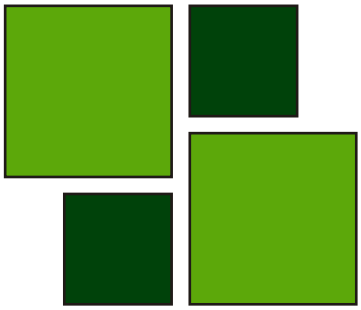


WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Recession Prediction Modeling

Vol. 4
No. 2
Summer
2009

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

JOURNAL OF APPLIED QUANTITATIVE METHODS

Contents

JOURNAL OF APPLIED QUANTITATIVE METHODS

Contents

AN ECONOMETRIC APPROACH REGARDING ANTICRISIS MEASURES

Eugen Stefan PECICAN[1]
PhD, University Professor, University of Economics,
Department of Statistics and Econometrics, Bucharest, Romania

E-mail: eugenpecican@yahoo.com

Abstract: *It is the purpose of this article to discuss certain econometric elements with a view to perceive the economic crisis signals and especially to reduce recession effects and overcome such a stage of the economic cycle. Together, economic theory, including policy measures and econometric methods provide certain chances to hurry up the end of depression. Thus, a simultaneous equation model adopted to simulate policy targets can be recommended in view to obtain new elements for an impact analysis and also for prognosis of main economic indicators.*

Key words: *economic cycle; depression; industrial output index; reference rate; fiscal policy; monetary policy; economic model; control variable; leader variable; econometric model with simultaneous equations*

1. Introduction

Certain significant indicators concerning Romanian economy and its evolution throughout the last years show positive trends until the second part of 2008. Afterwards the economic activity in many sectors stayed on a significant downward trend (fig.1.1. and 1.2.).

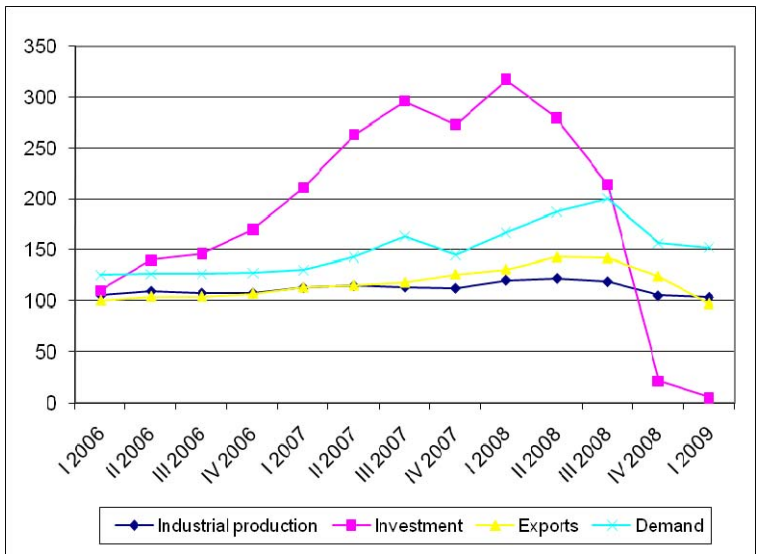


Figure 1.1.

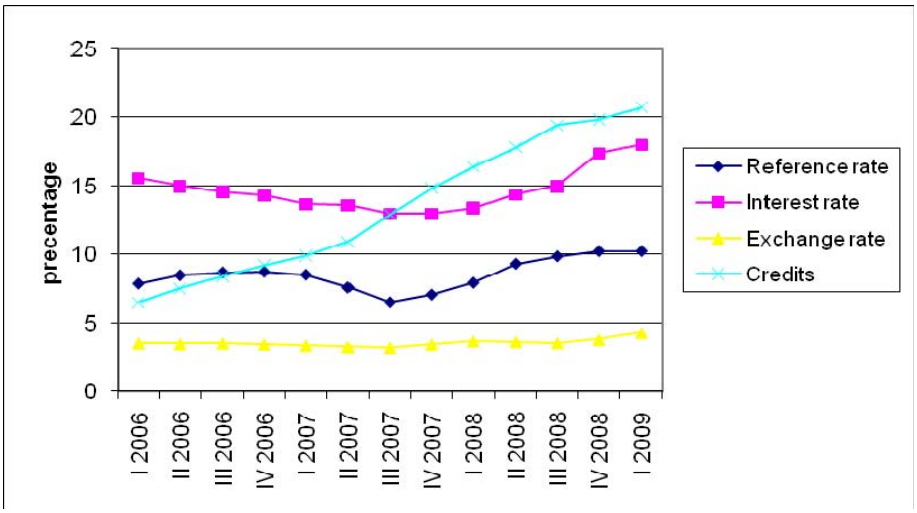


Figure 1.2.

In what follows we refer to certain interesting variables specific to economic cycle analysis[2].

- BET index, which registered 6586 points in 2005, 8050 points in 2006, 9825 points in 2007, reached approximately 5000 points in 2008 Q3, 2900 points in 2008 Q4, 2100 points in 2009 Q1;
- a significant decline of turnover from approximately 2000 (lei mill.) in 2008 Q2 to 1570 of 2008 Q3, 1000 of 2008 Q4 and 680 in 2009 Q1;
- credit risk ratio, kept stationary around 3.5 - 4.5 until August 2008, began a monthly increase as follows: 5.29; 5.38; 5.8; 6.52 in the interval Sept.- Dec., 2008, respectively 7.6 ; 8.42 in the interval Jan. – Feb., 2009;
- overdue and doubtful loans (net value), also placed at stationary levels (around 0.25) along the time, began an increase in the final part of 2008 and the first part of 2009 (from 0.35 in August 2008 to 0.63 in February 2009).

Similarly it comes out that the past-due debts;

- discount treasury certificates (nominal value- lei mill.), registered an accelerated increase (from 1769 in September, 2008 until 7480 in March, 2009);

All these changes in the last succession of time periods, pointed out that an economic recession has been emerging.

2. Economic reasons

The economic crisis, a phase of the economic cycle, has its issue, according to the monetarist hypothesis, in the expansion of money supply and credit (Hartley-1198, Friedman-1963) but also in the overestimating of expectations regarding demand and unlimited profit, that generate supply surplus (Beveridge-1981). Politics and electoral cycle can also influence the electors' expectations regarding income increases, thus generating some disturbances of market mechanism.

We can recognize certain reasons mentioned above if we take into account the preliminary period of the current crisis as follows: a large expanding of credit (however, such a phenomenon was slowing down in Romania due to NBR measures), a significant increase in profit rate especially on the real market, a macroeconomic government policy aiming to maintain optimistic expectations. The latter grew in importance in Romania on the background of Nov, 2008 elections, with negative consequences for the budget deficit, and at the same time due to quiet news about the possibility to appear an internal economic crisis.

If we refer to forecasting signals of recession, interesting especially from with a view to avoid in advance similar situations, according to economic theory, there are certain conspicuous trends such as: a major increase in house and house-lot prices associated to a decrease in interest rates; an raise in people naivety regarding the permanence of big incomes together with the pleasure of taking certain risks and luxury temptations (Juglar, 1862).

In Romania house prices and house-lot prices have increased in certain areas of economic importance especially in 2007-2008 interval. Another emergency signal was given by NBR as a preventive measure, namely the increase in the reference rate on the basis of a limited growth of credits. Other main macroeconomic indicators (inflation, interest rate of credit institutions, exports, industrial output and, partially, monthly average wage and exchange rate) have continued their positive trends throughout 2006, 2007 and 2008 (except for the last quarter of 2008), in spite of bad news about the economic crisis in the USA and many EU countries.

The recession in progress is described in economic theory by decreases in main economic indicators as: industrial output index, capital market indicators, investment, income, etc. along two or three successive periods ("three months rule" or more successive quarters) and simultaneous increases in unemployment, interest rate etc. It is easy to establish such evolutions (trends) if we follow all these indicators for Romanian economy beginning with 2008, Q4 (fig.1.1, 1.2).

An approach to crisis typology shows us an interesting opinion about the interval 2008-2010 according to which it is considered a final interval of a decreasing phase of a Kondratiev cycle. Moreover, in economic theory it is mentioned that "no cycle is exactly the same as another cycle". This is valid especially for the beginning of the recession in Romania as the graphical representation shows (fig.1.1, 1.2) and, as well, as signals from other sources (media channels) point out.

Both sources suggest that the major deficiencies in Romanian economy have appeared due to a significant fall in the demand on external market with immediate consequences on national industrial output, together with the liquidity retreat by mother banks having visible effects on credit interest rate as well as on crediting conditions in Romanian money market.

So, a crisis through contamination (somewhat natural in a globalized economy) overlapping a local structural crisis in the final phase, both marked by conducts specific to an election year. It can be said that misfortune hard to avoid appeared at the wrong time and in the wrong place.

Although we cannot say that there is a time and a place for the emergence of economic recession, the overcoming of recession is easier in an economy relying on significant reserves, able to support a fiscal relaxation policy in a regular time (non election year) in which the government is less interested in populist measures.

Mitigating the effects of recession and its overcoming are done through anti cyclical policies consisting of:

- public expenditure policy characterized by increasing this kind of expenditure in order to raise aggregate demand (socio-cultural investment, high social protection) and by stimulating production through fixed capital renewal;
- monetary and credit policy reflected in the reduction of key interest rate, of commercial banks obligatory reserves, incentives aimed to encourage lending, all these with a view to stimulate investment, production revival, employment, consumption;
- fiscal policy oriented towards reducing economic agents' taxation and towards encouraging demand.

Whatever the dilemma which economic theory is dealing with on the effectiveness of these policies, the important issue is the option for those measures which, considering the specificity of crisis evolution in a certain country, manage to reduce the disaster and hurry economic recovery.

In 2009 Romania measures related to both public spending (investment in infrastructure, reinitiating the car park renewal facilities) as well as to monetary policy (cutting key interest rate, diminishing obligatory reserves, State guarantees for housing loans)have been put forward and partially applied. But less emphasis was placed on the tax relaxation. The effect of such measures can be topics of study for Econometrics.

A summary of the above may be represented as an economic model. In a graphic expression, such a model looks like this:

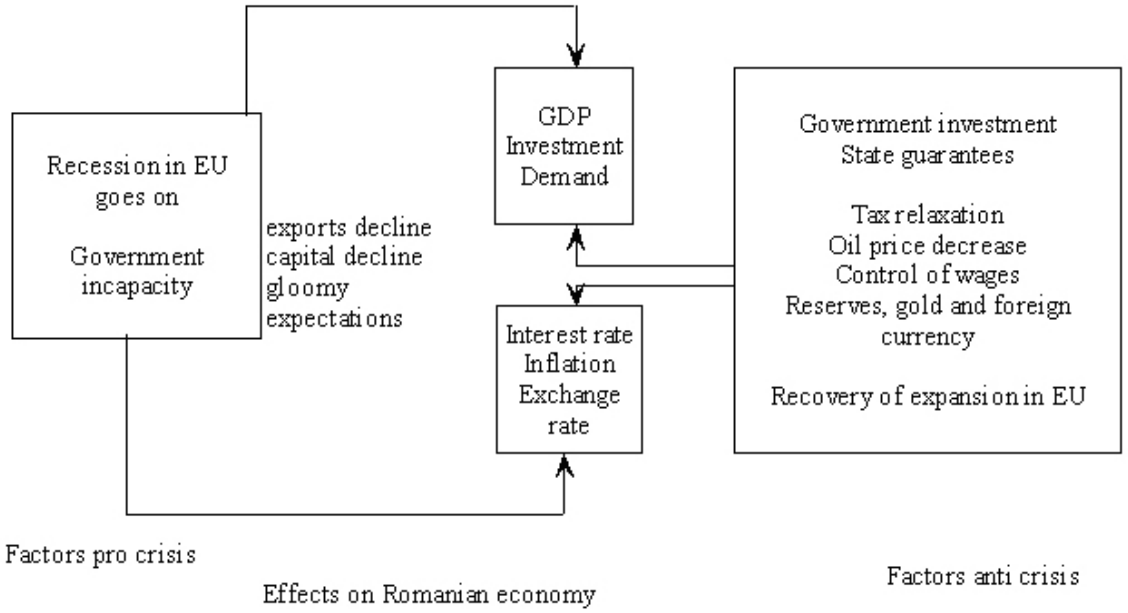


Figure 2.1.

The economic model is represented by an ensemble of cause-effect relationships including a low number of exogenous variables, recognized as determinant for certain economic sectors, but also control variables corresponding to the kind of economic policy preferred by government[3].

GNP= f (Investment, Demand, Exports)
Income (budget) =f (GNP, Fiscal policy, Debt (external))
Investments = f (Interest r., Monetary policy, Public expenditure policy)
Employment= f (GNP, Public expenditure policy)
Demand (aggregate)= f (Employment, Public expenditure policy, Monetary policy, Fiscal policy, Inflation)
Inflation = f (Money (narrow), Exchange rate, Fuel price, Demand (aggreg.))
Interest r. = f (Reference r., GNP, Monetary policy)
Exchange r. = f (Inflation, Investment (direct))

3. Statistical and econometric reasons

At least two aspects of recession can be considered suitable targets for econometric studies: (a) establishing leader variables whose evolutions can point out depression approaching, including the interval between the signal and the recession materialization; (b) selecting and promoting such processes (policy measures) in view to diminish effects and shorten duration of economic crisis.

For the first target (a), we consider certain elements of economic theory but also practical aspects regarding recent Romanian economy evolution. We can consider house price index and investment index as leader variable. The economic situations (represented by industrial output index) in certain EU countries (especially external business partners) can also represent significant signal-variables for the Romanian government.

The second target (b) is more important, and in view to analyze economic cycle, several macroeconomic models were created such as Samuelson's accelerator, and as well Metzler's, Hicks' or Frisch's.

We consider suitable for the present situation, at least for Romanian economy, a simultaneous equations model in which control variables are predominant.

There are same preliminary stages necessary to be solved out, in view to get our representation closer to reality by practical utilization. Thus the economic policy usually implies certain "new" variables (considered useful for diminishing crisis effects but for which there is no previous statistical data). For example new taxes or new government guarantees. Possible solutions: including expected values of some of the variables that describe similar situations in the past; choosing certain proxy variables with continuity along the entire interval in focus.

Another problem is represented by qualitative characteristics. For example, in the case in which fiscal relaxation for economic agents and state-guarantees for a limited period are considered as being determinant. Possible solutions: including adequate proxy variables, introduce binary variables, measuring on a numerical scale.

The structural form of economic model:

$$GNP_t = a_0 + a_1 INVEST_{t-1} + a_2 DEMAND_{t-1} + a_3 EXPORT_t^{(expected)} + u_1$$

$$INCOME_t = b_0 + b_1 GNP_t + b_2 FISC.POL._t + b_3 DEBT.EXT._t + u_2$$

$$INVEST_t = c_0 + c_1 INTEREST\ R._t + c_2 PUBLIC\ EXPEND.POL._t + u_3$$

$$EMPLOYMENT_t = d_0 + d_1 GNP_t + d_2 PUBLIC\ EXPEND.POL._t + u_4$$

$$DEMAND_t = e_0 + e_1 INFLATION_t + e_2 EMPLOYMENT_t + e_3 FISC.POL._t +$$

$$+e_4 \text{ MONETARY POL.} + \upsilon_5$$

$$\text{INFLATION}_t = f_0 + f_1 \text{MONEY(narrow)}_t + f_2 \text{EXCHANGE R.}_t + f_3 \text{PRICE(oil)}_t +$$

$$+f_4 \text{WAGE(adm.)}_{t-1} + \upsilon_6$$

$$\text{INTEREST R.}_t = g_0 + g_1 \text{REF.rate}_t + g_2 \text{GNP}_{t-1} + g_3 \text{MONETARY POL.}_t + \upsilon_7$$

$$\text{EXCHANGE R.}_t = h_0 + h_1 \text{INFLATION}_{t-1} + h_2 \text{INVEST(direct)}_t + \upsilon_8$$

The reduced form of such model would be a suitable representation intended to simulate economic policies in view to reduce and eliminate economic crisis effects.

References

1. Arrous, J. **Les theories de la croissance,** Ed.du Seuil, Paris, 1999
2. Bacescu, M. and Bacescu-Carbunaru, A. **Macroeconomie si politici macroeconomice,** Ed. All, Bucharest, 1998
3. Dobrescu, E. **Tranzitie in Romania. Abordari econometrice,** Ed. Economica, Bucharest, 2002
4. Griffiths, W., Judge, G. and Carter, H. **Learning and Practicing Econometrics,** Wiley, 1993
5. Hartley, J. E., Hoover, K. J. and Salyer,K. **Real business cycle,** Routledge, London, 1998
6. Mendoza, E. G. **The terms of trade, the real exchange rate and economic Fluctuation,** International Economic Review, 36, 1995
7. Mester, I. T. **Contributii privind modelarea ciclurilor economice,** PhD Thesis, Univ. Babes-Bolyai University, Cluj Napoca, 2006
8. Pecican, E.S. **Macroeconometrie - Politici economice guvernamentale si Econometrie,** Ed. Economica, Bucharest, 1996
9. Stiglity, J. E. and Walsh, C. E. (2002) *Economie*, Ed. Economica, Bucharest, 2002
10. Zarnovitz, V. **Business cycles: theory, history, indicators and forecasting,** The Univ. of Chicago Press, Chicago, 1996
11. * * * **Dictionar Macmillan de economie moderna,** Ed. Codecs, 1999
12. * * * **National Bank of Romania - Monthly Bulletin**, Feb., 2009, Year XVII, No.184

[1] Eugen Şt. Pecican is PhD professor at the Academy of Economic Studies of Bucharest where he teaches Econometrics and Statistics. Along the time he also taught Macro econometrics at Babes Bolyai University in Cluj and Oradea University, as well as Marketing and Econometric methods in Marketing at UPG Ploiesti. He was also invited to teach Applied Econometrics at National Bank of Romania in Bucharest. He was member of the editing team of Romanian Statistical Review. He was visiting fellow at the University of Reading, UK, where he studied and done research in Econometrics. Professor Pecican is well known for his papers on time series analysis, factor analysis, models and quantitative methods applied in economics. He also published books on Statistics and Econometrics.

[2] **Source:** * * * **National Bank of Romania - Monthly Bulletin**, Feb., 2009, Year XVII, No.184

[3] Control variables are not specified in a concretely form so that it is possible to choice from each categorical, an adequate variable.

THE BEHAVIOR OF CREDIT RISK EVALUATION MODELS UNDER RECESSION AND THE INTRODUCTION OF A GENERAL MODEL BASED ON SEMANTIC INTEROPERABILITY AND NOMOGRAMS[1]

Laura POPESCU
PhD Candidate, University of Economics, Bucharest, Romania
Developer, BRD Groupe Societe Generale, Bucharest, Romania

E-mail: laura.popescu@brd.ro

Andreea DIOSTEANU
PhD Candidate, Instructor, University of Economics, Bucharest, Romania
.NET Programmer, Totalsoft, Bucharest, Romania

E-mail: andreea.diosteanu@ie.ase.ro

Anca Nicoleta POPESCU
Master Student, SIMPRE, University of Economics, Bucharest, Romania
IT analyst, Outsource Bazaar, Bucharest, Romania

E-mail: anca.popescu@outsourcebazaar.com

Abstract: *The article analysis the old credit risk evaluation models performance and highlights the failure of complex econometric models to predict recession. Furthermore, this article is intended to propose a software solution for implementing a general, orientative consumption credit risk evaluation scoring model based on semantic web. The use of semantic web enables us to discover the common features from analyzing the evaluation forms used by several banks. These characteristics are taken into account when designing the model.*
The application will have a web interface and the weights and the cut-off value will be represented with the help of nomograms. This software product is intended to offer clients guidance and to provide, with a certain amount of risk, the information related to the chances that client has for obtaining a credit.

Key words: *credit risk; nomogram; semantic interoperability; OWL; recession*

Introduction

The new economic conditions under which we activate prove that we pass through an authentic recession period that has represented, metaphorically speaking, an "earthquake" for the traditional evaluation instruments. The theory has proved its consistency with reality, emphasizing that when environmental conditions change, classical models may lose their efficiency and lead to important losses.

The importance of the decision of granting credits has been confirmed by many specialists and business analysts as being one of the main sources of risk in the banking system. Therefore, it is recommended to create and implement efficient evaluation instruments that overcome the recession particularities and introduce an efficient manner of evaluating and granting credits with a minimized risk.

After presenting the results of a practical study carried out by the authors in order to prove that classical statistical models that are used by the great majority of banks in the financial system are less efficient during *recession*, the paper proposes web semantics and document mining as a solution for identifying a set of common features required by different banks in order to build a general credit scoring model. Furthermore, some theoretical aspects related to the visual instruments called *nomograms* are introduced. The authors present some aspects related to the creation of a general orientative *credit risk* evaluation model that brings the advantage of a visual instrument and of taking into account all the details related to a particular customer.

After building the nomogram – profile for a particular client and discussing his chances of being granted a credit, we emphasize the facilities offered by a visual decision tool. Finally, the authors summarize their main conclusions.

Romanian credit risk evaluation models

An evaluation of the indicators concerning credit institutions computed by the National Bank of Romania (BNR) brings to attention the fact that the percentage of defaulted credits as part of the total amount of existing credits at national level has clearly increased during the last quarter of 2008. If the percentage of "bad" credits was 0.24 in September 2008 (at the end of the third quarter in 2008 when the effects of the economic crisis were not perceived yet), the percentage increased at 0.35 at the end of 2008. This increase of 0.11 percentage points is a big increase if we take into account the relative stability of this indicator's values during the previous 4 quarters (0.22, 0.21, 0.30 and 0.24).

Table 1. The evolution of defaulted credits of credit institutions over 2008

Indicators concerning credit institutions	UM	Dec/2007	Mar/2008	Jun/2008	Sep/2008	Dec/2008
Defaulted credits/Total net value of granted credits	%	0.22	0.21	0.30	0.24	**0.35**

*) Indicators are computed based on the provisions data reported by credit institutions

**) Indicators comprise only commercial banks and Creditcoop as foreign banks branches do not report their solvability.

NB: Indicators refer to all the credit institutions in Romania, including commercial banks, foreign banks branches, Creditcoop.Starting from January 1, 2008, indicators are computed based on the financial reports FINREP and COREP of credit institutions. Source: National Bank of Romania (Banca Națională a României)

These results confirm the general expectations that the repayment ability of credit customers would decrease during recession. This feature is one of the first indicators of the economic recession.

Another observation is that the gross value of granted credits is still higher than the total value of deposits as the percentage of credits related to deposits is 122% (we can notice a small decrease in comparison to the third quarter of 2008: 124.71%) (Fugure 1). Many analysts talk about the need of increasing the amount of liquidities of every financial institution during the crisis.

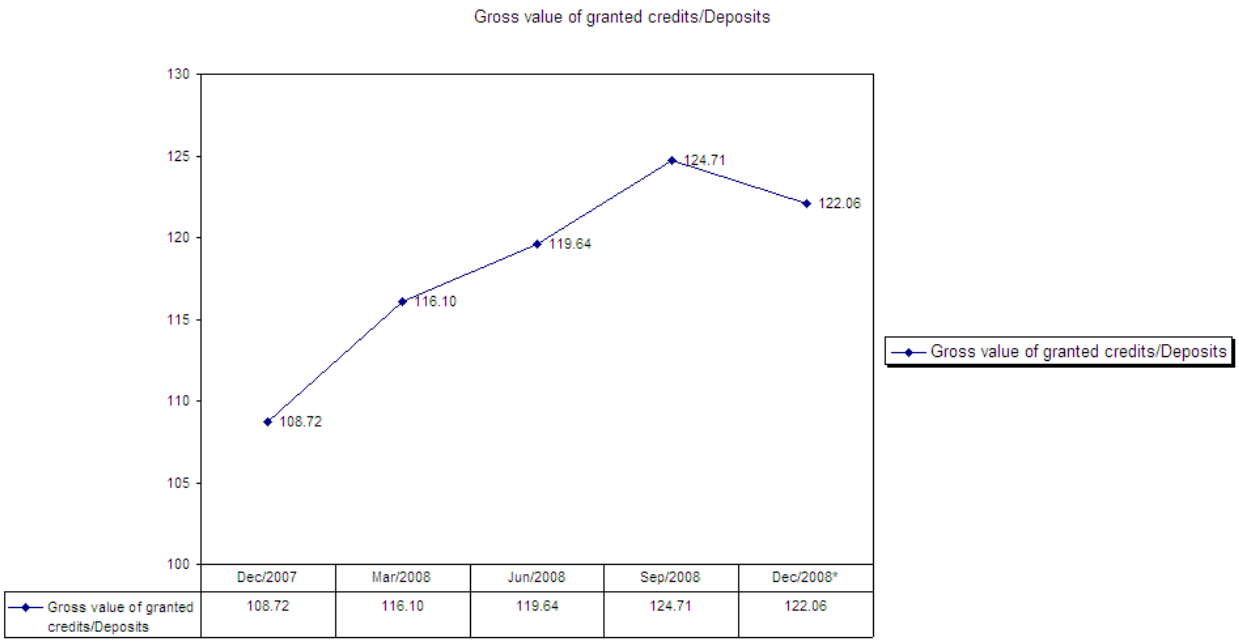


Figure 1. The evolution of granted credits related to deposits over 2008
Data Source: National Bank of Romania (Banca Nationala a Romaniei)

Most of the Romanian financial institutions use very simple methods of computing the credit score by taking into account several features and using weights which are established by experts without making usage of statistical, mathematical or informatic instruments. Consequently, Romanian banks and credit institutions still prefer to use traditional statistical methods (models that are determined based on regressions) as they are less expensive and easy to build and to understand and their coefficients can be rapidly adapted by internal risk analysts to new changes in customers' profiles identified in the annual credit reports.

An experiment carried out by the authors proved that, although very easy to build, use and adjust, those tools may prove less efficient under crisis conditions. The experiment used two data sets that contain information concerning several features recorded for 62 companies. The data source was represented by the financial statements available on Bucharest Stock Exchange or on the KTD Invest and the values reflect the status at the half of 2008 (June 30, 2008) and at the end of the year (December 31, 2008). Same features are taken into account for both periods in order to allow a comparison between the two moments in time. The net income was assumed to estimate a company's financial stability and ability to repay its loans.

The identification of the most significant characteristics is the first step in the process of creating a statistical credit scoring model (Piramuthu, 2006) as the different scores associated to variables in the credit scoring model are derived from the coefficients of the most representative variables estimated through regression.

Different regression models were built through the usage of Eviews 6.0, by applying Ordinary Least Squares and the process was reiterated as several characteristics were repeatedly eliminated and introduced under different scenarios. Finally, the most significant regression model (Dardac, 2005) was obtained when choosing capital, total debts and turn over as independent variables and total debts as dependent variable. The regression model was applied for both data sets and the following results were obtained:

Results for June 30, 2008

NET_INCOME=6,335.501+0.000751*CAPITAL+0.051040*TURNOVER - 0.096196*TOTAL_DEBTS
(0.0000) (0.0157) (0.0010)

Results for December 31, 2008

NET_INCOME=418.9257+0.001783*CAPITAL-0.043378*TURNOVER -0.141287*TOTAL_DEBT
(0.0000) (0.0195) (0.0110)

The fact that the p-values (probabilities that express the significance of the influence of selected variables over the dependent variable) are still quite high is determined by the inclusion of only three independent variables in the regression model. We are aware that there are other characteristics that influence the financial ability of paying the loan and that they may lead to a decrease of the p-values and an increase of the model's efficiency.

The most important conclusion that can be drawn by analyzing the results is that the relevancy of the selected variables decreases during the second half of the year. The capital remains the most significant feature for both data sets, while the p-value for turnover increases from 0.0157 to 0.0195 and the one for total debts increases from 0.0010 (30.06.2008) to 0.0110 (31.12.2008). Therefore, factors tend to lose their significance and the model proves to be less efficient, leading to a need to redesign the credit scoring models.

Semantic Web and customer features' selection

The novelty of the model that we propose is represented by the fact that we implement semantic web and document mining in order to analyze several banks' client profile evaluation forms and to determine a set of common features that are taken into account. The determined features will represent the points for nomogram graphic that is used to calculate the scoring. In the end, we will obtain a general minimalistic and orientative model for clients.

The semantic web is used to process the knowledge that is contained in the evaluation forms and also to facilitate the information integration and long time analysis. The semantic web is based on formal ontologies that structure data in order to represent messages across World Wide Web in a machine processable manner. The ontologies offer the possibility of representing knowledge in an easily interpreted language called *OWL* (Web Ontology Language). OWL also facilitates the implementation of a wide area of classes of semantic applications. The organization that is preoccupied with developing the semantic

web proposed the RDF meta language to implement and design semantic messages in an easily machine interpreted language (W3C).

According to Anubhav Madan, the use of semantic web in business applications is very important because it offers *semantic interoperability* possibilities and improves document analysis and research. Business interoperability is one very important factor in assuring economic performance. Furthermore, semantic web is applied in knowledge management.

Nowadays, high quality knowledge management is a key factor in assuring competitive advantage and risk reduction by efficiently exploiting the intellectual assets. Our article introduces the semantic web for knowledge management with impact in credit risk reduction in the actual economic context. Semantics are targeted to improve the traditional knowledge management models of the credit risk evaluation that on one hand were unable to predict the economic crisis and on the other hand proved their inefficiency in evaluating the risk. The consequences were very bad because many banks went bankrupt due to the clients' impossibility of covering their debts.

The model that we propose represents the starting point for a software application that will have two major components: a bank side component and a client side component

The bank side component has banks as main beneficiaries. The application will be based on semantic web services and will be composed of the bellow enumerated modules:

- **a semantic web service** that will implement semantic search for clients' characteristics and ontology development tools for structuring information into machine processable messages
- **a knowledge database** that contains complex credit necessary information, clients' detailed profiles obtained through semantic web search and document mining, and a set of structured rules that will be used by a expert system that will offer support in credit evaluation activity.
- **an expert system** that is designed to offer support for credit officers based on past decisions and a complex set of rules.
- **semantic web service description development tool**

The working scenario for the module will be the following:

Step 1: each bank will expose service that has a specific web service description referring to the functions that access certain databases containing non confidential information and a complete list of clients that were granted credits

Step 2: the semantic web service will provide tools for service discovery and semantic search for both lexical and semantic similarities. With the help of these tools the banks may collaborate in finding the common clients, their profile, and to validate the accuracy of the information the customer provided when applying for a certain amount of money.

