (i.e., spread *d*) of the primary interval can be achieved by the choice of σ (cf. Fig. 1).

Moreover, what has been said for the normal distribution could also be applied (maybe with minor modifications) with any family of two-parameter probability distributions in which one parameter represents a location parameter (to place the distribution at a location on a parameter axis) and the other a shape parameter (to control the shape/mass of the distribution around that location). Finally, these alternatives, and the three-interval-uniform-sampling approach, can be further generalized to respectively include a normal distribution (a two-parameter probability distribution) and a three-interval-uniform-sampling design for each model parameter separately. In this case, constants of the sampling designs are indexed by the model parameters.

We have also outlined an analysis of the different factors varied in this paper. Future work may address systematic variations of other factors as well; for instance, in the case of empirical functions, variations of sample size (especially for small sample sizes) and model complexity, or other PRAXIS settings (than *maxfun*).

Conclusion

The EPM strategy represents a wrapper around the function minimization routine PRAXIS. It effectively improves the original PRAXIS implementation by techniques that are applicable with other routines as well. Future work could include implementing this strategy for other routines in use; in particular, it would be interesting to see whether it provides similar advantages with, for instance, the functions for minimization in Mathematica and Matlab.

Availability

The C (ANSI C 99) source files, including EPM, PRAXIS, and MT19937, for double as well as long double versions, are freely available from the authors. Electronic mail may be sent to ali.uenlue@math.uni-augsburg.de or michael.kickmeier@uni-graz.at. The original PRAXIS source by Karl Gegenfurtner can be found at http://archives.math.utk.edu/software /msdos/numerical.analysis/praxis/.html (retrieved July 19, 2005).

The original source of the MT19937 pseudorandom number generator, which is a component of the EPM implementation, can be found at http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html (retrieved July 19, 2005). The C source file for simulating data using the classical unrestricted 2-classes latent class model and the data set simulated for the analyses in this paper can also be freely obtained from the first author by electronic mail.

References

1. Back, T. and Schwefel, H.-P. **An overview of evolutionary algorithms for parameter optimization,** Evolutionary Computation, 1,1993, pp.1-23
2. Brent, R.P. **Algorithms for minimization without derivatives,** Englewood Cliffs, NJ, Prentice-Hall, 1973
3. Carcione, L., Mould, J., Pereyra, V., Powell, D. and Wojcik, G. **Nonlinear inversion of piezoelectrical transducer impedance data,** Journal of Computational Acoustics, 9, 2001, pp. 899-910
4. Doignon, J.-P. and Falmagne, J.-Cl. **Spaces for the assessment of knowledge,** International Journal of Man-Machine Studies, 23, 1985, pp. 175-196
5. Doignon, J.-P. and Falmagne, J.-Cl. **Knowledge spaces,** Berlin, Springer, 1999
6. D'Zmura, M., Rinner, O. and Gegenfurtner, K.R. **The colors seen behind transparent filters,** Perception, 29, 2000, pp. 911-926
7. Erdfelder, E. and Buchner, A. **Process-dissociation measurement models: Threshold theory or detection theory?,** Journal of Experimental Psychology: General, 127, 1998, pp. 83-96
8. Fries, S. **Empirical validation of a Markovian learning model for knowledge structures,** Journal of Mathematical Psychology, 41, 1997, pp. 65-70
9. Gegenfurtner, K.R. **PRAXIS: Brent's algorithm for function minimization,** Behavior Research Methods, Instruments, and Computers, 24, 1992, pp. 560-564
10. Hadgu, A. and Qu, Y. **A biomedical application of latent class models with random effects,** Applied Statistics, 47, 1998, pp. 603-616
11. Hartinger, A. **Do generalized expected utility theories capture persisting properties of individual decision makers?**, Acta Psychologica, 102, 1999, pp. 21-42
12. Heller, J. **On the psychophysics of binocular space perception,** Journal of Mathematical Psychology, 41, 1997, pp. 29-43
13. Heller, J. **Mittenbildung bei achromatischen Farben: Das klassische Experiment von Plateau [Bisection with achromatic colors: Plateau's classical experiment],** Zeitschrift für Experimentelle Psychologie, 48, 2001, pp. 259-271

14. Hui, S.L. and Zhou, X.H. **Evaluation of diagnostic tests without gold standards,** Statistical Methods in Medical Research, 7, 1998, pp. 354-370
15. Hwang, J.S. and Tien, C. **XRAYSCAN: An indexing program considering dense spurious peaks in an optimization method,** Chinese Journal of Physics, 34, 1996, pp. 47-57
16. Leroy, A., Mozer, P., Payan, Y. and Troccaz, J. **Rigid registration of freehand 3D ultrasound and CT-scan kidney images,** in MICCAI 2004, PT 1, "Proceedings Lecture Notes in Computer Science", Vol. 3216, 2004, pp. 837-844
17. Matsumoto, M. and Nishimura, T. **Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator,** ACM Transactions on Modeling and Computer Simulation, 8, 1998, pp. 3-30
18. McClelland, J.L. and Chappell, M. **Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory,** Psychological Review, 105, 1998, pp. 724-760
19. McCullough, B.D. and Vinod, H.D. **The numerical reliability of econometric software,** Journal of Economic Literature, 37, 1999, pp. 633-665
20. Pintér, J.D. **Global optimization in action,** Dordrecht, Kluwer, 1995
21. Powell, M.J.D. **An efficient method for finding the minimum of a function in several variables without calculating derivatives,** Computer Journal, 7, 1964, pp. 155-162
22. Price, W.L. **Global optimization by controlled random search,** Journal of Optimization Theory and Applications, 40, 1983, pp. 333-348
23. Qu, Y. and Hadgu, A. **A model for evaluating sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test,** Journal of the American Statistical Association, 93, 1998, pp. 920-928
24. Qu, Y., Tan, M. and Kutner, M.H. **Random effects models in latent class analysis for evaluating accuracy of diagnostic tests,** Biometrics, 52, 1996, pp. 797-810
25. Regenwetter, M., Falmagne, J.-Cl. and Grofman, B. **A stochastic model of preference change and its application to 1992 presidential election panel data,** Psychological Review, 106, 1999, pp. 362-384
26. Ren, H.P., Chen, S.Z., Wu, W.K. and Yang, H. **PET and MR three-dimensional image registration based on second-order mutual information,** World Journal of Nuclear Medicine, 1(suppl. 2), 293, 2002
27. Ren, H.P., Wu, W.K., Yang, H. and Chen, S.Z. **High accuracy registration for three-dimensional brain images of PET and MRI,** Chinese Journal of Nuclear Medicine, 22, 2002, pp. 168-170
28. Törn, A. and Zilinskas, A. **Global optimization,** Berlin, Springer, 1989
29. Tubic, D., Zaccarin, A., Beaulieu, L. and Pouliot, J. **Automated seed detection and three-dimensional reconstruction. II. Reconstruction of permanent prostate implants using simulated annealing,** Medical Physics, 28, 2001, pp. 2272-2279
30. Ünlü, A. **Estimation of careless error and lucky guess probabilities for dichotomous test items: A psychometric application of a biometric latent class model with random effects,** Journal of Mathematical Psychology, 50, 2006, pp. 309-328
31. Wichmann, B.A. and Hill, I.D. **Algorithm AS 183: An efficient and portable pseudo-random number generator,** Applied Statistics, 31, 1982, pp. 188-190
32. Wichmann, B.A. and Hill, I.D. **Correction: Algorithm AS 183: An efficient and portable pseudo-random number generator,** Applied Statistics, 33, 1984, p.123,
33. Winker, P. **Optimization heuristics in econometrics,** Chichester, Wiley, 2001
34. Winker, P. and Gilli, M. **Applications of optimization heuristics to estimation and modelling problems,** Computational Statistics & Data Analysis, 47, 2004, pp. 211-223
35. Woelk, K. **Torus factor - the relationship between radiofrequency field and radial position in toroid cavity probes,** Journal of Magnetic Resonance, 146, 2000, pp. 157-164

---

[1] This work was supported by the Austrian Science Fund (FWF) project grant P17071-N04 to Dietrich Albert.

[2] Correspondence to: Ali Ünlü, Institute of Mathematics, University of Augsburg, D-86135 Augsburg, Germany Phone: +49 821 598-2236, Fax: +49 821 598-2200, Email: ali.uenlue@math.uni-augsburg.de

[3] Dr. Ali Ünlü is an assistant professor in the Department of Computer Oriented Statistics and Data Analysis of the Institute of Mathematics at the University of Augsburg, Germany. In the period 1993-2000, he studied mathematics and physics at the University of Technology Darmstadt, Germany. He received his diplom in mathematics in 2000

from the University of Technology Darmstadt. In the period 2000-2006, he studied quantitative psychology (primarily specializing in probabilistic knowledge space theory and item response theory) as a postgraduate and postdoctoral research assistant in the Cognitive Science Section of the Institute of Psychology at the University of Graz, Austria. In the period 2002-2004, he completed a doctoral study in applied mathematics at the Graz University of Technology, Austria, receiving the doctor of technical sciences with an interdisciplinary dissertation on statistical model selection in the psychometric theory of knowledge spaces.

His research focuses on the mathematical and statistical foundations in the quantitative behavioral sciences. His research interests are in the areas of knowledge space theory, nonparametric item response theory, latent class analysis, model selection, numerical optimization, data analysis, information visualization, and software development.

In October 2007, he has been appointed associate editor of the European Journal of Pure and Applied Mathematics.

Representative publications:

Ünlü, A. **Nonparametric item response theory axioms and properties under nonlinearity and their exemplification with knowledge space theory,** Journal of Mathematical Psychology, 51, 2007, pp. 383–400

Ünlü, A. **Estimation of careless error and lucky guess probabilities for dichotomous test items: A psychometric application of a biometric latent class model with random effects,** Journal of Mathematical Psychology, 50, 2006, pp. 309–328

Ünlü, A., & Albert, D. **The Correlational Agreement Coefficient CA - a mathematical analysis of a descriptive goodness-of-fit measure,** Mathematical Social Sciences, 48, 2004, pp. 281–314

Ünlü, A. (in press). **A note on monotone likelihood ratio of the total score variable in unidimensional item response theory,** British Journal of Mathematical and Statistical Psychology

Ünlü, A., & Sargin, A. (in press). **Interactive visualization of assessment data: The software package Mondrian,** Applied Psychological Measurement

[4] Michael D. Kickmeier-Rust is a psychologist and software developer. He is a senior project coordinator in the Cognitive Science Section of the Department of Psychology at the University of Graz, Austria. In the period 1998-2005, he studied psychology at the University of Graz and received his master degree in cognitive psychology and industrial psychology.

His research and development activities are concentrating primarily on intelligent, adaptive educational systems. In particular, he is working on the evolution of psycho-pedagogical frameworks and models of adaptivity on the macro and micro levels. This work includes individualized navigation within learning environments, individualized presentation of educational material, as well as non-invasive assessment and interventions. Since 2001, he is with the Cognitive Science Section at the Department of Psychology, University of Graz and working in international, cross-discipline research and development projects. Since 2006, his very focus is on the psychological aspects of learning with immersive digital computer games and the integration of intelligent educational technology in game-based learning. Since 2008, he is senior project coordinator of the FP7 ICT project 80Days that is dealing with game-based learning. Related fields of interest are human-computer or man-machine interaction and specific fields of aviation psychology, dealing with education and training.

Since 2004 he is vice-chair of the Austria Aviation Psychology Association.

Representative publications:

Ebner, M., Kickmeier-Rust, M., & Holzinger, A. **Utilizing Wiki-Systems in higher education classes: a chance for universal access?**, Universal Access in the Information Society, 2008

Kickmeier-Rust, M.D., Peirce, N., Conlan, O., Schwarz, D., Verpoorten, D., Albert, D. **Immersive Digital Games: The Interfaces for Next-Generation E-Learning?** In C. Stephanidis (Ed.), "Universal Access in Human-Computer Interaction. Applications and Services" (pp. 647-656). Lecture Notes in Computer Science, 4556/2007. Berlin: Springer, 2007

Kickmeier-Rust, M.D., Albert, D., Hockemeyer, C., & Augustin, T. **Not breaking the narrative: Individualized Competence Assessment in Educational Games,** In "Proceedings of the European Conference on Games-based Learning (ECGBL)", October 25-26, Paisley, Scotland, 2007

Kickmeier-Rust, M.D., Ebner, M., & Holzinger, A. **Wikis: Do they need usability engineering?,** In M. Pohl, A. Holzinger, R. Motschnig, & C. Swertz (Eds.), "M3 – Interdisciplinary aspects on digital media & education" (pp. 137-144). Vienna: Österreichische Computer Gesellschaft, 2006

Kickmeier-Rust, M.D., & Vormayr, E.M. (Eds.). **Lecture Notes of ISAP'5,** International Summer School on Aviation Psychology (Vol. 2), 2005

[5] The double data type has a range of $\pm$ 1.798E+308 and a precision of 15 decimal places. The long double data type has a range of $\pm$ 1.1E+4932 and a precision of 19 decimal places.

[6] PRAXIS settings: *prin* controls the printed output from the routine; *tol* is the tolerance used for the default break-off criterion; *ktm* specifies the number of times the default criterion must be fulfilled to stop the minimization process; *step* is a step-size variable; *scbd* is a scaling variable; *illc* specifies whether the problem is difficult to minimize (ill conditioned) – PRAXIS automatically sets *illc* to true if it finds the problem to be ill conditioned; *maxfun* specifies the maximum number of internal calls to the objective function – a value 0 indicates no limit on the number of function calls. For a more detailed explanation of the PRAXIS settings, see Gegenfurtner (1992).

CORRELATIONS BETWEEN CAPITAL MARKET DEVELOPMENT AND ECONOMIC GROWTH: THE CASE OF ROMANIA[1]

Laura OBREJA BRASOVEANU
PhD, Associate Professor, Finance Departament,
University of Economics, Bucharest, Romania

E-mail: laura_obreja@yahoo.com

Victor DRAGOTA[2]
PhD, University Professor, Finance Departament,
University of Economics, Bucharest, Romania

E-mail: victordragota@yahoo.com, victor.dragota@fin.ase.ro

Delia CATARAMA
PhD Candidate, Assistant Professor, Finance Departament,
University of Economics, Bucharest, Romania

E-mail: delia.catarama@fin.ase.ro

Andreea SEMENESCU[3]
PhD Candidate, University Assistant, Finance Departament,
University of Economics, Bucharest, Romania

E-mail: asemenescu@yahoo.com

Abstract: *In the literature on endogenous growth, the link between capital markets development and economic growth has received much attention. Although there are many studies regarding this aspect, approaches on emergent ex-communist countries' economies, especially for Romania, are very few comparatively to the general cases.*
Our paper examines the correlation between capital market development and economic growth in Romania using a regression function and VAR models. The results show that the capital market development is positively correlated with economic growth, with feed-back effect, but the strongest link is from economic growth to capital market, suggesting that financial development follows economic growth, economic growth determining financial institutions to change and develop.

Key words: *time-series; political economy; economic growth; capital market development*

1. Introduction

In the recent financial literature on endogenous growth, the relationship between capital markets development and economic growth has received much attention (see King and Levine, 1993; Levine, 1997; Rajan and Zingales, 1998; Filler, Hanousek, and Campos, 1999; Arestis, Demetriades, and Luintel, 2001; Calderon and Liu, 2002, Carlin and Mayer, 2003). In this context, King and Levine (1993) state that the level of financial intermediation is a good predictor for economic growth rate, capital accumulation and productivity. In the same context, Carlin and Mayer (2003) concluded that there is a strong relationship between the structure of countries' financial system and economic growth.

Garretsen, Lensink and Sterken (2004) found out a causal relationship between economic growth and financial markets development: a 1% improvement of economic growth determines a 0.4% rise of market capitalization/GDP ratio. Yet, according to their results, market capitalization/GDP ratio does not represent a significant determinant of the economic growth.

Beck, Lundberg and Majnoni (2006), also, found a positive correlation between capital market development (measured by a dummy variable computed to reflect if the market capitalization exceeds 13,5% of GDP) and economic growth.

Bose(2005) offers a theoretical financial model that explains the positive correlation between stock market development and economic growth; the model is based on the hypothesis that for levels of GDP per capita higher than a certain threshold the information costs become lower than bankruptcy costs, determining the development of capital markets. Hence, it is explained why stock markets appeared late after banks.

Beckaert, Harvey and Lundblad (2005) analyzed financial liberalization as a special case of capital market development and determined that equity market liberalizations, on average, led to a 1% increase in annual real economic growth.

Studying the link between domestic stock market development and internationalization, Claessens, Klingebiel and Schmukler (2006) using a panel data technique concluded that domestic stock market development as well as stock market internationalization are positively influenced by the log of GDP per capita, the stock market liberalization, the capital account liberalization and the country growth opportunities and negatively influenced by the government deficit/GDP ratio.

Minier (2003) analyzed the influence of the stock market dimension on economic development by regression tree techniques; he found evidence that the positive influence of stock market development on economic growth held only for developed stock markets in terms of turnover, in the case of underdeveloped stock markets the influence is negative.

Ergungor (2006) analyzed the impact of financial structure on the economic growth on the period 1980-1995; he concluded that in countries with inflexible judicial systems the stronger impact on economic growth is generated by the development of the bank-system, whereas in countries with greater flexibility of judicial systems the development of the capital market had a stronger influence.