Step3: after all necessary information is provided, the expert system will determine a set of structured rules based on previous similar situations and past experiences.

Step 4: in the end a client evaluation scoring model is implemented based on the rules determined. The result will be a percent that measures the client's associated risk level.

Step 5: judging by the determined percentage the credit officers will decide whether the bank will grant a credit for the analyzed client request and the level of warranty the client has to pay or guarantee with.

However, this article will present a detailed analysis of the client side component while the bank's module will be presented in our future articles.

Client semantic component

The client semantic component is designed for clients that intend to apply for a credit and are willing to determine the chances they have to obtain the sum of money they need.

This application proposes a general minimalistic and orientative credit scoring model based on the common features that banks are taking into account when designing the client's profile.

The application will have the following components:

- **a semantic web service** that will perform semantic search and document analysis on the banks evaluation forms and ontologie building tools.
- **a business logic web service** that will contain the general scoring model, based on the features discovered by the semantic web service (the model will be presented in "The usage of nomograms in determining a customer's credit score" subchapter)
- **a user web interface** that will display required information in nomogram graphics, with the help of which the scoring series will be developed, and also client solicited information in a user friendly and easy to understand format

Our model is based on analyzing the evaluation forms from three important Romanian banks: BRD, BCR and Piraeus Bank. For each bank taken into account we identify a set of general terms which represent the first level in the semantic network. Afterwards, we implemented the associated semantic models corresponding to each bank evaluation form.

The words having similar meanings in a standard semantic network, are organized into groups or clusters called synsets. There are two kinds of relations represented by pointers: lexical and semantic. The lexical relations refer to the word forms, while the semantic relations refer to word meaning. This lexical model represents the support for automatic text, document analysis and management. The type of semantic relationships established between terms in a semantic network are the following (grouped by parts of speech) :

Nouns: *hypernym*: B is a hypernym of A if every A is a (kind of) B**,** *hyponyms*: B is a hyponym of A if every B is a (kind of) A**,** *coordinate terms*: B is a coordinate term of A if A and B share a hypernym**,** *holonym*: B is a holonym of A if A is a part of B**,** *meronym*: B is a meronym of A if B is a part of A

Verbs: *hypernym*: verb Y is a hypernym of verb X if activity X is a (kind of) Y**,** *troponym*: verb Y is a troponym of verb X if activity Y is doing X in some manner**,** *entailment*: verb Y is entailed by X if by doing X you must be doing Y, *coordinate terms*: those verbs sharing a common hypernym

Adjectives: related nouns and participle of verb

Adverbs: root adjectives

For each bank we develop the semantic network of the general term "Criterii de evaluare" ("Evaluation Criteria") which represents the root of an arborescent structure. Words having similar meanings are grouped into synsets and correspond to the evaluation criteria in the client evaluation form. Between the synsets we can identify relations such as those mentioned above. However, there might be situations in which terms are expressed differently (for different banks) but have similar meanings.

In order to identify similarities between semantic networks, we take into account some parameters that are related to the type of relationships in the the semantic network (hypernyms, hyponyms, coordinate terms, holonyms, meronyms). For example, "Situatie locuinta"(Client's home situation) (BRD),"Locuinta clientului" (Client's Home) (BCR) and "Tipul locuintei"(Home type) (PIRAEUS) is a meronym of "Criterii de evaluare"(Evaluation criteria) and have similar meanings. We do this for all the elements of the semantic network. Furthermore, we can apply specific algorithms to determine the similarities such as (Frankel, 2007):

- Path Finder
- Depth Finder
  - Wup: Shortest path by scaling sum of values between node and root
  - Lch: (Leacock and Chodrow) Shortest path by scaling the max path
- Path: Inverse of the Shortest Path measures
- Information Content Finder
  - Resnik: Max Distance b/w concepts of both words
  - Jcn (Jiang and Conrath): Inverses the difference between Sum and LCS
  - Lin: Scales LCS IC with the description
- Gloss Finder
  - Lesk (Banerjee and Pederson) finds and scores overlaps between glosses
  - Vector creates a co-occurrence matrix with glosses in vectors
- Hso (Hirst and St-Onge): Specifies Direction between Words

The application based o semantic networks has to have a knowledge base that contains terms and semantic relationships as they are presented in DEX(Explicative Dictionary) for the Romanian language, or an WordNet database for English.

Our pilot solution (the first stage for developing a complex client module) computes semantic models similarities based on LIN algorithm and with a database composed of around 1000 terms and definitions from DEX.

A part of the semantic generated OWL Lite code can be seen below:

```
<rdf:Description rdf:about="VenituriNete">
<rdf:type><rdf:Description rdf:about="http://www.w3.org/2002/07/owl#Class"/></rdf:type>
<rdfs:subClassOf><rdf:Description rdf:about=" CriteriiEval"/>
</rdfs:subClassOf><owl:unionOf rdf:parseType="Collection">
<rdf:Description rdf:about=" VenitNesalarNetLunar"/>
<rdf:Description rdf:about=" VenituriDinDobLaDep"/>
<rdf:Description rdf:about=" VenitDinDreptDeAutor"/>
<rdf:Description rdf:about=" VenituriSocialeLunare"/>
<rdf:Description rdf:about=" SalariuNetLunar"/>
<rdf:Description rdf:about=" ChiriiIncasateLunar"/>
<rdf:Description rdf:about=" SalariuNetLunar"/>
<rdf:Description rdf:about=" PensiiIncasateLunar"/>
<rdf:Description rdf:about=" ChiriiIncasateLunar"/>
</owl:unionOf></rdf:Description>
```

The formula used for computing semantic similarities based on Lin algorithm is the following: $$sim(A,B) = \frac{\log P(common(A,B))}{\log P(description(A,B))}$$

Nomograms – Basic concepts

The nomogram is a statistic tool for concepts and complex data set visual representation. This type of graphic is also known as Kiviat diagram, radar diagram or spider. The great advantage brought by this type of graphic is that of offering a clear representation of the client scoring process with an emphasis on the elements with a higher risk. (Ivan, 2007)

If we consider a P_i client that is part of a set of n persons $\{ P_1, P_2, ..., P_n \}$ and we also create the set of scoring features made up of *m* features $\{ C_1, C_2, ..., C_m \}$ for risk analysis represented by the client when granting him a credit.

After specific credit scoring computations, for each feature scoring indicators are being determined $\{ I_1, I_2, ..., I_m \}$.

The scoring indicators representation by using a nomogram for a certain client (P_i) is developed by following the bellow presented steps:

- the base circle is being drawn;
- the circle is divided into m-1 sectors, each radius representing a certain feature
- for each characteristic C_j, j=1..m a corresponding indicator is defined and calculated I_j, j=1..m. The values are normalized into the [0,1] interval . The 0 value corresponds to the circle's centre and 1 value is represented on the circle's circumference
- each indicator I_j for C_j characteristic is afterwards set on the circle radius.
- the points represented on the radiuses are connected and a specific client's scoring characteristics associated polygon is obtained

In order to simplify the representation, the number of the scoring characteristics that are taken into consideration is m=6, and the selected characteristics are $\{ C_1, C_2, C_3, C_4, C_5, C_6 \}$. The associated weights for each characteristic are calculated and we consider that each characteristic has the same level of importance. A circle is drawn and the radiuses corresponding to each identified scoring characteristic. After following these steps six equal circle sectors are obtained (the angle for each sector has sixty degrees).

Figure 2. The base circle for a nomogram

Each I_i indicator is represented as a point on the radiuses and finally the points are connected and the client's characteristics' polygon inside of the circle is obtained. Consequently, the nomogram is obtained.

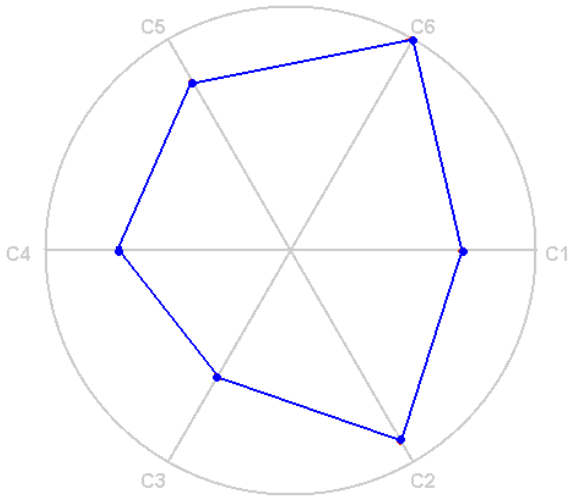


Figure 3. Client specific nomogram

The nomogram is developed so that to cover the general number of scoring characteristics, n. The formula that is used to determine the circle sector's angle is presented below: $teta = \frac{360^{\circ}}{n}$ where:

teta =circle sector angle;
n = scoring characteristics angle.

The n-regular polygon's points inscribed in the base circle is determined by using the following formulas:

$$X_i = \cos(i * teta) * R + X_{centru}$$
$$Y_i = \sin(i * teta) * R + Y_{centru}$$

where,

X_i = i point's abscise i=1..n;
Y_i = i point's ordinate , cu i=1..n;
X_{centru}= circle's centre point abscise;
Y_{centru}= circle's centre point ordinate;
R = base circle radius.

The usage of nomograms in determining a customer's credit score

After determining the common characteristics of the credit scoring applications, we obtained the results presented in Table 2. Consequently, we followed several steps in order to represent a particular customer's answers as points on a nomogram. As the goal of the process is to determine the client's risk degree based on the obtained scoring model, aggregated nomograms are used.

The representation of the aggregated nomogram is obtained by overlaying the client's scoring images and the bank's scoring requirements for the characteristics that are taken into account.

Table 2. The characteristics of the credit scoring model

FEATURE	OPTIONS	PTS	Normed indicators	Bank etalon
C1 = AGE	20 – 50 years (men) or 20-45 years (women)	6	1 (6/6)	0.66
	50 – 60 years (men) or 46-55 years (women)	4	0.66 (4/6)	
	Over 60 years (men) or over 55 years (women)	2	0.33 (2/6)	
C2=RESIDENCE TYPE	The person owns the place	8	1	0.5
	The person rents the place	4	0.5	
	The person doesn't have a house	2	0.25	
C3=YEARS AT CURRENT WORKPLACE	Retired	1	0.08	0.5
	Less than 1 year	3	0.25	
	Between 1 and 2 years	6	0.5	
	Between 2 and 5 years	9	0.75	
	More than 5 years	12	1	
C4=FAMILY MEMBERS	Married and both of them have a job	6	1	0.66
	Married, but only one of them has a job	4	0.66	
	Single	2	0.33	
C5=NUMBER OF PERSONS TO FEED	Less than 2	12	1	0.5
	Exactly 2	9	0.75	
	3-4	6	0.5	
	5 or more	3	0.25	
C6=NET INCOME	Less than 500 RON	1	0.11	0.66
	500-1000 RON	3	0.33	
	1000-2500 RON	6	0.66	
	More than 2500 RON	9	1	
C7=INDEBT DEGREE	Less than 10% of his net income	12	1	0.5
	10-20% of his net income	9	0.75	
	20-30% of his net income	6	0.5	
	More than 30% of his net income	3	0.25	

Figure 4 presents the aggregated nomogram for the number of characteristics m=6 and the characteristics $C_1, C_2, .., C_6$. The client's characteristics' polygon is blue shaded, and the banks' polygon is green shaded. By overlaying the two polygons we can highlight the positive and negative differences between client's self calculated scoring model and the one banks use.

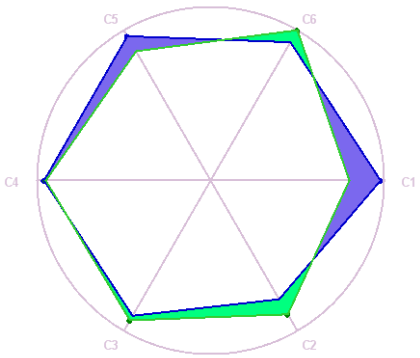


Figure 4. Aggregated nomogram

where,

S_{poz} positive diferences surface area

S_{neg} negative diferences surface area

In order to decide whether the analyzed client respects the imposed scoring legislation imposed by the etalon bank's scoring. The bank defines a quality evaluation indicator IC which is determined by dividing the positive and negative differences that were calculated according to the above presented model of the aggregated nomogram.
The following formula is used:

$$IC = \frac{S_{neg}}{S_{poz}}$$

where,

IC = analyzed client's scoring evaluation indicator;
Sneg = negative diferences surface area ;
Spoz = positive diferences surface area ;

The indicator can be interpreted as it follows: a value equal to 1 indicates a good scoring level (we can state that the client approaches the cut-off value), a value above 1 indicates a bad performance and might lead to the rejection of the customer's loan application, while a value below one indicates a very good performance and the fact that the client is a credible one and can be granted a loan.

For example, let's analyze the case of customer A that is 30 years old and lives in a rented apartment, has been working in the same workplace for less than one year and he is married to another employed person. The customer doesn't have other children and his net income is of 2,700 RON. His indebt degree is between 20-30%. Consequently, customer's A normalized values of the characteristics (as established in Table 2) are the following ones and his nomogram is pictured in Figure 5.

C1 = 1, C2 = 0.5, C3 = 0.25, C4=1, C5=1, C6=1, C7=0.5

It has to be mentioned that the blue line in Figure 5 represents the client's nomogram and the green area represents the characteristics for which the client outperforms the bank's requirement's. the red area represents the risk area and it is determined by customer's A low performance for the third characteristic.

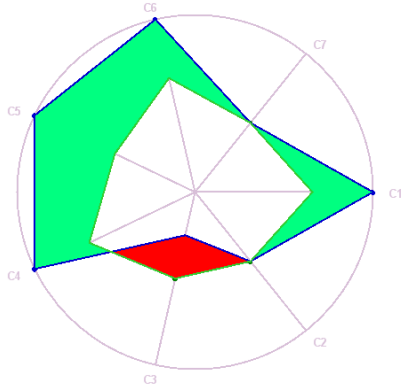


Figure 5. Customer's A nomogram

On the other hand, customer B is 52 years old and he lives in his own apartment. He works in the same place for 20 years and he is married, but has a retired wife. He has two children and his net income is of 5,000 RON, while his indebt degree is less than 10%. Therefore, customer's B characteristics are:

C1=0.66 ; C2=1; C3=1; C4=0.66; C5=0.75; C6=1; C7=1

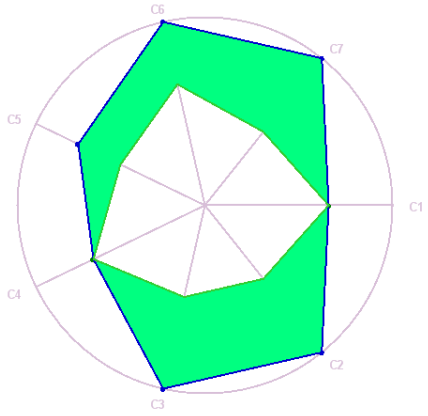


Figure 6. Customer's B nomogram

Conclusions

During economic crisis, the behavior of the credit institutions is different than in a steady or economic growth period. Statistics presented by The National Bank of Romania reveal an increased restrictiveness of credit conditions during the last quarter of 2008. After a short analysis of some general indicators concerning the national status of credit institutions, the authors present the results of a study which proves that, in the case of regression models, coefficients are less significant (higher p-values) during recession (the second half of 2008) when compared with the first half of the year when the economic crisis impact was not perceived in Romania. Therefore, a conclusion is drawn that under economic crisis conditions companies tend to have an abnormal behavior which leads to the necessity of redesigning classical credit scoring models.

The authors propose semantic networks and document mining as a solution for determining common features that can be included in a general, orientative credit scoring model. Finally, some basic aspects related to nomograms are introduced as they can be used as a visual tool for building a credit scoring models. The utility of a user-friendly credit scoring model is proved by taking some particular examples of loan applicants and drawing and discussing their nomograms.

Bibliography

1. Dardac, N. and Moinescu, B. **A quantitative credit risk evaluation from the BASEL II perspective,** Economie teoretica si aplicata, 2005, pp. 41-46
2. Frankel, D. **Business Network Transformation and Semantic Interoperability,** SAP Company, December 2007
3. Ivan, I., Boja, C. and Ciurea, C. **Metrici ale sistemelor colaborative,** Editura ASE, Bucharest, 2007

4. Madan, A. **WordNet: An Overview,** Tutorial, http://www.scribd.com/doc/6890426/tutorialWordNetSimilarity
5. Piramuthu, S. **On preprocessing data for financial credit risk evaluation,** Expert Systems with Applications, 30, 2006, pp. 489–497
6. * * * **An inquiery concerning the creditation of nonfinancial organizations and population,** The National Bank of Romania: www.bnr.ro, November 2008
7. * * * Bucharest Stock Exchange, http://www.bvb.ro
8. * * * KTD Invest, http://www.ktd.ro
9. * * * **W3C Semantic Web Activity,** http://www.w3.org/2001/sw/

[1] **Acknowledgements**
This article is a result of the project „Doctoral Program and PhD Students in the education research and innovation triangle". This project is co funded by European Social Fund through The Sectorial Operational Programme for Human Resources Development 2007-2013, coordinated by The Bucharest Academy of Economic Studies.

ANALYSIS OF THE RELATIONSHIPS OF FACTORS AFFECTING RICE CONSUMPTION IN A TARGETED REGION IN EKITI – STATE, NIGERIA

Joseph Omoniyi BASORUN[1]

PhD, University Lecturer, Department of Urban and Regional Planning,
Federal University of Technology, P M B 704, Akure, Ondo State, Nigeria

E-mail: nbason2003@yahoo.com

Abstract: *In Nigeria today, there is the growing concern for food security, with particular focus on rice being a staple food. If effective policies must be made to impact rice production, it is important to appreciate the phenomena underlying the consumption in the first place. This study, therefore, examines the rice consumers in a rice producing region in Ekiti State Nigeria vis-à-vis the interrelated factors influencing consumption. Five hundred (500) respondents were randomly selected and interviewed in Eighteen (18) of the sixty nine (69) residential quarters that constitute the six (6) Local Government Areas (LGAs) which represent the study region. The study employed the Pearson's Correlation test of the consumers variables. Findings discovered strong relationship between; (i) Sex of consumers, (ii) their level of education, (iii) income, (iv) household size, (v) the source of the local rice consumed, (vi) availability of the rice, (vii) how regularly if is consumed, (viii) the quality, (ix) price, and (x) the need to strengthen production at source. Suggestions for policy measures were offered based on research findings.*

Key words: *Rice; Rice consumers; region; Ekiti State; Nigeria*

1. Introduction

Food security is a sensitive problem affecting many nations, especially the Developing Countries (DC) due to their ever increasing population (Ajake, 2003). The situation is indifferent in Nigeria; the most populous country in Africa, having over 130 million people (Doney 2005). Government target for now is ensuring adequacy of food in quantity and quality to meet the nutritional needs of the teaming population. However, only regionally coordinated agricultural and food security policies may offer good approaches in this regard, as considerable additional public and private investment would be necessary to accelerate the pace of agricultural growth (Otzen, 2002).

Of all the food items, rice is the most widely consumed. It is a staple diet in Japan and the principal crop in India. It is equally the mainstay of the economy of Thailand and Vietnam (Encarta, 2004). Rice is a staple crop throughout West Africa, especially in Cote d'Ivoire, the Gambia, Guinea, Guinea Bissau, Liberia, Burkina Faso, Senegal and Sierra Leone (NISER, 2002). Rice is cultivated in virtually all the agro-ecological zones in Nigeria (Akande 2001). Since the mid-1970s, rice consumption in the country has risen tremendously (10.3% and above per annum) as a result of accelerating population growth

rate (+2.8% per annum) and increasing per capital consumption (+7.3% per annum) due to changing consumer preference (Akpokoje *et al.*2001).

The average Nigerian consumes 21 Kilograms of rice per year (WARDA, 2002). Unfortunately, the increase in demand in recent times has not be accompanied with corresponding rise in production in Nigeria. The core problems of production include: wide spread poverty, "dominance of the nation's agriculture by small holders" (Daramola, 2004), the use of relatively primitive tools for farm operations (Fakorede, 2001), lack of exposure to improved agricultural technology (improved seeds, fertilizers, pesticides etc) and farm mechanization aids by government. Highlighting the comments of Osinami (Head of the Africa Rice Centre in Nigeria), Nigeria produces two million tones and consumes about five million tones annually, thus, expends $800 million yearly on importing the deficit of about three million tones (Awe, 2006).

Nigeria still ranks third with Iraq (after Philippines and China) in the group of major rice importing countries in the world (Awe,2006). Reasons for the increase in Rice consumption in the country are numerous. They include: rapid urbanization, ease of preparation that fits easily into urban lifestyle of workers, and its general availability among food vendors and restaurants located in work places, especially urban areas (NISER, 2002). Nigerians prefer the local rice because of its taste and sometimes even its smell; good processing may make it compete favourably with imported rice (Longtau, 2000). Ekiti state is one of the country's sub-regions where rice is prominent. Over the years, its production has also expanded as a result of vast increase in land area put under cultivation. However, this was still considered insufficient to match their consumption increase as rice import makes up the shortfall (NISER, 2002).

2. The Rationale for Targeting the Study Region

Regional planning problem emerges from multi-dimensional issues. Regional needs, resource distribution, social priorities and taste, form the essentials of an overall framework for its analysis for effective planning. Hall (1994) observes that throughout the period of 1945 to 1980, planning policy at the regional level in Great Britain was targeted at the less prosperous regions of the North Scotland and Wales for the creation of factory jobs, unlike the more prosperous regions of the South and Midlands which were already commercial centres and seats of expanding service industries. Right from the 1960s, the growth of urban population has been accompanied by high levels of concentration of urban dwellers in a few large cities (Whang, 1988), majority of who are the absolute poor. This situation prompted the Korean government to target its cities for anti-poverty programmes in the 1980s. In 1983, the District Focus for Rural Development (DFRD) was initiated in Kenya, targeting the rural districts for allocation of resources that will enhance utilization of local resources, thus increasing employment opportunities (Schall, 2000).

In the South-Western part of Nigeria, Ekiti State remains one of the least developed economically,. About 87.0 percent of the State population are farmers, of which "70.0 percent are actively engaged in rice production" (NISER, 2002). Igbemo – Ekiti, in the region of study has national reputation for producing rice. Out of the major rice producing towns (Ikole, Ijero and Igbemo) in Ekiti, Igbemo controls a relatively large market area and enjoys the easiest accessibility to the State capital where it attracts high patronage. Other communities in the region have direct linkage with the town through trading and

consumption of its rice. Many rice processing and marketing activities complement rice farming to transform this centre to an agropole. It thus, presents the build-up of a system involving links between processes of activities on rice production and form of human communities in a regional setting. Targeting, therefore, appears a relevant approach in this study as it relies on specific criteria to identify the targeted location. The major utility of targeting is that, it assists in identifying the existence of large differences in the physical and living conditions between geographic areas (Okoko, 2004).

A major constraint in realizing the goals of targeting is determining the target group. Such indicators as location of residence of rice consumers, source of local rice consumed, distance to the source and major rice shopping centres can be used in delineating the targeted region. Another problem is how to select the local rice consumer and Igbemo catchment area for targeting. In this regard, the Access Opportunity (AO) model was employed. It assists in assessing the location of Igbemo in relation to the region of influence in terms of patronage. Accessibility in this case is estimated using the formular:

$$A0_i = \frac{S_j}{t_{ij}^b} \quad (1)$$

where $A0_i$ = access opportunity index, S = size of item required at centre (j), t = time taken to travel from residence 'i' to acquire item at centre 'j' and b = exponent describing distance-decay effect (Smith 1977, Morenikeji, 2006).

The sixteen (16) LGAs that constitute Ekiti State are involved in the estimate. Undoubtedly, their locations from Igbemo vary with distance which is a determinant of their A0. The higher the A0, the higher the access level of a particular LGA. The mean A0 for the entire region, therefore, is obtained by multiplying each LGA's A0 by its population, summing the values and dividing with total population thus:

$$A_i = \frac{\sum A0_i x N_{ij}}{\sum N_j} \quad (2)$$

The higher the value of Aj, the better. The number of LGAs that fall below the mean A0 value, however, represent the sub-regions of the state that are of less interaction in terms of access opportunity to the local rice.

It is possible to determine major areas of influence of the local rice using this model. But the usual constraint is determining the application to spatial dimensions. Whether it is better applied to the city or at regional level remains a crucial question. What is important, however, is the clear specification of the variables at any given level that reflects the function of a service centre.

3. Materials and Methods

Research Locale: The targeted region comprises of six (6) LGAs in Ekiti State. These are: Ado, Irepodun/Ifelodun, Ido/Osi, Gboyin, Oye and Ikole LGAs. The region lies on the rain-fed upland environment in the South-Western Nigeria. The region locates in the North-eastern part of Ekiti State where it shares boundaries with Ilejemeje and Moba LGAs as well as Kwara in the North; Ekiti East LGA, Kogi and Ondo State in the East; Emure, Ise-Orun, Ikere and Ekiti-South West in the South, and Ekiti –West and Ijero in the West (Fig. 1). Out of the total sixteen (16) LGAs, the study region which comprises only six (6) LGAs,

accommodates about 44.2% of the state population. The six LGAs surround Igbemo with the potential for strong inter-connections and well-established ties on rice business.

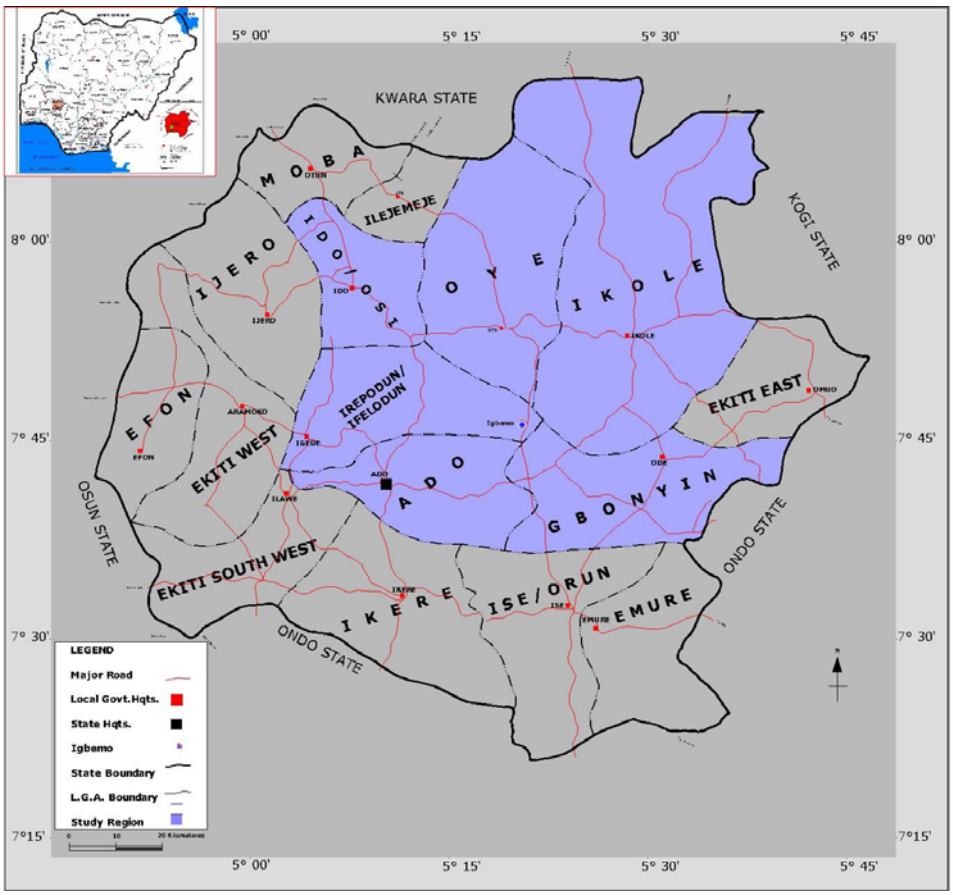


Figure 1. Igbemo and the Region of Study in Ekiti State
Source: Author's Finding, 2007

Database Description: In this study, a random selection of 500 respondents was undertaken in eighteen (18) residential quarters which are coterminous with the independent National Electoral Commission (INEC) political wards in the region. The sample (500 respondents) was disaggregated into the 18 wards according to the number of buildings which were used as proxy population for consumers' population in each ward. Thereafter, samples were randomly chosen from the wards which represent well-defined Data Delineation Areas (DDA). In a targeted house, only one household was interviewed, and the respondent was the household head of age 18 and above. A structured questionnaire was prepared and administered on the household heads (respondents).

About 418 out of the 500 respondents (representing 83.6%) in the study region consume the local rice. Subsequent analysis focused on this group being the targeted population. The variables that were employed in the correlation analysis are: sex of the consumer (SEX), the level of education (EDU), the household, size (HOLD), the level of income (INCOME), and the source of local rice consumed (SOURCE). Others include; the level of availability of the rice (AVAIL), how regularly it is consumed (REGULAR), the quality (QUALITY), the price (PRICE), and the need to strengthen production at source (NEEDS).

These variables have been selected because of their likelihood to impact rice consumption with distance from the production centre (Igbemo).

4. Discussions of Empirical Findings

It is possible to determine the characteristics and behaviours of the consumers through field observation. Another option, though less reliable, is to present only the opinions of this group on rice consumption. The logic in the cross-tabulation of the specific variables is that, the willingness of the consumer to opt for rice is determined by certain socio-economic factors (sex, income, quality, household size, price etc) which satisfy the quantity and quality needs in consumption.

Table 1 (see Appendix 1) reveals that level of education of the rice consumer (EDU) has significant relationship with his sex (SEX). The reality in the region is that, more females (59.8%) seek the rice than males (40.2%). These consumers are fairly educated with 64.1% having secondary education and above, of which 74.1% are under age 45. The flair for the local rice by the females, however, is symptomic of their relatively low income, occasioned by low education. Generally, they occupy lower-paid and lower-status jobs than men; their unemployment rates are higher than men's and are far more involved in the 'informal sector' occupations like street vending and market work (Ashford, 2001).

Coincidentally, income of the consumers (INCOME), is associating with EDU. No doubt, majority (78.2%) are literates, but quite a significant proportion (70.9%) earn below ₦80,000 ($640.0) annually. This is suggestive of the prevalence of poverty and preference for the local source which the low income can sustain. In Ekiti, agriculture predominates the occupational structure of the population (72.6%), followed by trading (12.2%); artisans/professionals (7.4%) and the civil servants (5.6%). Only have educational qualification (National Certificate in Education and above) that can attract high income. As expected in this region, most women, who are major consumers of the rice, are less educated than man – a situation that perpetuates their early entry into motherhood, low education and a continued circle of poverty.

Availability of the rice (AVAIL), is of significant relationship with SEX, EDU and source of the local rice (SOURCE). Field survey shows that most consumers (56.0%) consider the rice to be fairly available. Logically, increase in the proportion of females will result in high reproduction rate and lead to increased consumption with corresponding reduction in quantity of rice that will be available. This is in close association with the their level of education earlier explained. In any case, AVAIL has depended heavily on seasonality and low level of production at source at source. At present, the degree of production in *Igbemo* (source) can still be regarded as peasant as the average per-capital area of land cultivated by the rice farmer (4 hectare) falls short of 5 hectares (Adeola, 2002).

Regularity of consumption (REGULAR) has a significant association with EDU, SOURCE and AVAIL. The strong relationship with AVAIL in connection with source is explanatory. During harvest (usually July to September), the rice floods the market, thus, attracting vast majority of the rural and urban consumers for regularity of consumption. As scarcity of the product sets in during off-season (usually March to June), AVAIL reduces only for REGULAR to diminish. Igbemo remains the major source (82.8%) of the local rice consumed in the region. Given that most consumers are literates (82.8%), there is the

awareness of a popular source which stimulates preference for *Igbemo* rice, and affects REGULAR.

The assessment of quality of the rice (QUALITY) by the consumers is strongly related with their household size (HOLD), AVAIL and REGULAR. Despite the relatively low quality of the local rice to imported rice, many of the consumers (64.4%) see it to be good. In Ekiti, about 73.9% of the consumers maintain household size of up to 8 persons (1- 4 = 22.4%, 4 – 8 = 49.7%). This implies 4 – 6 children including dependants in some cases. Apparently, the consumers associate with large families in which food expenses are moderated by low income. At the market places, the local rice is cheaper than the imported rice while the marketers and consumers tend to have some relationships that promotes sales of he local rice. The most important are credit sales and quantity discount. These clandestine marketing factors account for the association between QUALITY and HOLD. The distribution channels of *Igbemo* rice are considerably wide, even beyond Ekiti. The existence of the neighbourhood markets makes AVAIL possible. The general acceptance of the quality by a diversity of household types therefore, facilitates REGUALR.

The price of the rice (PRICE) exhibits significant relationships with EDU, INCOME, SOURCE, AVAIL, REGULAR and QUALITY. Further analysis of data reveals that the average consumer in the region has secondary education. Subject to this qualification, income would be low. As a consequence, preference for a local source of rice (staple meal) with moderate price is likely to increase. This may have led to the interest in the local (₦250.00 or $2.0 per Kilogram) which is far cheaper than the imported rice (₦350.00 or $2.8 per Kilogram). With increased demand for the rice at cheap price, marketers strive to increase AVAIL to upgrade their enterprises and extend their services to aid REGULAR. The cultural preference for the rice is, therefore, not unconnected with the low QUALITY which reduces price.

Interestingly, the need to strengthen production, indicates very high relationship with EDU, HOLD, QUALITY and PRICE. The general response on NEEDS is high (97.2%), suggesting a strong linkage between production and consumption of the rice generally. Given that majority of the consumers are literates who are conscious of a regular source with large families (HOLD); low production at *Igbemo* may not offer the supply that will meet the regional needs. For the regional economies to gain momentum, therefore, a high level of sufficiency must be guaranteed. It is a clear fact that the consumers (usually of low-income) prefer the low-quality rice. This canvasses support services and finance for easier access to supply. The rice marketers enjoy freedom from restrictive local mores (market association that control entry into market and fixing prices) which enable them to operate with relatively stable price. The attitudes of the consumers in relation to PRICE reveals that costs of daily consumption of the local rice is tolerable, thus, indicating the need to strengthen production at source. As noted by Abubakar *et al*. (2008), rice represents Africa's best opportunity for reduction of food imports.

5. Conclusion

It is evident from this study that some associated factors affect rice consumption in the study region in Ekiti State. The study has used the questionnaire to source for data among 500 rice consumers in the region. It has also adopted the Random Sampling Technique to select cases investigated. Data analysis reveals that:

- Sex of the consumer is significant to the level of education.

- Annual income of the consumer has a close relationship with the level of education.
- Level of availability of the rice is dependent upon the sex and level of education of the consumer, as well as source of supply.
- The educational status of the consumer, the source of the rice and level of availability determine regularity of consumptions.
- Quality assessment of the rice is closely associated with the consumer's household size, the level of availability of the rice and regularity of consumption.
- Price of the rice, is of significant relationship with the level of education and income of the consumer, source of rice supply, level of availability, regularity of consumption as well as quality of the rice.
- The need to strengthen production is strongly related with sex and level of education of consumers, the quality and price of the rice.

Of special interest in this analysis are factors like: source of the rice, the level of availability, regularity of consumption, level of income and household size of the consumer and the rice quality. Majority of the regional population (83.6%) consume Igbemo rice. In view of the inter-relationships of these variables in consumption, special regional planning policies by the Federal and Ekiti State Government are required to enhance consumption of the local rice while increasing production. In this regard, a State District Development Commission (SDDC) is proposed to take charge of planning and management of specific projects in the region.

Two major problems limit rice production at source (Igbemo). First, is the predominant use of local implements by the rice producers. Second, is the seasonability of production on account of climatic changes. On this note, the SDDC should see to the possibility of mechanizing rice farming operations at Igbemo. Apart from encouraging modern farming technologies, the irrigated rice system should be introduced through the Project Planning and Implementation Unit of the commission to ensure dependable supply of irrigation water to the rice field during dry season. A collaborative effort of the SDDC and the national agency in the state (Benin–Owena River Basin Development Authority – BORBDA) is a viable option. This will strengthen production at source while creating employment opportunities (in rice farming, processing and marketing) to elevate income for regular consumption.

The spatial distribution of the rice, in most cases, reflects the level of availability – a determinant of price at various destinations. Observation reveals the important roles of the poor conditions and network of roads in this case. The newly constructed Ado-Afao road which links the old Afao-Igbemo roads, is the only route on which most traders prefer to travel. Although it is the shortest link to Ado (the state capital), it turns out to be the longest route to Ido-Osi, Ode-Aye and Ikole – which are major sub-regional centres. Improvement of the transportation infrastructure in the region (particularly around Igbemo) by the State government through the SDDC, therefore, becomes a priority. Fine-grained modified grid of lanes should be introduced to link these sub-regional centres with Igbemo for free flow of traffic. This will generate more trips by traders and make availability of the rice increasingly possible for consumption at different destinations, while reducing post-harvest losses.

A high level of poverty is prevalent among the consumers in the region, majority of who are farmers (72.6%), earning below ₦80,000($640.0) annually (70.9%). Traditionally, only self-financing options characterize all enterprises (farming, trading etc). This calls for a change of orientation to 'Competition fund' (CF) to restructure the consumers' enterprises

financing. By this, 4 – 5 people in the same trade should be encouraged by the SDDC to co-invest on projects of interest with a sharing formular for proceeds therefrom. Granting of soft-loans by government to the consumers through the SDDC under a Poverty Alleviation Programme for small-scale businesses will help to enhance their income level for effective rice demand.

The average consumer in the region maintains a large household size, ranging from 6 to 8 people. As a way of satisfying regularity of consumption by the households, an efficient Rice Marketing Centre (RMC) by the SDDC is recommended at source (Igbemo). This will represent a designated depot for receiving the milled rice in the area and discharging to organized marketing cooperatives that eventually pass to sub-regional markets for the urban and rural markets (UMs and RMs). It will also assist in regulating price and minimize the negative influence of emergency rice traders.

The average rice handler (farmer, processor, marketer and vehicle driver) in the study area, lacks necessary knowledge of food hygiene because of low level of education. Government, through the SDDC should collaborate with stakeholders (Ministry of Commerce and Industry, Agriculture, Local Government) and organize regular public enlightenment and training programmes (workshops, seminar etc) for the producers and other handlers of the product for adherence to acceptable standards of food production.

References

1. Abubakar, M., Olayinka, C., and Shady, Q. **NEPAD, Japan to Raise Rice Production in Africa,** The Guardian, 1, 2008, June 2
2. Adeola, C. **Analysis of Economic Efficiency of Male and Female Rice Farmers in Ekiti State, Nigeria,** Nigeria Institute of Social and Economic Research- NISER, Monograph Series, 15, 2002
3. Ajake, A.O. **Population Changes and Food Insecurity in Niger Delta**, Global Journal of Social Sciences, 2(1), . 2003, pp. 45 – 53
4. Akande, T., **An Overview of the Nigerian Rice Economy**, Monography published by the Nigerian Institute of Social and Economic Research (NISER) Ibadan, 2001, 11pp.
5. Akpokodje, G., Lancon, F, and Erenstein, O. **Nigeria's Rice Economy: State of the Art,** Draft Report, WARDA-NISER Collaborative Study, 2001
6. Ashford, L. **New Population Policies: Advancing Women's Health and Right**, Population Bulletin; Population Reference Bureau, Washington D.C. USA, 2001, 44 pp.
7. Awe, O. **Ban on Rice Importation Depresses Global Trade**, Punch, 2006, May 20
8. Daramola, A.G. **Competitiveness of Nigeria Agriculture in a Global Economy: Any Dividends of Democracy?** Inaugural Lecture series 36, Federal University of Technology Akure, 2004, 36 pp.
9. Doney, M. **Nigeria by Numbers**, Development, 2005, 32, 9
10. Fakorede, M.A.B. **Revolutionalizing Nigerian Agriculture with the Golden Seed**, Inaugural Lecture, Obafemi Awolowo University, Ile-Ife, 2001
11. Hall, P. **Urban and Regional Planning**, London, Routledge, 1992
12. Longtau, S. R. **Multi-Agency Partnership in West African Agriculture: A Review and Description of Rice Production System in Nigeria**. Monograph published by Eco-System Development Organization, Jos, 2001, 47 pp.
13. Morenikeji, W. **Research and Analytical Methods for Social Scientists, Planners and Environmentalists**, Jos University Press Ltd, 106, Jos, 2006
14. Okoko, E. **Tenants' Willingness to pay for Better Housing in Targeted Core Area Neighbourhoods in Akure, Nigeria,** Habitat International, 28, 2004, pp. 317– 332
15. Otzen, U. **Strengthening the Peasant Farming Economy for Poverty – Oriented Growth, Agriculture and Rural Development**, First edition, 2002, pp. 62 – 64
16. Schall, N., **Regional Development Planning in Kenya and Nepal, Agriculture and Rural Development**, 2002, 2, 50

17. Smith, D.M., **Human Geography: A Welfare Approach**, London. Edward Arnold.West African Rice Development Association – WARDA 2002. Nigeria Potential in the Rice Sector, WARDA and the Role of Rice in Nigeria, News Release, 4, 1977
18. Whang, In-Joung **Social Services Programme for the Poor in a newly Industrializing Country: Experience in South Korea,** in Rondwell and Cheema (eds) "Urban Services in Developing Countries", London, Macmillan Press, 1988. pp. 147-174
19. * * * **Encarta Premium Suite 2004**. Planting Rice, Retrieved on June 15, 2006
20. * * * Nigerian Institute of Social and Economic Research – NISER 2002. **Assessment of the Economic, Social and Environmental Impacts of Rice Production in Nigeria within the Trade Libralisation Framework**, A Research paper 93 pp
21. * * * Population Reference Bureau, **Demographic Data and Estimates for the Countries and Regions of he World, 2001** World Population Data Sheet USA, Washington, D.C., 2001

[1] J. O. Basorun is a lecturer in the Department of Urban and Regional Planning and currently the Acting Associate Director, University Diploma in Science Programme, at the Federal University of Technology, Akure, Nigeria. He had his Master's and PhD degree in Urban and Regional Planning from Enugu State University of Science and Technology and the Federal University of Technology Akure, Nigeria respectively. He is a member of Nigerian Institute of Town Planners and Registered of Town Planners Registration Council of Nigeria. He consults widely for many National and International Agencies and Goverments. He is well publishe with many and scholarly journals at local and international level. His research and teaching interests are in Urban Design, Quantitative Methods, Community Development and Regional Development and Planning.

Appendix 1.

Table 1. Correlation Matrix of Consumers Variables

Variable Code	SEX	EDU	HOLD	INCOME	SOURCE	AVAIL	REGULAR	QUALITY	PRICE	NEEDS
SEX	1.000	-0.111*	0.000	-0.047	-0.065	0.129**	-0.010	-0.029	-0.019	0.014
EDU		1.000	0.069	0.116*	-0.094	-0.118*	-0.117*	-0.088	-0.204**	0.147**
HOLD			1.000	-0.002	-0.042	0.020	0.077	0.147**	.0.057	0.234**
INCOME				1.000	-0.052	-0.045	-0.067	-0.065	-0.105*	-0.010
SOURCE					1.000	-0.123*	0.097*	0.076	0.113*	0.029
AVAIL						1.000	0.412**	0.189**	0.232**	.0.017
REGULAR							1.000	0.627**	0.506**	-0.082
QUALITY								1.000	0.565**	-0.157**
PRICE									1.000	-0.122*
NEEDS										1.000

* Correlation is significant at 0.05 level (P < 0.05)

** Correlation is significant at 0.01 level (P < 0.01)

Source: Author's Fieldwork, 2007

THE VOLATILITY PREMIUM RISK: VALUATION AND FORECASTING[1]

Bogdan NEGREA[2]
PhD, University Professor, Department of Money,
Faculty of Finance, Insurance, Banking and Stock Exchange, University of Economics,
010374, 6 Romana Square, Bucharest, Romania
Associate Researcher, University Paris 1 (Panthéon-Sorbonne), France

E-mail: negrea@univ-paris1.fr

Abstract: *Empirical studies, such as Lamoureux and Lastrapes (1993), Guo (1998), Fouque et al. (2000) show that the market price of volatility risk is nonzero and time varying. This paper provides a theoretical investigation of the market price of volatility risk. We consider that the market price of volatility risk is a function of two variables: the price of underlying asset and its volatility. We suggest a closed-form solution for the price of volatility risk under the conditions of stochastic volatility and of correlation between the underlying asset price and its volatility. This formula involves in a direct way the unobservable market price of volatility risk. We prove that the correlation between underlying price and its volatility has no impact on the price of volatility risk. Finally, we present empirical results using the prices of CAC 40 index and of CAC 40 index call options from January 2006 to December 2007.*

Key words: *Premium risk; Volatility risk; Stochastic volatility; Option pricing; Risk forecasting*

1. Introduction

Option prices observed on a liquid option market contain rich information about the market's expectation of the future distribution of the underlying asset, and the risk premium for unhedgeable distribution risk.

If the market is complete, the derivative's risk can, theoretically, be perfectly hedged by the underlying asset and there is no volatility risk premium to be estimated. We consider that the volatility is stochastic and therefore not constant, the market is incomplete and consequently the price of volatility risk will be non-zero.

The assumption of zero correlation between volatility changes and aggregate consumption changes is often invoked to justify a zero risk price for volatility risk (Hull and White 1987). This assumption was challenged by Melino and Turnbull (1990), who found that a stochastic volatility model with no positive price of volatility risk explains observed prices better than the constant volatility models. Lamoureux and Lastrapes (1993) empirically showed that the market price of volatility risk is nonzero and time varying.

Moreover, a negative risk price parameter is consistent with the belief that volatility changes are negatively correlated with the aggregate consumption growth, and the investors prefer to pay a risk premium to hedge the volatility risk.

Various stochastic volatility option pricing models have been developed over the past few years. Many researchers, like Hull and White (1987), Stein and Stein (1991), Heston (1993), and Bakshi et al. (1997, 2000) concluded that the volatility of an asset's return could be itself a random variable describing a specific process. The model of Heston (1993) allows for the systematic volatility risk to be specified, whereas Hull and White (1987) and others have to assume a zero price of volatility risk in order to obtain a tractable option pricing model. Heston (1993) provided a closed-form solution for the price of a European-style option on an asset with mean-reverting square-root stochastic volatility. In the model of Heston, the dynamics of the underlying asset and the volatility are:

$$dS_t = \mu S_t dt + \sigma_t S_t dB_t \quad (1)$$

$$d\sigma_t^2 = k\left(\theta - \sigma_t^2\right)dt + \sigma_v \sigma_t dW_t \quad (2)$$

where dB_t and dW_t are Wiener processes with instantaneous correlation ρ. The instantaneous variance $\sigma^2(t)$ follows a square-root process that was used by Cox, Ingersoll, and Ross (1985) to model the instantaneous interest rate process. The volatility diffusion process is a mean-reverting process with long-term mean parameter of θ, mean-reversion speed parameter of k, and the volatility of volatility parameter of σ_v.

Because there is no traded security than can be used to hedge the risk of volatility, it is difficult to form a risk-free portfolio. In this case, the option valuation is no longer preference free, and the market price of volatility risk needs to be determined. A formal treatment of the risk price requires an equilibrium model. In rational expectation equilibrium models of asset markets (Breeden 1979), a state variable's risk price is a function of its covariance with the representative investor's marginal utility of consumption, so $\gamma\left(S,\sigma^2,t\right) = \lambda COV\left(d\sigma^2, dC/C\right)$, where $C(t)$ is the consumption rate and λ is the coefficient of relative risk aversion of the representative investor. If percentage changes in consumption have constant covariance with the volatility changes, the price of the volatility risk can be represented as being proportional to volatility. Hence, Heston used the following condition: $\gamma\left(S,\sigma^2,t\right) = \gamma\sigma_t^2$. A negative price of volatility risk arises if this covariance is negative ($\gamma < 0$).

The market price of the volatility risk is an unobservable variable. We have only one piece of information about it: the volatility risk premium depends on the stock price and its volatility. In this paper we consider that the market price of volatility risk follows a diffusion process and depends on the underlying price and the volatility of underlying returns.

2. The Volatility Premium Risk Valuation

We consider an option pricing model with two state variables. Accordingly, the price of a European option depends on the price of the underlying asset and on the volatility, which describes the following stochastic process:

$$dS_t = \mu S_t dt + \sigma_t S_t dB_t \quad (3)$$

$$d\sigma_t^2 = \varphi\sigma_t^2 dt + \sigma_v \sigma_t^2 dW_t \quad (4)$$

where the current values of S and σ^2 are known. The same stochastic definitions of the state variables are used by Hull and White (1987). The return drift, μ, is a parameter which depends on S (the price of the underlying asset), σ (the volatility), and time. The Brownian motions B_t and W_t are correlated ($dB_t dW_t = \rho dt$) and ρ is the correlation coefficient ($-1 < \rho < 1$). The volatility of the volatility (σ_v) and the drift of the variance (φ) are assumed to be constant, therefore implying that the non-anticipated variations of the variance are stationary.

Using the Girsanov's Theorem, under the risk-neutral probability equivalent with the real probability, the diffusion process described by the price of the underlying asset is given by:

$$dS_t = rS_t dt + \sigma_t S_t dB_t^* \quad (5)$$

where r is the risk-free return, B_t^* is the equivalent Brownian motion and $dB_t^* dW_t = \rho dt$. We consider that the market price of volatility risk is a function of two state variables: the price of underlying asset and the volatility, $\gamma\left(S, \sigma^2, t\right)$. The price of volatility risk describes the following diffusion process:

$$d\gamma_t = \sigma_t dB_t^* + \sigma_v dW_t \quad (6)$$

Using the Itô's lemma, the dynamic of the market price of the volatility risk is given by:

$$d\gamma = \left[\frac{\partial\gamma}{\partial t} + \frac{\partial\gamma}{dS} rS + \frac{\partial\gamma}{\partial\sigma^2}\varphi\sigma^2 + \frac{1}{2}\frac{\partial^2\gamma}{\partial S^2}\sigma^2 S^2 + \frac{1}{2}\frac{\partial^2\gamma}{\partial\left(\sigma^2\right)^2}\sigma_v^2\sigma^4 + \frac{\partial^2\gamma}{\partial S\partial\sigma^2}\rho\sigma_v\sigma^3 S\right] + \frac{\partial\gamma}{\partial S}\sigma S dB + \frac{\partial\gamma}{\partial\sigma^2}\sigma_v\sigma^2 dW \quad (7)$$

Taking into account the stochastic process followed by the market price of the volatility risk, we obtain the following conditions:

$$\frac{\partial\gamma}{\partial t} + \frac{\partial\gamma}{dS} rS + \frac{\partial\gamma}{\partial\sigma^2}\varphi\sigma^2 + \frac{1}{2}\frac{\partial^2\gamma}{\partial S^2}\sigma^2 S^2 + \frac{1}{2}\frac{\partial^2\gamma}{\partial\left(\sigma^2\right)^2}\sigma_v^2\sigma^4 + \frac{\partial^2\gamma}{\partial S\partial\sigma^2}\rho\sigma_v\sigma^3 S = 0 \quad (8)$$

$$\frac{\partial\gamma}{\partial S}\sigma S = \sigma \quad (9)$$

$$\frac{\partial \gamma}{\partial \sigma^2}\sigma_v\sigma^2 = \sigma_v \tag{10}$$

This allows us to compute the partial derivatives: $\frac{\partial \gamma}{\partial S} = \frac{1}{S}$, $\frac{\partial \gamma}{\partial \sigma^2} = \frac{1}{\sigma^2}$, $\frac{\partial^2 \gamma}{\partial S^2} = -\frac{1}{S^2}$, $\frac{\partial^2 \gamma}{\partial \left(\sigma^2\right)^2} = -\frac{1}{\sigma^4}$ and $\frac{\partial^2 \gamma}{\partial S \partial \sigma^2} = 0$. Finally, we obtain an ordinary differential equation by substituting the partial derivatives into the equation (8).

$$\frac{d\gamma}{dt} = -r - \varphi + \frac{1}{2}\sigma_v^2 + \frac{1}{2}\sigma^2(t) \tag{11}$$

with the initial condition: $\gamma(0) = 0$.

The solution of this equation gives the following formula for the valuation of the market price of volatility risk:

$$\gamma_t = \frac{1}{2}\left(\sigma_v^2 + \bar{V}\right)t - (r+\varphi)t \tag{12}$$

where $\bar{V} = \frac{1}{t}\int_0^t \sigma^2(u)\,du$, $\forall 0 < u < t$, represents the mean of historical variances. Moreover, the annualized market price of the volatility risk is:

$$\gamma_{at} = \frac{1}{2}\left(\sigma_v^2 + \bar{V}\right) - (r+\varphi) \tag{13}$$

Switching from t (the current date) to T (the future date), we can obtain the market price of volatility risk at the maturity of an option. The future price of volatility risk can be written as:

$$\gamma_T = \gamma_t + \frac{1}{2}\left(\sigma_v^2 + \tilde{\bar{V}}\right)\tau - (r+\varphi)\tau \tag{14}$$

where γ_t is known. The parameter τ is the time to maturity T, $\tau = T - t$. The mean of the futures variances between t and T is a random variable: $\tilde{\bar{V}} = \frac{1}{\tau}\int_t^T \sigma^2(u)\,du$, $\forall t < u < T$.

The obtained formula (14) can be used to forecast the market price of the volatility risk. Therefore,

$$E[\gamma_T] = \gamma_t + \frac{1}{2}\left(\sigma_v^2 + E\left[\tilde{\bar{V}}\right]\right)\tau - (r+\varphi)\tau \tag{15}$$

After some stochastic calculus, the expected value of the mean futures volatilities is given by:

$$E\left[\tilde{\bar{V}}\right] = \frac{e^{\varphi\tau} - 1}{\varphi\tau}\sigma^2 \tag{16}$$

The forecasting volatility premium risk is therefore given by:

$$E[\gamma_T] = \gamma_t + \frac{1}{2}\left(\sigma_v^2 + \frac{e^{\varphi\tau} - 1}{\varphi\tau}\sigma^2\right)\tau - (r+\varphi)\tau \tag{17}$$

We note that the volatility premium risk is inversely proportional to the instantaneous risk free return and no depends on the correlation between the asset price and its volatility.

3. Empirical Results

CAC 40 index series used in our studies were extracted from the Bourse de Paris database from July 1996 to December 1998. The database includes a time-stamped record of every trade occurred on the CAC 40 index. Our study focuses on the volatility premium risk of CAC 40 index traded in 2006 and 2007. On the sample, more than 35,000 quotations are reported.

The EURIBOR 3 months, which was the most liquid maturity on the French market, serves as a proxy for the risk free interest rate and is obtained from Datastream.

Option series used in these studies were extracted from the Bourse de Paris database from January 2006 to December 2007. Our study focuses on CAC 40 index call options (PXL contracts) traded at the MONEP in 2007 and 2008.

Options written on CAC 40 index are the most actively traded of the MONEP on the sample period. In 2006 and 2007, an average of 745,309 option contracts were traded monthly and 35,775 daily. Option premium reached EUR 750 million a month and EUR 35 million a day.

Over these years, both equity and index options accounted for a roughly similar proportion of lots traded, but equity options represented less than 25% of the total amount corresponding to the sum of premium.

A. Estimation of the volatility premium risk using historical volatility.

We estimate the market price of volatility risk of CAC 40 index from January 2006 to December 2007. Let x_i be the log-return on the index price defined as:

$$x_i = \ln\left(\frac{S_i}{S_{i-1}}\right) \tag{18}$$

where S_i and S_{i-1} are respectively the index price at the end of the day i and of the day $i-1$, and x_i is the continuously compounded return (not annualized) of the index in the day i. Figure 1 shows the evolution of the continuously compounded return from July 2006 to December 2007.

We compute the historical volatility from January 2006 to December 2007 using the closing index prices from daily data over the recent 90 days. The historical volatility s_t is the standard deviation of the x_i's:

$$s_t = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \tag{19}$$

where $n = 90$ days, and $\overline{x}$ is the mean of index returns.

Assuming that time is measured in trading days and that there are 252 trading days per year, the volatility per annum at the date t, is:

$$\sigma_t = s_t\sqrt{252} \tag{20}$$

Figure 2 shows the evolution of historical volatility from October 2006 to December 2007.

We compute the mean of historical volatilities ($\overline{V}$) which occur in the last 90 days.

$$\overline{V} = \frac{1}{90}\sum_{i=1}^{90}\sigma_i^2 \tag{21}$$

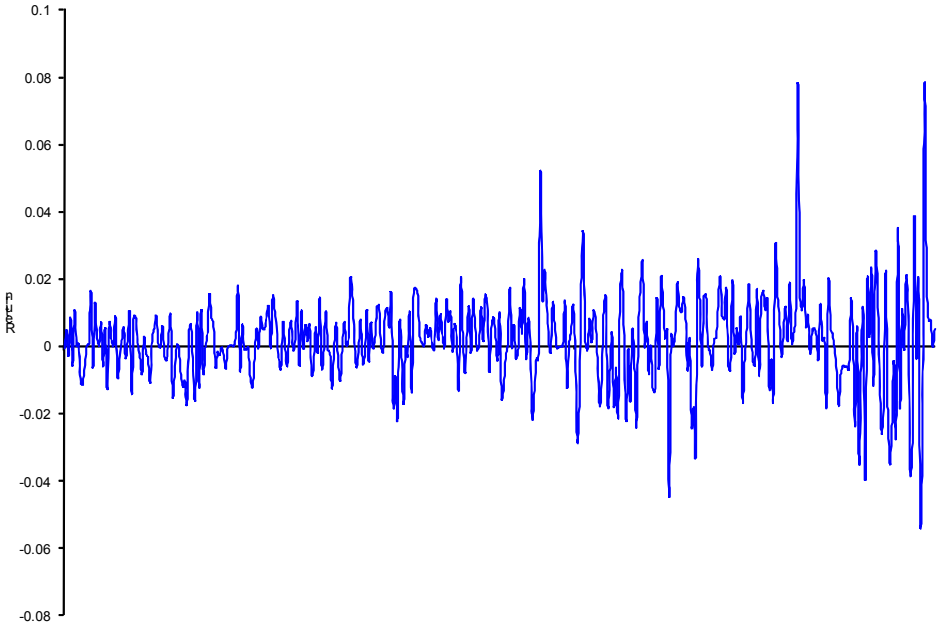


Figure 1. The Continuously Compounded Return

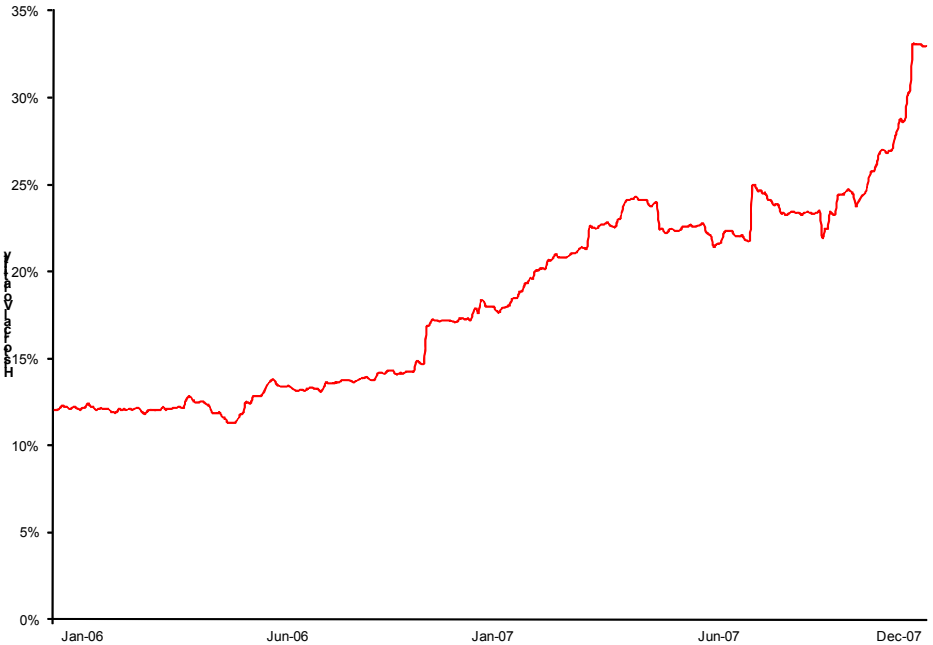


Figure 2. The Historical Volatility

It remains to estimate the volatility diffusion parameters, φ and σ_v, in order to apply the formula of the current price of volatility risk. We use the maximum likelihood method and the GMM estimation. The log-likelihood function is:

$$L = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_v^2 h - \frac{1}{2}\frac{\left[\ln \frac{\sigma_t^2}{\sigma_{t-1}^2} - \left(\varphi - \frac{\sigma_v^2}{2}\right)h\right]^2}{\sigma_v^2 h} \tag{22}$$

In the case of GMM estimation we have the following conditions:

$$E_{t-1}\left[\varepsilon_t\right] = 0 \quad \text{and} \quad E_{t-1}\left[\varepsilon_t^2 - \sigma_v^2 h\right] = 0 \tag{23}$$

with

$$\varepsilon_t = \ln \frac{\sigma_t^2}{\sigma_{t-1}^2} - \left(\varphi - \frac{\sigma_v^2}{2}\right)h \tag{24}$$

We obtain the value of volatility diffusion parameters, φ and σ_v, at each date t between January 2006 and December 2007 using the time series of historical volatilities σ_t. Figure 3 shows the estimated volatility drift (φ) of CAC 40 index from January 2006 to December 2007.