Studies on the relation between capital market development and economic growth in different countries were performed. Nieuwerburgh, Buelens and Cuyvers (2006) analysed the long-run relationship between stock market development (measured by market capitalization and number of listed shares) and economic growth (measured as a logarithmic difference of GDP per capita) in Belgium. They performed Granger causality tests and emphasized that stock market development determined economic growth in Belgium especially in the period 1873-1935, but also on the entire analyzed period (1800-2000) with variations in time dues to institutional changes affecting the stock exchange.

Hondroyiannis, Lolos and Papapetrou (2005) studied the case of Greece (1986-1999); they found out that the relationship between economic growth and capital market development is bi-directional.

Studying the effect of different components of financial systems on economic growth in Taiwan, Korea and Japan, Liu and Hsu (2006) emphasized the positive effect of stock market development (measured by market capitalization as percentage of GDP, turnover as percentage in GDP and stock return) on economic growth.

Bolbol, Fatheldin, and Omran (2005) analyzed the effect of financial markets (measured by the ratio of market capitalization on GDP and the turnover ratio) on total factor productivity and growth (the per capita GDP growth rate) in Egypt (1974-2002); they demonstrated that capital market development had a positive influence on factor productivity and growth.

Ben Naceur and Ghazouani (2007), studying the influence of stock markets and bank system development on economic growth on a sample of 11 Arab countries, concluded that financial development could negatively influence the economic growth in countries with underdeveloped financial systems; they stressed the role of building a sound financial system.

In the context of UE enlargement, an analysis of the relationship between capital markets development and economic growth could explain why different countries reach different economic growth rates, and could find solutions in order to stimulate the process of economic growth through capital market using public policy instruments. Related to this issue, even there are many studies regarding developed countries, approaches on East-European ex-communist countries' economies are very few relatively to developed countries cases.

Romanian capital market had developed slowly starting from 1995. Moreover, several years after 1989 Romania had negative economic growth rates (the real rate of GDP growth). Only since 2000 Romania had positive economic growth rates accompanied by the development of the financial system; these particular aspects could alter the relationship between economic growth and capital market development, and more specifically the conclusion on whether capital market development is a good predictor for economic growth rates. This is the reason why the starting point of our study is the year 2000.

Our paper examines the correlation between capital market development and economic growth for Romanian case, considering quarterly data for the period 2000-2006, using a regression function and VAR models for explaining the relationship between market development (size and liquidity of the capital market) and economic growth. This paper tests if the predicts of endogenous growth theory that capital market development positively affects the rate of economic growth is also true in case of Romania.

This paper is structured as follows. In Section 2, we define the measures for capital markets development and economic growth, relatively to the mainstream of financial literature. Section 3 presents the database and methodology that we used. Section 4 presents the main empirical results and Section 5 contains the concluding remarks.

2. Theoretical analysis

There are several discussions about the relationship between the development of the financial system and the economic growth. The literature focuses on the financial system's components, the banking sector or the capital market, that influence economic growth.

Graff (1999) stated that there are four possibilities concerning the causal relationship between financial development and economic growth:

(1) financial development and economic growth are not causally related. An example of this type of relation could be found in the development of modern economy, in Europe, in the 17th Century. In this case, the economic growth was the result of real factors, while the financial development was the result of financial institutions development.

(2) financial development follows economic growth. In this context, economic growth causes financial institutions to change and to develop, so as both the financial and credit market grow.

(3) financial development is a cause of economic growth. In this case, there could be identified two possibililies, respectively: (a) financial development is a precondition for economic growth; (b) financial development actively encourages economic growth (see, e.g. Thornton, 1995). Provided that there are no real impediments to economic growth, mature financial systems can cause high and sustained rates of economic growth (see, Rousseau and Sylla, 2001).

(4) financial development is an impediment to economic growth. Similar to the previous posibility, causality runs from financial development to real development, but the focus lies on potentially destabilizing effects of financial overtrading and crises (see, e.g. Stiglitz, 2002) rather than on the efficient functioning of the financial system. This view considers the financial system as inherently unstable.

The economic growth is a complex process that is influenced by more factors, other than the capital market development. Moreover, capital market development is the results of many influence factors. There are several interdependencies between these factors, which makes it difficult to establish and isolate the causal relation between the economic growth and the capital market development.

There are several empirical studies that analyse the correlation between the economic growth and the financial development. Calderon and Liu (2002), studying the direction of this causality, conclude that, as a general trend, the financial development generates economic growth, the causal relation being stronger in the emergent countries and being explained by two channels: the fast capital accumulation and the growth of productivity. Rajan and Zingales (1998) emphasize that the financial development is a prediction element for the economic growth, because the capital market reflects the present value of the future growth opportunities. The ex-ante development of the financial markets facilitates the ex-post economic growth of the external financing dependent sectors. Levine (1997) and Levine and Zevros (1998) consider that the capital market's liquidity is a good predictor of the GDP per capita growth and of the physical capital and productivity growth, but other indicators of the capital market development such as volatility, size and international integration are not significant for explaining economic growth. Carlin and Mayer (2003) analyse the link between financial systems and economic growth for developed countries and reveal a link between financial system and type of economic activities which can influence the economic growth. Arestis, Demetriades and Luintel (2001), use the autoregressive vector for an empirical analysis on five developed economies; their study concludes that the capital market has effects on the economic growth, but the impact of the banking sector is stronger. Filer, Hanousek and Campos (1999) notice that capital markets include the future growth rates in current prices, especially in the developed countries, which is a result consistent with the efficient markets hypothesis.

Although in cross-country analyses it can be depicted a correlation between the financial development and the economic growth, we can question if, in the emergent countries, an active capital market is a stimulating factor for the economic growth. An affirmative answer would imply an important role of the public policies and international aid targeted at introducing and maintaining the capital market structures (see Filer, Hanousek and Campos, 1999).

The previous empirical studies assessed and quantified the correlation between capital market development and economic growth using different techniques[4]. The variables used in these studies can be grouped in the following categories:

(A) Capital market variables:

- **size variables**: market capitalization/GDP ratio (used by Filer, Hanousek and Campos, 1999), the logarithm of the stock market capitalization ratio[5] (used by Arestis, Demetriades and Luintel, 2001);
- **liquidity variables**: turnover ratio[6] and value traded ratio[7] (used by Levine and Zevros, 1998)
- **volatility variables**: eight-quarter moving standard deviation of the end-of-quarter change of stock market prices (used by Arestis, Demetriades and Luintel, 2001)

(B) Economic growth variables: logarithm of real GDP (used by Arestis, Demetriades and Luintel, 2001), GDP growth rate (used by Baier, Dwyer Jr. and Tamura, 2004), GDP per capita growth rate (used by King and Levine, 1993).

In this article, we aim at realizing a country-case study for Romania for the period 2000 – 2006. We use quarterly data to identify the existence of a correlation between the development of the capital market and the economic growth.

3. Data and Methodology

We analyze the link between capital market and the economic growth in Romania on quarterly data from 2000: 1 to 2006: 2, meaning 26 observations. In Table 1 we present the variables used to characterize the Romanian capital market and economic growth.

Table 1. Variables

Variables		**Symbol**
Capital market variables	Market capitalization	MCN
		MCR
	Number of listed shares	NLS
	Trading volume	TVN
		TVR
	Liquidity proxy	NLI
		VLI
	Bucharest Stock Exchange Index	BET
		BET-C
		BETCR
Economic growth variables	Gross Domestic Product	GDP
		GDPR
		GDPRG

The capital market indicators can be classified in three categories:
(i) **size indicators**: market capitalization in nominal (MCN) and real (MCR) values, number of listed shares (NLS);
(ii) **liquidity indicators**: trading volume in nominal (TVN) and real (TVR) values, number of traded shares (TLS) and two proxies for liquidity (NLI and VLI);
(iii) **return indicators**: the BET and BET-C indexes, which are the Bucharest Stock Exchange official indexes.

We selected for our study the following indicators of capital market:
(i) the real market capitalisation, computed in order to eliminate the inflation effect;
(ii) the real trade volume, reflected by the indicator log(TVR) as a measure of the market size, but also of the market liquidity;
(iii) a liquidity proxy, computed as the number of traded shares divided by the number of listed shares.

In order to analyze the correlations between economic growth and capital market development we used linear regression and vector autoregressive methods.

Several regressions were selected:

(R 1) $\log(GDPR_t) = a + b\cdot\log(MCR_t) + c\cdot D_{1t} + \varepsilon_t$

(R 2) $\log(GDPR_t) = a + b\cdot\log(TVR_{t-2}) + c\cdot D_{1t} + \varepsilon_t$

(R 3) $\log(TVR_t) = a + b\cdot\log(GDPR_t) + c\cdot\log(GDPR_{t-1}) + d\cdot D_{2t} + \varepsilon_t$

(R 4) $\log(GDPR_t) = a + b\cdot\log(GDPR_{t-2}) + c\cdot\log(TVR_{t-2}) + d\cdot\log(TVR_t) + e\cdot D_{1t} + \varepsilon_t$

(R 5) $\log(TVR_t) = a + b\cdot\log(GDPR_t) + c\cdot\log(GDPR_{t-2}) + d\cdot\log(TVR_{t-2}) + e\cdot D_{1t} + \varepsilon_t$

Several forms of vector autoregressive were selected:

(VAR 1) log(GDPR), log(TVR), with 2 lags for endogenous variables

(VAR 2) GDPRG, NLI, with 2 lags for endogenous variables.

The lags are determined using the Haugh statistic.
In Section 4 the main numerical results are presented.

4. Empirical results

First, we tested regressions (R1), (R2), (R3). In Table 2 we present the results.

Table 2. Results for regressions (R1), (R2), (R3)

Independent variables	log(GDPR)	log(GDPR)	log(TVRt)
	Regression (R 1)	**Regression (R 2)**	**Regression (R 3)**
log(MCRt)	0.153471** [7.573879]		
$\log(TVR_{t-2})$		0.133338** [5.890239]	
log(GDPRt)			1.624155** [3.537126]
log(GDPRt-1)			4.411427** [8.284857]
D_1t	-0.345995** [-6.937547]	-0.296155** [-5.202606]	

Independent variables	log(GDPR)	log(GDPR)	log(TVRt)
	Regression (R 1)	Regression (R 2)	Regression (R 3)
D_2t			1.651389** [5.947379]
C	20.47147** [45.46441]	21.96865** [66.83687]	-129.4007** [-9.624874]
R^2	0.819541	0.763110	0.848617
R^2 adjusted	0.803849	0.740549	0.826990
Durbin-Watson test	1.825150	1.734727	2.11
Jarque-Bera probability	0.70	0.63	0.51
Critical p-value(1%)	2.779	2.797	2.787

t-statistic in [ ]
**All the coefficients in the table are significant at 1% level.

(R 1) $\log(GDPR_t) = 20.47147 + 0.153471 \cdot \log(MCR_t) - 0.345995D_{1t}$

(R 2) $\log(GDPR_t) = 21.96865 + 0.133338 \cdot \log(TVR_{t-2}) – 0.296155 \cdot D_{1t}$

(R 3) $\log(TVR_t) = -129.4007 + 1.624155 \cdot \log(GDPR_t) + 4.411427 \cdot \log(GDPR_{t-1}) + 1.651389 \cdot D_{2t} + \varepsilon_t$

According to regression (R1), the indicator used for quantifying the economic growth (log(GDPR)) is positively correlated to capital market development, measured by log(MCR), R^2 for the equation is 0.81, reflecting that the market capitalisation and the economic growth are strongly correlated. This result is consistent with the developed countries case where the structure of market index is similar to the GDP structure and capital market is efficient. It is interesting to find that this situation applies to Romania, an emergent country, even though in this case the market index structure does not follow the GDP structure and the market is not efficient.

The regressions (R1) and (R2) show the relation between the economic growth and the capital market development. The relation between the trade volume on the capital market and the real GDP reflects a „feed-back effect”.

We performed a VAR model for log(GDPR) and log(TVR) in order to find which of these two variables influence the other. A vector autoregressive analysis with two lags was performed, which proved to be the most suitable. The results are listed in Table 3.

Table 3. Results for (VAR 1)

Independent variables	log(GDPR)	log(TVR)
	(VAR 1) – equation 1	(VAR 1) – equation 2
$\log(GDPR_{t-2})$	-0.461722** [2.96094]	1.607671* [2.67358]
$\log(TVR_{t-2})$	0.210746** [5.76643]	0.642622** [4.55985]
C	31.75561** [9.32489]	-32.81263* [-2.49869]
R^2	0.617306	0.769426
R^2 adjusted	0.580859	0.747467

t-statistic in [ ]
**The coefficients in the table are significant at 1% level
*The coefficients in the table are significant at 2% level

The trade volume growth influences GDP growth with two lags, and the GDP growth influences the trade volume growth with two lags. That means that the relation between economic growth and development of the capital market is bi-directional.

After this VAR analysis we improved the regressions (R2) and (R3) and tested the regressions (R4) and (R5). The results are presented in Table 4.

Table 4. Results of regressions (R4) and (R5)

Independent variables	log(GDPR)	log(TVRt)
	Regression (R 4)	Regression (R 5)
log(TVR_{t-2})	0.087074* [2.667285]	0.122043 [0.615192]
log(TVRt)	0.120594** [3.548049]	
log(GDPRt-2)	-0.404943** [-3.610031]	2.178879** [3.784749]
log(GDPR)		3.304636** [3.548049]
D_1t	-0.27880 ** [-5.832966]	1.061820** [3.126049]
C	30.48892** [12.67927]	-117.8590** [-4.252721]
R^2	0.878540	0.865139
R^2 adjusted	0.852970	0.836747
Durbin-Watson test	1.945242	1.851539
Jarque-Bera probability	0.654	0.62

t-statistic in [ ]
**The coefficients in the table are significant at 1% level
*The coefficients in the table are significant at 2% level

(R 4) $\log(GDPR_t) = 30.48892 - 0.404943 \cdot \log(GDPR_{t-2}) + 0.087074 \cdot \log(TVR_{t-2}) + 0.120594 \cdot \log(TVR_t) - 0.27880 \cdot D_{1t}$

(R 5) $\log(TVR_t) = -117.8590 + 3.304636 \cdot \log(GDPR_t) + 2.178879 \cdot \log(GDPR_{t-2}) + 0.122043 \cdot \log(TVR_{t-2}) + 1.061820 \cdot D_{1t}$

These results confirm the conclusion above on the feed-back effect between economic growth and capital market development. The main conclusions are the same: the economic growth influences the capital market development, but the influence is more important with two lags and the trade volume is a good predictor of the economic growth.

Regarding the market liquidity, we performed a VAR model for economic growth and a liquidity proxy computed as the number of traded shares divided by the number of listed shares.

Table 5. Results for (VAR 2)

Independent variables	GDPRG	NLI
	(VAR 2) equation 2	(VAR 2) equation 2
$GDPRG_{t-1}$	-0.476538* [-2.56392]	0.133181* [2.64883]
$GDPRG_{t-2}$	-0.746704† [-3.33507]	-0.049297 [-0.81393]

Independent variables	GDPRG	NLI
	(VAR 2) equation 2	(VAR 2) equation 2
NLI_{t-1}	1.222409 [1.45959]	0.623408† [2.75166]
NLI_{t-2}	-1.127060 [-1.50759]	0.171445 [0.84775]
C	0.037609 [0.14554]	0.162183* [2.32009]
R^2	0.437066	0.815404
R^2 adjusted	0.311970	0.774383

t-statistic in [ ]
*The coefficients in the table are significant at 2% level
† The coefficients in the table are significant at 10% level

The test revealed that the liquidity of the market, from the point of view of the number of traded companies, is not a determinant factor of the economic growth. However, as shown above, the trade volume is a determinant of the economic growth; this means that on the Romanian market the volume of trading counts in generating economic growth, and not the number of traded companies. Hence, the speculative transactions on the capital market are important for generating economic growth.

5. Conclusions

This study analyses the dependence between economic growth and capital market development for Romanian case. We found that there is a feed-back effect between capital market trade volume and economic growth; our results are similar to the findings of Hondroyiannis, Lolos and Papapetrou (2005) for Greece.

The regressions and vector autoregressive suggest that the capital market development is positively correlated with economic growth, with feed-back effect, but the strongest link is from economic growth to capital market, suggesting that financial development follows economic growth, economic growth determining financial institutions to change and develop. This results is consistent with the second posibility of a causal relationship between financial development and economic growth stated by Graff (1999). However, the lack of information, only 25 periods could question the validity of these conclusions for a long period analysis.

Bibliography

1. Arestis, P., Demetriades, P. and Luintel, K. **Financial Development and Economic Growth: The Role of Stock Markets,** Journal of Money, Credit and Banking, 33, 2001, pp. 16-41
2. Baier, S. L., Dwyer Jr, G. P. and Tamura, R. **Does Opening a Stock Exchange Increase Economic Growth?,** Journal of International Money and Finance, 23/3, 2004, pp. 311-331
3. Beck, T., Lundberg, M. and Majnoni, G. **Financial intermediary development and growth volatility: do intermediaries dampen or magnify shocks?,** Journal of International Money and Finance, 25, 2006, pp. 1146-1167

4. Beckaert, G., Harvey, C. and Lundblad, C. **Does financial liberalization spur growth?,** Journal of Financial Economics, 77, 2005, pp. 3-55
5. Ben Naceur, S., and Ghazouani, S. **Stock markets, banks, and economic growth: Empirical evidence from the MENA region,** Research in International Business and Finance, 21, 2007, pp. 297–315
6. Bolbol, A., Fatheldin, A., and Omran, M. **Financial development, structure, and economic growth: the case of Egypt, 1974-2002,** Research in International Business and Finance, 19, 2005, pp.171-194
7. Bose, N. **Endogenous Growth and the Emergence of Equity Finance,** Journal of Development Economics, 77, 2005, pp. 173– 188
8. Calderon, C. and Liu, L. **The Direction of Causality between Financial Development and Economic Growth,** Central Bank of Chile. Working Paper, 2002, http://www.bcentral.cl/estudios/documentos-trabajo/pdf/dtbc184.pdf, accessed 19 July 2007
9. Carlin, W. and Mayer, C. **Finance, Investment, and Growth,** Journal of Financial Economics, 69, 2003, pp. 191-226
10. Claessens, S., Klingebiel, D. and Schmukler, S. **Stock market development and internationalization: Do economic fundamentals spur both similarly?,** Journal of Empirical Finance, 13, 2006, pp. 316–350
11. Dragota, V. **Minority shareholders' protection in Romanian capital markets: evidence on dividends,** Euro-Mediterranean Economics and Finance Review, Vol. 1, No.1, 2006, pp. 76-89
12. Ergungor, E. **Financial system structure and economic growth: Structure matters**, International Review of Economics and Finance, 2006, available online 13 November 2006
13. Filer R., Hanousek, J. and Campos, N. **Do Stock Markets Promote Economic Growth?,** CERGE-EI Working Papers wp151, The Center for Economic Research and Graduate Education - Economic Institute, Prague, 1999, www.cerge-ei.cz/pdf/wp/Wp151.pdf accessed 19 July 2007
14. Garretsen, H., Lensink, R. and Sterken, E. **Growth, financial development, societal norms and legal institutions,** Journal of International Financial Markets, Institutions and Money, 14, 2004, pp. 165-183
15. Graff, M. A. **Financial Development and Economic Growth - A New Empirical Analysis,** Dresden Discussion Papers in Economics No. 5, 1999, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=258928, accessed 19 July 2007.
16. Hondroyiannis, G., Lolos, S. and Papapetrou, E. **Financial market and economic growth in Greece, 1986-1999,** Journal of International Financial Markets, Institutions and Money, 15, 2005, pp. 173-188
17. King, R. G. and Levine, R. **Finance and Growth: Schumpeter Might Be Right,** The Quarterly Journal of Economics, MIT Press, vol. 108 (3), 1993, pp. 717-737
18. Levine, R. **Financial Development and Economic Growth: Views and Agenda,** Journal of Economic Literature, 35, 1997, pp. 688-726
19. Levine, R. and Zervos, S. **Stock Markets, Banks, and Economic Growth,** American Economic Review, 88, 1998, pp. 537-558
20. Liu, W. and Hsu, C. **The role of financial development in economic growth: The experiences of Taiwan,** Korea, and Japan. Journal of Asian Economics, 17, 2006, pp. 667–690
21. Minier, J. **Are small stock markets different?,** Journal of Monetary Economics, 50, 2003, pp. 1593–1602
22. Nieuwerburgh, S., Buelens, F. and Cuyvers, L. **Stock market development and economic growth in Belgium,** Explorations in Economic History, 43, 2006, pp. 13–38

23. Rajan, R. and Zingales, L. **Financial Dependence and Growth,** The American Economic Review, 88, 1998, pp. 559-586
24. Rousseau, P. L. and Sylla, R. **Financial Systems, Economic Growth, and Globalization,** Working paper 8323, 2001, http://www.nber.org/papers/w8323, accessed 19 July 2007
25. Stiglitz, J. **Globalization and its Discontents,** New York, WW Norton Company, 2002
26. Thornton, J. **Financial Deepening and Economic Growth inDeveloping Countries,** Economia Internationale, 48(3), 1995, pp. 423-430

Appendix

Appendix 1: Data description

For market capitalization and trading volume we used quarterly average of daily data, provided by the Bucharest Stock Exchange Research Department. As the annually rate of inflation in Romania during 2000-2006 was between 9% and 45.8%, we deflated the values using the quarterly average consumer price index (fix basis: first quarter, 2000). The Romanian National Institute for Statistics and Economic Studies (INSSE) reports, in its annual reports, the monthly consumer price index, using the same basis (October 1990). Based on it, it was calculated the quarterly average consumer price index with the same basis (October 1990 = 100%), and then the quarterly average consumer price index with the basis in the first quarter 2000.

Market capitalization is an absolute measure of the market size and the trading volume is a measure of the financial redistribution made by the capital market in the economy, demonstrating the importance of Bucharest Stock Exchange in the Romanian financial system.

The number of listed shares (NLS) is determined for the last trading day of the quarter. It reflects the role of the capital market both for companies and investors; listed companies have the possibility of obtaining resources, and investors can invest their economies in the traded shares. The number of traded shares (NTS) is calculated as the quarterly average of the daily data. It reflects the market liquidity. Both, NLS and NTS were provided by the Bucharest Stock Exchange Research Department.

For assessing the market liquidity we have calculated two proxies:

1) the ratio between the number of traded shares and the number of listed shares:

$$NLI = \frac{NTS}{NLS} \tag{1}$$

2) the ratio between the trading volume and the market capitalization:

$$VLI = \frac{TVR}{MCR} \tag{2}$$

The BET and BET-C indexes quantify the evolution of the market portfolios, being general indicators for Romania's capital market. For the same reasons as above we deflated the BET-C index, obtaining a BETCR index.

The economic growth is measured by the real gross domestic product growth rate. The nominal quarterly values (GDPN) were taken from the Romanian National Institute for Statistics and Economic Studies (they are computed based on the ESA '95 methodology for the national accounts). The values were then deflated, using the quarterly average consumer price index with the basis in the first quarter, 2000. Thus, we obtained real values for the gross domestic product (GDPR). Based on the real values it was determined the growth rate of the real gross domestic product (GDPRG), which is a measure of the Romanian economic development, during the analyzed period.

The variables of the economic growth that have been selected for regression were the real gross domestic product by the indicator of log(GDPR) and the economic growth calculated as the growth rate of the real GDP (GDPRG).

We noticed a seasonal evolution of the quarterly real gross domestic product. In order to eliminate the seasonal effect for the 4th quarter, we introduced dummy variables (D_1, D_2).

[1] **Acknowledgements:**
Authors are grateful for the comments provided by participants to 39th Meeting Euro Working Group on Financial Modelling, CERAM Sophia Antipolis, November 16-17, 2006, and especially to Michel H. Bouchet, Michel M. Dacorogna and Yue Qi.
Authors are thankful for the comments provided by participants to 3rd Iberian International Business Conference, Aveiro, 19th-20th October 2007.
Also, authors are grateful for the comments provided by Bogdan Negrea and Andreea Stoian for their useful comments. The remaining errors are ours.
This research was supported by Romanian Ministry of Education and Research – National Authority for Scientific Research (NASR) through RTD National Programme CEEX, Grant No. 1505 / 2006 (Module 2).

[2] Victor Dragota holds a PhD in Finance granted in 2003 by University of Economics, Bucharest. He graduated the Faculty of Finance from Bucharest University of Economics.
He is (co)author of papers and books like:
Dragota, V., Dumitrescu, D., Ruxanda, Gh., Ciobanu, A., Brasoveanu, I., Stoian, A. and Lipara, C. **Estimation of Control Premium: The Case of Romanian Listed Companies,** Economic Computation and Economic Cybernetics Studies and Research, Vol. 41, No. 3-4, 2007, pp. 55-72
Dragota, V. and Dragota, M. **Considerations on Passive Portfolio Management,** Economic Computation and Economic Cybernetics Studies and Research, Vol. 40, No. 1-2 / 2006, pag. 203-208.
Dragota, V. **Minority shareholders' protection in Romanian capital markets: evidence on dividends,** Euro-Mediterranean Economics and Finance Review, Vol. 1, No.1, Jan. 2006, pp. 76-89.
Dragota, V., Dragota, M., Obreja Brasoveanu, L., Stoian, A., Lipara, C. and Cetean, E. **Piete si sisteme financiare,** Ed. ASE, Bucharest, 2008
Dragota, V. **Evaluarea actiunilor societatilor comerciale,** Coeditie Ed. Economica si Ed. IROVAL, Bucuresti, 2006
He is member in the scientific board of Economic Computation and Economic Cybernetics Studies and Research and The Valuation Journal.
The main research fields are: market efficiency, dividend policy, corporate finance, investments, business valuation.
He is teaching the following courses: corporate finance, investments, corporate governance.

[3] Born in 1980, she graduated Bucharest University of Economics in 2003 and a Banks and Insurance master program at Orleans University in 2004. Since then she works in the Finance Department of Bucharest University of Economics. Her competence domains are: corporate finance, capital markets, public finance, direct investments. Her doctoral research program is "The application of methods for determining the efficiency of public investment projects".

[4] Regression, VAR, Granger Causality, Event Studies etc.

[5] Stock market capitalization ratio equals market capitalization over GDP.

[6] Turnover ratio equals trading volume over market capitalization.

[7] Value traded ratio equals trading volume over GDP.

ISSUES AND PROCEDURES IN ADOPTING STRUCTURAL EQUATION MODELING TECHNIQUE

Siu Loon HOE[1]
PhD, Adjunct Research Fellow, UWA Business School,
The University of Western Australia, Australia

E-mail: hoesl@graduate.uwa.edu.au

Abstract: *When applying structural equation modeling (SEM) technique for analytical procedures, various issues are involved. These issues may concern sample size, overall fit indices and approach. Initiates of SEM may find it somewhat daunting in resolving these technical issues. The purpose of this paper is to highlight key issues in adopting SEM technique and various approaches available. This paper provides a discussion on the sample size, fit indices, standardized paths, unidimensionality test and various approaches in relation to SEM. It is hoped that having reviewed the paper, new researchers can devote more time to data analysis instead of procedural issues involved.*

Key words: Structural equation modeling; fit indices; unidimensionality; eigenvalue

Introduction

The modern positivist paradigm for conducting scientific research rests on developing sound theoretical frameworks followed by rigorous testing of these theories. One often adopted technique is structural equation modeling (SEM). SEM is a powerful statistical technique that combines measurement model or confirmatory factor analysis (CFA) and structural model into a simultaneous statistical test.

SEM is particularly valuable in inferential data analysis and hypothesis testing where the pattern of inter-relationships among the study constructs are specified a priori and grounded in established theory. It has the flexibility to model relationships among multiple predictor and criterion variables, and statistically tests a priori theoretical assumptions against empirical data through CFA (Chin, 1998). In most cases, the method is applied to test 'causal' relationships among variables.

In applying SEM technique for analytical procedures, many issues are involved. These issues may concern various overall fit indices and selection of the appropriate approach (Lei & Wu, 2007). Initiates of SEM may find it somewhat daunting in resolving these issues. The purpose of this paper is to highlight key issues in adopting SEM technique and various approaches so that researchers can devote more time to data analysis instead of dealing with procedural issues. This paper provides a discussion on the sample size, fit indices, standardized paths, unidimensionality test and various approaches in relation to SEM. Wherever appropriate, numerical examples were provided to illustrate the issues and procedures highlighted.

Structural equation modeling and sample size

Typically, a hypothesized model is tested with a linear equation system through SEM. This method of study investigates the extent to which variations in one variable corresponded to variations in one or more variables based on correlation co-efficient. SEM is usually used because it permits the measurement of several variables and their inter-relationships simultaneously. It is more versatile than other multivariate techniques because it allows for simultaneous, multiple dependent relationships between variables.

The hypothesized causal relationships can be tested among the theoretical constructs using software programs such as EQS (Bentler, 2002) to estimate and evaluate the structural portion of the model. The raw data for the variables are input into the software to generate the iterations, goodness-of-fit indices and standardized paths. The various variables are usually summated scales where the attributes measuring a common underlying construct are summed and divided by the number of items.

McQuitty (2004) suggested that it is important to determine the minimum sample size required in order to achieve a desired level of statistical power with a given model prior to data collection. Schreiber et al (2006) mentioned that although sample size needed is affected by the normality of the data and estimation method that researchers use, the generally agreed-on value is 10 participants for every free parameter estimated. Although there is little consensus on the recommended sample size for SEM (Sivo et al, 2006), Garver and Mentzer (1999), and Hoelter (1983) proposed a 'critical sample size' of 200. In other words, as a rule of thumb, any number above 200 is understood to provide sufficient statistical power for data analysis.

Fit indices

There are several indicators of goodness-of-fit and most SEM scholars recommend evaluating the models by observing more than one of these indicators (Bentler & Wu, 2002; Hair et al. 1998). Marsh, Balla and McDonald (1988) proposed that the criteria for ideal fit indices are relative independence of sample size, accuracy and consistency to assess different models, and ease of interpretation aided by a well defined pre-set range. Based on this stated criteria, Garver and Mentzer (1999) recommended the nonnormed fit index (NNFI); the comparative fit index (CFI), and the root mean squared approximation of error (RMSEA). Therefore, the commonly applied fit indices are NNFI and CFI (>0.90 indicates good fit), RMSEA (<0.08 indicates acceptable fit), and commonly used χ^2 statistic (χ^2/ d.f. ratio of 3 or less).

The NNFI, also known as the Tucker-Lewis index, compares a proposed model's fit to a nested baseline or null model. Additionally, NNFI measures parsimony by assessing the degrees of freedom from the proposed model to the degrees of freedom of the null model. NNFI also seems resilient against variations in sample size and, thus, is highly recommended. An acceptable threshold for this index is 0.90 or greater. Bentler (1990) developed the CFI as a noncentrality parameter-based index to overcome the limitation of sample size effects. This index ranges from 0 to 1, with 0.90 or greater representing an acceptable fit. RMSEA is an extremely informative criterion in evaluating model fit. The RMSEA index measures the discrepancy between the observed and estimated covariance matrices per degree of freedom (Steiger, 1990). It measures the discrepancy in terms of the population and not the sample. Thus, the value of this fit index is expected to better approximate or estimate the population and not be affected by sample size. Again, values run on a continuum from 0 to 1. Values less than 0.05 indicate good fit, values up to 0.08 reasonable fit and ones between 0.08 and 0.10 indicate mediocre fit.

Chi-square (χ^2) is the most common method of evaluating goodness-of-fit. A low χ^2 value, indicating nonsignificance, would point to a good fit. This is because chi-square test is used to assess actual and predicted matrices. Thus, non-significance means that there is no considerable difference between the actual and predicted matrices (Hair et al., 1998). Therefore, low χ^2 values, which result in significance levels greater than 0.05 or 0.01, indicate that actual and predicted inputs are not statistically different. The significance levels of 0.1 or 0.2 should be exceeded before nonsignificance is confirmed (Fornell, 1983).

In terms of a model's goodness-of-fit, p-values indicate whether the model is significantly different than the null model. In statistics, the null is usually '0'. This, however, is not necessarily so in SEM. The null hypothesis is the hypothesized model in which the parameters were set up for the hypothesized model, indicating whether a path should exist or not between variables. A high ρ-value, or a value larger than zero, would mean that the null hypothesis is rejected leading to a high probability that it would be wrong in doing so (MacLean & Gray, 1998). Thus, a high 'ρ' is good as it indicates that the observed model is not significantly different from what was expected. Conversely, a low ρ-value, or one close to zero, implies a 'bad model' because the null hypothesis is rejected with a low probability of being wrong in reaching that conclusion.

There is a limitation to the chi-square test. The χ^2 is highly sensitive to sample size especially if the observations are greater than 200. An alternate evaluation of the χ^2 statistic is to examine the ratio of χ^2 to the degrees of freedom (d.f.) for the model (Joreskog & Sorbom, 1993). A small χ^2 value relative to its degree of freedom is indicative of good fit. Kline (1998) suggested that a χ^2/ d.f. ratio of 3 or less is a reasonably good indicator of model fit.

As an example, fit indices were generated for a hypothesized model using SEM technique and presented in Table 1.

Table 1. Examples of fit indices of a hypothesized model

Model	d.f.	χ^2	NNFI	CFI	RMSEA
Hypothesized Model	18	416.69	-0.043	0.330	0.321

As indicated in these results, the goodness of fit measures for the hypothesized model came nowhere near the minimum requirements for the benchmark fit indices. The χ^2 value is 416.69 based on 18 d.f. and probability value (ρ) for χ^2 statistic is less than 0.001. The NNFI = -0.043, CFI = 0.330 and RMSEA = 0.321 which indicate a 'bad fit' for the hypothesized model (Bentler, 1990). Since the hypothesized model did not have a 'good fit', it was rejected.

Standardized paths and test for unidimensionality

Besides the 'goodness-of-fit' indices, SEM may also be used to look at paths among the variables. The causal paths can be evaluated in terms of statistical significance and strength using standardized path coefficient that range between -1 and +1. Based on α of 0.05, the test statistic generated from the EQS output should be greater than $\pm$ 1.96 to indicate that the null hypothesis can be rejected. The rejection of the null hypothesis means that the structural coefficient is not zero (Bentler, 2002; Byrne, 1994). After reviewing the statistical significance of the standardized paths, the strength of relationships among the variables can then be reviewed. According to Chin (1998), standardized paths should be at least 0.20 and ideally above 0.30 in order to be considered meaningful for discussion.

As another example, the standardized paths of a hypothesized model were computed and shown in Table 2.