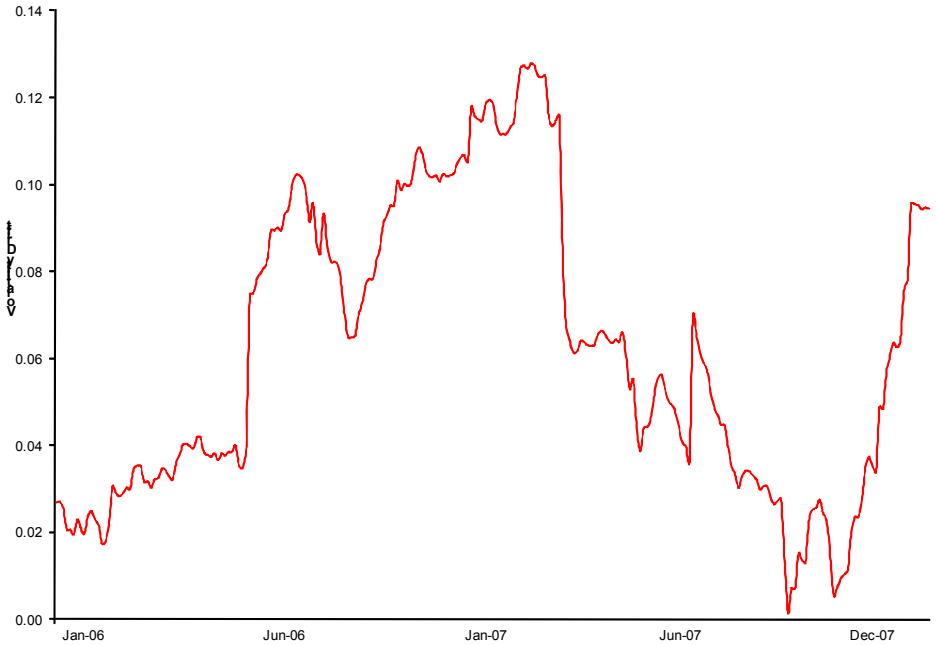


Figure 3. The Volatility Drift

Using the estimated parameters and applying the obtained formula, we compute the volatility premium risk. Its evolution is described in the Figure 4. We notice that the price of volatility risk is almost perfectly negatively correlated with the volatility drift. One unit of volatility premium decrease corresponds almost to one unit of volatility drift increase. If one buys calls with high volatility premium, the price paid for those options is generally higher. If a move up occurs, there is a chance that the underlying may not be able to move up fast enough to compensate for the decrease in volatility premium. We observe that the volatility premium may decrease as the price of the underlying increases.

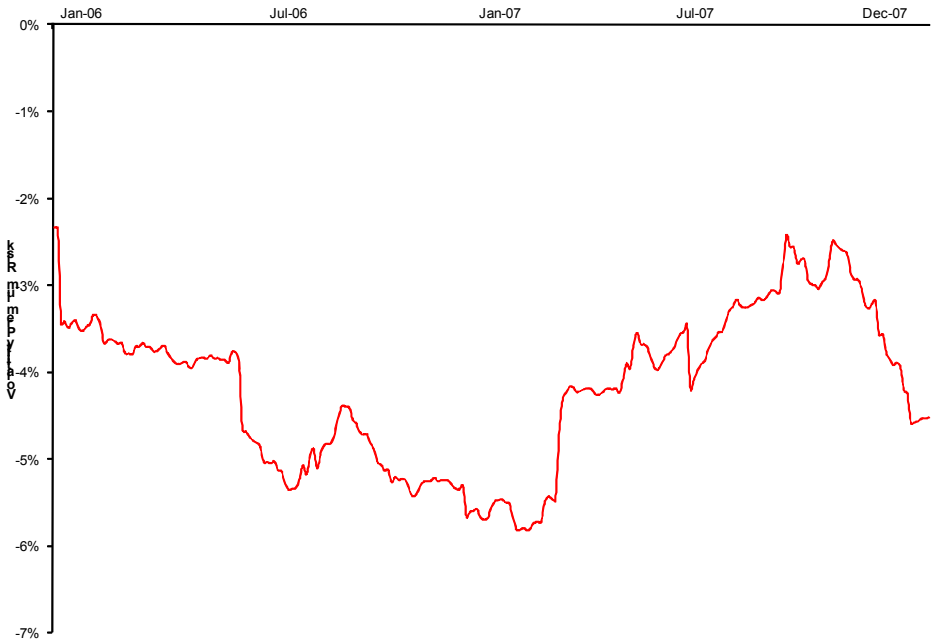


Figure 4. The Volatility Premium Risk Computed with Historical Volatility

B. Estimation of the volatility premium risk using implied volatility.

In order to compute the implied volatility we apply the Black-Scholes formula to the pricing of CAC 40 index call options traded at the MONEP. Reversing the Black-Scholes formula, we estimate the implied parameter (i.e. the implied volatility) for options negotiated on the MONEP.

The theoretical price of an option contract depends on four observable parameters: the strike price, the maturity, both specified in the contracts, the underlying price, and the risk free interest rate, which can be taken from public market data, as well as some non-observable parameters describing the risk-neutral density function. One parameter is required by the Black and Scholes model: the volatility. We use a non linear least squares procedure in both cases where n is the number of all call options j, $j=[1,...,n]$, available on a given day t, for a given maturity τ. The Black-Scholes call price is defined as $C_t^* = C_t(\hat{\sigma})$, where $\hat{\sigma}$ is the solution of the following minimization problem:

$$\hat{\sigma}_t = Arg\left\{ \underset{\hat{\sigma}}{Min}\left[\sum_{j=1}^{n} \left(C_j^{obs} - C_j^{BS}(\hat{\sigma}) \right)^2 \right] \right\} \qquad (25)$$

Backing out implied parameter from all option prices on a day, we allow parameter to vary daily, which is inconsistent with the models assumption if parameter proved to be truly variable. Several empirical studies use this procedure (see, for instance, Bakshi et al., 1997 and Dumas et al., 1998).

Following the same technique described above, the volatility drift and the volatility of the volatility are computed using the implied volatility time series. Therefore, using the theoretical formula, we compute the market price of the volatility risk. Figure 5 shows the evolution of the volatility premium risk during the study period.

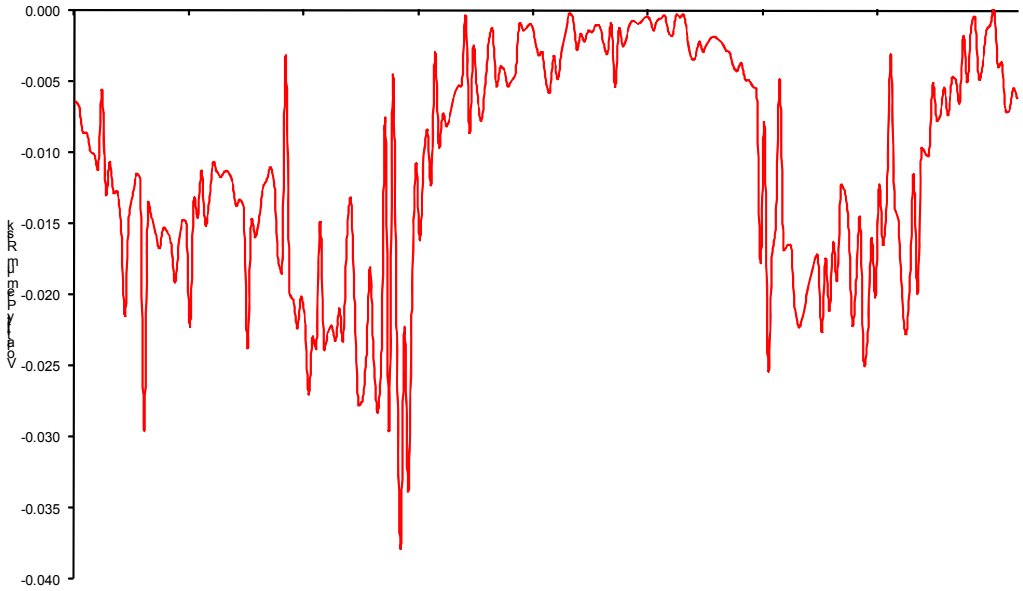


Figure 5. The Volatility Premium Risk Computed with Implied Volatility

We compute the forecasting price of volatility risk at each date t until date T using the expression (17).

The forecasted market price of volatility risk ($E[\gamma_T]$) of CAC 40 index is compared with the current market price of volatility risk (γ_t). The figure 6 describes the evolution of the current and of the forecasting prices of volatility risk from December, 22 2006 to Mars, 31 2007 while the figure 7 exhibits data from June, 24 2007 to September, 30 2007. The forecasted price of volatility risk and the current price of volatility risk are both negative.

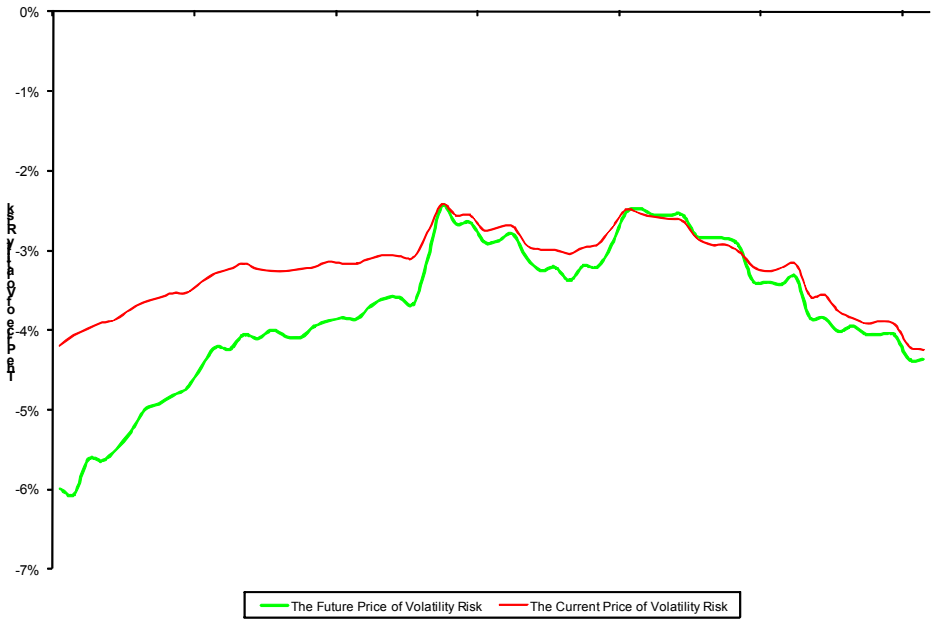


Figure 6. Forecasting versus Current Volatility Premium Risk

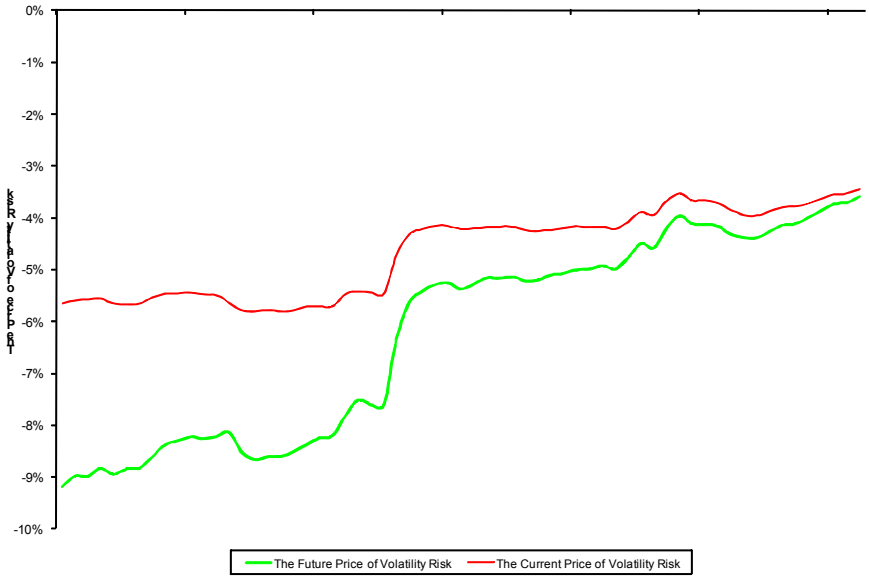


Figure 7. Forecasting versus Current Volatility Premium Risk

The forecasted price of volatility risk is lower than the current price of volatility risk. This explains the market's "crash-o-phobia". The forecasted volatility risk premium gives the possibility to quantify at each date the degree of "crash-o-phobia" of the market. In the first case, the forecasted price of volatility risk is -9.19% predictable for the 90th day in the future while the current price of volatility risk is -5.66%. In the second case, the forecasted price of volatility risk is -5.99% while the current price of volatility risk is -4.19%. The degree of market's "crash-o-phobia" was most important during the winter - spring period of 2007 than during the summer - autumn period of 2007.

4. Conclusions

We have shown that a model characterized by two state variables, the stock price and its volatility, allows us to obtain the volatility risk premium. By modeling the volatility and stock price processes, we obtain a formula for computing the price of volatility risk. This formula involves in a direct way the unobservable market price of volatility risk. We conclude that the volatility risk is determined by the volatility and by the diffusion parameters corrected by the risk-free interest rate. Hence, the historical data can be used to estimate the volatility risk premium.

The market price of volatility risk is negative. If one buys calls with high volatility premium, the price paid for those options is generally higher. If a move up occurs, there is a chance that the underlying may not be able to move up fast enough to compensate for the decrease in volatility premium. We may therefore notice a decrease in the volatility premium although the price of the stock goes up.

Lastly, the forecasted price of volatility risk can be used to measure the degree of market's "crash-o-phobia".

References

1. Bakshi, G., Cao, C. and Chen, Z. **Empirical Performance of Alternative Option Pricing Models**, Journal of Finance, 52 (5), 1997, pp. 2003-2049
2. Bakshi, G., Cao, C. and Chen, Z., **Pricing and Hedging Long-Term Options**, Journal of Econometrics, 94, 2000, pp. 277-318
3. Bakshi, G., Cao, C. and Chen, Z. **Do Call Prices and the Underlying Stock Always Move in the Same Direction?**, Review of Financial Studies, 13 (3), 2000, pp. 549-584
4. Black, F. **Studies of stock price volatility changes**, in "Proceedings of the 1976 meetings of the American Statistical Association, Business and Economics Statistics Section", 1976
5. Campbell, J., Lo, A. and MacKinley, C. **The Econometrics of Financial Markets**, Princeton University Press, 1996
6. Cox, J.C., Ingersoll, J. and Ross, S.A. **A Re-examination of Traditional Hypotheses about the Term Structure of Interest Rate**, Journal of Finance, 36, 1981, pp. 769-799
7. Christie, A. **The stochastic behavior of common stock variances: value, leverage, and interest rate effects**, Journal of Financial Economics, 10, 1982, pp. 407-432
8. Fouque, J.-P., Papanicolaou, G. and Sircar, R. **Mean-Reverting Stochastic Volatility**, International Journal of Theoretical and Applied Finance, Vol.3, No 1, 2000, pp. 101-142
9. Guo, D. **The Risk Premium of Volatility Implicit in Currency Options**, Journal of Business & Economic Statistics, 16 (4), 1998, pp. 498-507
10. Heston, S. **A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options**, The Review of Financial Studies, 6 (2), 1993, pp. 327-343
11. Hull, J. and White, A. **An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility**, Advances in Futures and Option Research, 3, 1988, pp. 29-61
12. Hull, J. and White, A. **The Pricing of Options on Assets with Stochastic Volatilities**, Journal of Finance, 42 (2), 1987, pp. 281-300
13. Karatzas, I. and Shreve, S. **Brownian Motion and Stochastic Calculus**, Second Edition, Springer Verlag, 1991
14. Lamoureux, C.G. and Lastrapes, W.D., **Forecasting Stock-Return Variance: Toward an Understanding of Implied Volatilities**, The Review of Financial Studies, 6 (2), 1993, pp. 293-326

---

[1] **Acknowledgement**
This paper is part of research project, no.169/2007, "Measuring the Amplitude of Financial Market Crisis Using an Index Following the Richter Scale from Seismology. An Application of the Econophysics Principles", financed by Academic Financing Institute (UEFISCSU).

[2] **Research interest field:** option valuation, risk valuation, market microstructure. **Publications:**
Articles:

- « A note on skewness and kurtosis adjusted option pricing models under the Martingale restriction », Quantitative Finance, vol.4, October 2004 (joint work with E. Jurczenko – ESCP-EAP and B. Maillet – Université Paris 1).
- « La volatilité des marchés augmente-t-elle ? », Revue d'Economie Financière, no. 74, May 2004 (joint work with T. Chauveau – Université Paris 1, B. Maillet – Université Paris 1, E. Jurczenco – ESCP-EAP, J. Héricourt – Université Paris 1, H. Raymond-Feingold – Université Paris 1, C. Lubochinsky – Université Paris 2, S. Friederich – London School of Economics, C. Moussu – ESCP-EAP).
- « Revisited Multi-moment Approximate Option Pricing Models (Part 1) », London School of Economics – Financial Markets Group (LSE-FMG), no. 430, 2002 (joint work with B. Maillet and E. Jurczenko - Université Paris 1).
- « Skewness and Kurtosis Implied by Option Prices: A Second Comment », London School of Economics – Financial Markets Group (LSE-FMG), no. 419, 2002 (joint work with B. Maillet and E. Jurczenko - Université Paris I).

Books:

- « Financial Assets Pricing: an introduction to the stochastic process theory », Economica Publisher, Bucharest, 2006 (in Romanian language).
- « Financial Engineering », ASE Publisher, Bucharest, 2006 (in Romanian language).
- « La volatilité des marchés financiers », Chapter 1: Mesurer l'instabilité financière, Editions AEF, Paris, 2004 (joint work with T. Chauveau – Université Paris 1, B. Maillet – Université Paris 1, E. Jurczenco – ESCP-EAP, J. Héricourt – Université Paris 1, H. Raymond-Feingold – Université Paris 1, C. Lubochinsky – Université Paris 2, S. Friederich – London School of Economics, C. Moussu – ESCP-EAP) (in French language).

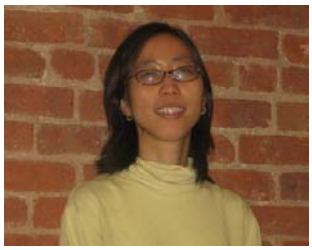
LOGISTIC REGRESSION TO DESCRIBE HOW THE RELATIONSHIP BETWEEN SOCIAL CONNECTION AND SELF-RATED HEALTH VARIES BY GENDER

Michael A. LEWIS[1,2]
Hunter College School of Social Work, New York, USA

E-mail: michael.a.lewis@hunter.cuny.edu

Eri NOGUCHI[3]
Columbia University School of Social Work, New York, USA

E-mail: en16@columbia.edu



Abstract: *Social Scientists and health researchers have allocated a great deal of time trying to understand the social correlates of health outcomes, some in the question of the association between social relationships and such outcomes. This paper builds on this work by exploring the association between how connected one feels to her or his neighborhood and the chance that one reported that she or he was in poor health. Using logistic regression on a data set of 1,899 females and 1,521 males from the United States we found that, for both males and females, those who reported that their neighbors give them a sense of community were less likely to report that they were in poor health, but this association was stronger for females than for males.*

Key words: *Logistic Regression; Social Connections; Self-rated Health; Gender*

1. Introduction

Social scientists and health researchers have done a great deal of quantitative work on the social correlates of health outcomes. Race (Gibbs, et al., 2006), gender (McDonough, Peggy and Vivienne, Walters, 2001), and socioeconomic status (Malat, Jennifer R.; van Ryn, Michelle; Purcell, David, 2006) are among these correlates. Another key correlate of health outcomes, one that will be the focus of this paper, is social connection. It has been found

that the degree of people's connections to others is negatively associated with the chance of neonatal death (Gayen and Raeside, 2007), degree of mental health problems (Fiori, Antonucci, and Cortina 2006; Ueno, 2005; and Cannuscio, Colditz, and Rimm, 2004), impairment of physical functioning (Unger, McAvay, and Bruce 1999), and the risk of developing a disability (Mendes de Leon, Glass, and Beckett, 1999). The degree of people's connections to others has been found to be positively associated with the chance of recovering from a disability (Mendes de Leon, Glass, and Beckett, 1999) and level of self-rated health status (Helweg-Larsen, Kjoller, and Thoning 2003; Zunzunegui, Kone, and Johri, 2004; and Veenstra, Luginaah, and Wakefield, 2005). Very little of this research, however, has explored the question of whether the association between social connection and health outcomes varies by gender.

There are good reasons, however, for thinking that the association between social connection and health outcomes might differ by gender. First, such a difference may be due to gender based differences in the composition of groups to whom people are connected, and these compositional differences may result in females being influenced in ways that lead them to behave in ways different from males. Second, even if group composition is the same, females, on average, may relate to people differently than men, leading to differences in the influence of social connection on health.

This paper builds on previous work by exploring the association between social networks and health outcomes. More specifically, we address the question of the degree of association between whether one reported that one's neighbors give them a sense of community and the likelihood that one reported that she or he was in poor health. We also assess whether this association varies by gender. We will proceed as follows.

First, we will provide more details regarding the theoretical model we explored. Second we will discuss the data set and methods we used. Third, we will provide an overview of our results. And fourth, we will conclude with a brief discussion of these results.

Theoretical Model

Let i represent a given individual and assume that she or he lives in a given neighborhood. Let $COMMUNITY_i$ be a variable representing whether i's neighbors give him or her a sense of connection to others or a sense of community. If i feels that his or her neighbors give him or her such a sense, this variable takes the value 1. If not, it takes the value 0. Let $POOR_i$ be a variable representing whether i feels he or she is in poor health. If i feels he or she is in poor health, this variable takes the value 1. If i does not feel this way, it takes the value 0.

Having set out these preliminaries, we propose the following model of self-rated poor health:

$$P(POOR_i = 1) = f(COMMUNITY_i, \mathbf{V}_i) \quad (1)$$

where $P(POOR_i = 1)$ refers to the probability that person i reported that she or he is in poor health, f refers to some mathematical function, and $\mathbf{V}_i$ represents a set of other factors that are thought to be associated with $P(POOR_i = 1)$.

Data and Methods

The data we used to examine the above model came from the Social Capital Community Benchmark Survey, which was designed by the Saguaro Seminar at Harvard University Kennedy School of Government. A sample of respondents from across the United States was obtained through random-digit-dialing (RDD) procedures. The actual telephone interviews were conducted by Intersearch, an international survey firm, and each one took an average of 26 minutes to complete (Roper Center, 2005). Our analysis is based on 1,899 employed female and 1,521 employed male respondents. Some were employed full-time (>=35 hours/week) and some part-time (< 35 hours/week).

Most of the survey questions required respondents to choose, from among pre-determined answers, the one that, in their judgment, best reflected them. For example, someone who was primarily a full-time homemaker, but who had worked a little for pay the previous week, would decide for herself whether she identified as a "part-time worker" or "full-time homemaker." Data are available, from this survey, on a variable referred to as "health." It is based on respondents' self-reported health statuses. Respondents were asked to rate their health as poor (coded 0), fair (coded 1), good (coded 2), very good (coded 3), or excellent (coded 4). Those who stated that they didn't know their health status were coded 8 and those who refused to offer a response were coded 9. For the analyses described below, we recoded health to $POOR_i$. 0 was recoded to 1, 8 and 9 were recoded to system missing (that is, they were not considered in the analyses, and 1, 2, 3, and 4 were recoded to 0. We didn't consider codes 8 and 9 because we were only interested in those respondents who actually reported a category to describe their health status.

$COMMUNITY_i$ was also based on a recoding of an original variable. In the original survey, this variable is called belnei and came from the question whether respondents felt that their neighbors give them a sense of community. Those who answered "no" were coded 1, those who answered "it depends" 2, those who answered "yes" 3, those who answered "does not apply" 4, those who answered "don't know" 8, and those who refused to answer 9. We were not interested in those who didn't know the answer to the question or who refused to answer, so 8 and 9 respondents were eliminated from the analysis. 1, 2, and 4 were recoded as 0, and 3 was recoded as 1. In order to determine if our data were consistent with our proposed theoretical model, we analyzed the data using logistic regression analysis (LRA).

Analysis

As stated above, we assumed that variables in addition to $COMMUNITY_i$ might be associated with $POOR_i$. These variables are contained in $\mathbf{V}_i$ and were included in our LRA models as covariates. In many cases, these variables were also based on recodes of original variables, and respondents who answered with "don't knows" or who refused to answer questions were not further considered in our analyses. We discuss more about the relevant aspects of these variables in the appendix.

As stated in equation (1), our interest was in modeling the probability that $POOR_i = 1$ as a function of $COMMUNITY_i$ and a set of covariates represented by $\mathbf{V}_i$. A standard

statistical approach used to test such models, is the Linear Probability Model (LPM), a variant of Least Squares Regression Analysis. In our case, LPM could have been used to test the model specified in equation (1). A key problem associated with LPM, however, is that one can end up with predicted probabilities outside the interval [0,1]. To address this problem, we used the alternative approach of LRA. LRA models were run with SPSS 13.0, with separate models run for males and females to determine if the association between $COMMUNITY_i$ and $P(POOR_i = 1)$ varied by gender, controlling for the variables included in $\mathbf{V}_i$. Unlike many works in the social sciences, logistic regression was used as a descriptive as opposed to inferential technique. This requires some explanation.

Researchers who analyze survey data invariably have to deal with the problem of respondents refusing to answer questions, giving incomplete answers, or in some other way providing less than useful answers. This is the missing data problem. One way of handling it is through listwise deletion. But, as is well known (Berk, 2004), if one is interested in generalizing from a sample to a population, there is a key shortcoming of listwise deletion. Cases with complete sets of values may be systematically different from cases with missing values. Thus, even if one started out with a probability sample, once one drops cases with missing values from the analysis, there is a very good chance that the sample one ends up with is no longer a probability sample. This is a serious problem, statistically, because it can lead to biased estimates of population parameters (Gelman and Hill, 2007 and Berk, 2004). Given the fact that, after listwise deletion, we were left with only 1,899 out of the 15,299 females in the overall data set and 1,521 out of 10,526 males, we were quite concerned that the biased estimates problem might apply to our situation.

There are other techniques designed to address the missing data problem that are often utilized by social scientists. These techniques often involve the imputation of values for cases with missing data. But imputing values either has the same potential to result in biased estimates as listwise deletion does or requires one to have good information about the mechanisms that account for why cases end up with missing data (Berk, 2004). Following the advice of Berk (2004), since we didn't feel that we had such information, we chose not to utilize any of these imputation methods.

A way to address our dilemma, which some statisticians regard as quite useful but underrated (Berk, 2004), is to use regression analysis to describe patterns in a data set, instead of as a method of inference. In our case, logistic regression could be used descriptively to determine if the relationship between $COMMUNITY_i$ and $P(POOR_i = 1)$ varied by gender for the set of cases we had after listwise deletion (Berk, 2004). This use of regression has much in common with similar methods, such as classification algorithms, found in computer science (Bramer, 2007). Since the approach dispenses with the notion of using the sample to test hypotheses about population parameters, the concern about biased estimates is no longer relevant. A limitation of this approach is that one isn't in a position to know the extent to which one's findings hold in other populations. But since the alternative was, in our view, a good chance of ending up with biased estimates, we were willing to accept this limitation. We were also willing to accept it because we realize that it can be addressed by other researchers going out and trying to replicate our findings in different populations, something we regard as one of the hallmarks of science.

Results

Table 1 contains summary statistics for the variables included in our Logistic regression models, separately for females and males.

Table 1. Summary Statistics for Variables Included in Logistic Regression Models by Gender

Variables	**Females N= 1899**		**Males N = 1521**	
	Mean/Mode	Standard Deviation/Range	Mean/Mode	Standard Deviation/Range
Income[m,r]	2	0-5	2	0-5
Religious[m,r]	1	1-5	1	1-5
Soctrst	.03	.67	.02	.68
Tvhrs	2.65	2.50	2.63	2.44
Age	36.74	9.04	37.62	9.51
Kids_5	.55	.75	.67	.90
Commute	.40	.39	.49	.49
PARTTIME[b]	.23	.42	.07	.25
FULLTIME[b]	.77	.42	.93	.26
BLACK[b]	.20	.40	.13	.34
ASIAN[b]	.02	.15	.03	.16
NATIVE[b]	.01	.12	.02	.13
OTHER[b]	.08	.27	.09	.29
WHITE[b,ref1]	.69	.46	.73	.44
SATISFIED[b]	.95	.21	.95	.21
CITIZEN[b]	.96	.19	.93	.26
HOME[b]	.73	.45	.76	.43
POOR[b]	.01	.11	.01	.10
NOTPOOR[b,ref2]	.99	.10	.99	.10
COMMUNITY[b]	.82	.38	.83	.38

[m,r] Modes and ranges were reported since these were technically ordinal variables.
[b] These were assumed to be Bernoulli variables and means and variances were calculated accordingly (see Wasserman, 2005).
[ref1] The reference category for the race/ethnicity variables above it.
[ref2] The reference category for the variable directly above it.

Table 2 contains our key results of interest, the female and male adjusted odds ratios for $COMMUNITY_i$. These ratios tells us how the odds of reporting that one is in poor health for those assigned a 1 for $COMMUNITY_i$ compares with the odds for those assigned a 0.

Table 2. Female and Male Odds Ratios

Variables	**Female Odds Ratio**	**Male Odds Ratio**
Income	.81	.89
Religious	.98	1.23
Soctrst	1.67	2.45
Tvhrs	1.12	1.09
Age	1.03	1.02
Kids_5	1.86	.84
Commute	1.41	1.92
PARTTIME	1.52	2.44
BLACK	.11	.52

ASIAN	3.21	4.27
NATIVE	1.85	4.33
OTHER	.00	1.45
SATISFIED	.36	.25
CITIZEN	$3*10^{7}$	4.54
HOME	.87	.65
COMMUNITY	.39	.69
Constant	.00	.00

For females, Table 2 shows that the adjusted odds ratio was about .39. Thus, the adjusted odds of reporting poor health, if their neighbors give them a sense of community, were about .4 of the adjusted odds for those whose neighbors do not give them such a sense. That is, for women, neighborhood networks seemed to be associated with a decrease in the chance of reporting poor health. For males, Table 1 indicates that the adjusted odds ratio was about .69. Thus, the adjusted odds of reporting that one is in poor health, if their neighbors give them a sense of community, were .7 of the adjusted odds for those whose neighbors do not give them such a sense. Thus, for males also, a sense of community is associated with a decrease in the chance of reporting poor health, although this decrease is smaller than was the case for females.

We should also say a bit about how well our models fit the data. There are many ways to communicate such fit, but we think the most intuitive is to report the accuracy of predictions using our models. One way to think of LRA is as a method of trying to predict membership in a given set. In our case, we were trying to predict whether people were in the set of those who regard themselves to be in poor health or in the set of those who don't. For both our male only and female only models, 99% of our predictions were correct, indicating excellent fit to our data.

Discussion

This paper has been concerned with the association between whether one gets a sense of community from her or his neighbors and the chance that one reports he or she is in poor health and whether this association varied by gender. Using logistic regression analysis to describe patterns in our data, we found an association between a sense of community and the chance of rating oneself in poor health, for both males and females, with the association for females being stronger than that for males. To our knowledge, this is the first paper to explore how the association between a sense of community and self-rated health varies by gender. This finding is important because it may lead to other research that explores how the association between social connection and other measures of social interest may depend on gender. Further research on this issue may deepen our understanding of the roles played by social connection and gender in social life.

References

1. Berk, R.A. **Regression Analysis: A Constructive Critique**, Sage Publications, California, 2004
2. Bramer, M. **Principles of Data Mining**, Springer, London, 2007
3. Cannuscio, C. C., Colditz, G. A. and Rimm, E. B. **Employment Status, Social Ties, and Caregivers' Mental Health**, Social Science and Medicine, vol. 58, no. 7, April 2004, pp. 1247-1256
4. Fiori, K., Antonucci, T. C. and Cortina, K. S. **Social Network Typologies and Mental Health Among Older Adults**, Journal of Gerontology, Series B: Psychological Sciences and Social Sciences, vol. 61B, no. 1, January 2006, pp. 25-32
5. Freedman, D., Pisani, R. and Purves, R. **Statistics,** W.W. Norton & Company, New York, 2007
6. Gayen, K. and Raeside, R. **Social Networks, Normative Influence and health Delivery in Rural Bangladesh,** Social Science and Medicine, vol. 65, no. 5, September 2007, pp. 900-914
7. Gelman, A. and Hill, J. **Data Analysis Using Regression and Multilevel/Hierarchical Models,** Cambridge University Press, New York, 2007
8. Gibbs, B. K., Nsiah-Jefferson, L., McHugh, M. D. and Prothrow-Stith, D. **Reducing Racial and Ethnic Health Disparities: Exploring an Outcome-Oriented Agenda for Research and Policy,** Journal of Health Politics, Policy and Law, vol. 31, no. 1, pp. 185-218
9. Helweg-Larsen, M. Kjoller, M. and Thoning, H. **Do Age and Social Relations Moderate the Relationship Between Self-Rated Health and Mortality Among Adult Danes?,** Social Science and Medicine, vol. 57, no. 7, October 2003, pp. 1237-1247
10. Kleinbaum, D.G. **Logistic Regression: A Self Learning Text,** Springer, New York, 1994
11. Malat, J. R., van Ryan, M. and Purcell, D. **Race, Socioeconomic Status, and the Perceived Importance of Positive Self-Preservation in Health Care,** Social Science and Medicine, vol. 62, no. 10, 2006, pp. 2479-2488
12. McDonough, P. and Walters, V. **Gender and health: Reassessing Patterns and Explorations,** Social Science and Medicine, vol. 52, no. 4, 2001, pp. 547-559
13. Mendes de Leon, C. F., Glass, T. A., Beckett, L. A. **Social Networks and Disability Transitions Across Eight Intervals of Yearly Data in the New Haven EPESE,** Journal of Gerontology, Series B: Psychological Sciences and Social Sciences, vol. 54 B, no. 3, May 1999, pp. S162-S172
14. Ueno, K. **The Effects of Friendship Networks on Adolescent Depressive Symptoms,** Social Science Research, vol. 34, no. 3, September 2005, pp. 484-510
15. Unger, J. B., McAvay, G. and Bruce, M. L. **Variation in the Impact of Social Network Characteristics on Physical Functioning in Elderly Persons: MacArthur Studies of Successful Aging,** Journal of Gerontology, Series B: Psychological Sciences and Social Sciences, vol. 54 B, no. 5, September 1999, pp. S245-S251
16. Veenstra, G., Luginaah, I. and Wakefield, S. **Who You Know, Where You Live: Social Capital, Neighborhood and Health,** Social Science and Medicine, vol. 60, no. 12, June 2005, pp. 2799-2818
17. Wasserman, L. **All of Statistics: A Concise Course in Statistical Inference,** Springer, New York, 2005
18. Zunzunegui, M. V., Kone, A. and Johri, M. **Social Networks and Self-Rated Health in Two French-Speaking Canadian Community Dwelling Populations Over 65,** Social Science and Medicine, vol. 58, no. 10, May 2004, pp. 2069-2081
19. * * * **Social Capital Community Benchmark Survey,** Roper Center, www.ropercenter.uconn.edu/scc_bench.html, retrieved on August 25, 2005

Appendix

This appendix focuses on the covariates in the logistic regression models we discussed in the body of the paper. This is so that researchers who'd like to replicate our findings will know precisely what we did. The table below contains the names of the original variables found in the survey, the original coding, the recoded variable names, and the recodes.

Original Variables/Coding and Recoded Variables/Recodes

Original Variable Name	Original Codes	Recoded Variable Name	Recodes
Wrktime	other = -9; don't know = -7 and 98; refused = -6 and 99; no answer = -5; blank = -4; number of hours, on average, one works in week	PARTTIME	average hours worked/week < 35hrs/week = 1; average hours >= 35 hrs/week = 0; don't know, refused, blank, and no answer not considered
happy	0 = not happy with one's life; 1 = not very happy with one's life; 2 = happy with one's life; 3 = very happy with one's life; 8 = don't know; 9 = refused	SATISFIED	0 and 1 recoded to 0; 2 and 3 recoded to 1; 8 and 9 not considered
Citizen	non-citizen = .00; citizen = 1.00	CITIZEN	non-citizen = 0; citizen = 1
Own	rent home = 0; own home = 1; don't know = 8; refused = 9	HOME	rent home = 0; own home = 1; don't know and refused not considered
race_all	white = 1; black = 2; Asian/Pacific Islander = 3; Alaskan Native/Native American = 4; Other = 5; don't know = 8; refused = 9	[1]BLACK; ASIAN; NATIVE; OTHER; WHITE was the reference group	black = 1 non-black = 0; Asian/Pacific Islander = 1 non-Asian/Pacific Islander = 0; Alaskan Native/Native American = 1 non-Alaskan Native/Native American = 0; Other =1 non-other = 0
income	income in [$0, $20K); in ($20K, $30K); in ($30K, $50K); in ($50K, $75K); in ($75K, $100K); income >= $100K	INCOME	income in [$0, $20K) = 0; in ($20K, $30K) = 1; in ($30K, $50K) = 2; in ($50K, $75K) = 3; in ($75K, $100K) = 4; income >= $100K = 5
Relaten2	< yearly attendance at religious service = .00; a few times/year = 1.00; 1-2 times/year = 2.00; almost weekly = 3.00; weekly or more = 4	RELIGIOUS	< yearly attendance at religious service = 1; a few times/year = 2; 1-2 times/year = 3; almost weekly = 4; weekly or more = 5
[2]soctrst	Index created by survey designers		
Age	measured in years		
kids_5	Number of kids under five for whom one is primary caretaker		
Commute	number of hours it takes to commute to work		
Tvhrs	number of hours of television watched on the average weekday		

[1]These were dummy variables; the coding scheme is 1 for a person who falls in a given category (such as the category of black persons and 0 for those who don't fall in the given category). White persons served as the reference category in our analysis.

[2]Where no variable appears in the third column of the table, this indicates that the original coding was used in our analysis.

---

[1] **Correspondence:** Michael A. Lewis, 129 East 79th Street, New York, NY 10075

[2]Michael Lewis holds a masters degree in social work From Columbia University and a doctorate in Sociology from the City University of New York Graduate Center. He teaches courses in social welfare policy and political economy at the Hunter College School of Social Work. Lewis' research interests are in poverty, social welfare policy, and quantitative methods. His work, some co-authored with Eri Noguchi, has appeared in a number of peer-review journals.

[3]Eri Noguchi holds a masters degree in social work, a masters degree in public administration, and a doctorate in Sociology, all from Columbia University, and currently teaches courses on human behavior, program evaluation, and administration at Columbia University and Hunter College Schools of Social Work. Her research interests include labor markets, poverty, social welfare policy, and civic engagement/participation. She also serves as the Chief Program Officer of the Association to Benefit Children - a New York-based non-profit organization dedicated to providing early childhood, youth, and wrap-around family support services to children and families.

RESEARCH ABOUT IMPLEMENTING E-PROCORD – NEW MEDICAL AND MODELING APPROACHES IN IT&C AGE APPLIED ON CARDIOVASCULAR PROFILE EVALUATION AT MOLECULAR LEVEL[1]

Dan–Andrei SITAR–TAUT[2]
PhD, Business Information Systems Department,
Babes - Bolyai University, Cluj Napoca, Romania

E-mail: dan.sitar@econ.ubbcluj.ro

Loredana MOCEAN[3]
PhD, Business Information Systems Department,
Babes - Bolyai University, Cluj Napoca, Romania

E-mail: loredana_mocean@yahoo.com

Adela–Viviana SITAR–TAUT[4]
PhD Candidate, Cardiology-Rehabilitation Department,
University of Medicine and Pharmacy Iuliu Hatieganu, Cluj Napoca, Romania

E-mail: adela.sitar@umfcluj.ro

Abstract: *This paper offers a new homogeneous approach and intends to provide a starting model with implemented elements useful in medicine and national health policies proposals. The final model will be entitled e-ProCord and it will represent one of the aims for an important research project in the field of cardiovascular evaluation at molecular level.*

Key words: *information system; database; cardiovascular risk; molecular; data minig*

1. Introduction

The EU is spending 50 million Euro annually from 2003 to 2008 to improve the collection of data, the exchange of information and to offer more information about primary and secondary prevention. The European Commission's major public health goals are: contribute to reducing the incidence of major diseases in the EU; contribute to the development of more effective and efficient health systems; providing medical information and analysis to support these goals. [1][5]

The e-Health project is one of the main priorities of the European Union. E-Health means better health as well as better ways of preventing illness through information

technology and communications (ITC). The e-Health plan of action explains the use of these technologies in order to provide better and cheaper health services while reducing waiting times and human error throughout Europe. The European plan is to develop electronic systems to store data regarding the population's health, identifying patients and last but not least broadband Internet transmissions.

We consider that at least part of the European Commission priorities for public health can be covered by the proposed project, considering the fact that it is attempting the identifying of various risk factors, the way in which these interact for the favoring for the appearance of clinical manifestations belonging to cardiovascular diseases, the way in which they can be influenced. We intend to create and validate a new cardiovascular risk scoring including not only the classic risk factors, but also markers of the endothelial dysfunction, that can be use as platform to create new policies of prevention.

The modern technologies will be applied to allow – through the use of WEB instruments – the anytime, anywhere secure access according to user categories to the available information. We will decide for an open source platform using MySQL as database server, and PHP as scripting programming language. We will use these technologies because they are free of charge. The software applications are cross-platform, they can be used not only on Windows operating systems, but also on Unix or Linux based ones. The database will suffer a solid design, using normalization as modeling tool. Normalization requires splitting the existing relations into atomic ones, which are free of redundancy and updating anomalies. It is also an answer for data confidence and security. The initial launch of the web site has more like a documentary and promoting goal; then it will become a powerful tool that can be used not only for our team's purposes, but also to assist the academic and non-academic communities. Mainly, in medicine there are huge amounts of unstructured data. The way how data is handled will be homogenized via XML. Simultaneously, XML will permit massive data updates with almost no effort from user's side. The data mining tools will offer a different modeling approach.

Databases represent an ordinary reality of the modern economic life. We find them everywhere, within prehistoric or advanced stages, simple or complex, small or colossal, locally accessed or by means of network technologies, being present in all fields. They represent collections, or integrated systems, coherent and shared system files [2], whose foundations were laid decades ago.

The relational model emerged as a viable alternative to previous and problematic hierarchical and network data models. The key paper on this revolutionary model [3] was published in ***1970***, the work of the mathematician ***E. F. Codd***. Even now, a large part of his ideas are still applicable, while other was validated in time. The process of logical data modeling is somewhat controversial and still debated. One of the most important aspects – if not crucial ones – of the logical data modeling within relational databases is the **normalization** process. Rigorous or flexible, mathematically sound or flimsily applied, either theoretical or pragmatic, the normalization built up its reputation as both "angel and demon" of the relational database design. It is generally defined as a process of elimination of redundancies within the entities. The normalization process is not part of the relational theory, rather completes it. The literature mentions 10 normal forms (and a few other less familiar), in the following relationship: [4]

1NF ⊆ 2NF ⊆ 3NF ⊆ EKNF ⊆ BCNF ⊆ 4NF ⊆3,3NF ⊆ 5NF ⊆ 6NF ⊆DK/NF

To compromise orthogonality and its disadvantages due to physical implementation, specialists agree that 3NF or BCNF are sufficient enough. We subscribe to this statement.

The various medical analyses we want to evaluate for our patients requires also one of the three supertype-subtype approaches. We will choose that one fitting better also for future processing (e.g. statistical), requiring minimum structural and content changes.

2. Actual knowledge stage in IT

The **Web** is believed to be *the largest federative databases system* worldwide. Although the Internet has skyrocketed with no apparent connection with the databases technology, nowadays, no relationship between the Web and database technologies is almost unconceivable. The web applications have passed the initial development stage and have evolved to a maturity stage characterized by complex applications. One of the great challenges that appear in this stage is the *communication between the applications and different types of platforms*. In the latest years this communication has become a basic condition for developing future systems, imposed by the necessity of a global environment and correlative virtual space. In these circumstances the interoperability of a system represents an important challenge, concerning the development of web applications and their necessity of interaction and correlation with one another. *XML represents in this context a viable solution for developing some web applications that will interact with informational systems that use different technologies and run on different platforms.*

Thus, XML has imposed as a metalanguage capable of offering the possibility of communication between applications running on different platforms and of creating a virtual correlative environment. XML permits the separation of information from its presentation, reaching in this way to create a self-descriptive document and permits the data transfer among applications. The advantage of this parting is the possibility of using also in other devices that can have internet connections. XML is independent of the platform, extensible and Unicode compliant. The impact of XML technologies can be considered a real evolution in the information technology domain. [5] In general web applications are based on the n-tier architecture in which the layer applications communicate between them, every layer being specialized and serving a certain functionality in the application. The advantages offered by such an architecture, flexibility, distributed components have imposed the n-tier architecture in realizing web applications. To the advantages offered by such an architecture are being added flexibility, independence towards the platform offered by XML documents. So, XML documents are used for transport of information between the layer of an application and towards other web applications no matter of the technologies used or the platforms that they run on. In actual fact it results a flexible, extendible and reusable structure, in which conditions a very important aspect must be taken into consideration: the application and data model can change over time and the advantages of a document structure that adjusts to different changes allows obtaining some important advantages.[6] As a format of database, XML has some advantages: is **self descriptive**, is **portable** (Unicode) and can **describe data in a branched structure**. A **drawback** of this language is the **difficult access to data as a result of the data processing need**.

The facilities offered by XML are:

★data storage (XML documents)

★descriptive language (DTS, XML Schema, etc.)
★query language (Xquery, Xpath, XQL, XML-QL)
★programming interface (SAX, DOM).

However, it has to be taken into account the fact that a series of *advantages* like the storage efficiency, security, indexation, transitioning, data integrity or the multi-user access are imposing the relational databases as a storage base for information [7], the XML documents being used for assuring the transfer of information from the database to different modules of an application or even to different applications.

One of the most prevalent standards in the medical field is Health Level 7 (HL7), which has now reached the 3rd version. Developed into a public-private partnership, the standard enjoys a wide acceptance by the American companies, being also supported by affiliated organizations from over 26 countries, the European ones inclusively. [8] HL7 has imposed as a standard for the electronic support exchange of information in the health field, in the clinical domain, as well as in the administrative one. The major objective of the HL7 standard constitutes the easy exchange of messages between the applications that manage the medical data. HL7 methodology is based on accepting as an event inside the medical procedure determines an exchange of messages between two or more applications. For example, hospitalizing a patient determines collecting information about him/her and transmitting them to other systems. The standard describes a Reference Informational Model (RIM) – fundamental model from which all HL7 messages derive. In 2000 as a component of RIM, CDA (the architecture of medical documents) – has been approved – as a set of XML specifications for the exchange of structured medical documents. CDA HL7 version 3 allows through a common form the exchange of medical data related to patients between different subsystems of the same hospital or even of different hospitals. This HL7 standard aims to constitute the base for the universal registration of medical data. The project frame for development (HDF) HL7 version 3 represents an evolved process that tries to develop specifications that facilitate interoperability between health caring systems.

Data mining, also known as knowledge discovery in databases, is the most recent technology for analyzing data, along with OLAP and data warehouses. Data mining refers to solving problems through already existing data present in the databases.

Data mining represents an analytical process that explores a very high number of data in searching for some patterns or relationships between variables, then generalizes these results into a model, formula or branched decision tree, and finally verifies the correctness of the generated model through testing it on the existing data set or of a new one. Initially data mining was a statistical term meaning overusing of data for deducing inferential invalids [9]. Bonferroni's theory warns us that in the situation that there are many possible conclusions some of them may be real only for pure statistics motives, without a physical validity. That is why the necessity of creating a new automatic instrument appeared, that should transform (through a corresponding procedure) different data contained in extra large databases into information and knowledge useful in the observing process as well as in decision making. In the data mining process, the data is stored electronically and the search is made automatically or at least completed with the help of computer [10] Defining for a data mining process is the fact that allows the discovery of some knowledge without prior formulation of some hypothesis. *Through this it is not aimed the check-up, confirmation/information of hypothesis, but it is aimed to discover some unexpected, unintuitive knowledge that can contradict the intuitive perception, being completely unknown*

at the moment of the process realization. The key elements that make Data Mining a distinct form of software are: [11] **Automatic analyze ✹Large or complex data sets.** One of the attractions of data mining is that it makes possible the analysis of extra large data sets in a reasonable period of time. Data mining is also suitable for complex problems that imply relatively small groups of data but in which there are a lot of fields or variables to analyze. **The second choice is the one that characterizes our study ✹The discovery of significant patterns or trends that otherwise would pass unnoticed.**

The objective in data mining is to discover the relationships between data that can provide useful meanings. The data mining instruments can scan databases and can identify patterns, previously hidden in a single step.

A successful example [9] is ✹ Comparing human kinds accomplishing or not a certain condition has allowed the discovery of a multitude of genes that together determine a lot of diabetes cases. This way of extracting knowledge from data becomes important in the moment of building human kind.

The process of data knowledge acquisition comprises the following steps:

✹selecting and processing data;

✹transforming data;

✹performing some methods and techniques for extracting knowledge from data (patterns, models);

✹validation and interpretation of results (consolidating results and obtaining the so called knowledge).

The process of acquisition of knowledge is iterative because, during this process, the mentions steps are executed repeatedly, through recommencing of some of them. Although the method and techniques for extracting knowledge are applied automatically, the process of acquisition of knowledge from data requires human effort.

The IT&C related elements mentioned above will be applied in all the stages of the project. They will be the platform for the advanced medical data processing (the discovery of new relationships between factors, new methods in statistical processing, etc.) in order to measure the cardiovascular risk factors.

3. Actual knowledge stage in medicine

The data accumulated in last years' medical literature outline new very interesting ideas concerning the general cardiovascular diseases, **in particular for the women**:

- Cardiovascular affections are the *main cause of mortality and main cause of hospitalization in both genders*
- Because of the amplitude of cardiovascular diseases (CVD), in 1997, the European Commission has started and published the **European Initiative for Cardiovascular Health**; also, the European Consensus for the European cardiology and preventive medicine has published in 1998 **recommendations and guides for cardiovascular prevention** with the intention of stimulating the development and reviewing of the national guides for CVD prevention. *The prevention and treatment of CVD is a national and international priority matter*.
- **Cardiovascular affections kill more women than men each year;**

♦ There are important differences between the appearance and the presentation way of CVD in women as compared to men (outlined by very large studies at Framingham, SAVE) and **still, there is a false perception that CVD are not a real threat for women.**

♦ Therefore, the women behave differently [12] [13] concerning the affection because of biological risk, a far more active health monitoring for men, studies conducted predominantly on males, the risk attained by the lifestyle, the way of reacting to the workplace, psychological status, because of the way in which they perceive the symptoms and the treatment of the disease [14] and the behavior related to the disease; the women's associated co-morbidities; previous medical assistance, the differences between women's cardiovascular system as compared with that of men's: physiological, in the CVD physiopathology – regarding the constituents of the atherosclerotic plaque, the endothelial dysfunction - as well as differences in the pharmakinetics and pharmadynamics;

♦ *For improving prevention and optimizing the CVD treatment for the women at menopause it is necessary to be conscious of the **cardiovascular risk factors**:*

♦ **age** and the cardiovascular affections history;

♦ **overweight or obesity** (a high corporal mass coefficient is usually associated with the dramatically increase in the cardiovascular risk factors [15];

♦ **arterial hypertension**;

♦ **smoking**;

♦ **the lipid profile** for women is affected by the hormonal status (therefore, the **estrogen** maintains a high level of HDL cholesterol, increase the level of triglycerides and decrease the level of LDL-cholesterol);

♦ **diabetes mellitus** as the enhancing factor for other risk factors (obesity, hypertension, and especially the hypertriglyceridemia, endothelial dysfunction), as well as an independent risk factor;

♦ **Physical activity** – physical inactivity is more frequent in women, being a reverse association between this, the obesity and cardiovascular events.

♦ **Menopause** – despite the increase in life expectancy to almost 80 years, the age of menopause has remained relatively constant. Also, there is an abrupt increase in the rate of mortality for post-menopause women. During a visible span of time, of almost several years, the estrogen deprivation effects appear, as well as the specific modifications for genital, osteoarticular, cardiovascular, psychosomatic, ocular, digestive and neurological systems. Although up to now the **cardiovascular risk determined by the menopause could not be quantified**, the alterations in vascular tonus, in the coagulation – fibrinolysis ratio for the seric lipoproteins are well known aterogenetic modifications. The Framingham study has shown that the **10 year cardiovascular disease incidence for women at post-menopause, 50 – 59 years has been four times greater than that of women at pre-menopause (women of the same age).**

♦ **Other biochemical factors** – like the cellular adhesion molecules (ICAM 1, VCAM 1), endothelin, plasminogen-activator inhibitor 1 (PAI 1), IL 6, MCP 1, E-selectin, P-selectin, myeloperoxidase, Lipoprotein-associated phospholipase A2 (Lp-PLA2), sCD40L, PCR–(far less dosed in our country) are associated with the increase in cardiovascular risk [16]. **Endothelial cells are not inert but rather are highly metabolically active** [17] **The endothelium plays an important role in many physiological functions**, including the control of vasomotor tone, blood cell trafficking, haemostatic balance, permeability, proliferation, survival, and innate and adaptive immunity. The vascular endothelium is a

highly active tissue regulating vascular tone, leucocytes and platelet activation, angiogenesis, inflammatory response, permeability and the metabolism of vascular mediators [17-19]. **There are many factors released by endothelium :** ★ vasoactive substances – **vasodilatating substances (**relaxing factor derived from endothelium, EDHF, prostacyclin, bradykinin, serotonin, P-substance**), but also vasoconstrictive (**endothelin, prostaglandin H2, reactive oxygen species ROS) ★ **growth's modulators – with positive** (Insulin-like growth factor-1– Interleukin 1, growth factor derived from platelets) or **negative influence** (heparin sulphate) ★inflammation's modulators – **adhesion molecule** (Endothelial leukocyte adhesion molecule –ELAM- , intracellular adhesion molecule –ICAM–, vascular adhesion molecule -VCAM-), P-selectin and **antibodies** (major histocompatibility complex) ★ **haemostatic and fibrinolysis factors .** [18,20]. Endothelial dysfunction is generally considered a key initial event in the atherosclerosis process, preceding the development of structural changes in the coronary arteries [19] The endothelial dysfunction (meaning **unbalance between vasodilators and vasoconstrictive factors, and factors produced by endothelial cells or affecting them** [21]) sets in before the atherosclerosis and is an essential factor in it progression. [20]. A current tendency is the early identification of the endothelial dysfunction, completing the characterization of the patients from a cardiovascular risk point of view. The endothelium is not amenable to traditional physical diagnostic maneuvers of inspection, palpation, percussion and auscultation. From a laboratory standpoint, the endothelium continues to elude convenient diagnostic interrogation. **The endothelium has enormous untapped potential as a therapeutic target** [17].

With all this exulted attention related to the link between endothelial dysfunction and atherogenesis, it is well known that the present studies and dates are not enough for a better therapeutically approach.

Due to the different impact of cardiovascular risk factors in men and women, the strategies for prevention should be treated differently for the two genders. [33].

4. Methodology of the research

The fast development of research imposes keeping a balance between the progress of science and aspects related to research ethics, or ethics in general.

The study will follow **more direction**. Subjects will be assigned to 1 of 2 groups:

1. patients **without cardiovascular disease** (men and women). The patients will be followed according to their age (under and over 55 years of age). Will no be included the patients in which we have no capacity of sure exclusion of this condition (using usual tests such as cardiologic exam, ECG, and sometimes, treadmill test or Holter ECG/24 hours).
2. patients **with coronary heart disease** (stable angina, old myocardial infarct, cardiac heart failure)- no matter gender and age.

We will **collect the information** using representative sample as method. The participation of patients to the research activities will be made only **after the individual informative agreement** will be signed**.**

The patients will **be prospectively followed** – successive clinical, paraclinical (classical cardiovascular risk factors, endothelial dysfunction markers), life's quality evaluation. Regarding endothelial dysfunction will be studied proinflamator status,

protrombtic status and vasomotor impairment. The **studies will be made in the same time** (the data from the first study will be used as starting point for the others-relation between cardiovascular risk factors and for the survival free of disease estimation). In the same time we will create the **informatic platform for efficient, secure, and orthogonal management of data**. It will be launched a new site not only for advertising purposes, but also with active communication features. The software that will support our research project are Windows (Vista, XP) as operating systems, MS Office (2007, 2003) package, MySQL, PHP, XML for databases and site.At the end of first medical and QoL related evaluation, we will start the **mathematical model conception and creation** in order to estimate cardiovascular risk score, as a function of endothelial dysfunction. The last part of project is represented by the estimation, on the basis of the previously made model, **of a cardiovascular risk score, taking** into account all these parameters.

The main purpose of the research project is cardiovascular risk evaluation, using new medical and modeling tools, but also creation of a cardiovascular risk score by introducing new variables.

5. Results

The big stages followed in the elaboration of our model are shown in figure 1.