Table 2. Examples of standardized paths of a hypothesized model

Hypothesis	Causal Path	Standardized Path Coefficient
H1	Informal knowledge acquisition → Market knowledge use	0.209***
H2	Informal knowledge dissemination → Market knowledge use	0.210**
H3	Shared vision → Informal knowledge acquisition	Nonsignificant
H4	Shared vision → Informal knowledge dissemination	0.452****
H5	Interpersonal trust → Informal knowledge acquisition	0.150*
H6	Interpersonal trust → Informal knowledge dissemination	Nonsignificant

Note: * $\rho < 0.05$
** $\rho < 0.01$
*** $\rho < 0.001$
**** $\rho < 0.0001$

In this hypothesized model, four of the paths were statistically significant. Comparing these results with the hypotheses, the standardized path coefficient of 0.209 seems to indicate that informal knowledge acquisition is positively associated with market knowledge use (H1). Also, the standardized path coefficient of 0.210 suggests that informal knowledge dissemination is also positively associated with market knowledge use (H2).

The path between shared vision and informal knowledge acquisition was not statistically significant indicating that shared vision is not positively associated with informal knowledge acquisition (H3). The standardized path coefficient between shared vision and informal knowledge acquisition was 0.452. This seems to suggest that shared vision is positively associated with informal knowledge dissemination (H4).

Although the results point to a significant association between interpersonal trust and informal knowledge acquisition (H5) with standardized path coefficient of 0.150, this path adds minimal value to the understanding of the relationship between interpersonal trust and informal knowledge acquisition. The reason is because the standardized path coefficient failed to meet the minimum benchmark for path strength. Chin (1998) has proposed that standardized paths should be at least 0.20 and ideally above 0.30 in order to be considered meaningful. The path between interpersonal trust and informal knowledge dissemination was nonsignificant suggesting that interpersonal trust is not positively associated with informal knowledge dissemination (H6).

Once the overall model fit has been evaluated, the variables can be assessed for unidimensionality. In accordance to accepted practice (Anderson, 1987; Churchill 1979; Gerbing & Anderson, 1988), the property of scales for unidimensionality was assessed. Unidimensionality is referred to as the existence of one construct underlying a set of items. Germain, Droge and Daugherty (1994) suggested the use of principal components analysis to test for unidimensionality. Based on this suggestion, each variable should be separately subject to principal components analyses to determine the eigenvalue. As a rule, eigenvalues that are greater than 1 provide support for the unidimensionality of these scales.

As an illustration, the eight variables in a particular study were separately subject to principal components analyses and the eigenvalues presented in Table 3.

Table 3. Examples of eigenvalues of measures

Measure	Component	Initial Eigenvalues Total	 % of Variance	 Cumulative %
Market Knowledge Use	1	4.048	57.824	57.824
	2	1.040	114.852	72.676
	3	0.726	10.365	83.041
	4	0.425	6.071	89.112
	5	0.327	4.664	93.776
	6	0.247	3.528	97.304
	7	0.189	2.696	100.00
Structural Knowledge Acquisition	1	1.925	64.155	64.155
	2	0.651	21.702	85.858
	3	0.424	14.142	100.00
Structural Knowledge Dissemination	1	3.094	61.888	61.888
	2	0.660	13.193	75.082
	3	0.461	9.214	84.296
	4	0.427	8.547	92.836
	5	0.358	7.164	100.00
Informal Knowledge Acquisition	1	2.620	52.396	52.396
	2	0.910	18.191	70.587
	3	0.627	12.549	83.136
	4	0.507	10.139	93.275
	5	0.336	6.725	100.00
Informal Knowledge Dissemination	1	1.786	59.539	59.539
	2	0.731	24.362	83.901
	3	0.483	16.099	100.00
Shared Vision	1	3.147	78.679	78.679
	2	0.317	7.936	86.615
	3	0.283	7.081	93.697
	4	0.252	78.679	100.00
Interpersonal Trust	1	3.272	81.804	81.804
	2	0.398	9.961	91.765
	3	0.198	4.959	96.725
	4	0.131	3.275	100.00
Perceived Importance of Market Knowledge	1	4.049	80.974	80.974
	2	0.473	9.462	90.436
	3	0.205	4.097	94.534
	4	0.156	3.122	97.656
	5	0.117	2.344	100.00

Except for market knowledge use, only the first eigenvalue was greater than 1 for all the rest of the scales. This provided support for the unidimensionality of these scales. For market knowledge use, two eigenvalues were greater than 1 but the second eigenvalue was only 1.04. Since second eigenvalue is close to 1 and this is a measure that has been used extensively in previous research, it is reasonable to accept the unidimensionality of this scale.

Various approaches

There are several approaches to weighing individual scale items in SEM. These approaches include total aggregation, total disaggregation, partial disaggregation, and partial aggregation (Bagozzi & Heatherton, 1994).

Under the total aggregation approach, all the items are summed into a single indicator or latent variable. The items are arbitrarily given the same weight as is traditionally done in non-SEM research. On the other hand, under the total disaggregation approach, each individual item is taken separately as an individual indicator. Under this approach, SEM weighs the individual items to optimize contribution to the latent variables. Bagozzi and Foxall (1996) mentioned that one of the main drawbacks of the total aggregation approach for CFA is that information is lost and the distinctiveness of the components is obscured. A disadvantage of the total disaggregation approach is it is very sensitive to measurement error which makes it more difficult to obtain satisfactory model fits. The partial aggregation approach retains the idea of a single underlying factor where dimensions of the construct are organized hierarchically as indicators of the factor. Similar to the total aggregation approach, the main drawback is that the unique dimension of the construct may be obscured.

To overcome such limitations, Bagozzi and Heatherton (1994) recommended that a partial disaggregation approach be used. Partial disaggregation is a practical SEM application that allows the use of a large number of indicators to represent a latent variable (Garver & Mentzer, 1999). It is an intermediary level of analysis between the total aggregation and total disaggregation approach. Unlike a total aggregation approach, the partial disaggregation approach helps to reduce the number of parameters to be estimated and to capitalize on the increase in reliability resulting when items on sub-scales are summed. Each dimension can be measured with two indicators wherein each indicator is itself the sum of multiple items.

Bagozzi and Heatherton (1994) suggested between five to seven items that can be randomly divided into two components for partial disaggregation. If there are more than nine items, then there could be three or more components. Based on the partial disaggregation procedures, there is no need to divide the aggregated total by the number of items so that each latent variable is an average score. Thus, the items for each of the latent variables can be randomly grouped into two components each based on odd and even sequence and the raw data input into the measurement model.

Conclusion

This paper has provided a short discussion on the sample size, fit indices, standardized paths, unidimensionality test and various approaches in relation to SEM. Some examples were provided to illustrate the issues highlighted. Through a better understanding of the issues and procedures in adopting the structural equation technique, researchers will be able devote more time to data analysis rather than resolving procedural issues.

References

1. Anderson, J. C. **An approach for confirmatory measurement and structural equation modeling of organizational properties,** Management Science, 33, 4, 1987, pp. 525-541
2. Bagozzi, R.P. and Heatherton, T.F. **A general approach to representing multifaceted personality constructs: Application to state self-esteem,** Structural Equation Modeling, 1, 1, 1994, pp. 35-67
3. Bagozzi, R. P. and Foxall, G. R. **Construct validation of a measure of adaptive-innovative cognitive styles in consumption,** International Journal of Research in Marketing, 13, 1996, pp. 201-213
4. Bentler, P. M. **Comparative fit indexes in structural models,** Psychological Bulletin, 107, 2, 1990, pp. 238-246
5. Bentler, P. M. **EQS Structural Equations Program Manual,** Encino, CA, Mulitvariate Software, 2002
6. Bentler, P. M. and Wu, E. J. C. **EQS 6 for Windows User's Guide,** Encino, CA: Mulitvariate Software, 2002
7. Byrne, B. M. **Structural Equation Modeling With EQS And EQS/Windows,** Newbury Park, Sage, 1994
8. Chin, W. W. **Issues and opinion on structural equation modeling,** MIS Quarterly, 22, 1, 1998, pp. 7-16
9. Churchill, G. Jr. **A paradigm for developing better measures of marketing constructs,** Journal of Marketing Research, 16, 1, 1979, pp. 64-73
10. Fornell, C. **Issues in the application of covariance structure analysis: A comment,** Journal of Consumer Research, 9, 4, 1983, pp. 443-448
11. Garver, M. S. and Mentzer, J.T. **Logistics research methods: Employing structural equation modeling to test for construct validity,** Journal of Business Logistics, 20, 1, 1999, pp. 33-57
12. Gerbing, D. W. and Anderson, J. C. **An updated paradigm for scale development incorporating unidimensionality and its assessment,** Journal of Marketing Research, 25, 1, 1988, pp. 186-192
13. Germain, R., Droge, C. and Daugherty, P. J. **The effect of just-in-time selling on organizational structure: An empirical investigation,** Journal of Marketing Research, 31, 4, 1994, pp. 471-483
14. Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. and Black, W. C. **Multivariate Data Analysis** (5th ed.), New Jersey, Prentice-Hall, 1998
15. Hoelter, D. R. **The analysis of covariance structures: Goodness-of-fit indices,** Sociological Methods and Research, 11, 1983, pp. 325–344
16. Joreskog, K. G. and Sorbom, D. **LISREL6 - Computer Program,** Moorseville, IN, Scientific Software, 1986
17. Kline, R. B. **Principles And Practice Of Structural Equation Modeling,** New York, Guilford Press, 1998
18. Lei, P. W. and Wu, Q. **An NCME Instructional Module on Introduction to Structural Equation Modeling: Issues and Practical Considerations,** Educational Measurement, Issues and Practice, 26, 3, 2007, pp. 33-44
19. MacLean, S. and Gray, K. **Structural equation modeling in market research,** Journal of the Australian Market Research Society, 1998, Retrieved September 26, 2001, from World Wide Web: http://www.smallwaters.com/whitepapers/marketing
20. Marsh, H. W., Balla, J. R. and McDonald, R. P. **Goodness-of-fit in confirmatory factor analysis: The effect of sample size,** Psychological Bulletin, 103, 3, 1988, pp. 391-410

21. McQuitty, S. **Statistical power and structural equation models in business research,** Journal of Business Research, 57, 2, 2004, pp. 175-183
22. Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. and King, J. **Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review,** The Journal of Educational Research, 99, 6, 2006, pp. 323-337
23. Sivo, S. A., Fan, X. T., Witta, E .L. and Willse, J. T. **The Search for 'Optimal' Cutoff Properties: Fit Index Criteria in Structural Equation Modeling,** The Journal of Experimental Education, 74, 3, 2006, pp. 267-289
24. Steiger, J. H. **Structural model evaluation and modification: an interval estimation approach,** Multivariate Behavioral Research, 25, 1990, pp. 173-180

---

[1] Siu Loon Hoe is an Adjunct Research Fellow at UWA Business School, The University of Western Australia. His research interests are focused on organizational learning and boundary spanning. Siu Loon holds a BSc from National University of Singapore, and an MBA (International Business) and PhD (Management) from The University of Western Australia. He is currently Head of Strategic Planning for a Pan-Asian consumer group. In recognition of his academic and professional achievements, Siu Loon was included in International Who's Who of Professionals 2001/02.

ADAPTABLE SOFTWARE SHELLS VERSUS MICROSOFT SOFTWARE SHELLS

Dumitru TODOROI[1, 2]
PhD, University Professor, Academy of Economic Studies, Kishinev, Moldova

E-mail: todoroi@ase.md

Abstract: *Development and evolution of Microsoft Office and Microsoft Windows shells is based in general on the special methodology of Software creation and implementation such as macros, subroutine, custom commands and specialized features. This methodology of Microsoft Software shells is analyzed. The universal methodology of Adaptable Software creation is proposed. Present result evaluates from [Tod-08.1]*[3] *which is an evolution of the Fulbright research project no. 22131 "Societal Information Systems' Adaptable Tools" performed in the University of Omaha at Nebraska, USA in the 1997/1998 academic year [Tod-98].*

Key words: *adaptable software; Microsoft software; software shells*

1. Introduction

Microsoft Office is an office suite from Microsoft for Microsoft Windows and Apple Mac OS X operating systems (http://en.wikipedia.org/wiki/Microsoft_Office). Along with core office applications, it includes associated servers and web-based services. Recent versions of Office are referred as "Office system".

The current versions are Office 2007 for Windows, launched on January 30, 2007, [4] and Office 2008 for Mac, released January 15, 2008. Office 2007/Office 2008 features a distinct user interface and a new XML-based primary file format.

Office was introduced by Microsoft in 1989 on Mac OS,[5] with a version for Windows in 1990[6]. Initially a marketing term for a bundled set of applications, the first version of Office contained Microsoft Word, Microsoft Excel, and Microsoft PowerPoint. Additionally, a "Pro" version of Office included Microsoft Access and Schedule Plus. Over the years, Office applications have grown substantially closer with shared features such as a common spell checker, OLE data integration and Microsoft Visual Basic for Applications scripting language. Microsoft also positions Office as a development platform for line-of-business software under the Office Business Applications (OBA) brand.

2. Microsoft Office & Windows 95 / 97 / 08

As was underlined in (http://www.zisman.ca/Articles/1995/Office95.html) Windows 95 is nice, but while it does a pretty good job of running the current generation of applications, it really needs new applications to take advantage of all its improvements. It really shouldn't come as much of a surprise, therefore, that Microsoft is revamping virtually its entire product line to run as native, Win95 applications.

The flagship of Microsoft's armada is the MS Office suite [1995-Office] - it should come as no surprise that MS Office 95 was released on August 24th - the same day as Windows 95 itself - might as well give everyone in those long lineups something else to buy!

In the past few years, software Suites have become big business - accounting for a majority of sales of word processors and spreadsheets, for example. And despite competition from Lotus SmartSuite and Novell (formerly Word Perfect) Perfect Office, MS Office has garnered over 70% of the suite sales. Where in the former DOS world, Microsoft seemed to always have the also-ran contender in word processor and spreadsheet sales, the combination of Windows and suites has pushed them into a comfortable sales lead.

All Win95 applications must be written as 32-bit software, rather than the 16-bit versions standard with Windows 3.x. This doesn't automatically make them run faster. The 32-bit Windows NT versions of MS Word and Excel, for example, ran slower than their 16-bit equivalents, because they left out the highly-optimized programming code of the older versions in order to be compatible with NT on a wide range of CPUs.

The newer Win95 versions of these programs, however, do feel snappier than the last generation.

Win95 allows new programs to be multi-threaded - to run separate processes in the background to improve performance. The Office applications don't make much use of multithreading, however, adding it only to printing.

A nice Win95 feature that may take getting used to is 'Scraps' - the user can highlight part of a document, and drag it to the desktop, where the scrap can sit and wait for you to drag it into a different document or even a different application. Use it, for instance, to drag an address from a letter, and drop it into a contact list... but like the other features, only usable with Win95-compatible applications, like all of Office 95. Similarly, right-clicking on the desktop gets you the standard Win95 popup menu - choose New to simply create a new Word or Excel or PowerPoint document. In fact, as with Win95 in general, right-mouse clicking is implemented throughout Office 95.

Office 95 applications also all make use of Win 95's Exchange client - the user can send e-mail from any of them, or use it to share information across a network. You can also use Word to edit your Exchange e-mail. Help items in all the applications can automatically connect the user to the correct forum on the Microsoft Network.

2.1. Working together

Schedule Plus [1995-Office], a group scheduler and personal information manager that debuted in Windows for Workgroups has been moved from **the core** Windows package into Office. As well, suite marketing has claimed that purchasers would find it easier working with a collection of applications that were sported the same interface.

Office 95 is better than previous versions in this way - Microsoft's programmers worked on using as much common code for all of the applications: they share File Open and Save dialogue boxes, for example, enhancing Win95's standard dialogues with Previews of the selected files and integrated search. While users of Word 6.0 could run AutoCorrect, now this feature appears across the Office, as does a new help extension, the Answer Wizard. A nice touch is the vertical scrollbars - you can now see what page number the user are scrolling to. The user can open these applications from Win95's Start Menu, but by default, he gets the Microsoft Office Shortcut Bar, replacing the previous version's anemic Microsoft Office manager (MOM). The user can have multiple Shortcut toolbars-- it's configurable enough that some users may find it preferable to Win95's Start Menu for most of their computing.

In addition, Office adds a new kind of document - a Binder file. This lets the user to combine data from any of the Office apps, or other Office-compatible software into a single, notebook-like setting... as easily as dragging them in. Because the Binder uses OLE 2, as user changes to a page created in a different application, the user don't open that application-- his tool bar and menus change to fit the data.

Several years ago, Microsoft promised a single, Visual Basic-based macro language, that would be supported across its applications. While Excel has supported Visual Basic for Applications, Word and PowerPoint still lack this support. Word and Excel continue to use the same file formats with the previous versions, but PowerPoint and Access produce **new, incompatible files.**

2.2. Working separately

The applications in Office95 (Table 1) and its successor Office97/98 (Table 2) benefit from being designed for Windows95, and to share the common features of Office. Otherwise, the actual feature-sets of the separate applications are not dramatically changed from the last generation. That's not a bad thing-- Word 6.0 and Excel 5.0 were both rich with features that many users still haven't had a chance to fully get used to.

Like other suites, Office 95 is a hefty program [1995-Office] that can demand a lot of ram and hard drive space. It can be installed in a minimal fashion, taking a 'mere' 27 megs, or can be set up to run off a CD-ROM drive (which still asks for 30 megs of hard drive space).