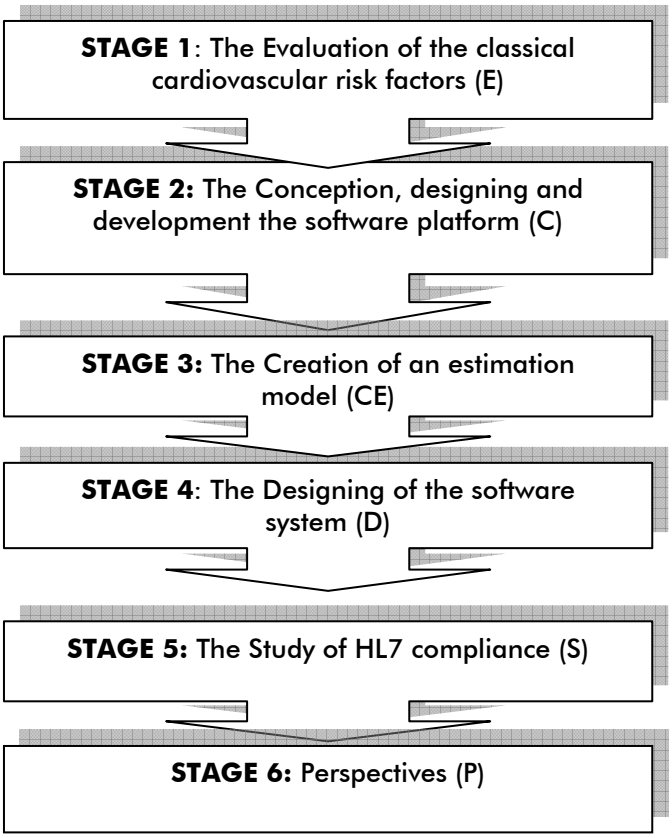


Figure 1. The levels followed in the elaboration of the model

In the following paragraph we detail the objectives followed at each phase:

Stage 1(E): The evaluation of the classical cardiovascular risk factors (overweight/obesity, smoking, dyslipidemia, hypertension, diabetes mellitus), the endothelial dysfunction (intracellular adhesion molecule ICAM 1, vascular adhesion molecule VCAM 1, P-selectin, plasminogen-activator inhibitor 1, endothelin, tumor necrosis factor TNF ά interleukin 6 IL-6) in:

- patients without cardiovascular disease (men and women), under and over 55 years of age;
- patients with coronary heart disease (stable angina, old myocardial infarct, cardiac heart failure)- no matter gender and age. A descriptive, observational and longitudinal study will be run in order to investigate the presence of risk factors, the modifications of the endothelial dysfunction markers.

A prospective, observational will be made in order to investigate cardiovascular risk factors and endothelial dysfunction. We will collect the information using representative sample as method.

Stage 2 (C): The conception, designing and development of a software platform for:

- The control of accuracy and confidentiality of the medical data
- The secure management and controlled disposal of study-specific data
- The promoting, through WEB technologies, of the goals and the achievements of this study
- The achievement of a communication bridge in the national and international medical academic community
- Discovering possible unknown patterns in cardiovascular diseases by using data mining methods or tools, data support and new approaches in modeling, homogenized data integration related to medical issues; Support for updating impressive data amount using almost automated procedures.

Stage 3 (CE): The creation of an estimation model for the cardiovascular risk

Stage 4 (D): The designing of a software system with a role of substitution for the medical experiment (by taking into account the cardiovascular risk score).

On the basis of previous studies will be estimated a cardiovascular risk score by taking into consideration various parameters (personnel and familial disease, used medication, age, smoking, weight, height, body mass index, blood pressure, glycemia, cholesterol, tri-glycerides, HDL, LDL, but also ICAM 1, VCAM 1, endothelin, PAI 1, P-selectin, tumor necrosis factor TNF ά, interleukin 6 IL-6)).

Stage 5 (S): The study of Health Level 7 (HL7) compliance in the perspective of integration in the European virtual space.

Stage 6 (P): Perspectives and new challenges occurrence identification, not only at the fundamental research stage, but also at applicative level.

The start of the project

e-ProCord starts in May 2009. Due to the financial and time constraints this year we will be able to evaluate a smaller patients sample than we proposed before. Also, the most expensive analysis will be delayed. In these circumstances the results will be used as the grounding for the initial modeling purposes considering also the theoretical and previous similar studies' results. The model will be confirmed or infirmed by the future investigations.

On the other hand, the obtained data will be considered as training tests for data mining tools as classifiers, clusterers, and associate makers. In our study we will compare the results provided by classifiers such Decision Table, NaïveBayse, ID3, and C4.5 (J48) [28], with different parameter settings, after data preprocessing and attribute quality evaluation (e.g. InfoGainAttributeEval or GainRatioAtributeEval from Weka) [10]. Then, the results will be compared by calculating the following performance indicators:

True Positive (TP) Rate $\frac{TP}{TP+FN}$ – proportion of patients classified as ill from all the persons who are proved being ill.

True Negative (TN) Rate $\frac{TN}{TN+FP}$ – proportion of patients classified as not being ill from all the persons who are not ill indeed.

The previous indicators show the proportion of the instances classified in a class from all the instances confirmed being in that class by the "gold standard". In medicine, the first one presents more interest (TP rate). It is also called **Sensitivity (Sensibility) –** probability of an ill subject to have a positive test – and the second one is **Specificity**, which is the probability for a not ill subject to have a negative result on examination.

False Positive (FP) Rate $\frac{FP}{FP+TN}$ – indicates how many healthy persons are considered ill using the test

False Negative (FN) Rate $\frac{FN}{FN+TP}$ – how many ill persons are considered healthy by our classifier.

Precision – proportion of the instances from a class which are really classified as that class.

Attributable risk (risk excess) – measures the connection's specificity between risk factor and disease, being calculated as difference between the risk of expose persons and those of non expose person

Success Rate (Accuracy) $\frac{TP+TN}{n}$ (n represents instances) – the number of correct classified instances from the total number of instances

ROC (Receiver-Operating-Characteristics) – represents a diagram on which every point has two coordinates (1-specify, sensibility). The indicator that resumes ROC is area under curve (the bigger values show that investigated test is more valuable, bringing more useful information).

Kappa Statistics (Cohen's Kappa) – measures the agreement between the prediction and reality:

$$K=\frac{(tp+tn)-[(tp+fn)*(tp+fp)+(fp+tn)*(fn+tn)]}{1-[(tp+fn)*(tp+fp)+(fp+tn)*(fn+tn)]},$$

$$tp=\frac{TP}{n}, tn=\frac{TN}{n}, fp=\frac{FP}{n}, fn=\frac{FN}{n}$$

For the beginning, we want to create an initial model, using data from an important study –INTERHEART **[0]**. The main purpose of this was to investigate the

relationship between cardiovascular risk factors and risk of myocardial infarction. In this study (which included approximately 20,000 people, mean age of 58 years), the investigators reported:

- the prevalence of cardiovascular risk factors in controls (subjects without myocardial infarction) and in cases (subjects with myocardial infarction)
- population attributable risk for every cardiovascular risk factor.

We used this information to simulate datasets in order to evaluate the prediction capacity in one subject the disease's presence or absence. Using Monte Carlo simulation, we generated tables with pseudo-random values, according to the risk factor's prevalence and attributable risk and with the respect of independency between attributes. We selected as risk factors current smoking, diabetes mellitus, hypertension, abdominal obesity, no diet with fruits or vegetables, no physical exercise, and no consume of alcohol in small quantities. We considered that a subject without risk factor's presence has an attributable risk equal to one, and a subject in which presence of risk factor is identified has a attributable risk equal with 1 plus the value reported (for the respective risk factor) in INTERHEART (transforming the percentage values in values between 0 and 1). After this, we calculated the individual aggregated attributable risks as a product of attributable risk for each factor. We created four tables with 100, 1,000, and 30,000 (two times) respectively instances. We computed for every dataset – using MedCalc 10.4.0.0 – the capacity of compound attributable risk factor to predict the disease presence in a period of time similar to referred study. The results are presented in Table 1.

Table 1. Testing results

	100 instances	***1,000 instances***	***30,000 instances V1***	***30,000 instances V2***
Sensibility (%)	82.7	58	60	61.4
Specificity (%)	52	71	65	65.5
Cut-off value	1.337	1.55	1.5	1.52
AUC (confidence interval)	0.689 (0.588-0.778)	0.676 (0.646-0.705)	0.675 (0.670-0.680)	0.677 (0.672-0.682)
p	**0.0003**	**0.0001**	**0.0001**	**0.0001**

As we can see, no matter the number of instances generated, the AUC was significantly different from 0.5. A cut-off value more than 1.5 is capable, in approximately 60 percent of the cases, to identify ill people. In approximately 65% of the cases with no disease the test will be negative. If we consider (on the third simulation), a cut-off value equal to 1.91, then the specificity is increasing to 90%.

We also compared the area under curve for different risk factors with those ones obtained for the compound attributable risk. As we can see in Figure 2 and in Table 2, compound attributable risk had a bigger area under curve. In the same time, significant differences were registered between AUC for compound attributable risk and AUCs obtained for others risk factors ($p<0.001$).

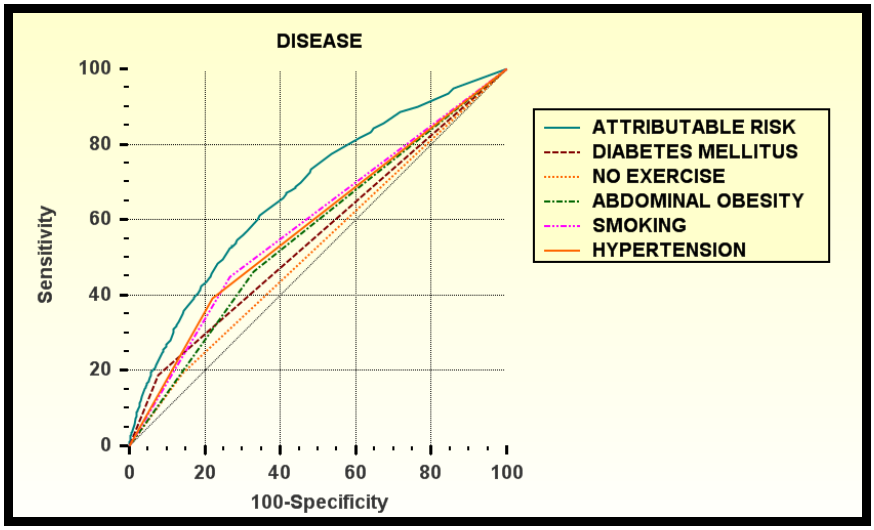


Figure 2. Comparative AUCs

Table 2. Risk factors comparison indicators (30,000 instances V2)

RISK FACTOR	AUC	STANDARD ERROR	CONFIDENCE INTERVAL
ABDOMINAL OBESITY	0.566	0.00332	0.560 to 0.572
DIABETES MELLITUS	0.555	0.00333	0.549 to 0.561
HYPERTENSION	0.586	0.00330	0.580 to 0.592
NO_EXERCISE	0.526	0.00334	0.520 to 0.531
SMOKING	0.590	0.00329	0.585 to 0.596
ATTRIBUTABLE RISK	**0.677**	**0.00311**	**0.672 to 0.682**

This model will be tested using the real data obtained during e-ProCord study.

Discussion

In Romania, not only that compared to the EU countries, it has a **higher mortality rate** due to the cardiovascular affections, but what is even more alarming is the fact that in other countries the rate is decreasing while **in our country is increasing**. We don't have real information about cardiovascular disease evolution and tendency.

National Program regarding Evaluation of Health Status in Romanian Population had as main targets to identify the prevalence of cardiovascular risk factors in general population, to diagnose and to monitories severe diseases in order to avoid premature deaths [30]. What results do we have? The survey showed that almost 15 percents of Romanian people are at risk for cardiovascular diseases development, that cardiovascular risk factors had been identified in a large number of subjects (for example 29.26% of subjects had been diagnosed with diabetes mellitus [31]). Unlike the National Program, which, due to the financial aspects, did not succeed to evaluate in a complete way the cardiovascular risk factors, our project has as the main goal to "heal this deficiency", even if using a small sample. In the same time, we don't have studies about relationships between molecular dysfunction and classical cardiovascular risk factors (depending on age, gender and pathology). Thus we have to try identifying the molecular differences between patients, in order to consider also the classical risk factors, to be able to implement an efficient preventive policy.

Conclusions

So, in consideration with Romanian presented situation, we need "aggressive" measures for cardiovascular disease's prevention, evaluation, control, and research projects like e-ProCord representing just one of the firs steps made in this field.

References

1. Aird, W.C. **Endothelium in health and disease,** Pharmacological reports, 2008, pp. 139-143
2. Albert, M., Glynn, R., Buring, J. E. and Ridker, P. M. **Relation between soluble intercellular Adhesion Molecule-1, homocisteine and fibrinogen levels and race/ethnicity in women without cardiovascular disease,** 99, Am J Cardiol, 2007, pp. 1246-1251
3. Albert, M.A. and Ridker, P. M. **Reactive Protein as a Risk Predictor - Do Race/Ethnicity and Gender Make a Difference,** Circulation, 114, 2006, pp. e67-e74
4. Anghel, T. **Programarea in PHP. Ghid practic,** Ed. Polirom, Iassy, 2005
5. Balaita, R., Hulea, M. and Olariu, C. **Data Mining - Inteligenta artificiala,** Universitatea Tehnica Gh. Asachi Iassy, Facultatea de Automatica si Calculatoare, 2003, Biblioteca virtuala de Inteligenta artificiala, http://eureka.cs.tuiasi.ro/~fleon/bvia.htm
6. Benıtez, M.B. *et all*. **Endothelial and leukocyte adhesion molecules in primary hypertriglyceridemia,** Atherosclerosis, 2007
7. Campisi, R. **Noninvasive assessment of coronary microvascular function in women at risk for ischaemic heart disease,** Int J Clin Pract, 62, 2, 2008, pp. 300–307
8. Codd, E. F. **A Relational Model of Data for Large Shared Data Banks,** CACM 13, 1970
9. Cooper, B., Burger, N., Toth, M., Cushman, M. and Sites, C. **Insulin Resistence with hormone replacement therapy: associations with markers of inflammation and adiposity,** Am J obstet Gynecol, 196, 2007, pp. 123.e1-123e7
10. Daconta M.C., Obrts, L.J. and Smith, K.T. **The Semantic Web - A Guide to the Future of XML,** Web Services and Knowledge Management, Wiley Publishing, 2003
11. Davis, S.R., Dinatale, I., Rivera-Woll, L. and Davison, S. **Obese women and the relation between cardiovascular risk profile and hormone therapy, glucose tolerance, and psychosocial conditions,** Diabetes Care, 29(11), 2006, pp. 2477-2482
12. De Ferrari, L. and Aitken, S. **Mining housekeeping genes with a Naive Bayes classifier,** BMC Genomics, 2006, 7:277
13. DeSouza, C. A., Dengel, D. R., Macko, R. F., Cox, K. and Seals, D.R. **Elevated Levels of Circulating Cell Adhesion Molecules in Uncomplicated Essential Hypertension,** Am J Hypertens, 10, 1997, pp. 1335–1341
14. Erl, Th. **Service-Oriented Architecture, A field guide to Integrating XML and Web Services,** Prentice Hall PTR, 2004
15. Kublickiene, K. and Luksha, L. **Gender and the endothelium,** Pharmacological reports, 60, 2008, pp. 49-60
16. Marchesi, S., Vaudo, G., Lupattelli, G., Lombardini, R., Roscini, A.R., Brozzetti, M., Siepi, D. and Mannarino, E. **Fat distribution and endothelial function in normal-overweigh menopausal women,** Journal of Clinical Pharmacy and Therapeutics, 32, 2007, pp. 477–482
17. Milner, K.A., Funk, M., Richards, S., Wilmes, R.M., Vaccarino, V. and Krumholz, H.M. **Gender differences in symptom presentation associated with coronary heart disease,** Am J Cardiol, 84, 1999, pp. 396-399
18. Nitchi, St. and Racovitan, D.M. **Bazele prelucrarii informatiilor si tehnologie informationala,** Ed. Intelcredo, Deva, 1996
19. Obrenović-Kirćanski, B. **Endothelial dysfunction reversibility,** Vojnosanit Pregl, 64 (5), 2007, pp. 337-343
20. Omran, S. and Al-Hassan, M. **Gender differences in signs and symptoms presentation and treatment of Jordanian myocardial infarction patients,** Int J Nurs Pract, August, 12 (4), 2006 pp. 198-204

21. Sitar-Taut, D.A. **Baze de date distribuite,** Ed. RISOPRINT, Cluj Napoca, 2005
22. Stollberger, C. **Gender and mortality in myocardial infarction -more similarities than differences,** Eur. Heart J, 22 (4), 2001, pp. 279 –80
23. Sumino, H., Ichikawa, S., Ohyama, Y., Takahashi, T., Saito, Y., Nakamura, T., Kanda, T. and Kurabayashi, M. **Effect of transdermal hormone replacement therapy on the monocyte chemoattractant protein-1 concentration and other vascular inflammatory markers an don endotelial function in postmenopausal women,** Am j Cardiol, 96, 2005, pp. 148-153
24. Ullman, J.D. **Data Mining Lecture Notes,** Stanford University, 2003, http://infolab.stanford.edu/~ullman/
25. Vitale, C., Miceli, M. and Rosano, C. **Gender-specific characteristics of atherosclerosis in menopausal women risk factors, clinical course and strategies for prevention,** Climacteric, 10, Suppl 2, 2007, pp. 16–20
26. Witten, I.H. and Frank, E. **Data mining: practical machine learning tools and techniques, 2nd edition,** Morgan Kaufmann series, Elsevier Inc., 2005
27. Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J. and Lisheng, L. **On behalf of the INTERHEART Study Investigators Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study,** Lancet, 364, 2004, pp. 937–952
28. * * * **Health Level 7 (HL7)**, http://www.hl7.org
29. * * * http://www.Europe.eu.int/scadplus/leg/en/lvb/l24226f.htm
30. * * * http://www.ms.ro/fisiere/programe_nationale/13_51_334_h.htm
31. * * * http://www.romedic.ro/diabetul-zaharat-afecteaza-cei-mai-multi-romani-0N5786

[1] **Acknowledgement**
The current research was financed from The National University Research Council*, project no.2246/2008, entitled "E-PROCORD – New Medical and Modeling Approaches in IT&C Age Applied on Cardiovascular Profile Evaluation at Molecular Level. Differences Implied by Gender, Age, and Existing Pathology"**, IDEI Program, PNCDI PN II.
*CNCSIS – Consiliul Național al Cercetarii Stiintifice din Învatamantul Superior (in Romanian)
**E-PROCORD – Noi abordari medicale si de modelare in era TI & C in evaluarea profilului cardiovascular la nivel molecular. Diferente determinate de sex, varsta si patologie prezenta (in Romanian)

[2] Dan-Andrei SITAR-TAUT has a Bachelor's degree in Business Information Systems from Faculty of Economics and Business Administration, Babeş-Bolyai University of Cluj-Napoca and a Master's degree in Informatics Strategies Applied in Economy and Business from the same educational institution. He also holds a PhD diploma in Cybernetics and Economic Statistics.
He is the author of 2 books, about 30 papers in the field of Databases, ERP, Data mining, and Web.

[3] Loredana MOCEAN has graduated Babes-Bolyai University of Cluj – Napoca, the Faculty of Computer Science in 1993, she holds a PhD diploma in Economics from 2003 and she had gone through didactic position of assistant and lecturer, since 2000 when she joined the staff of the Babes-Bolyai University of Cluj- Napoca, Faculty of Economics and Business Administration.
She is the author of more than 10 books and over 35 journal articles in the field of Databases, Data mining, Web Services, Web Ontology, ERP Systems and much more.

[4] Adela-Viviana SITAR-TAUT has a Bachelor's degree in Business Information Systems from Faculty of Economics and Business Administration, Babeş-Bolyai University of Cluj-Napoca and a Bachelor's degree in General Medicine from "Iuliu Haţieganu" Medicine and Pharmacy University from the same city. Currently she is a resident medical staff in Internal Medicine and also Ph. D. candidate in Cardiology.
She participated and published papers to many national and international events on Medicine, Business Information Systems, and Bioinformatics related topics.

[5] Codification of references:

[1]	* * * http://www.Europe.eu.int/scadplus/leg/en/lvb/l24226f.htm
[2]	Nitchi, St. and Racovitan, D.M. **Bazele prelucrarii informatiilor si tehnologie informationala,** Ed. Intelcredo, Deva, 1996
[3]	Codd, E. F. **A Relational Model of Data for Large Shared Data Banks,** CACM 13, 1970
[4]	Sitar-Taut, D.A. **Baze de date distribuite,** Ed. RISOPRINT, Cluj Napoca, 2005
[5]	Daconta M.C., Obrts, L.J. and Smith, K.T. **The Semantic Web - A Guide to the Future of XML,** Web Services and Knowledge Management, Wiley Publishing, 2003

[6]	Erl, Th. **Service-Oriented Architecture, A field guide to Integrating XML and Web Services,** Prentice Hall PTR, 2004
[7]	Anghel, T. **Programarea in PHP. Ghid practic,** Ed. Polirom, Iassy, 2005
[8]	* * * **Health Level 7 (HL7)**, http://www.hl7.org
[9]	Ullman, J.D. **Data Mining Lecture Notes,** Stanford University, 2003, http://infolab.stanford. edu/~ullman/
[10]	Witten, I.H. and Frank, E. **Data mining: practical machine learning tools and techniques, 2nd edition,** Morgan Kaufmann series, Elsevier Inc., 2005
[11]	Balaita, R., Hulea, M. and Olariu, C. **Data Mining - Inteligenta artificiala,** Universitatea Tehnica Gh. Asachi Iassy, Facultatea de Automatica si Calculatoare, 2003, Biblioteca virtuala de Inteligenta artificiala, http://eureka.cs.tuiasi.ro/~fleon/bvia.htm
[12]	Milner, K.A., Funk, M., Richards, S., Wilmes, R.M., Vaccarino, V. and Krumholz, H.M. **Gender differences in symptom presentation associated with coronary heart disease,** Am J Cardiol, 84, 1999, pp. 396-399
[13]	Stollberger, C. **Gender and mortality in myocardial infarction -more similarities than differences,** Eur. Heart J, 22 (4), 2001, pp. 279 –80
[14]	Omran, S. and Al-Hassan, M. **Gender differences in signs and symptoms presentation and treatment of Jordanian myocardial infarction patients,** Int J Nurs Pract, August, 12 (4), 2006 pp. 198-204
[15]	Davis, S.R., Dinatale, I., Rivera-Woll, L. and Davison, S. **Obese women and the relation between cardiovascular risk profile and hormone therapy, glucose tolerance, and psychosocial conditions,** Diabetes Care, 29(11), 2006, pp. 2477-2482
[16]	Albert, M.A. and Ridker, P. M. **Reactive Protein as a Risk Predictor - Do Race/Ethnicity and Gender Make a Difference,** Circulation, 114, 2006, pp. e67-e74
[17]	Aird, W.C. **Endothelium in health and disease,** Pharmacological reports, 2008, pp. 139-143
[18]	Kublickiene, K. and Luksha, L. **Gender and the endothelium,** Pharmacological reports, 60, 2008, pp. 49-60
[19]	Campisi, R. **Noninvasive assessment of coronary microvascular function in women at risk for ischaemic heart disease,** Int J Clin Pract, 62, 2, 2008, pp. 300–307
[20]	Obrenović-Kirćanski, B. **Endothelial dysfunction reversibility,** Vojnosanit Pregl, 64 (5), 2007, pp. 337-343
[21]	Vitale, C., Miceli, M. and Rosano, C. **Gender-specific characteristics of atherosclerosis in menopausal women risk factors, clinical course and strategies for prevention,** Climacteric, 10, Suppl 2, 2007, pp. 16–20
[22]	Albert, M., Glynn, R., Buring, J. E. and Ridker, P. M. **Relation between soluble intercellular Adhesion Molecule-1, homocisteine and fibrinogen levels and race/ethnicity in women without cardiovascular disease,** 99, Am J Cardiol, 2007, pp. 1246-1251
[23]	Sumino, H., Ichikawa, S., Ohyama, Y., Takahashi, T., Saito, Y., Nakamura, T., Kanda, T. and Kurabayashi, M. **Effect of transdermal hormone replacement therapy on the monocyte chemoattractant protein-1 concentration and other vascular inflammatory markers an don endotelial function in postmenopausal women,** Am j Cardiol, 96, 2005, pp. 148-153
[24]	Cooper, B., Burger, N., Toth, M., Cushman, M. and Sites, C. **Insulin Resistence with hormone replacement therapy: associations with markers of inflammation and adiposity,** Am J obstet Gynecol, 196, 2007, pp. 123.e1-123e7
[25]	Marchesi, S., Vaudo, G., Lupattelli, G., Lombardini, R., Roscini, A.R., Brozzetti, M., Siepi, D. and Mannarino, E. **Fat distribution and endothelial function in normal-overweigh menopausal women,** Journal of Clinical Pharmacy and Therapeutics, 32, 2007, pp. 477–482
[26]	Benıtez, M.B. *et all.* **Endothelial and leukocyte adhesion molecules in primary hypertriglyceridemia,** Atherosclerosis, 2007
[27]	DeSouza, C. A., Dengel, D. R., Macko, R. F., Cox, K. and Seals, D.R. **Elevated Levels of Circulating Cell Adhesion Molecules in Uncomplicated Essential Hypertension,** Am J Hypertens, 10, 1997, pp. 1335–1341
[28]	De Ferrari, L. and Aitken, S. **Mining housekeeping genes with a Naive Bayes classifier,** BMC Genomics, 2006, 7:277
[29]	Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J. and Lisheng, L. **On behalf of the INTERHEART Study Investigators Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study,** Lancet, 364, 2004, pp. 937–952
[30]	* * * http://www.ms.ro/fisiere/programe_nationale/13_51_334_h.htm
[31]	* * * http://www.romedic.ro/diabetul-zaharat-afecteaza-cei-mai-multi-romani-0N5786

RELIABILITY SAMPLING PLANS: A REVIEW AND SOME NEW RESULTS

Alexandru ISAIC-MANIU

PhD, University Professor, Department of Statistics Econometrics
University of Economics, 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[1]

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 work we present a large area of aspects related to the problem of sampling inspection in the case of reliability. First we discuss the actual status of this domain, mentioning the newest approaches (from a technical view point) such as HALT and HASS and the statistical perspective. After a brief description of the general procedure in sampling inspection, we offer what we did call here as „personalized procedures": this means that we take into consideration the specific statistical law for time-to-failure.*
The original part refers to the construction of (n, 0) sampling plans and the use of ISO standard 2859 (MILSTD 105 E, the American original) in order to derive sampling plans by linking the AQL indicator (a defective fraction, in fact) to the hazard rate function.
Some examples are given and necessary tables are provided also.

Key words: *reliability sampling plans; AQL; HALT; HASS; personalized procedures; (n,0) – plans; power distribution; Rayleigh law*

1. Preliminaries: the actual status of the problem

Reliability and its main component – durability – are considered as **dynamical** quality characteristics in the sense that possible nonconformities of the underlying entities (products, components, systems) are put into light in the process in which these ones perform their functions. For instance, we may observe the exploitation of the object by a client or may organize specifical lab durability/reliability tests.

A truck company for instance, records the mileage of its „moving units" up to their first failure and these records may constitute a database for the reliability analysis of that peculiar truck brand.

Laboratory tests are more specific since the experimenter must choose a certain strategy for them: to use complete or (multi) censored samples, a time-truncation procedure (all performed under normal or accelerated conditions) a.s.o.

The cost element must be involved too in such test designs.

A huge monograph of Blischke and Murthy (2000, [3][2]) describes new research directions in this field and provides a detailed list of ISO, NATO and from other professional associations (e. g. SAE – Society of Automotive Engineers, U.S.A.) documents regarding reliability and related domains (see also an earlier work of Kochendarfer and Pabst, Jr. (1971) [18]).

It is important to mention that in the TQM (Total Quality Management) frame work or in the recent SIX SIGMA approach (see Gupta, 2004 [12]) reliability is often considered as a specific technical aspect since:

(i) does not appear directly in the manufacturing process;

(ii) is imbedded in the design stage of an entity in a provisional form;

(iii) the acceptance (receiving inspection) of final products is usually performed on the so-called static characteristics such as length, volume, pressure, voltage, hardness a.s.o.;

(iv) to test and to evaluate reliability requires an adequate logistics and a documented statistical know-how (that is a qualified personnel).

It is also necessary to remember that most of the complex products generated by nowadays industries are „mecatronic entities" – that is their structure is a combination between mechanical, electrical and electronic parts each one with its own degree of automation and self-adjustment: this structural and functional complexity needs high engineering techniques for testing as well as an appropriate methodology for data analysis since the components of a mecatronic system have their specific behaviour in exploitation.

In this sense, the classical techniques – such as running in of various vehicles at the highest level of stress intensity allowed by specifications – are combined now with two new procedures known as HALT and HASS. Proposed by dr. Gregg Hobbs in 1995 at QUALMARK Corp. in Santa Clara (California) and registered as a **federal mark** these represent advanced methodologies for accelerated life-testing in order to determine as quick as possible the item reliability: this „quick evaluation" is needed to evaluate the possibility of a rapid intervention in the design stage, the goal being functional improvement (see also Dowling, 1999, [7]).

HALT or **H**ighly **A**ccelerated **L**ife **T**est means an experiment in which the objects are submitted to a highly accelerated test in order to estimate the durability of that kind of object that is its failure-free working period.

The method is used in the design stage in order to detect weak points of the intended item to by mass produced, such as those probable nonconformities be eliminated before launching: the corrective action consists here in **re-designing** the product.

HASS – or **H**ighly **A**ccelerated **S**tress **S**creening means an experiment in which an item is submitted under accelerated stress in the manufacturing stage in order to detect and eliminate the so-called „hidden defects", thus, preventing them to be genetically transmitted by the process to future batches.

The elements (stresses) used in HALT and HASS procedures are rapid transitions of temperature and omni axial vibrations (with six degrees of freedom) – this combination of random vibrations and thermo-cycles being the most performing way to accelerate the failure mechanisms of any kind.

The idea to combine various types of stresses in not new: the engineers working in the „resistence des materiaux" (T. Albert, 1829 and J. V. Poncelet, 1839 in France, P. Hodge, 1850 in England or A. Z. Wöhlen and Z. Bauwesen, 1858, in Germany) have used this combination of random mechanical stresses with various temperatures in order to investigate the durability of metallic materials employed in railroad industry (details in Cioclov, 1975 [5] page 5 – 10).

Some authors believe that this study of material strength face to various random mechanical and thermo stresses can be considered as **the embryo of the reliability theory** (see Bârsan-Pipu et al, 1999, [2]).

A last remark in this context: accelerated tests are needed especially if we deal with very high reliable objects for which is almost impossible to wait their natural failure in order to evaluate numerically the underlying reliability. The exclusive usage of complete samples and normal conditions has proved to be in many cases time consuming, extremely expensive or even not very significant from a technical viewpoint (examples of such pitfalls are given in Meeker and Hamada, 1995, [20]).

As regarads the receiving inspection of products for which the characteristics of interest is durability/reliability, we distinguish several approaches which are quite different since are based on distinct principles. These are:

a) Attributive approach: viewed as the oldest and simplest one, it considers the durability/reliability as a measurable characteristic which can be attributivisable (an **attribute**, as for instance conforming or nonconforming). In our framework, we submit to a reliability test a witness batch of size N_0 (non-repairable items, for example), during a fixed period of time (T_0); then, record the number (d) of failed elements over that period and compute the estimated defective fraction $\hat{p} = d/N_0$ of the lot. This value $(\hat{p})$ may by taken as the desired AQL needed by the ISO standard 2859 or its American variant MIL STD 105 E in order to employ for instance single sampling plans $(n, A \mid R = A + 1)$, where A and R are acceptance and rejection numbers (see Iliescu, 1982, [13]).

The major advantage of attributive method is its simplicity: one has to compare the number (d) of failed items in a sample of size (n) – this one given by the standard, with the acceptance number A. If $d \le A$ the lot is accepted – otherwise, that is $d \ge R = A + 1$, the lot is rejected (this is the case of single sampling plans).

The main disadvantage in this case is the reductionism itself of the method: it does not take into account the specific law for **failure behaviour**, which is a key element in **reliability evaluation** (Cătuneanu – Mihalache, 1990, [4]).

As a consequence, the so-called personalized procedures have been devised and these acceptance schemes use effectively the statistical distribution of time-to-failure.

b) Average operating time approach: this procedure takes into consideration a specific time-to-failure law and establishes acceptable and unacceptable values for mean operating time. Using a sample of size (n) and an acceptable number A (which may be a minimum member of hours of failure-free operation, considered acceptable), the lot will be accepted or rejected in comparison with this number.

c) Hazard rate approach: the procedure is similar to the above one, the fixed values being of hazard (failure) rate.

d) Sequential approach: we have in mind Abraham WALD'S procedure which can be applied in both cases – attributive and variable ones (see Wald, 1973, [23]).

In this paper we shall propose another procedure, namely linking the fraction defective (p) with the hazard rate, obviously taking into account the specific form of time-to-failure distribution.

2. The general procedure

Attribute sampling inspection may be performed regardless the very nature of quality characteristic tested: static or dynamical one. There are considered two proportions P_1 and P_2 of product units, first acceptable with $1-\alpha=0,95$ probability (that is $\alpha=0,05$) and the second one acceptable with a lot smaller probability, usually $\beta=0,10.$ We may have also $\alpha=\beta$. The problem consists in determining the sampling plan (n, A) where n is the sample size and A is the acceptance number. To find n and A we proceed as follows:

(i) from the relationship (see Baron et al, 1988, [1]).

$$\frac{P_2}{P_1}=\frac{\chi^2_{1-\beta;m}}{\chi^2_{\alpha;m}} \qquad (1)$$

where $\chi^2_{\varepsilon;m}$ is the ε - quantile of the chi-square distribution with m degrees of freedom (see [13]), we approximate „the best m" for which (1) is fulfilled.

(ii) the acceptance number A is given by

$$A=\left[\frac{m}{2}-1\right], \text{ where [ ] is the integer part} \qquad (2)$$

(iii) the sample size (n) is furnished as

$$n=\frac{\chi^2_{\alpha;\,2(A+1)}}{2\,P_1} \text{ (rounded to the nearest integer)} \qquad (3)$$

Example: if we take P_1 = 0,25% and P_2 = 1,75% and $\alpha=\beta=0,05$, based on chi-square tables, we get

A	1	2	3	4	5	6	7	8
P_2/P_1	13	7,5	5,7	4,6	4,0	3,6	3,3	3,1

Since P_2/P_1 = 1,75/0,25 = 7, hence the nearest value is 7,5 for which we extract A = 2. If m = 2 (A + 1) = 6, we shall obtain

$$n=\frac{\chi^2_{0.05;6}}{2\times 0.0025}=\frac{1.635}{0.005}=327 \text{ (units)} \qquad (4)$$

Therefore, the sampling plan is (n = 327, A = 2) which seems to be not very economical (since n is quite large) for an expensive testing.

3. Personalized procedures

We shall consider now the time-to-failure model, namely the mathematical object

$$\left\{T\middle|f(t;\theta),\ t\in[0,+\infty),\ \theta=(\theta_1,\theta_2,\ldots\theta_n),\ \theta_j\in R,\ j=\overline{1,k}\right\} \quad (5)$$

we have $f(t;\theta\geq 0,\ \forall t\geq 0),\quad \int_0^{\infty} f(t;\theta)dt=1.$

Here, T is the continuous random variable which represents the time-to-failure behaviour. The simplest model is considered:

3.1. The exponential model

This is described by the density

$$T: f(t;\theta)=(1/\theta)\exp(-t/\theta),\quad t\geq 0,\quad \theta>0 \quad (6)$$

where θ has the significance of average durability, since $E(T)=\theta$ and its inverse is just the hazard rate $h(t;\theta)=f(t;\theta)/R(t;\theta)$ where $R(t;\theta)=1-F(t;\theta)=\exp(-t/\theta)$ is the reliability function.

In this case we shall fix two values for E(T), namely θ_1 (acceptable mean-life with $1-\alpha$ associated probability, $0<\alpha<1$) and θ_2 (undesirable mean-life with β associated probability $0<\beta<1$) where usually $\alpha=0{,}05$ and $\beta=0{,}10$.

If we will employ an „r out of n" durability test, that is we will obtain a censored sample

$$t_{1n}\leq t_{2n}\leq\ldots\leq t_n \qquad r<n \quad (7)$$

the average life estimate is

$$\hat{\theta}_{rn}=\frac{1}{r}\left[\sum_{i=1}^{r} t_{in}+(n-r)t_{rn}\right] \quad (8)$$

(see [1], vol. I, page 532).

The acceptance numbers are as below

$A_1=\dfrac{\theta_1\cdot\chi^2_{\alpha;2r}}{2r}$, if we adopt producer's risk variant or

$A_2=\dfrac{\theta_2\cdot\chi^2_{1-\beta;2r}}{2r}$, if we prefer customer's risk variant.

The decision to accept the lot is taken if $\hat{\theta}_{rn}\geq A_1$ or $\hat{\theta}_{rn}\geq A_2$.

Example: Let $\theta_1=1000$ (hours), $\alpha=0{,}05$ and r = 4 when the sample size in n = 26 units (we work using producer's variant). From chi-square tables ([1]) we find $\chi^2_{0.05;8}=2.733$ and hence $A_1=1000\times 2.733/8\approx 342$.

Suppose that the average mean-life was $\hat{\theta}_{rn}=960$ (hours). Since $\hat{\theta}_{rn}>A_1$, the underlying lot is accepted with confidence level of 95%.

3.2. The Weibull model

Proposed in 1951 (see [24]) by a Swedish military engineer, Wallodi WEIBULL (1887 – 1979), this model is considered as a generalization of exponential and Rayleigh laws (see Isaic-Maniu, 1983, [14]). Its density is

$$T: f(t;\theta,k) = (k/\theta)t^{k-1}\exp\left(-t^k/\theta\right), \quad t \geq 0, \quad k > 0 \tag{9}$$

and has the following average value

$$E(T) = \theta^{1/k} \cdot \Gamma(1+1/k) \text{ where from } \theta = \left[\frac{E(T)}{\Gamma(1+1/k)}\right]^{1/k} \tag{10}$$

Here $\Gamma(u) = \int_0^{\infty} e^{-t} \cdot t^{u-1} dt$ is the well-known Gamma function (see Dorin et al, 1994 [6], page 239 – 244).

If $p = p_0$ is the proportion of nonconforming objects, then

$$p_0 = 1 - \exp\left\{-\left[\frac{t \cdot \Gamma(1+1/k)}{E(T)}\right]^k\right\} \tag{11}$$

where θ has been replaced by its expression from (10).

Hence, we deduce

$$\frac{t}{E(T)} = \frac{[-\ln(1-p_0)]^{1/k}}{\Gamma(1+1/k)} \tag{12}$$

and taking $t = T_0$ as testing time (expressed usually in hours) we can compute $T_0/E(T)$ if k and p_0 are known.

As it has been shown in [22, page 112] this approach leads often to large sample sizes, which is not always convenient.

This handicap could be eliminated by constructing (n, 0) type sampling plans, that is plans where the acceptance number is zero.

3.3. The Gamma model

A random variable T has a Gamma density function if

$$T: f(t;\theta,k) = \frac{1}{\theta\Gamma(k)} \cdot \left(\frac{t}{\theta}\right)^{k-1} \exp\left(-t^k/\theta\right), \quad t \geq 0, \quad \theta, k > 0 \tag{13}$$

It is also a generalization of the exponential one (for k = 1), we get $f(t;\theta,1) = (1/\theta)\exp(-t/\theta), \quad t \geq 0, \quad \theta > 0$).

The average life is in this case $E(T) = k\theta$. If the shape parameter k is known, then if we fix an acceptable mean-value $k\theta_0$ and a testing time T_0, such that with a given probability P, the average durability be al least $k\theta_0$, using the ratio $T_0/k\theta_0$ and P, one may deduce the needed sample size to perform the test (details in Gupta and Groll, 1961 [11] or Vodă, 1981 [22], page 116 – 119).

Example: Suppose that we give A = 0 (the acceptance number) P = 0.95, k = 2 and the average-life $k\theta_0 = 10.000$ (hours) – that is $\theta_0 = 5.000$. If T_0 (testing time) is 1.000 hours, then $T_0/k\theta_0 = 0.10$. From the below table (reproduced from [22] page 119) we detect n = 170 units. Therefore

a) a sample of n = 170 item are submitted to a test over the period of T_0 = 1.000 hours;

b) if there are no failures during this testing period, the lot is accepted;

c) if there exists at least one failure, the lot is rejected, since the acceptance number is zero.

$T_0/k\theta_0$ / 4	1.0	0.05	0.10	0.05
0	4	10	170	639
1	6	16	269	1013
2	9	22	358	1444
3	11	27	411	1655

4. (n, 0) – type sampling plans

These plans, no matter which are the other input elements are invariantly based on the zero acceptance criterion, that is always A = 0.

A way to construct such plans is to fix in advance the following elements: testing time T_0, lower bound for reliability R(t) evaluated for $t = T_0$ or a lower bound for the average lifetime E(T) and the consumer risk β.

This procedure has been described in a series of research reports of SVÚSS (Státny Výzkumný Ústav pro Stavbu Strojů/Běchovice, ČSSR – State Research Institute for Machine Construction/ Běchovice, former Czechoslovakya – see Drimlová (1970 [8], 1973 [10]) and Drimlová and Žaludová (1971) [9]) and it was applied for the exponential law. In this case a lower bound for the reliability $R(t;\theta) = c^{-t/\theta}, \quad t \geq 0, \quad \theta > 0$ may be easily obtained from a limit fixed for the mean durability, since $E(T) = \theta$.

The sample size is given as

$$n = \frac{-\ln \beta}{\exp(-T_0/\theta_0)} \tag{14}$$

where θ_0 is the acceptable value for E(T).

Example: Take T_0 = 100 (hours) and R_{lower} (100) = 0.90 and $\beta = 0.10$ (or 10%). Since - $\ln\beta \approx 2.302$, we get easily $n \approx 2.3/0.10 = 23$ units. Hence, the sampling plan is (n = 23, A = 0).

For the Weibull law, taking account the results from § 3.2. and § 3.3., we get

$$n = \frac{-\ln \beta}{\left[\frac{T}{E(T_{lower})} \cdot \Gamma(1+1/k)\right]^k} \tag{15}$$

where $E(T)_{lower}$ is the mean lifetime which we wish to accept with β probability. If the sample exhibits a smaller value for E(T) then $E(T)_{lower}$, the lot is rejected with $(1-\beta)$ probability (in this model, the shape parameter k is assumed to be known).

Example: Let us consider a Rayleigh distribution (which is a Weibull one for k = 2 – see for details the present authors 1998 [16]). We wish to reject batches with a mean

durability less than 2000 (hours) with a 90% probability (that is $\beta = 0.10$). The testing time was fixed as T_0 = 500 (hours) and obviously, A = 0.

In the next table, we present some values of (n) for various β and $T_0/E(T)$, using formula (15).

β	$T_0/E(T)$			
	0.1	0.2	0.4	0.8
0.05	382	95	24	6
0.10	293	73	18	5

In our case, $T_0/E(T)$ = 500/2000 = 0.4 and we have for $\beta = 0.10$ the sample size n = 18. units. Therefore the sampling plan is (n = 18, A = 0).

For the power distribution, namely

$$T : F(t;\delta,b) = \left(\frac{t}{b}\right)^{\delta}, \quad 0 \le t \le b, \quad \delta > 0 \tag{16}$$

which is a generalization of the uniform one (which is obtained for $\delta = 1$) or a peculiar case of Sedrakian's one (see Sedrakian, 1968 [21] or the present authors, 1995 [15]):

$$T : F(t;b,c,\delta,k) = 1 - \left[1 - \left(\frac{t-c}{b-c}\right)^{\delta}\right]^{k} \tag{17}$$

where $0 \le c \le t \le b, \delta, k > 0$ (the power form is recovered if c = 0 and k = 1), we have immediately for (16):

$$E(T) = \frac{\delta b}{\delta + 1} \text{ and } Var(T) = \frac{\delta b^2}{(\delta+1)^2(\delta+2)} \tag{18}$$

Hence, if δ is known, if we fix a lower acceptable bound for E(T), we get

$$b_{lower} = \frac{1+\delta}{\delta} \cdot E(T)_{lower} \tag{19}$$

and taking into account the results from [22 page 114], we finally obtain

$$n = \frac{\ln\beta}{\ln\left[1 - \left(\frac{\delta\Delta}{\delta+1}\right)^{\delta}\right]} \tag{20}$$

where $\Delta = T_0/E(T)_{lower}$

In the below table we offer some values for n given some Δ and $\beta = 0.01; 0.05; 0.10$ and $\delta = 3/2; 2$.

α	β	$\Delta = T_0/E(T)_{lower}$			
		0.2	0.4	0.6	0.8
3/2	0.01	54	18	9	5
	0.05	70	23	12	7
	0.10	108	36	18	11
2	0.01	128	31	13	6
	0.05	167	40	17	8
	0.10	256	62	26	13

Example: Assume that $\delta = 2$ and we with to reject batches with a mean lifetime less than 3000 (hours) with a 99% probability (hence $\beta = 0.01$). The testing time T_0 is 600 (hours).

We have $\Delta = 600/3000 = 0.2$ and for $\beta = 0.01$ and $\delta = 2$ we read in the table n = 128. Therefore the plan is (n = 128, A = 0).

5. The use of MILSTD 105 E

As we said in § 1, the attributive sampling in practice makes use of the well-known document MILSTD 105 E or ISO variant ISOSTD 2859 (see Iliescu, 1982, [13, page 162 – 180]).

This standard does not refer to purely reliability elements such as testing time, failure/hazard rate, MTBF (Mean Time Between Failures), EOT (Effective Operating Time) a.s.o. (for specific English acronyms used in reliability theory, see Kovalenko, 1975 [19, page 437 – 466]). The items are simply divided into two classes: conforming and nonconforming/defective ones – no matter what indicator is considered.

In this paragraph we shall construct sampling plans by linking the lot defective fraction (p = AQL) with the specific hazard rate of those objects, using some input elements of MILSTD 105 E such as lot size (N), code-letter (CL) which will lead to the sample size (n).

Our approach assumes that the failure model is of the type

$$f(t;\theta) = \varphi'(t;\theta)\exp[-\varphi(t;\theta)] \tag{21}$$

where $\varphi(t;\theta) \geq 0$ for every $t \geq 0$ and $\theta > 0$ and

$$\int_0^{\infty} f(t;\theta)dt = 1 \tag{22}$$

The form (21) provides the distribution function

$$F(t;\theta) = 1 - \exp[-\varphi(t;\theta)] \tag{23}$$

the defective fraction p being just $p = 1 - e^{-\varphi(t;\theta)}$.

Taking logarithms, we have

$$-\ln(1-p) = \varphi(t;\theta) \tag{24}$$

and since the hazard rate is $h(t;\theta) = f(t;\theta)/R(t;\theta)$

where $R(t;\theta) = 1 - F(t;\theta) = \exp[\varphi(t;\theta)]$, we obtain $h(t;\theta) = \varphi'(t;\theta)$ and therefore (24) becomes

$$-\varphi'(t;\theta)\ln(1-p) = h(t;\theta)\varphi(t;\theta) \tag{25}$$

If we take now $-\varphi(t;\theta) = \theta t^2$, then $h(t;\theta) = 2\theta t, \quad t \geq 0 \quad \theta > 0$ which is just the classical Rayleigh hazard rate (see Bârsan-Pipu et al, 1999 [2, page 91]).

In this case (25) has the form

$$-2\ln(1-p) = t \cdot h(t;\theta)$$

Consider now the simplest situation when we know N (the lot size), T_0 (the testing time – assumed to be the life span of the items) and the acceptable hazard rate $h(t;\theta)$ for t = T_0, expressed in failures per hour.

In order to ease the computations, we present some values of $100\,t \cdot h(t;\theta)$ linked with some **preferential AQL values** listed in MILSTD 105 E (table II A) – see also Kirkpatrick, 1970 [17, page 363]).

Table (*)

AQL (%)	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0
$100\,t \cdot h(t;\theta)$	0.30	0.50	0.80	1.30	2.01	3.02	5.6	8.16

In these instances, the sampling procedure is the following:

1) Knowing N (lot size – let's say N = 930 units) and using the general inspection level II (suggested in most of the cases by the standard – see Table I „Sample size and code letters", [17, page 362]), we find the code letter CL = J; we shall denote **IL** as inspection level.

2) from Table II A „Single sampling plans for normal inspection" (already mentioned) we have to draw from the lot a sample of size n = 80 units which will be submitted to the test.

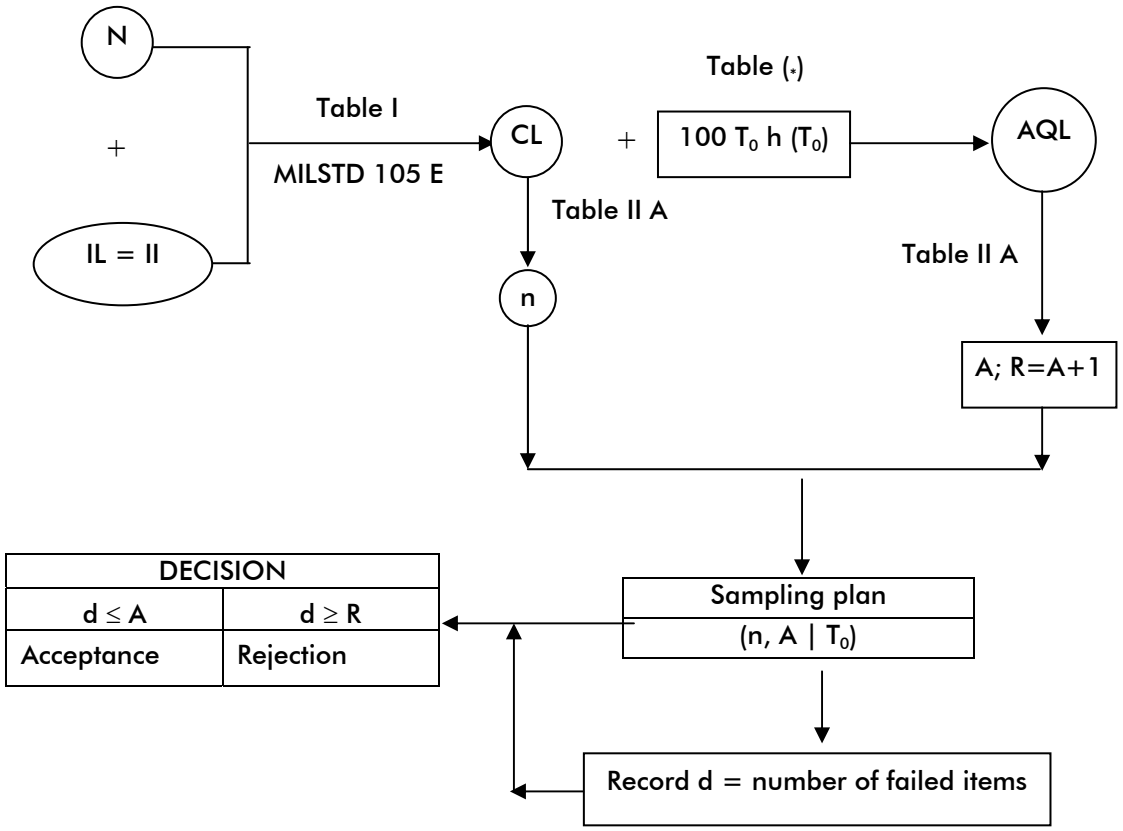
3) assuming that the testing time is T_0 = 1000 (hours) and the acceptable hazard rate for T_0 is h (1000) = 0.000003 failures/hour, we deduce

100 T0 h (T0) = 100 ×1000 × 0.000003 = 3.0

and from the table (*) we see that the nearest value is 3.02 wich indicates AQL = 1.5%.

4) with CL = J and AQL = 1.5%, from the same table II A we read the acceptance number A = 3 (consequently, the rejection one is R = A + 1 = 4).

Therefore, during the testing period T_0 = 1000 (hours) we shall observe the number (d) of failed elements from the sample (n = 80 units). If $d \le A = 3$, the lot is accepted – otherwise (that is $d \ge R = 4$) the lot is rejected. See the below scheme:



References

1. Baron, T., Tovisi, L., Isaic-Maniu, Al. *et all*. **Calitate si fiabilitate. Manual Practic, vol. I – II**, Ed. Tehnica, Bucharest, 1988
2. Barsan-Pipu, N., Isaic-Maniu, Al. and Voda, V., Gh. **Defectarea. Modele statistice cu aplicatii**, Ed. Economica, Bucharest, 1999
3. Blischke, W., R. and Murthy, D., N., P. **Reliability. Modeling, Prediction and Optimization**, John Wiley and Sons Inc., New York, 2000
4. Catuneanu, V., M. and Mihalache, A., N. **Reliability Fundamentals,** Elsevier, Amsterdam, 1989, (Fundamental Studies in Engineering, No. 10 – translated from Romanian by A. N. Mihalache)
5. Cioclov, D. **Rezistenta si fiabilitate la solicitari variabile**, Ed. Facla, Timisoara, 1975
6. Dorin, Al. C., Isaic-Maniu, Al. and Voda, V., Gh. **Probleme statistice ale fiabilitatii**, Ed. Economica, Bucharest, 1994
7. Dowling, N. E. **Mechanical Behaviour of Materials (2nd Edition),** Prentice-Hall, Upper Saddle River, New Jersey, 1999
8. Drimlová, B. **Odhady parametrů některých důležitnych modelů životnosti a spolehlivosti strojirenských výrobků,** SVÚSS, Výzkumná Zpráva, č. SVÚSS (Bechovice), 1970
9. Drimlová, B. and Žaludova, A. **Některé základni přejimaci postupy vzhledem ke spolehlivosti a životnosti strojirenských výrobků,** SVÚSS, Výzkumná Zpráva, č. SVÚSS (Bechovice), 1971
10. Drimlová, B. **Smérnice pro přejimani dilů a zařizeni s ohledem na životnost**, SVÚSS, Výzkumná Zpráva, č. SVÚSS (Běchovice), 1973
11. Gupta, S. S. and Grall, Ph. A. **Gamma distribution in acceptance sampling**, J. Amer. Statist. Assoc. nr. 56, 1961, pp. 942–970
12. Gupta, P. **The Six Sigma Performance Handbook. A statistical Guide to Optimizing Results,** Mc.Graw Hill Book Co., New York (Six Sigma Operational Methods Series, No. 7), 2004
13. Iliescu, D. V. **Controlul calitatii loturilor de produse,** Ed. Tehnica, Bucharest, 1982
14. Isaic-Maniu, Al. **Metoda Weibull. Aplicatii**, Ed. Academiei Republicii Socialiste România, Bucharest, 1983
15. Isaic-Maniu, Al. and Voda, V. Gh. **O noua generalizare a repartitiei putere,** Stud. Cerc. Calc. Econ. Cib. Econ., vol. XXIX, no. 1, 1995, pp. 19– 26
16. Isaic-Maniu, Al. and Voda. V. Gh. **Aspecte privind repartiția Rayleygh**, Stud.Cerc.Calc.Econ.Cib.Econ., vol. XXXII, no. 1, 1998, pp. 5– 13
17. Kirkpatrick, E. G. **Quality Control for Managers and Engineers**, John Wiley and Sons. Inc., New York, 1970
18. Kochendarfer, D. C. and Pabst, Jr., W. R. **MILSTD 883 and related documents**, Journal of Quality Technology, vol. 3, no. 3, 1971, pp. 129– 137
19. Kovalenko, Ye., G. **English – Russian Dictionary of Reliability and Quality Control (about 22.000 terms),** Russian Language Publishers, Moscow, 1975
20. Meeker, W. Q. and Hamada, M. **Statistical tools for the rapid development and evaluation of high – reliability products**, IEEE Transactions on Reliability, R-44, 1995, pp. 187–198
21. Sedrakian, L. G. **K statisticeskoi teorii procinosti. Izdatelstvo Armianskogo Instituta Stroimaterialov i Sooruțenii (AFSS)**, Yerevan, Armenia, 1968
22. Voda, V. Gh. **Controlul durabilitatii produselor industriale**, Ed. Tehnica, Bucharest, 1981
23. Wald, A. **Sequential Analysis,** Dover Publ. Inc., New York, 1973 (republication of the work originally published by John Wiley and Sons Inc., New York, 1947)
24. Weibull, W. **A statistical distribution of wide applicability**, Journal of Applied Mechanics, vol. 18, 1951, pp. 293 – 297

[1] With deep regrets we announce that our colleague Viorel Gh. VODĂ passed away in the last part of May 2009.