Home users and some small offices running Win95 may find their needs better served with one of the new generation of all-in-one programs, such as Microsoft Works-95, or the soon-to-be-released Claris Works 4.0.

Even though the biggest improvements users will find with this suite come from Windows 95 (like long file names) rather than from the core applications, at the moment, this is the product of choice for business users running Win95. Even when Lotus and Novell release their Win95 competitors, in a few months, this will remain the one to beat.

Table 1. Microsoft Office95 edition

Standard	Professional
Word	Word
Excel	Excel
PowerPoint	PowerPoint
Schedule+	Schedule+
	Access

Table 2. Microsoft Office 97 Powered by Word 98 editions

Standard	Professional	Small Business	Developer
Word	Word	Word	Word
Excel	Excel	Excel	Excel
Outlook	Outlook	Outlook	Outlook
PowerPoint	PowerPoint		PowerPoint
	Access		Access
		Publisher	Developer Tools and SDK

3. Microsoft Office 2000

In [2000-MsOffice] the author underlined that "… some people mistakenly think that Office is a single program. The confusion is understandable because Microsoft markets Office in this manner. Although the Office 2000 (Table 3) programs and its successor Office XP (Table 4) are designed to almost seamlessly together, Office 2000 still consists of several individual programs. There is a smart strategy to learn and use the Office suites of programs. Each Office program works best for a different task: you can best compose text with Word, you can best crunch numbers with Excel, you can best manage tables of information with Access…Given this, what you want to do is learn to use the Office programs that best handle the sort of data that you work with…".

The programs in Office 2000 include integrated features; for example [2000-MsOffice] "… Word includes a rather crude spreadsheet features (called tables) that you can use to make calculations – such as for the budget. Excel's worksheets, however, are both easier to use and more powerful in what they can do. Similarly, you can use Excel to create lists of information, but for working with lots of data pieces, Access's tables feature is both easier to use and more powerful than Excel". Regardless on which Office features the user uses most often, he/she performs many tasks in exactly the same way in all Office programs.

Table 3. Microsoft Office 2000 edition

Standard	Small Business *	Professional	Premium	Developer
Word	Word	Word	Word	Word
Excel	Excel	Excel	Excel	Excel
Outlook	Outlook	Outlook	Outlook	Outlook
PowerPoint		PowerPoint	PowerPoint	PowerPoint
	Publisher	Publisher	Publisher	Publisher
		Access	Access	Access
			FrontPage	FrontPage
			PhotoDraw	PhotoDraw
	Small Business Tools			Developer Tools and SDK

Table 4. Microsoft Office XP edition

Standard	Professional	Professional with FrontPage *	Professional Special Edition	Developer
Word	Word	Word	Word	Word
Excel	Excel	Excel	Excel	Excel
Outlook	Outlook	Outlook	Outlook	Outlook
PowerPoint	PowerPoint	PowerPoint	PowerPoint	PowerPoint
	Access	Access	Access	Access
		FrontPage	FrontPage	FrontPage
			Publisher	Developer Tools and SDK
				SharePoint Team Services

4. Microsoft Office 2003

Microsoft Office 2003 (Table 5) introduces the user to the new, far-reaching features of Microsoft Office 2003 and shows how you can put them to work in your business, organization, or home. If the user works with ideas…his imagination will kick into high gear as he see how Office 2003 can support the total lifespan of an idea, from that first scribble on the back of his business card, to the final Web publication produced by his management team, and the translation and distribution of the smart document to his satellite offices in 14 countries around the globe.

The user [2003-MsOffice] "…don't have to be part of a large organization to get maximum benefit from the new features in Office. Employees of small to mid-size business, entrepreneurs, and independent contractors – anyone who exchanges ideas and data with someone else – will find features in Office that enhance communication and collaboration; make project management simpler then ever; capture innovative ideas from everyone on the team; and easily produce and change documents, Web pages, reports, and presentations based on data saved in structured formats. It's the "create-it-once, used-it-many-times" idea, which allows the user to work smarter and faster by streamlining the creation process and reducing the margin for error among different versions of the same document".

Table 5. Microsoft Office 2003 edition

Basic	Standard	Small Business	Professional Edition	Professional Enterprise Edition
Word	Word	Word	Word	Word
Excel	Excel	Excel	Excel	Excel
Outlook	Outlook	Outlook with Business Contact Manager	Outlook with Business Contact Manager	Outlook with Business Contact Manager
	PowerPoint	PowerPoint	PowerPoint	PowerPoint
		Publisher	Publisher	Publisher
			Access	Access
				InfoPath

5. Microsoft Office 2007

In [2007-MsOffice] the authors underlined that in Microsoft Office 2007 "…the entire user interface has been redesigned to be more intuitive, easier to navigate, and better suited to the task at hand…the developers pf Office 2007 …decided to go back to the drawing board and create an interface based on the way people use their computers today. The result is a simplified, smart system that brings you just the tools you need, when you need them. No more clicking through menus, submenus, and nested dialog boxes. Now the command you need come to you, depending on the type of object you select and the application you are using". All those features are represented in the Table 6.

The newest version of Microsoft Office is 2007, which was released at the same time as Windows Vista (on January 30, 2007). The different editions of Microsoft Office 2007 are:[7]

- Microsoft Office Home and Student 2007 (not for use by commercial entities)
- Microsoft Office Standard 2007 (Retail and volume license)
- Microsoft Office Small Business 2007 (Retail and volume license)
- Microsoft Office Professional 2007 (Retail only)
- Microsoft Office Ultimate 2007 (Retail only)

Limited Availability of Microsoft Office 2007 are:

- Microsoft Office Basic 2007 (Available only through OEMs)
- Microsoft Office Professional Plus 2007 (Available only through volume licensing)
- Microsoft Office Enterprise 2007 (Available only through volume licensing)

Table 6. Microsoft Office 2007 edition

Basic	Home and Student	Standard	Small Business	Professional	Ultimate	Professional Plus	Enterprise
No box shot available						No box shot available	No box shot available
Word	Word	Word	Word	Word	Word	Word	Word
Excel	Excel	Excel	Excel	Excel	Excel	Excel	Excel
	PowerPoint	PowerPoint	PowerPoint	PowerPoint	PowerPoint	PowerPoint	PowerPoint
Outlook		Outlook	Outlook	Outlook	Outlook	Outlook	Outlook
			Accounting Express	Accounting Express	Accounting Express		
			Publisher	Publisher	Publisher	Publisher	Publisher
				Access	Access	Access	Access
					InfoPath	InfoPath	InfoPath
					Groove		Groove
	OneNote				OneNote		OneNote
						Communicator	Communicator

6. Microsoft Software perspectives

6.1. Common features of Microspoft-Office-2007-Windows-Vista

A major feature of the Office suite is the ability for users and third party companies to write **add-ins** that extend the capabilities of an application by adding **custom commands and specialized features**. The type of add-ins supported differ by Office versions:

- Office 97 onwards (standard Windows DLLs i.e. Word WLLs and Excel XLLs)
- Office 2000 onwards (COM add-ins) [8]
- Office XP onwards (COM/OLE Automation add-ins)[9]
- Office 2003 onwards (Managed code add-ins) [10]

These programs are included in most editions of Microsoft Office.

6.2. Microsoft Office 2007

The newest version of Microsoft Office is 2007, which was released at the same time as Windows Vista (on January 30, 2007). The different editions of Microsoft Office 2007 are:[11]

- Microsoft Office Home and Student 2007 (not for use by commercial entities)
- Microsoft Office Standard 2007 (Retail and volume license
- Microsoft Office Small Business 2007 (Retail and volume license)
- Microsoft Office Professional 2007 (Retail only)
- Microsoft Office Ultimate 2007 (Retail only)

6.3. Microsoft Office 2008 for Mac

Microsoft Office 2008 for Mac is available in three editions. All three editions include Entourage 2008, Excel 2008, PowerPoint 2008 and Word 2008.

- Office 2008 for Mac Home and Student Edition
- Office 2008 for Mac Standard Edition *(adds Microsoft Exchange support)*
- Office 2008 for Mac Special Media Edition (Microsoft Exchange support and Expression Media)

6.4. Prior editions for Microsoft Windows

Prior Editions for Microsoft Windows are:

- August 30, 1992: Office 3.0 (CD-ROM version: Word 2.0c, Excel 4.0a, PowerPoint 3.0, Mail): (repackaged as Office 92).
- January 17, 1994: Office 4.0 (Word 6.0, Excel 4.0, PowerPoint 3.0).
- July 3, 1994: Office for NT 4.2 (Word 6.0 [32-bit, i386, MIPS, PowerPC, and Alpha], Excel 5.0 [32-bit, i386, MIPS, PowerPC, and Alpha], PowerPoint 4.0 [16-bit], "Microsoft Office Manager").
- June 2, 1994: Office 4.3 (Word 6.0, Excel 5.0, Power Point 4.0, Mail 3.2 and in the pro version, Access 2.0.). This is the last 16-bit version. This is also the last version to support Windows 3.x, Windows NT 3.1 and Windows NT 3.5 (Windows NT 3.51 was supported up to and including Office 97).
- August 30, 1995: Office 95 (7.0) (Word 7 for Windows 95, etc.) - coincided with the Windows 95 operating system release.
- December 30, 1996: Office 97 (8.0) (Word 97, etc.) (was published on CD-ROM as well as on a set of 45 3½-inch floppy disks), was Y2K safe with Service Release 2. Last version to support Windows NT 3.51 on i386 and Alpha.
- June 20, 1998: Office 97 Powered by Word 98 (8.5) was released only in Japanese and Korean editions. First version to contain Outlook 98 in all editions and Publisher 98 in the Small Business Edition. The only way to get Word 98. And also the first version of Office 97 to support Windows 98.

- January 27, 1999: Office 2000 (9.0) (Word 2000, etc.). Last version to support Windows 95, and last version that does not use Product Activation.
- May 31, 2001: Office XP (10.0) (Word 2002, etc.). Last version to support Windows 98/Me/NT 4.0. Improved support for working in restricted accounts under Windows 2000/XP.
- November 17, 2003: Office 2003 (11.0) (Word 2003, etc.). Last version to support Windows 2000.
- January 30, 2007: Office 2007 (12.0) (Word 2007, etc.). Broadly released alongside Windows Vista, Microsoft's newest operating system.
- Unknown (possibly the first half of 2009):[12] Office 14. There will be no Microsoft Office 13 due to superstition. [13]

6.5. Windows Lifecycles

Beginning in 2002, Microsoft instituted a policy of Support Lifecycles including [4], [5]:

- Earlier versions than Office 97 (including Outlook 97) are no longer supported.
- Office 97 (including Outlook 98) – Mainstream hotfix support ended on August 31, 2001. Extended hotfix support ended on February 28, 2002. Assisted support ended on January 16, 2004.
- Office 2000 – Mainstream support ended June 30, 2004. Extended support is available through July 14, 2009.
- Office XP – Mainstream support ended July 11, 2006. Extended support will be provided until July 12, 2011.
- Office 2003 - Mainstream support will end on April 14, 2009. Extended support will end at April 8, 2014.
- Office 2007 - Mainstream support will end on April 10, 2012. Extended support will end on April 11, 2017.

Current and future versions - Mainstream support will end 5 years after release, or 2 years after the next release, whenever time is later, and extended support will end 5 years after that.

6.6. Difficulties in porting Office

Microsoft develops Office for Windows and Mac platforms. Beginning with Mac Office 4.2, the Mac and Windows versions of Office share the same file format. Consequently, any Mac with Office 4.2 or later can read documents created with Windows Office 4.2 or later, and vice-versa. Microsoft Office 2008 for the Mac drops VBA support. [14], Microsoft has replaced VBA with support for AppleScript. As a result, macros created with Office for Windows will not run on Office for the Mac, and vice versa. In addition, Microsoft has also ceased development on Microsoft Virtual PC. [15].

There were efforts in the mid 1990s to port Office to RISC processors such as NEC / MIPS and IBM/ PowerPC, but they met problems such as memory access being hampered by data structure alignment requirements. Difficulties in porting Office may have been a factor in discontinuing Windows NT on non-Intel platforms.

6.7. Criticisms of Microsoft Office

Microsoft Office is commonly criticized for its security issues and infections from macro viruses.[16] Secunia reports that out of the 15 vulnerabilities reported in 2006 for Microsoft Office 2003 (Standard Edition), 20% are unpatched, 33% are marked as Extremely Critical and 53% are marked as Highly Critical.[17]

Another common criticism of Microsoft Office is its preference of proprietary formats over open standards to store data, which is often intended to be shared with other users, hence forcing them into adoption of the same software platform. [18] However, Office

Open XML, the document format for the latest versions of Office for Windows and Mac, is an ECMA standard and open for implementation to anyone. Microsoft has freely published the complete format documentation under the Open Specification Promise [19] and has made available free downloadable converters for previous versions of Microsoft Office including Office 2003, Office XP, Office 2000 and Office 2004 for the Mac. Implementations of Office Open XML exist on the Mac platform (iWork 08) and Linux (OpenOffice Novell Edition).

Microsoft Office for Mac has for long been criticized for its lack of support of Unicode and BiDi languages, notably Arabic and Hebrew. This has not changed in the 2008 version. [20,21]

7. Adaptable Software

The Adaptable Software can be created on the base of adaptable computing theory [Tod-07.1] which is based on the notion of adaptable tools.

ADAPTOR is a meta-system tool which supports the Adaptable Software creation and application. This meta-system tool support Adaptable Software flexibility (extension and reduction):

Adaptable language	A	Adaptable processor
New data	D	New actions
	A	(operators, statements,
	P	controls)
	T	
Extension definition	O	Extension call
	R	

ADAPTOR is composed from the pragmatic, syntactic, semantic, environment, and examples element's component parts:

_BL_ <element's pragmatics>
_SY_ <element's syntax>
_SE _ <element's semantics>
_CO_ <element's usage context>
_EX_ <element's examples call>
_EL_

The ADAPTOR' s component parts support both the process of adaptation of languages and of processors as component parts of Adaptable Systems. ADAPTORs permit the definition, modification, and usage processes of derived data, operations, instructions, and controls Software' elements.

Each ADAPTOR is represented by the corresponding EXTENDER and REDUCTOR. Both, the EXTENDER and REDUCTOR permit to obtain special and universal Software for human-machine communication in the Information and Knowledge based Societies.

The ADAPTORs permit the Bottom-Up, Top-Down, and Horizontal Adaptable Software' Development.

Adaptable Software is represented by Adaptable Software's Basis and its flexible environment.

7.1. Adaptable Software's Basis

Adaptable Software's Basis is represented by the Definition, Fixation, Calling, and Reduction Adaptable sub-systems. The Definition sub-system implements the extension definition: Fixation sub-system fixates the extension definition in the Adaptable Software; Calling sub-system implements the extension call in the Adaptable Software; Reduction sub-system creates the individual Adaptable Software.

7.2. First level of Adaptable Software

There are distinguished three different types of the first complexity level of Adaptable Software. They are based on three types of the Adaptable Software creation methods: (1) the Extension's Time Implementation Method (E-T-I-M), (2) the Extension's Level Implementation Method (E-L-I-M), and (3) the Processor's Type Implementation Method (P-T-I-M).

7.3. Second level of Adaptable Software

There are distinguished three different types of the second complexity level of Adaptable Software: the ELIM-PTIM type, the ETIM-PTIM type, and the ELIM-ETIM type.

The second complexity level ELIM-PTIM type of Adaptable Software, for example, is represented by the L-L-Preprocessors, L-D-Preprocessors, and L-L-D-Preprocessors. The second level ELIM-PTIM type of Adaptable Software is created on the base of Extension's Level Implementation and of Processor's Type Implementation Methods.

The second complexity level ETIM-PTIM type of Adaptable Software is created on the base of Extension's Time Implementation and of Processor's Type Implementation Methods.

The second complexity level ELIM-PTIM type of Adaptable Software is created on the base of Extension's Level Implementation and of Processor's Type Implementation Methods.

It is demonstrated [Tod-08.2] the possibility to realize the second complexity level's Adaptable Software on the base of translation interactions of the first complexity level's Adaptable Software.

7.4. Third level of Adaptable Software

Third complexity level of Adaptable Software is represented by such types of processors as pre-processor-L-L-compiler, inter-processor -L-D-compiler or post-processor-L-L-D-compiler.

The demonstrations of automatically creation of Adaptable Software of the third complexity level is based on the corresponding demonstrations of automatically creation of the Adaptable Software of the first and of the second levels of translation complexity.

7.5. Adaptable Software advantage

It is demonstrated [Tod-08.2] that Adaptable Tools as base for creation, application, and development of Adaptable Software are characterized by a set of advanced linguistics' and processors' features.

It is demonstrated [Tod-07.2] that Adaptable Languages as part of Adaptable Software integrate such linguistics' features as:

- Multilanguageability,
- Universality,

- Speciality,
- Exensibility,
- Dialectability,
- Compactibility of Basis,
- Reducability,
- Effectivity of modification,
- Continuity of Human and Machine experience,
- Touchability to the Formal Natural Language level of Human-Machine Interactions.

It is demonstrated [Tod-07.2] that Adaptable Processors as part of Adaptable Software integrate such processors' features as:

- Universality,
- Mobility,
- Transferability,
- Cognisability,
- Specializability,
- Minimizing of Processors' Quantity, and
- Raising the level of Adaptable Sofrtware to the level of Problem Formulation.

8. Conclusion

Human social and economic demand and supply for Adaptable Software in the Information and Knowledge Based Societies is too important.

Different types of Adaptable Software will have different domains of its applicability in the process of computerized human-machine intelligent interaction. This process conducts to develop human-machine interaction on the base of Natural Language Processing Adaptable Software.

The Adaptable Software forms new industry branch of Informational technologies of the Information and Knowledge Based Societies.

Bibliography

1. Cox, J., Frye, C., Dow, L. M. III, Lambert, S., Preppernau, J. and Murray. K. **Step by Step 2007 Microsoft Office System,** Microsoft Press, Redmond, Washington 98052-6399, 2007, pp. I-CII, pp. 1-641
2. Murray, K. **First Look Microsoft Office 2003,** Microsoft Press, Redmond, Washington 98052-6399, 2003, pp. I-XV, pp. 1-216
3. Nelson, S. L. **Office 2000: The Complete Reference,** Osborne/McGraw-Hill, Berkeley, Toronto, 1999, pp. I-XXXVII, pp. 1-1040
4. Todoroi, D. and Micusha, D. **Adaptable Software,** 2008, to be published
5. Todoroi, D. **Sisteme adaptabile de programare. Generalizari,** Analele ASEM, Ed. ASEM, 2007, p. 210 – 221
6. Todoroi, D. **The Software' Adaptable shells creation models,** in **"**Proceedings of the 32nd Annual Congress of ARA", July, 22th – 26th, 2008, Boston, USA, 2008, to be published
7. Todoroi, D., Micusha, D., Todoroi, Z. and Spataru, S. **From Recursive programs to the Adaptable processors construction,** in "Proceedings of the 31th Annual Congress of ARA "Pro-Active Partnership in Creativity for the Next Generation", Presses Internationales Politechnique, (Transilvania University of Brasov, Romania, July, 31-August, 5, 2007), Quebec, Canada, Vol. I, 2007, pp. 445-448
8. Todoroi, D., Nazem, S., Jucan, T. and Micusha, D. **Transition to a Full Information Society: Stage Development,** Working Paper No. 98-2, March, UNO, College of Business Administration, Omaha, Nebraska, USA, 1998, 38 p.
9. Zisman, A. **MS Office 95: This Suite smells of success,** First published in Our Computer Player, October 1995; http://www.zisman.ca/Articles/1995/ Office95.html
10. * * * **Microsoft Office,** Wikipedia, the free encyclopedia, Part I -X; http://en.wikipedia.org/wiki/Microsoft_Office

[1] Dumitru Todoroi holds a PhD granted by Ukraine AS Cybernetic Center, Kiev in 1982. He also gained a MSc from USSR AS Computer Center, Moscow in 1968. He currently is University Professor, PhD Adviser, at the Academy of Economic Studies of Moldova, Kishinev and, since 2002, Director of "Economic Research Center" from the same university. He was a member of The Moldova Parliament in the period from 1990 to 1994.
Last 10 year he taught courses like: "XXI Century Societal Information Systems", "Adaptable Computing", "Adaptable Data & Knowledge bases", "Object-Oriented programming", "High-performance Compilers and Run-time Systems", "Integrated Systems" with students of the Academy of Economic Studies of Moldova and the Romania "Al. I. Cuza" University at Iassy.
His research and teaching interests include Include: XXI Century Societal Information Systems, Adaptable Computing, Natural language processing, Data structures, Software engineering, Data & Knowledge bases management systems, Program development and software design, Natural language processing, Object & Action - oriented programming, Software development, Computer aided education systems, Intelligent computer communications, Multimedia, networking and visualization.
He published over 270 papers, 11 Scientific Books and 18 Educational Books.

[2] Address for correspondence:
Banulescu-Bodony str., 69, 607"B", Chisinau MD 2005, Republic of Moldova

[3] Codification of references:

[2007-MsOffice]	Cox, J., Frye, C., Dow, L. M. III, Lambert, S., Preppernau, J. and Murray. K. **Step by Step 2007 Microsoft Office System,** Microsoft Press, Redmond, Washington 98052-6399, 2007, pp. I-CII, pp. 1-641
[2003-MsOffice]	Murray, K. **First Look Microsoft Office 2003,** Microsoft Press, Redmond, Washington 98052-6399, 2003, pp. I-XV, pp. 1-216
[2000-MsOffice]	Nelson, S. L. **Office 2000: The Complete Reference,** Osborne/McGraw-Hill, Berkeley, Toronto, 1999, pp. I-XXXVII, pp. 1-1040
[Tod-08.2]	Todoroi, D. and Micusha, D. **Adaptable Software,** 2008, to be published
[Tod-07.1]	Todoroi, D. **Sisteme adaptabile de programare. Generalizari,** Analele ASEM, Ed. ASEM, 2007, p. 210 – 221
[Tod-08.1]	Todoroi, D. **The Software' Adaptable shells creation models,** in **"**Proceedings of the 32nd Annual Congress of ARA", July, 22th – 26th, 2008, Boston, USA, 2008, to be published
[Tod-07.2]	Todoroi, D., Micusha, D., Todoroi, Z. and Spataru, S. **From Recursive programs to the Adaptable processors construction,** in "Proceedings of the 31th Annual Congress of ARA "Pro-Active Partnership in Creativity for the Next Generation", Presses Internationales Politechnique, (Transilvania University of Brasov, Romania, July, 31- August, 5, 2007), Quebec, Canada, Vol. I, 2007, pp. 445-448
[Tod-98]	Todoroi, D., Nazem, S., Jucan, T. and Micusha, D. **Transition to a Full Information Society: Stage Development,** Working Paper No. 98-2, March, UNO, College of Business Administration, Omaha, Nebraska, USA, 1998, 38 p.
[1995-Office]	Zisman, A. **MS Office 95: This Suite smells of success,** First published in Our Computer Player, October 1995; http://www.zisman.ca/Articles/1995/ Office95.html
[1990-Office]	* * * **Microsoft Office,** Wikipedia, the free encyclopedia, Part I -X; http://en.wikipedia.org/wiki/Microsoft_Office

[4] http://en.wikipedia.org/wiki/Microsoft_Office#_note-8#_note-8

[5] http://en.wikipedia.org/wiki/Microsoft_Office#_note-1#_note-1

[6] http://en.wikipedia.org/wiki/Microsoft_Office#_note-2#_note-2

[7] http://en.wikipedia.org/wiki/Microsoft_Office#_note-8#_note-8

[8] http://en.wikipedia.org/wiki/Microsoft_Office#_note-3#_note-3

[9] http://en.wikipedia.org/wiki/Microsoft_Office#_note-4#_note-4

[10] http://en.wikipedia.org/wiki/Microsoft_Office#_note-5#_note-5

[11] http://en.wikipedia.org/wiki/Microsoft_Office#_note-8#_note-8

[12] http://en.wikipedia.org/wiki/Microsoft_Office#_note-9#_note-9

[13] http://apcmag.com/microsoft_to_skip_unlucky_office_13.htm

[14] http://en.wikipedia.org/wiki/Microsoft_Office#_note-13#_note-13

[15] http://www.macnn.com/articles/06/08/07/ms.kills.virtualpc/

[16] http://en.wikipedia.org/wiki/Microsoft_Office#_note-14#_note-14

[17] http://en.wikipedia.org/wiki/Microsoft_Office#_note-15#_note-15

[18] http://www.gnu.org/philosophy/no-word-attachments.html

[19] http://en.wikipedia.org/wiki/Microsoft_Office#_note-16#_note-16

[20] http://en.wikipedia.org/wiki/Microsoft_Office#_note-17#_note-17

[21] http://en.wikipedia.org/wiki/Microsoft_Office#_note-18#_note-18

ACADEMIC RESEARCH INTEGRATION SYSTEM

Iulia SURUGIU[1]
PhD Candidate, University of Economics, Bucharest, Romania

E-mail: : iulia_surugiu2003@yahoo.com

Manole VELICANU
PhD, University Professor, Department of Computer Science
University of Economics, Bucharest, Romania

E-mail: mvelicanu@ase.ro

Abstract: *This paper comprises results concluding the research activity done so far regarding enhanced web services and system integration. The objective of the paper is to define the software architecture for a coherent framework and methodology for enhancing existing web services into an integrated system. This document presents the research work that has been done so far in this aspect by applying the proposed architecture for system integration in the academic field. The basics concepts used are Enterprise Application Integration (EAI) and Service Oriented Architecture (SOA), as the most commonly used approaches to information systems integration lately.*

SOA is an Information Technology (IT) architectural style that supports the transformation of businesses into a set of linked services or repeatable business tasks that can be accessed when needed over a network. This may be a local network, it may be the Internet, or it may be geographically and technologically diverse, combining services in New York, Paris, and Beijing as though they were all installed on a local desktop. These services can coalesce to accomplish a specific business task, enabling businesses to adapt quickly to changing conditions and requirements.

Key words: *Service Oriented Architecture; Enterprise Application Integration; Business Process Management*

1. Introduction

One of the most important activities inside a university is scientific research. Academy accreditation is one of the main goals for university, as it attests the quality and performances of the academic activities and its position through similar educational institutions. I consider tremendously important to mention here a phrase we read once in Lewis Carroll's book, *Alice in Wonderland*, which we found interesting to apply in academic field: "*If you want to stay where you are, you must keep moving*". In my opinion, the management of information is the only key factor capable to assure success continuality and real performances. Education and research are fields where science facts grow with astonishing dynamicity and flexibility and adaptive characters become vital for surviving.

Research activities grave the name and fame of the university on educational international context through its member submitted, accepted and pleaded papers, approved and appreciated research projects, the quality and properness of scientific research results, the viability of proposed solutions.

The first step in organizing research is being informed about conferences, symposiums, workshops, scientific research magazines and ISI quoted publications that allow an efficient dissemination of the results of research and attract investors and research financing, another important piece in the activity of research.

Given the previously mentioned context, my proposal is to create an integrated system that allows educational and scientific research services recovery. The target user group of the system is academic environment and research institutes that need quick and organized information on scientific research projects and related research topics.

Wouldn't be useful for a university to have access to a system that, given the present situation of the academy, to come up with a complete and coherent program about all the activities that need to be done in order to obtain an accreditation? All that the user of the system must do is to authenticate with the system, specify the university traits and its educational background features, maybe other distinctions that the academy already has gained, then choose a program that the academy wants to attend, eventually a commission for evaluation or an accreditation institute. The result will be a list of steps to make in order to achieve the goal, which could be: leadership of European research projects, involvements in educational projects financed by governments, usage of required standardization politics, papers and documents given to evidence to certain international educational and/or research organisms.

All these tasks and activities will be identified using Web Services, the existing ones or some others wrapped onto existing applications that have the permission to integrate with our system. Service recovery and integration will be achieved by creating a web services repository that will serve client requests by wrapped web responses.

2. System Architecture

We will present further a detailed architecture of a Web Services Bus, proposed as a whole in another personal paper called "*Health Information Service Bus*", being in course of publishing to "*Informatica Economica*" journal**.**

In general, a UDDI registry is quite focused on Web Services. Users or applications describe Web Services inside the registry using the UDDI interface (based on Web Services). The description of a Web Service inside a UDDI registry covers different aspects such as where is located its WSDL and/or its access point, or what metadata is attached to a service or to its access points.

The Virtualization Engine extends the concept of services to other middleware than just Web Services. It is possible to define services that can be accessed through FTP, XFB, EJB, JMS, SOAP/HTTP … The ways to access it are still the UDDI Web Services and the GUI. The major goal of SOA is service reusability. A described service can be reused if people know about it and if they can find it inside the registry. UDDI allows attaching metadata to the objects managed by the registry in order to ease the search operations. The Virtualization

Engine comes with a pre-defined set of categorization identifiers. There identifiers can be used to describe in a standardized way the metadata that can be attached to a service (for example its status or the version of one of its instances). They can allow the users to tell if a particular service is in production or not, what is its version number, what kind of middleware it uses and many other information.

As described inside the UDDI standard, the categorization data can be added to any level of the UDDI structure. The Virtualization Engine will use specific metadata that will be inserted at the Service level and at the Instance (access point) level.

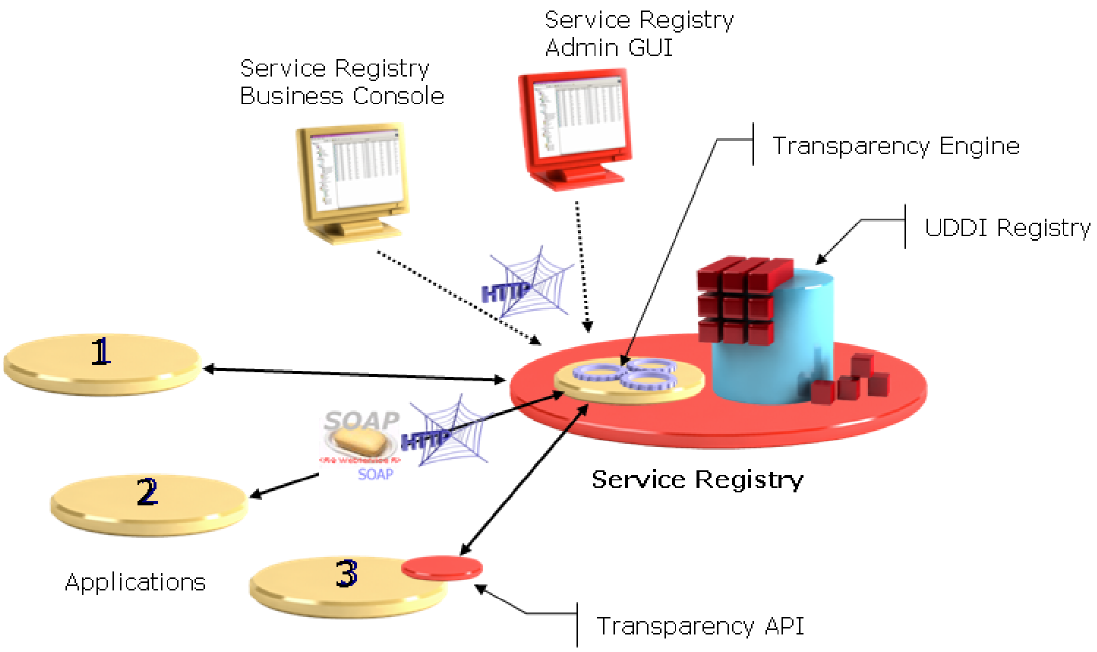


Figure 1. Virtualization Engine

2.1. Architectural overview

The proposed solution is the most obvious for the situations where there are already implemented web services for education and research topics, because it will require minimal but strict modifications on the existing web services. Still, some Web Services (WS) standards might not be directly supported and other problems with standardization could appear and must be solved by modifying existing service structure.

The main disadvantage of the integration solution would be that:

- Not all services fulfill framework requirements.
- Requires implementation effort to cope with unsupported standards.

An architectural overview of the solution proposed would be the following:

- The framework integrates e-Research and e-Learning services and applications through the Enhanced Web Service Framework
- Each e-Research service is published into the UDDI Registry which is running on the integration server
- The UDDI Registry stores metadata into the UDDI Repository along with state
- When a request is made, the public portal calls the front-end Web Services Engine using its service client mode

- The Web Services Engine calls the Virtualization Engine to get the best instance for the requested service
- This returns an endpoint using registry metadata or raises an exception if none was found
- The front-end server forwards the initial call to the given back-end instance which has another Web Services Engine working as a service provider
- If the service is fully compliant, then it will be invoked directly
- If it is partially compliant, then it will be invoked using the Relay Proxy to insure compliancy
- If it is a legacy application, first it will be published as a Web service using the Service Integration Layer and then invoked through the Relay Proxy
- Each Web Services Engine will log its service activity using the Auditing and Monitoring Engine, which updates the global UDDI Repository and its Registry
- Both the front-end and the back-end servers are controlled by the Management Engine, which is used also to manage their Web Services Engines and Auditing and Monitoring Engine

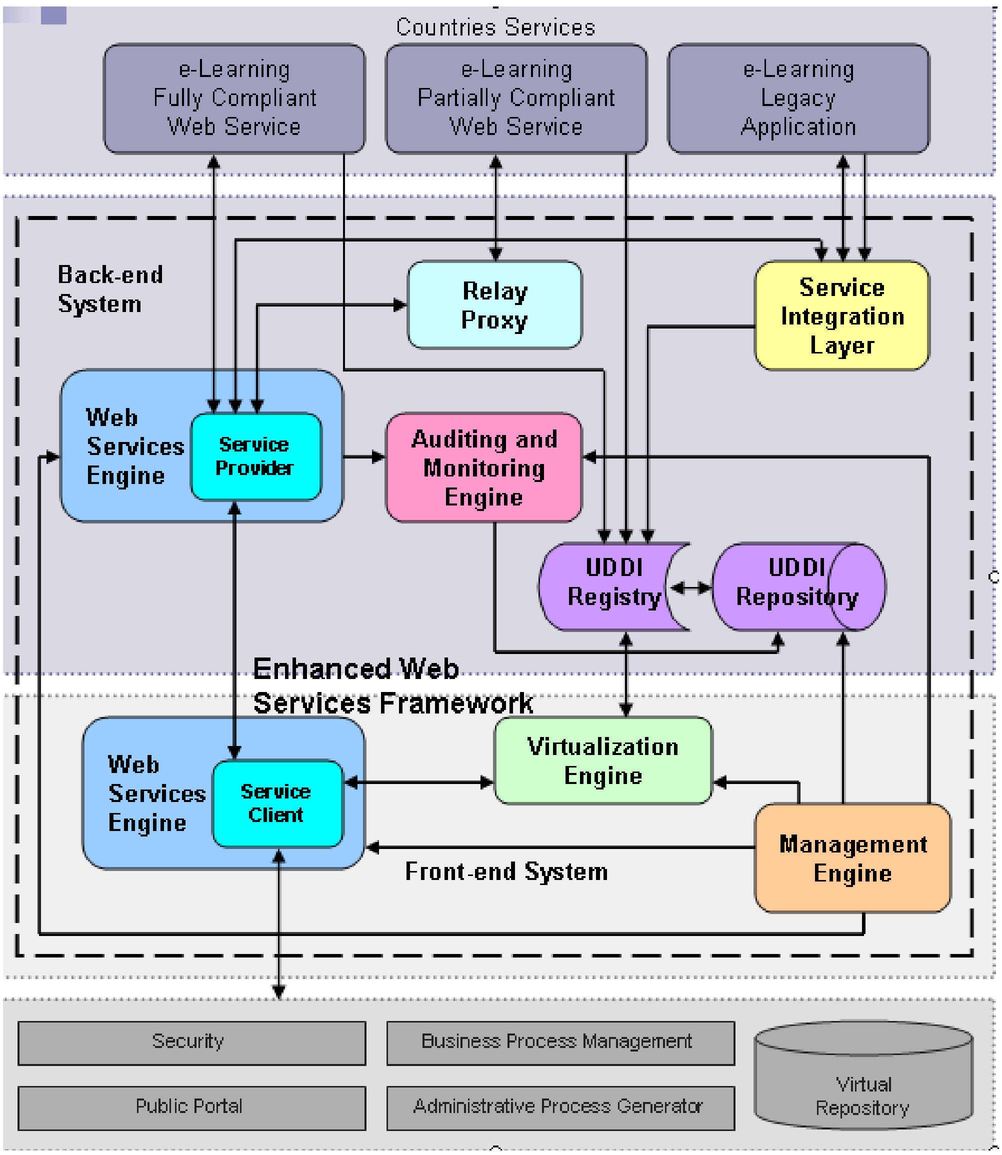


Figure 2. Architectural overview

2.2. Securing Communication with Server Relay Mode

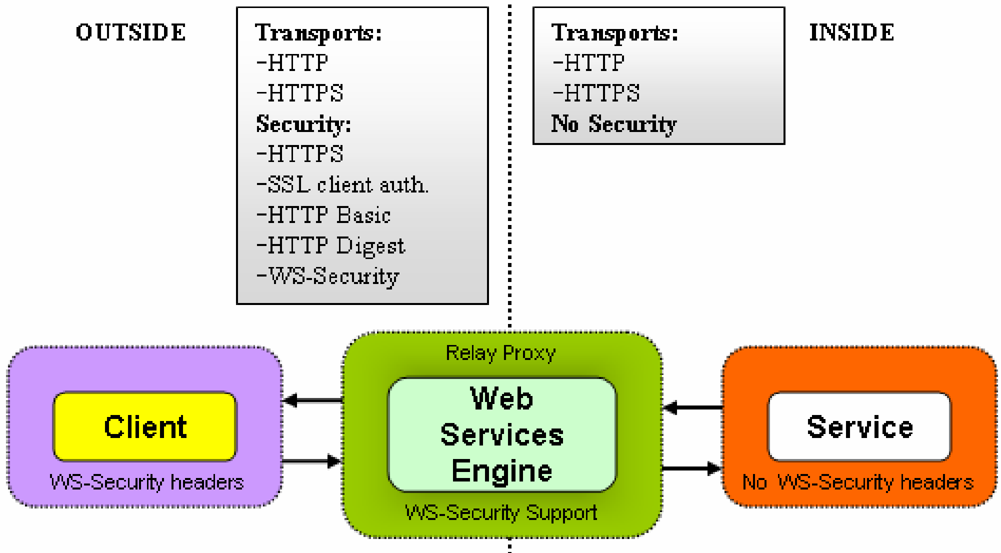


Figure 3. Security with Server Relay Mode

- The client produces the SOAP message with security headers and sends it to the Service Security Proxy.
- The Service Security Proxy strips out security headers from the message and forwards pure SOAP message to the destination Service, injecting security tokens into custom headers.
- When the response arrives, the server extracts custom headers and applies security policy.
- Then, the server delivers the response message to the client.

2.3. Securing Communication with Client Relay Mode

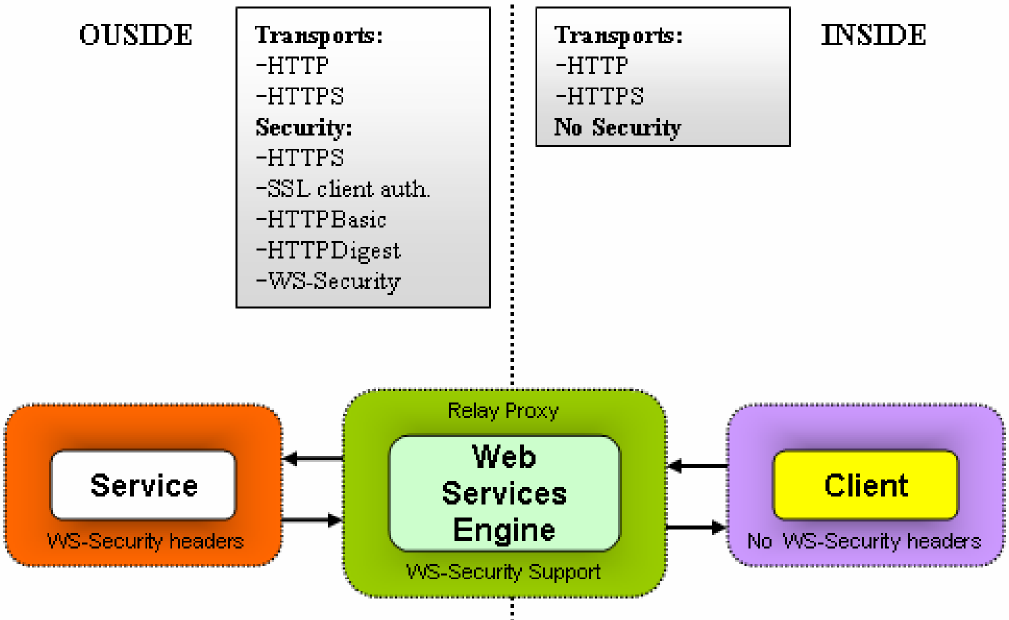


Figure 4. Security with Client Relay Mode

- The client produces the SOAP message with custom headers containing security tokens and sends it to the Service Security Proxy.
- The Service Security Proxy extracts custom headers from the message, applies security policy and forwards SOAP message to the destination service
- When the response arrives, the server strips out security headers and checks security policy
- Then, the server delivers the response to the client

2.4. Managing Communication with Server Relay Mode

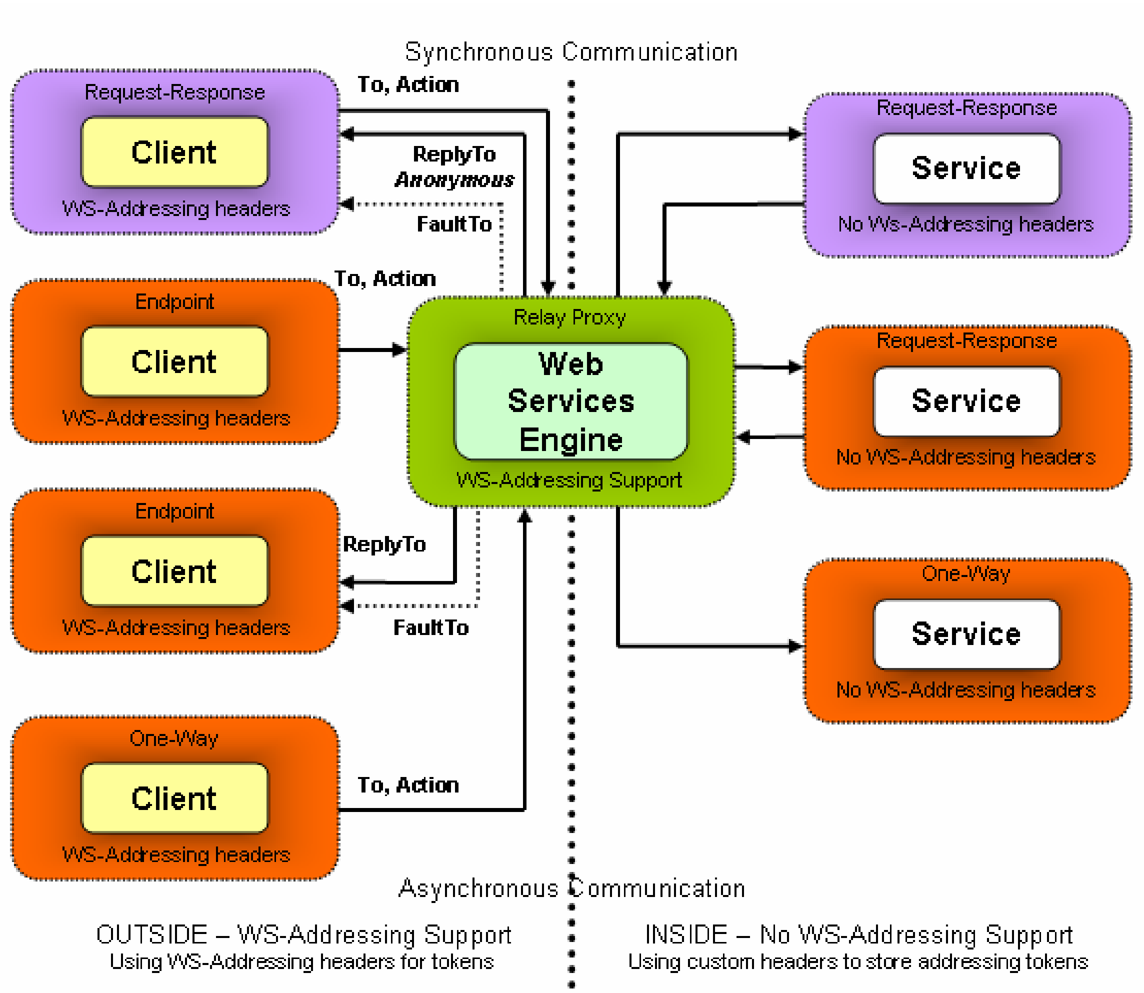


Figure 5. Communication with Server Relay Mode

Asynchronous Communication:

- Client sends a SOAP message to the server containing addressing headers, having a *ReplyTo* other than the anonymous URI
- Server translates addressing headers into custom headers, and then invokes the service
- Service sends back a response message to the server forwarding custom headers with addressing tokens
- Server extracts addressing tokens and uses them to delivery the SOAP message to ReplyTo endpoint
- If the response was a SOAP fault, then the server delivers a it to the *FaultTo* endpoint
- Client can send a One-Way message containing addressing headers

- Server strips out the addressing headers and forwards the One-Way call to the service

Synchronous Communication:

- Client sends a SOAP message to the server containing addressing headers, including an anonymous URI for ReplyTo
- Server translates addressing headers into custom headers, and then invokes the service
- Service sends back a response message to the server forwarding custom headers with addressing tokens
- Server extracts addressing tokens and uses them to deliver the SOAP message to ReplyTo, the same with the initial client endpoint
- If the response was a SOAP fault, then the server delivers it to the FaultTo endpoint

2.5. Managing Communication with Client Relay Mode

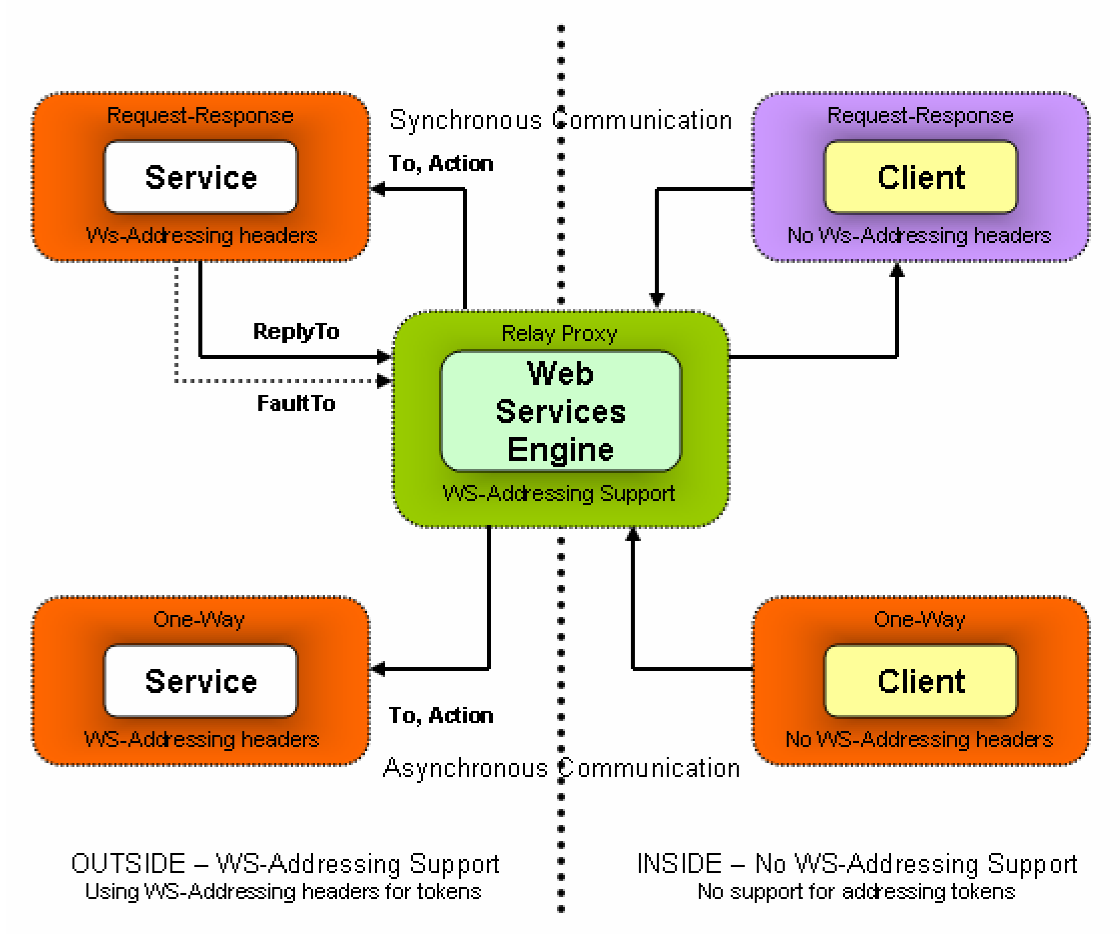


Figure 6. Communication with Client Relay Mode

Asynchronous Communication:

- Client sends a One-Way message to the server without addressing headers
- Server injects addressing headers and forwards the One-Way message to the service

Synchronous Communication:

- Client sends a Request-Response message to the server without addressing headers
- Server injects addressing headers, setting its endpoint as the target of the response and invokes the service

- Service sends back the response message to the server, being the reply endpoint
- Server delivers the response to the client endpoint
- If the response was a SOAP fault, then the server delivers it to the client endpoint

3. Conclusions

As different academic research activities are spread all over the Internet through Web sites, Web Services and portals, we found a way to bring all this information together inside an integrated system in order to be used by universities when involving into scientific research activities. In order to achieve that, it is needed help and support from the research service providers, that must allow access to their services and cooperation in case already existing systems need to be modified for being integrated.

The main objectives set for the proposed WS integration architecture are the following:

- Combining several Web Services additional standards and ad hoc data structures to ensure management of contextual information in order to provide Business Process Management (BPM) support, audit trail and error management.
- Asynchronous behaviour of the Web Services
- Defining a balance between Web Services security capabilities and the underlying security architecture
- Providing location transparency of Web Services in order to provide redundancy and scalability
- Proposing new ways to present Web Services inside a service registry in order to make them easier to understand to business analysts
- Checking with selected technology providers that their implementation of the Web Services standards (basic & additional) allow perfect interoperability within the functional perimeter defined for Web Services usage inside the proposed architecture.

As research and educational services are very heterogeneous and universities disconnected, the architecture proposed by this paper comes as a lien between universities and research institutions and commissions for educational evaluation. The combination of information technologies and concepts such as Web Services, Service Oriented Architecture, Business Process Management, in the way presented, leads to the integration of all those services in order to provide academies with easier and efficient access to a unified educational and research system, no matter the position from which they use system's services.

4. References

1. Sadter, C, Huber, P. and Sangmin, Y. **Enabling SOA Using WebSphere Messaging**, IBM, March 2006
2. Stamos, A. and Stender, S. **Web Services Security – Scenarios, Patterns, and Implementation Guidance for Web Services Enhancements (WSE) 3.0**, Microsoft, November 2005
3. Surugiu, I. **Health Information Service Bus**, sent to publication to "Informatica Economica" Journal, University of Economics, Bucharest, Romania

4. * * * **Business Process Management,** 4th International Conference, BPM 2006, Vienna, Austria
5. * * * **Patterns: SOA Foundation Service Creation Scenario,** IBM publication
6. * * * **Web Service Atomic Transaction Protocol (WS-AtomicTransaction)**, http://www-128.ibm.com/developerworks/library/specification/ws-tx/#atom
7. * * * **Web Services Addressing Protocol (WS-Addresing)**, http://www.w3.org/Submission/ws-addressing
8. * * * **Web Services Secure Conversation Protocol (WS-SecureConversation),** http://specs.xmlsoap.org/ws/2005/02/sc/WS-SecureConversation.pdf
9. * * * **Web Services Security Protocol (WS-Security)**, http://www.verisign.com/wss/wss.pdf
10. * * * **Web Services Trust Protocol (WS-Trust)**, http://www.verisign.com/wss/WS-Trust.pdf

[1] Iulia Surugiu, Software Engineer, TotalSoft S.A., Bucharest, Romania

Gheorghe NOSCA[1]

PhD, Association for Development through Science and Education, Bucharest, Romania

E-mail: r_g-nosca@yahoo.com

PROFESSOR MARCEL DRAGOS STOICA AT 75th ANNIVERSARY

Professor Marcel Dragos Stoica has graduated the Faculty of Civil Buildings, Bucharest, in 1956 and and the Techno-Economics Faculty, Technical University Bucharest in 1966. After 13 years of practical and reasearch activity in the construction domain, he has joined to the staff of the Bucharest Academy of Economic Studies in 1969, starting a very rich activity both in teaching and in scientific research in many fields like economics computing, economicsl processes simulating, operational research, marketing, fuzzy systems, building project management, building management, cybernetics systems.

He holds a PhD diploma in Economics, Economic Cybernetics specialty from 1975.

Currently he is Consulting Professor and a PhD coordinator within the Academy of Economic Studies. He is the author of 14 books, more than 15 monographs and syntheses, over 30 journal articles and more than 45 conferences in Romania and abroad. For his laborious research and teaching activities Prof Stoica has received numerous diplomas. For his entire activity, the National University Research Council granted him in 2005 with the national diploma, Opera Omnia. In 1998 he was nominated Man of the Year 1998 (ABI), 20th Century Award for Achievement (IBC, Cambridge). Prof Stoica is corresponding member of The Romanian – American Academy.

He has, also, received multiple grants for research, documentation and exchange of experience. Professor Stoica is a distinguished member of the scientific board for the magazines and journals like:

- *Fuzzy Economic Review*
- *Economic Computation and Economic Cybernetics Studies and Research*
- *Romanian Review of Informatics and Automatics*

In the last period his work has focused on of the subtle sets theory, and its applications in many fields. In this transdisciplinary area professor Stoica is a pioneer in Romania.

The Editorial Board takes advantage of this opportunity to wish to Professor Stoica all the best, every succes and good health.

Happy Birthday, Professor!

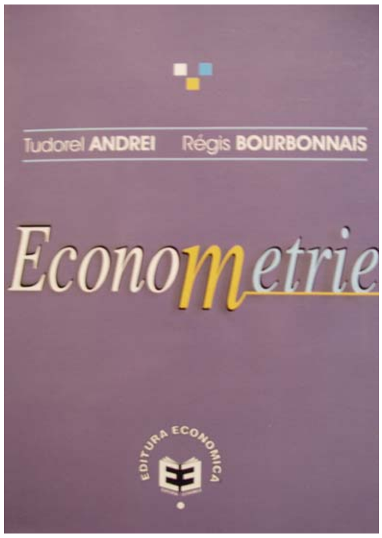
[1] Gheorghe Nosca graduated Mechanical Faculty at Military Technical Academy in 1981, and Cybernetics, Statistics and Informatics Economics Faculty at Academy of Economics Studies in 1992.
He obtained his PhD degree in Economics, Cybernetics and Statistics Economics specialty in 2003.
He is currently researcher at Association for Development through Science and Education.
He has published (in co-operation) 3 books, 16 articles in informatics journals.
He has taken part in about 20 national and international conferences and symposiums.
His research interests include data quality, data quality management, software quality cost, informatics audit, and competitive intelligence.

Daniel Traian PELE
PhD, Assistant Professor, Department of Statistics and Econometrics
University of Economics, Bucharest, Romania

E-mail: danpele@ase.ro, danpele@yahoo.com, **Web page:** http://www.danielpele.ase.ro

Key words: *econometrics; econometrie; Tudorel Andrei, Régis Bourbonnais*

Book Review on
ECONOMETRICS ("ECONOMETRIE")
by Tudorel ANDREI and Régis BOURBONNAIS
Editura Economica, Bucharest, 2008



The domain of economic phenomena shaping through quantitative methods is one of extreme current interest, from the perspective of theoretical approaches as well as from the one of applications in real life.

In a world continuously expanding, where the explosion of information is overwhelming, there is an acute need for a synthetic spirit who systematically organizes the methods, techniques and instruments necessary to approach a concrete issue from the econometric perspective.

From this viewpoint, the book of Tudorel Andrei and Regis Bourbonnais (Econometrie, Editura Economica, Bucharest, 2008) is a welcome item on the market of specialty publications.

Benefitting from the fortunate collaboration of the two authors, both working in the higher education system (the first at the Academy of Economic Studies Bucharest and the second at the University Paris-Dauphine), the outcome resulted in a well structured paper, covering the classical issues of econometrics as well as the regression models, or the analysis of time series and other modern issues such as the Tobit models.

The book is outstanding by the harmonious joining of theoretically describing models and practically applying them through statistic software programs. Thus, each chapter has numerical examples accompanied by the Eviews program, necessary to resolve them.

A special attention is paid to data, its quality being essential to the result of any econometric process. Consequently, the authors provide a complete list of electronic data sources, useful to analyse at macroeconomic levels.

Moreover, the book can easily represent a useful instrument in the research work of postgraduates in economic science, and of those who use quantitative methods in general. It is to be appreciated the fact that the authors emphasise the methodology of quantitative methods, starting from the accuracy of data series, passing to assessing the hypotheses, up to the degree of trustworthiness of the conclusions. This process must be present in the list of research instruments of any person with skills in quantitative shaping.

To conclude, the book is recommended as referential to the young researchers in econometrics.

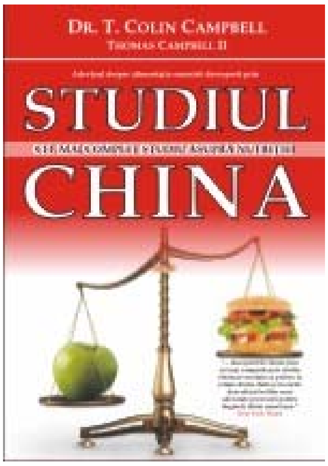
Claudiu HERTELIU
PhD, Assistant Professor, Department of Statistics and Econometrics,
University of Economics, Bucharest, Romania

E-mail: claudiu.herteliu@gmail.com, **Web page:** http://www.hertz.ase.ro

Key words: *China Study critique; nutrition; life style; quantitative; T. Colin Campbell, Thomas Campbell II*

CRITICAL OBSERVATIONS UPON THE METHODOLOGICAL ASPECTS IN THE "CHINA STUDY" VOLUME

THE CHINA STUDY – The most comprehensive study of nutrition ever conducted and the startling implications for diet, weight loss and long-term health ("STUDIUL CHINA. Cel mai complet studiu asupra nutritiei")
by Dr. T. Colin Campbell and Thomas Campbell II
Casa de Editura Advent, Rm. Valcea, 2007[1]



A book of great international prestige, considered to be a best-seller (with over 100.000 readers worldwide), recommended by many remarkable persons[2], declared the book of the month or the book of the year[3] by some organisations, was also published in Romania in April 2007. The book gained success[4] and excited the interest of Romanian readers[5]. In the followings we wish to make a brief critical presentation, highlighting the quantitative elements included in the book.

Written by father-son collaboration, combining the vast experience of a famous American scientist[6] with a career of over 40 years in nutrition, with the youth and written communication skills of his young follower[7], the book claims to be a complete guide in nutrition and its implications upon health. **The main idea presented by this book is that "western" diseases (especially cardiovascular diseases, diabetes, coronary diseases, and the various types of cancer) can be fought against in a drastic manner, preventively, trough diet means. T**he book is structured into four sections: the first is dedicated to actually presenting the China Study; the second presents the so-called wealth-diseases; the third presents a short guide of correct nutrition, and the last section (the fourth) is dedicated to vast critical constructions about the fact that

> *"the entire system – government, science, medicine, industry and media – promotes profits in the detriment of health,*

technology in the detriment of nutrition and confusion in the detriment of transparency."[8]

Practically, the study that entitles the book only represents a quarter of it (approximately 100 pages) the rest representing additional elements of comparison and analysis.

In order to avoid confusions, in the text we shall use words like book, volume, paper when referring to the entire publication, and the expression "The China Study" when referring strictly to the biomedical research conducted in China.

The China Study is an experimental research starting from an idea resulted after the studies performed in China in the early '70s; the results were published in the paper "Cancer Atlas Survey"[9]. The team was formed of experienced[10] researchers, including two Chinese[11]. The methodology of gathering information was based on a sample of 65 districts located in 24 of the 27 regions of China. Out of each district included in the sample 2 villages were selected, and biological tests[12] were taken from each village, from 100 volunteers (50 men and 50 women) with ages from 35 to 64. To summate, tests were taken from 13.000 persons. Because of the fact that the blood samples could be analysed, due to small quantities, in a reduced number of tests the blood samples from persons of same gender and village were combined, thus creating the possibility (by the considerable enhancement of blood quantity) of collecting information for 367 variables. Several times[13], the book mentions that *"over 8.000 statistical significant associations"* were obtained through the analysis of the 367 variables. Far from bringing accusations to the intellectual honesty and the efforts made by all researchers committed to "The China Study" experimental research, some critical comments are worth mentioning:

- First of all, we must observe that there is no mention of the manner of sample selection (both the villages sample and the sample of persons whom were biologically tested). Obviously, if the selection was not at random, an eventual statistical inference in the conclusions for the entire population of China is misplaced.
- Secondly, although the experimental study was realized by collecting biological samples from 13.000 persons, a big question mark appears as a consequence of the combination of all blood samples from persons of same gender and village. This fact has implications such as:
    i. Through this procedure, the volume of cases analysed by statistic means was drastically reduced from 13.000 to 260 (2 sets of blood samples –one for male one for female – from every village, each multiplied by 2 – two villages of every district – and then by 65). Changing the size class, in a descendent way, leads to a serious alteration in applying the statistical inference;
    ii. By this combination, the operations were performed with a medium level for each gender from each village, implicitly supposing that there are no individual differences as against to the average of the village. This assumption was never proved. Ulterior, on the base of the associations between these medium levels at the range of village, conclusions referring to individual diet behaviour were drawn. Although a village is composed of individuals, their behaviour and genetic inheritance is not identical. In this context, any conclusion of this study applied to an individual level is not appropriate.
- Third of all, the 8.000 statistically significant associations are impressive as volume. Still, the maximum number of combinations in which the 367 variables

could be analyzed is 67.161 (C_{367}^{2}). The 8.000 statistically significant associations represent under 12% of the total number of possibilities. There are no details upon how many of the 8.000 associations were in the direction claimed by the authors. It is yet hard to believe that absolutely all were positive; there surely must have also been statistically significant associations but which infirmed the hypotheses of the authors.

In conclusion, the volume, despite the over 700 bibliographic references, is one that treats the subject descriptively, with the purpose of rather spreading the science than purely scientific. This fact is stated by the authors in the note at the beginning of the VIIIth page:

> *"No information in this book should be looked upon as a substitute of medical consultation; also, no change in diet, physical exercise, or in lifestyle is to be made without a previous advice of the personal doctor, especially if there is a treatment for reducing heart-disease risk, arterial pressure, or any form of adult diabetes."* [14]

All these technical observations relatively to the methodological aspects specific to quantitative methods don't seriously diminish the value of the book. The volume remains one of reference in Nutrition, having as author an American scientist with an impeccable reputation, whom has dedicated his entire career to researches in this domain. The fact that it is easier to prevent than to treat is well known, and although there are some inconsistencies from the point of view of quantitative methodology standards, the book offers adequate advices and directions taking into account the behavioural features of the world we live in.

---

[1] Translation from English original version: Dr. T. Colin Campbell and Thomas Campbell II **The China Study – the most comprehensive study of nutrition ever conducted and the startling implications for diet, weight loss, and long-term health**, BenBell Books, Dallas, Texas, 2004

[2] A sort list of considerations is presented on cover IV and in the beginning of the book including important people diverse background (from Nobel Prize laureates –in Physics – to former officials of the World Bank).

[3] It received, for example, the award for the book of the year-2005 on behalf of the editors of VegNews Magazine, and the book of the month on behalf of the site: www.1001bacau.ro

[4] Despite the fact that the publishing house where it appeared is not among the important names in the field.

[5] A simple web search on www.google.com about the "China Study" offers over 135 thousant results (only in Romanian, because in English it offers over 247 million results)

[6] Dr. T. Collin Campbell graduated the university studies (undergraduate, master and PhD program) at Cornell University. He then became associate researcher at Massachusetts Institute of Technology, and he had an university career for 10 years at Virginia University. Afterwards he was in charged with a didactic position at Cornell University.

[7] As his father did, Thomas Campbell II graduated Cornell University and presently attends the Medical School of Buffalo University- after reorienting his career with this book. Thomas Campbell II 's hobbies are soccer, ski, hiking, marathons, writing and acting.

[8] The text is our translation of the paragraph from Romanian version which can be founded at page 270: "intregul sistem – guvern, stiinta, medicina, industrie si mass-media – promoveaza profituri in defavoarea sanatatii, tehnologiei in defavoarea alimentatiei si confuzie in defavoarea transparentei."

[9] *** **Atlas of Cancer Mortality in the People's Republic of China,** China Map Press, Shanghai, China, 1979

[10] We mention Professor Sir Richard Peto: Professor of Medical statistics & Epidemiology at the University of Oxford, United Kingdom.

[11] It is about one of the co-authors of "Cancer Atlas Survey": Dr. Junyao Li and Dr. Junshi Chen Assistant Deputy Chief of the main – at the time - laboratory of research in health and nutrition in China.

[12] Blood from all the persons included in the sample and urine from 50% of them.

[13] For example: page 7 or page 77

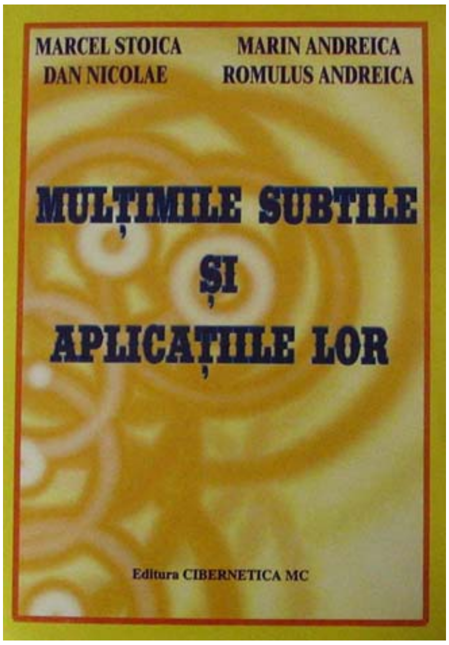
[14] The text is our translation of the disclaimer from Romanian version which can be founded at page VIII[th]: "Nici o informatie din aceasta carte nu va trebui privita ca un substituient al consultatiei medicale . De asemenea, nici o schimbare in dieta, exercitii sau stil de viata nu va fi facuta fara o consultare prealabila a medicului personal, in special daca este in curs un tratament impotriva reducerii riscului de boala de inima, tensiune arteriala sau vreo forma de diabet la adulti."

Gheorghe NOSCA[1]
PhD, Association for Development through Science and Education, Bucharest, Romania

E-mail: r_g-nosca@yahoo.com

Key words: *subtle sets, fuzzy; socio-economic indicators, indicators aggregation*

Book Review on
SUBTLE SETS AND THEIR APPLICATIONS ("MULTIMI SUBTILE SI APLICATIILE LOR")
by Marcel STOICA, Marin ANDREICA, Dan NICOLAE and Romulus ANDREICA
CIBERNETICA MC Publishing HOUSE, Bucharest, 2008



Subtle Sets and Their Applications is an interdisciplinary work, the first publication in this field in Romania, opening a new research domain.

The book has six chapters: The Subtle Sets in Economy, Using the Subtle Sets to Evaluate Socio – Economics Phenomena, Theorems Regarding The Aggregations of The Socio – Economics Indicators, The Progresses Facilitated by The Subtle Set in the Socio –Economic Processes Investigation, Applications in Welfare Partnership Building, The Possibilities to Modeling Socio - Economics Systems Evolution through Sublimation.

The profound transformation that have taken place in the human society structure, in the last decades has determined a conceptual rethinking of the way to achieve the micro-, medium- , and macro – economics systems studies. It is necessary to take into consideration the inter-human relationship and the morality aspects that have been neglected till now. In this stage of the technical – scientific revolution it is compulsory to take into consideration that information which allows describing the functions of the structure determining components. It is possible in this way to pass from the superficial notice of the apparent to the profound study of the subtle.

The subtle sets both from practical and theoretical points of view offer the possibility to unify the conception to investigate very different domain of the science. In this way the medicine will play a central role having in view its millenaries experience.

The book is addressed both to the researchers in the interdisciplinary domains, and to students who intend to develop such kind of researches.

[1] An author's short presentation is available at page 105.