[2] Codification of references:

[1]	Baron, T., Tovisi, L., Isaic-Maniu, Al. *et all*. **Calitate si fiabilitate. Manual Practic, vol. I – II**, Ed. Tehnica, Bucharest, 1988
[2]	Barsan-Pipu, N., Isaic-Maniu, Al. and Voda, V., Gh. **Defectarea. Modele statistice cu aplicatii**, Ed. Economica, Bucharest, 1999
[3]	Blischke, W., R. and Murthy, D., N., P. **Reliability. Modeling, Prediction and Optimization**, John Wiley and Sons Inc., New York, 2000
[4]	Catuneanu, V., M. and Mihalache, A., N. **Reliability Fundamentals,** Elsevier, Amsterdam, 1989, (Fundamental Studies in Engineering, No. 10 – translated from Romanian by A. N. Mihalache)
[5]	Cioclov, D. **Rezistenta si fiabilitate la solicitari variabile**, Ed. Facla, Timisoara, 1975
[6]	Dorin, Al. C., Isaic-Maniu, Al. and Voda, V., Gh. **Probleme statistice ale fiabilitatii**, Ed. Economica, Bucharest, 1994
[7]	Dowling, N. E. **Mechanical Behaviour of Materials (2nd Edition),** Prentice-Hall, Upper Saddle River, New Jersey, 1999
[8]	Drimlová, B. **Odhady parametrů některých důležitnych modelů životnosti a spolehlivosti strojirenských výrobků**, SVÚSS, Výzkumná Zpráva, č. SVÚSS (Bechovice), 1970
[9]	Drimlová, B. and Žaludova, A. **Některé základni přejimaci postupy vzhledem ke spolehlivosti a životnosti strojirenských výrobků,** SVÚSS, Výzkumná Zpráva, č. SVÚSS (Bechovice), 1971
[10]	Drimlová, B. **Smérnice pro přejimani dilů a zařizeni s ohledem na životnost**, SVÚSS, Výzkumná Zpráva, č. SVÚSS (Běchovice), 1973
[11]	Gupta, S. S. and Grall, Ph. A. **Gamma distribution in acceptance sampling**, J. Amer. Statist. Assoc. nr. 56, 1961, pp. 942–970
[12]	Gupta, P. **The Six Sigma Performance Handbook. A statistical Guide to Optimizing Results,** Mc.Graw Hill Book Co., New York (Six Sigma Operational Methods Series, No. 7), 2004
[13]	Iliescu, D. V. **Controlul calitatii loturilor de produse,** Ed. Tehnica, Bucharest, 1982
[14]	Isaic-Maniu, Al. **Metoda Weibull. Aplicatii**, Ed. Academiei Republicii Socialiste România, Bucharest, 1983
[15]	Isaic-Maniu, Al. and Voda, V. Gh. **O noua generalizare a repartitiei putere,** Stud. Cerc. Calc. Econ. Cib. Econ., vol. XXIX, no. 1, 1995, pp. 19– 26
[16]	Isaic-Maniu, Al. and Voda. V. Gh. **Aspecte privind repartiția Rayleygh**, Stud.Cerc.Calc.Econ.Cib.Econ., vol. XXXII, no. 1, 1998, pp. 5– 13
[17]	Kirkpatrick, E. G. **Quality Control for Managers and Engineers**, John Wiley and Sons. Inc., New York, 1970
[18]	Kochendarfer, D. C. and Pabst, Jr., W. R. **MILSTD 883 and related documents**, Journal of Quality Technology, vol. 3, no. 3, 1971, pp. 129– 137
[19]	Kovalenko, Ye., G. **English – Russian Dictionary of Reliability and Quality Control (about 22.000 terms),** Russian Language Publishers, Moscow, 1975
[20]	Meeker, W. Q. and Hamada, M. **Statistical tools for the rapid development and evaluation of high – reliability products**, IEEE Transactions on Reliability, R-44, 1995, pp. 187– 198
[21]	Sedrakian, L. G. **K statisticeskoi teorii procinosti. Izdatelstvo Armianskogo Instituta Stroimaterialov i Soorujenii (AFSS)**, Yerevan, Armenia, 1968
[22]	Voda, V. Gh. **Controlul durabilitatii produselor industriale**, Ed. Tehnica, Bucharest, 1981
[23]	Wald, A. **Sequential Analysis,** Dover Publ. Inc., New York, 1973 (republication of the work originally published by John Wiley and Sons Inc., New York, 1947)
[24]	Weibull, W. **A statistical distribution of wide applicability**, Journal of Applied Mechanics, vol. 18, 1951, pp. 293 – 297

ESTIMATING THE PROJECT'S UTILITY BY MEANS OF HARMONIZATION THEORY

Avner BEN-YAIR[1,2]
PhD, Department of Industrial Engineering and Management,
Sami Shamoon College of Engineering, P.O.Box 950, 84100, Beer-Sheva, Israel

E-mail: avnerb@sce.ac.il

Dimitri GOLENKO-GINZBURG[3]
PhD, Department of Industrial Engineering and Management (Professor Emeritus),
Ben-Gurion University of the Negev, P.O.Box 653, 84105, Beer-Sheva, Israel
& Department of Industrial Engineering and Management,
Ariel University Center in Samaria, Ariel, 44837, Israel

E-mail: dimitri@bgu.ac.il

Doron GREENBERG[4]
PhD, Department of Economics and Business Administration,
Faculty of Social Science, Ariel University Center in Samaria, Ariel, 44837, Israel

E-mail: dorongreen@gmail.com

Zohar LASLO[5]
PhD, Department of Industrial Engineering and Management,
Sami Shamoon College of Engineering, P.O.Box 950, 84100, Beer-Sheva, Israel

E-mail: zohar@sce.ac.il

Abstract: *This paper is a generalization of our previous publications [2,19-20] where some newly developed models of the harmonization theory, together with various practical applications, have been outlined. The goal of the paper under consideration is to describe in depth both the general concepts of the harmonization theory and the fitness of the latter by citing an example from a widely used class of stochastic network projects (PERT-COST models).*

Harmonization theory is a multi-parametric optimization model in order to maximize the system's utility as a generalized quality measure of the system's functioning. We have implemented all the principles of the developed theory in a PERT-COST project and have outlined all the mechanisms to evaluate the project's utility. A numerical example is presented.

Key words: *System's utility; Multi-parametrical harmonization theory; PERT-COST network projects with random activity durations; Independent and dependent basic parameters; Two-level optimization algorithm; Partial harmonization model*

1. Introduction

In recent years problems associated with developing various quality concepts have been discussed extensively in scientific literature. However, numerous publications refer mostly to quality control, which is usually applied to products and services. As a matter of fact, the existing quality techniques, including the developed utility theory [13-15, 21][6], are not applicable to technical and organization systems, which are actually supervising and monitoring the process of the systems' functioning: all those models are restricted to solving market competitive problems alone. Thus, nowadays, the existing utility theory centers on analyzing the competitive quality of organization systems' outcome products rather than dealing with the quality of the systems' functioning, i.e., with organization systems in their entirety. This may result in heavy financial losses, e.g. when excellent project objectives are achieved by a badly organized project's realization.

Thus, a conclusion can be drawn that the existing utility theory cannot be used as the system's quality techniques. In order to fill in the gap, we have undertaken research in the area of estimating the quality of the system itself, e.g. the system's public utility. We will consider a complicated organization system which functions under random disturbances. Such a system usually comprises a variety of qualitative and quantitative attributes, characteristics and parameters, which enable the system's functioning. The problem arises to determine a generalized (usually quantitative) value which covers all essential system's parameters and can be regarded as a system's qualitative estimate. We will henceforth call such a generalized value the system's utility.

The backbone of this paper is:

- *to formalize the multi-parametric harmonization model in order to maximize the system's utility as a generalized quality measure of the system's functioning, and*
- *to develop the techniques of the harmonization problem to estimate the stochastic network project's utility.*

To develop the corresponding techniques we suggest to take into account the basic parameters, which actually form the utility of the system - validity, reliability, flexibility, cost, sensitivity, forecasting (timeliness), etc. We suggest subdividing the basic parameters into two sub-sets:

- independent parameters, where for each parameter its value may be preset and may vary independently on other parameters' values, and
- dependent parameters whose values may not depend uniquely on the values of independent parameters.

We suggest a multi-stage solution of harmonization problems. At the first stage a look-over algorithm to examine all feasible combinations of independent basic values, is implemented. The independent parameters' values obtained at that stage are used as input

values at the second stage where for each dependent parameter a local subsidiary optimization problem is solved in order to raise the system's utility as much as possible. Solving such a problem enables the solely dependence of the optimized value on any combination of independent input parameters. At the next stage the system's utility value is calculated by means of basic parameters' values obtained at the previous stages, with subsequent search for the extremum in order to determine the optimal combination of all basic parameters' values delivering the maximum to the system's utility.

The structure of the paper is as follows. In *Section 2* the general concepts and definition of the harmonization theory are presented, while *Section 3* considers the harmonization's model optimality. *Section 4* presents the PERT-COST project's description. In *Section 5* harmonization model for PERT-COST projects is formulated, while *Sections 6* and *7* provide both the model's heuristic solution and the partial harmonization model. In *Section 8* a step-wise procedure to control a PERT-COST network project by means of harmonization is presented. *Section 9* presents applications in a real design office while in *Section 10* conclusions are outlined. In the Appendix, terms to be used in the paper, are presented.

2. General Concepts and Definitions

Consider a complicated organization system which functions under random disturbances. Such a system usually comprises a variety of qualitative and quantitative attributes, characteristics and parameters, which enable the system's functioning. The problem arises to determine a generalized (usually quantitative) value which covers all essential system's parameters and can be regarded as a system's qualitative estimate, namely, the system's utility.

We will require some new definitions.

Definitions

I. Call the system's model M_s a formalized description of the system's structure as well as the system's functioning. M_s usually comprises the logical links between the system's elements, decision-making rules, various random parameters, etc. For project management systems various M_s may be used, e.g. PERT-COST models [6, 9-12], GANTT chart models [21], CPM models [4, 6], GERT models [6], etc. PERT-COST network models which are widely used in project management [11-12], are used as M_s in our paper. Such a network model is actually a graph type simulation model comprising activities with random durations. The p.d.f. of each activity duration depends parametrically on the budget value assigned to that activity.

M_s usually comprises all the basic parameters (see below) which have an influence on the system's utility.

II. Call a quantitative parameter entering the system a basic parameter on condition that changes in the parameter result in changing the system's utility. Note that the

restriction value for any basic parameter is, actually, the worst permissible value that may be implemented into the system. The set of basic parameters, together with the corresponding restriction values, are externally pregiven.

III. Call the system's utility which corresponds to the pregiven restriction values for all basic system's parameters, the basic utility. Denote henceforth the basic utility by U_0. Value U_0 is externally pregiven as well.

IV. Call the direction of changing a basic parameter's value which results in increasing the system's utility, a positive direction, and vice versa. Call the change of the system's utility caused by altering a parameter by its unit value in the positive direction, a local parametric utility. Denote henceforth the additional local parametric utility for the k-th basic parameter by $\alpha_k > 0$. Parametric utility values are also pregiven externally.

Denote henceforth the pregiven restriction values for each basic k-th parameter R_k, $1 \le k \le n$, by R_{k0}, correspondingly.

V. Note that to solve the harmonization problem, we need to define for each k-th basic parameter its best values which by no means can be refined. Denote those values which are externally pregiven, by R_{k00}, correspondingly.

VI. Call the basic n_1 system's parameters which can be pregiven independently from each other, independent basic parameters.

VII. Call other $n_2 = n - n_1$ basic system's parameters dependent basic parameters. Thus, the basic parameters can be subdivided into two groups: independent and dependent parameters. The latter do not depend uniquely on the preset values of independent parameters. Moreover, a combination of independent parameters may correspond to numerous different values (sometimes to an infinite number) of a certain dependent parameter. If, for example, a PERT-COST network project is carried out under random disturbances, setting the cost value (assigned for the project) and the time value (in the form of the project's due date) does not define solely the value of the project's reliability, i.e., its probability to meet the deadline on time. This is because the budget value C assigned to the project has to be reallocated beforehand among the project activities in order to start processing the latter. Each budget reallocation results in a certain project's reliability and, thus, different feasible (but non-optimal!) reallocations correspond to different non-optimal reliability values. However, for the same preset independent basic parameters - cost and time values - it is possible to maximize the project's reliability by means of optimal budget reallocation among the project's activities.

Thus, we suggest to implement a solely dependency of each dependent basic

parameter on the combination of independent input values by means of a subsidiary optimization procedure (heuristic, simulative, approximate) in order to maximize the system's utility for the fixed combination of independent parameters and the optimized dependent parameter.

VIII. Call a partial harmonization problem PHM_j an optimization problem (analytic, simulative, heuristic) which on the basis of preset independent basic parameters delivers an optimum value to a dependent basic parameter R_j in order to maximize the conditional system's utility. Thus, a PHM enables the solely dependence of a dependent parameter from independent ones.

IX. We suggest to calculate the system's utility by

$$U = \sum_{i=1}^{n_1} \alpha_i^{(ind)} \cdot R_i^{(ind)} + \sum_{j=1}^{n_2} \beta_j^{(dep)} \cdot R_j^{(dep)}, \qquad 1 \le i \le n_1, \tag{1}$$

$$1 \le j \le n_2 = n - n_1,$$

where

$R_1^{(ind)},..., R_{n_1}^{(ind)}$ - independent basic parameters;

$R_1^{(dep)},..., R_{n_2}^{(dep)}$ - dependent basic parameters.

Denoting by $PHM_j\left\{\vec{R}_i^{(ind)}\right\} = R_j^{(dep)}, \quad 1 \le j \le n_2,$ a partial harmonization model, we finally obtain

$$U = \sum_{i=1}^{n_1} \alpha_i^{(ind)} \cdot R_i^{(ind)} + \sum_{j=1}^{n_2} \beta_j^{(dep)} \cdot PHM_j\left\{\vec{R}_i^{(ind)}\right\}. \tag{2}$$

Value U may comprise both analytic PHM_j as well as PHM_j based on simulative modeling. In some cases PHM_j can be based on subjective decision-making.

3. Harmonization's Model Optimality

The harmonization problem is as follows: determine optimal values R_k, $1 \le k \le n$, to maximize the system's utility

$$\underset{\{R_k\}}{Max}\ U = U_0 + \sum_{k=1}^{n} \alpha_k \cdot \left|R_k - R_{k0}\right| \tag{3}$$

subject to

$$Min\ \{R_{k0}, R_{k00}\} \le R_k \le Max\ \{R_{k0}, R_{k00}\}. \quad (4)$$

Since U_0 remains constant, the objective can be simplified as follows

$$\underset{\{R_k\}}{Max} \sum_{k=1}^{n} \{\alpha_k \cdot |R_k - R_{k0}|\} \quad (5)$$

subject to (4).

Problem (3-5) is a very complicated optimization problem which usually does not provide analytical estimates.

Let us analyze the general harmonization problem in greater detail. Since independent basic parameters $R_i^{(ind)}$ serve as input values which can be optimized by means of a search algorithm, the harmonization problem's solution suggests itself as a combination of two sequential problems:

- to determine an optimal combination of independent basic values $\{R_i^{(ind)(opt)}\}$ by means of a lookover algorithm that checks the feasibility of each possible combination (*Sub-Problem I*),
- to solve all the partial harmonization problems by means of $PHM_j\{R_i^{(ind)}\}$ (*Sub-Problem II*), and
- to facilitate a search for the extremum in order to maximize utility value (2).

Theorem

Optimal values $R_k^{(opt)}$, $1 \le k \le n$, in problem (4-5) satisfy

$$\{R_k^{(opt)}\} \equiv \{R_i^{(ind)(opt)}\} \cup PHM_j\{R_i^{(ind)(opt)}\}. \quad (6)$$

Proof

Assume that $\{R_k^{(opt)}\}$ dose not satisfy (6), i.e., there exists a combination

$$\{R_k'\} \equiv \{R_i^{(ind)'}\} \cup \{R_j^{(dep)'}\} \quad (7)$$

which satisfies (5) and does not coincide with (6). Note, first, that relation

$$\{R_j^{(dep)'}\} \equiv PHM_j\{R_i^{(ind)'}\} \quad (8)$$

holds, otherwise the combination $\{R_k'\}$ may be improved by substituting $R_j^{(dep)'}$ for $PHM_j\{R_i^{(ind)'}\}$. This, in turn, contradicts relation (5). Secondly, relation

$$\{R_i^{(ind)'}\} \equiv \{R_i^{(ind)(opt)}\} \quad (9)$$

holds as well, since values $\{R_i^{(ind)(opt)}\}$ have been obtained by means of an optimal lookover algorithm which checks all possible combinations $\{R_i^{(ind)}\}$, including $\{R_i^{(ind)'}\}$. Thus, our assumption proves to be false and combinations (6) and (7) fully coincide.

The proved theorem enables solution of problem (4-5) by means of a sequential solution of *Sub-Problems I* and *II*. However, if, due to the high number of possible combinations $\{R_i^{(ind)}\}$, solving both problems on a lookover basis requires a lot of computational time, we suggest a simplified heuristic algorithm as follows.

Since practically most partial harmonization models PHM_j (see, e.g. [1]) for organization systems are complicated non-linear functions of independent parameters $\{R_i^{(ind)}\}$, determining the optimal system's utility results in implementing the theory of unconstrained optimization for non-linear problems. As outlined in [18], the most effective and widely known methods for maximizing a non-linear function of several variables, e.g. the gradient method, the Newton's method, the conjugate direction method, etc., cannot be carried out without determining the gradient vector at each search step. However, solving the gradient equation for partial harmonization problems based on simulation models comprising stochastic programming constraints leads usually to futile computational efforts.

Thus, a conclusion can be drawn that more attractive and at the same time more realistic approximated algorithms have to be implemented. According to the general recommendations outlined in [18] we have replaced the precise lookover algorithm (*Sub-Problem I*) by the cyclic coordinate search algorithm (CCSA). The latter optimizes the non-linear function of independent parameters cyclically, with respect to coordinate variables. The cyclic coordinate algorithm has been applied in many Production and Project Management problems [1, 11-12, 19].

4. The PERT-COST Project's Description

A PERT-COST project $G(N, A)$ [1, 9, 11-12, 21] is characterized by the following parameters:

- the budget C assigned to the project which has to be redistributed among the project's activities;

- the due date D for the project to be accomplished;
- the project's reliability R, i.e., the probability of meeting its due date on time subject to the pre-assigned budget C.

It can be assumed that for each activity time duration its density function depends parametrically on the budget which is assigned to that activity.

A conclusion can be drawn from various studies in PERT-COST [3-7, 9, 12, 16-17, 23-25] that for most activities (i, j) entering the network model, their random time duration t_{ij} is close to be inversely proportional to the budget c_{ij} which is assigned to that activity. Thus, three different distributions may be considered:

- random activity durations are assumed to have a beta-distribution, with the probability density functions (p.d.f.) as follows:

$$p_{ij}(t) = \frac{12}{(b_{ij} - a_{ij})^4}(t - a_{ij})(b_{ij} - t)^2, \tag{10}$$

where $b_{ij} = \frac{B_{ij}}{c_{ij}}$ and $a_{ij} = \frac{A_{ij}}{c_{ij}}$, A_{ij} and B_{ij} being pregiven constants for each activity (i, j) entering the PERT-COST network model.

- random activity durations are assumed to be normally distributed with the p.d.f. $N(a, \sigma^2)$

$$p_{ij}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{2(x-a)}{2\sigma^2}} dx, \tag{11}$$

where the mean value a and the variance σ^2 are calculated by

$$a = 0.5 \cdot \frac{A_{ij} + B_{ij}}{c_{ij}}, \quad \sigma = \frac{B_{ij} - A_{ij}}{6c_{ij}}. \tag{12}$$

- random activity durations are assumed to be distributed uniformly in the interval $\left(\frac{A_{ij}}{c_{ij}}, \frac{B_{ij}}{c_{ij}}\right)$, with the p.d.f.

$$\frac{c_{ij}}{B_{ij} - A_{ij}} = \frac{1}{b_{ij} - a_{ij}}. \tag{13}$$

In the problem under consideration all those cases will be examined.

The following restrictions will be implemented in the model:

- $C \le C_0$, where C_0 is the maximal permissible budget to be assigned to the project;
- $D \le D_0$, where D_0 is the maximal permissible due date to be accepted by the project management;
- $R \ge R_0$, where R_0 is the least permissible reliability of meeting the project's deadline on time, i.e., the minimal probability of accomplishing the project before its due date.

Besides those worst permissible pregiven values C_0, D_0 and R_0, one can define the best pregiven possible correspondent values - the minimal budget C_{00} to be assigned to the project, the earliest due date D_{00} (there is no need in accomplishing the project before D_{00}), and the maximal reliability value R_{00} (usually $R_{00} = 1$). It can be well-recognized that any project values C, D and R satisfy

$$\begin{cases} C_{00} \le C \le C_0 \\ D_{00} \le D \le D_0 \\ R_0 \le R \le R_{00}. \end{cases} \tag{14}$$

5. Harmonization Model for PERT-COST Projects

Using general relations (1-5) and (14) for the case of a PERT-COST project, we suggest to evaluate the project's utility by

$$U = \alpha_C \cdot [C_0 - C] + \alpha_D \cdot [D_0 - D] + \alpha_R \cdot [R - R_0], \tag{15}$$

where C_0, D_0 and R_0 are the least permissible budget, due date and reliability values which can be implemented in a PERT-COST project, while values C, D and R are the corresponding current values for a project under consideration. Linear coefficients α_C, α_D and α_R define additional partial utilities which the project obtains by refining its corresponding parameter by a unit's value. Note that parameters C and D are

independent parameters since they can be preset beforehand independently on each other, while parameter R is practically defined by values D and C and, thus, is a dependent parameter.

The multi-parametrical harmonization model is as follows: determine optimal non-contradictive project parameters $C^{(opt)}$, $D^{(opt)}$ and $R^{(opt)}$ resulting in the maximal project's utility

$$\underset{\{C,D,R\}}{Max}\ U(G) = \underset{\{C,D,R\}}{Max}\ \{U_0 + \alpha_C(C_0 - C) + \alpha_D(D_0 - D) + \alpha_R(R - R_0)\} \quad (16)$$

subject to

$$C_{00} \le C^{(opt)} \le C_0, \quad (17)$$

$$D_{00} \le D^{(opt)} \le D_0, \quad (18)$$

$$R_{00} \ge R^{(opt)} \ge R_0. \quad (19)$$

Note that since the basic utility U_0 is a constant value which remains unchanged, it may be canceled and, thus, the harmonization model satisfies

$$\underset{\{C,D,R\}}{Max}\ U(G) = \underset{\{C,D,R\}}{Max}\ \{\alpha_C(C_0 - C) + \alpha_D(D_0 - D) + \alpha_R(R - R_0)\} \quad (20)$$

subject to (17-19). Values C, D and R are called non-contradictive if budget C can be reassigned among the project activities to satisfy

$$Pr\left\{T(G)_{c_{ij}} \le D\right\} = R \quad (21)$$

subject to

$$\sum_{(i,j)} c_{ij} = C. \quad (22)$$

6. The Model's Solution

Solving problem (17-20) can be carried out by solving two sequential problems: to determine an optimal budget value C and an optimal due date D (*Sub-Problem 1*) and to carry out the PHM (*Sub-Problem 2*).

Sub-Problem 1 centers on determining an optimal couple $\left(C^{(opt)}, D^{(opt)}\right)$ by means of a look-over algorithm that checks the feasibility of each possible combination (C, D). If the number of combinations is high enough and taking into account that:

- each combination requires a PHM solution, and
- *Sub-Problem 1* is a NP-complete one [8, 22],

- solving both problems on a look-over basis requires a lot of computational time. To avoid this obstacle, we suggest a two-level high-speed approximate heuristic algorithm. At the upper level a heuristic simplified search procedure, e.g. a cyclic coordinate search algorithm (CCSA) [1, 18], has to be carried out in the two-dimensional space in order to determine an optimal combination (C, D). At the bottom level, a heuristic high-speed procedure to optimize the partial harmonization model $PHM/_{C,D}$ with independent input values C and D, has to be implemented. Thus, we substitute objective (20) by

$$\underset{C,D}{Max}\left\{CCSA\{C,D\} \cup PHM/_{C,D} \Rightarrow U(C,D,R)\right\}, \tag{23}$$

where $\cup$ stands for a unification sign.

7. Partial Harmonization Model

As outlined above, parameters C and D are input values of *PHM Problem 2* as well as values $c_{ij\ min}$, $c_{ij\ max}$, A_{ij} and B_{ij}, $(i,j) \subset G(N,A)$. The problem is as follows: determine optimal reassigned budget values c_{ij} for each activity $(i,j) \subset G(N,A)$, to maximize the project's conditional reliability, i.e.,

$$\underset{\{c_{ij}\}, \sum_{(i,j)} c_{ij} = C}{Max} \left[\Pr\left\{T\left(G_{c_{ij}}\right) \le D\right\}\right] \tag{24}$$

subject to

$$C_{ij\ min} \le C_{ij} \le C_{ij\ max}, \tag{25}$$

$$\sum_{(i,j) \subset G(N,A)} c_{ij} = C. \tag{26}$$

Model (24-26) is outlined in several previous publications [1, 9, 11, 19-20]. The corresponding algorithm can be easily programmed on PC. Thus, the system's model (see

Section 2) for the case of a PERT-COST network project together with the corresponding harmonization model can be represented on *Table 1*.

8. Harmonization Model for Managing Stochastic Network Projects

In case when project $G(N, A)$ is represented in a formalized shape and activities $(i, j) \subset G(N, A)$ do not bear any engineering definitions and have an abstract meaning, we suggest to use harmonization modeling as the project's planning and control technique. Note that undertaking harmonization modeling for the project under consideration results in optimal budget reallocation among the project's activities. This basic assertion will be used later on, by implementing the project's on-line control.

We suggest a step-wise procedure to control the PERT-COST network project by means of harmonization as follows:

Table 1. System's model and PHM for project management systems

System's model	Parameters		Partial harmonization models
	Indep.	Dep.	
$G(N, A)$ - PERT-COST network; (i, j) - activity, $(i, j) \subset A \subset G(N, A)$; c_{ij} - budget assigned to (i, j); $c_{ij\ min}$, $c_{ij\ max}$ - lower and upper c_{ij} bounds; Total budget $C \geq \sum_{(i,j)} c_{ij\ min}$; Due date D; p.d.f. $t_{ij}(c_{ij}) = \frac{12}{(b_{ij} - a_{ij})^4}(t - a_{ij})(b_{ij} - t)^2$; $a_{ij} = \frac{A_{ij}}{c_{ij}}$, $b_{ij} = \frac{B_{ij}}{c_{ij}}$, A_{ij}, B_{ij} - const.; $T\{G \mid (c_{ij})\}$ - random project duration with assigned c_{ij}.	BUDGET C DUE DATE D	RELIABILITY R	Determine $c_{ij}^{(opt)}$ to $\underset{\{c_{ij}\}}{Max} R = \underset{\{c_{ij}\}}{Max}\left[Pr\{T\{G \mid (c_{ij})\}<$ subject to $c_{ij\ min} \leq c_{ij} \leq c_{ij\ max}$; $C = \sum_{\{i,j\}} c_{ij}^{(opt)}$; $R^{(opt)} \mid c_{ij}^{(opt)} = R^{opt} = PHM(C$.

Step 0. Given the input information:

- PERT-COST project $G(N, A)$;
- pregiven values $c_{ij\ min}$, $c_{ij\ max}$, A_{ij} and B_{ij} for each activity $(i, j) \subset A \subset G(N, A)$;
- pregiven partial utilities α_C, α_D and α_R;
- pregiven admissible intervals $[C_{00}, C_0]$, $[D_{00}, D_0]$ and $[R_0, R_{00}]$.

Step 1. Undertake harmonization modeling for $G(N, A)$ beforehand, i.e., before the project actually starts to be carried out. Denote the corresponding optimized values which define the maximal project's utility, by C^*, D^* and R^*. Note that restrictions

$$\begin{cases} C_{00} \le C^* \le C_0 \\ D_{00} \le D^* \le D_0 \\ R_0 \le R^* \le R_{00} \end{cases} \tag{27}$$

hold, otherwise harmonization cannot be accomplished.

Step 2. If budget value C^* is accepted, reassign C^* among the project's activities according to values $c_{ij}^{(opt)}$ obtained in the course of undertaking harmonization at *Step 1*. Afterwards the project starts to be carried out.

Step 3. In [12], a control model for PERT-COST projects is outlined. The model determines planned trajectories, observes at each control point the progress of the project and its deviation from the planned trajectory, and establishes the next control point. This control model has to be implemented at *Step 3*, in order to determine the routine control point $t > 0$.

Step 4. At each control point t the progress of the project is observed, i.e., network graph $G(N, A)$ has to be updated at point t, as well as the remaining budget C^*. Denote those values by $G_t^*(N, A)$ and C_t^*, correspondingly.

Step 5. At each routine control point $t > 0$ solve harmonization problem in order to reallocate later on the remaining budget C_t^* among remaining activities $(i, j) \subset A_t \subset G_t(N, A)$. Denote the corresponding optimal budget values by $c_{ij\ t}^{(opt)}$.

Reallocate, if necessary, budget C_t^* among activities $(i, j) \subset A_t$ according to

Step 6. the results of *Step 5*. Note that implementing numerous budget reallocations is actually the only control action in the course of performing on-line control. Go to *Step 3*.

Step 7. The algorithm terminates after inspecting the project at the due date D, i.e., at the last control point.

It can be well-recognized that, besides undertaking on-line procedures, the suggested step-wise algorithm comprises both harmonization modeling and risk analysis models. Indeed, the latter are not similar to traditional risk management methods which involve technological risks, uncertainties in products' marketing, etc. However, optimal budget reallocation serves actually as a regulation model under random disturbances and can be regarded as a risk analysis element.

Note that in the course of the project's realization certain parameters entering the input information may undergo changes, e.g. restriction values C_0, C_{00}, R_0, R_{00}, D_0, D_{00}, as well as partial utility values α_C, α_D and α_R. New values have to be implemented in the harmonization model in order to facilitate optimal budget reallocation among the remaining project's activities at Step 5 of the algorithm. If problem (17-22) has no solution, the decision-making to be undertaken at the company level results either in obtaining additional budget value ΔC or in increasing the due date by ΔD. Both values can be determined by means of harmonization.

9. Practical Applications

This *Section* refers to considering practical achievement on the basis of implementing harmonization models for monitoring various PERT-COST network projects. The experimental design has been taken from a real design office [1].

We will henceforth consider a PERT-COST type project with random activity durations and p.d.f. satisfying (10), (11) or (12). The project's initial data is presented in *Table 2*. The basic project's parameters are as follows: project's budget C, due date D and reliability R. Partial utility coefficients are $\alpha_C = 1.0$, $\alpha_D = 0.5$ and $\alpha_R = 1.0$, while the initial search steps (first iteration) for $CCSA$ are $\Delta C = 4$ and $\Delta D = 2$. The number M of simulation runs for the PHM is taken $M = 2{,}000$. Computer program for the PHM algorithm is written in Borland C++ language on Pentium-IV PC. Other project's parameters are as follows: $R_0 = 0.7$, $R_{00} = 0.95$, $C_0 = 250$, $C_{00} = 230$, $D_0 = 95$, $D_{00} = 85$, $\delta C = 10$, $\delta D = 2.0$, $\delta R = 0.1$ and $\varepsilon = 0.001$.

The second iteration for the $CCSA$ is carried out with $\Delta C = 2.0$ and $\Delta D = 1.0$, while all further iterations, $v \geq 2$, are realized with $\Delta C = 1.0$ and $\Delta D = 1.0$.

The performance of the harmonization model's algorithm is illustrated on *Tables 3-5* (for the case of p.d.f. satisfying (10), (11) and (12), correspondingly). It can be well-recognized that:

1. The cyclic coordinate search algorithm for determining the optimal utility of a medium-size project requires only four iterations to carry out the optimization process. The increase of the project's utility parameter after completing the fourth iteration (14 search points), as compared with the initial search point, shows utility improvement of approximately 45%. Thus, the two-level heuristic algorithm to optimize the project's harmonization model performs well.

2. Using the beta-distribution function results in obtaining the highest values for the project's utility parameter. This stems from the obvious fact that the mean value $\mu = 0.6a + 0.4b$ for beta-distribution p.d.f. within the distribution range (a,b) is closer to the lower bound a, than in case of normal and uniform distributions with symmetrical mean values $\mu = 0.5(a+b)$. It goes without saying that lower mean activity – time values result in higher reliability estimates. Since values of the truncated normal distribution concentrate closer to the mean value, than uniformly distributed values, the corresponding project's utility estimates are slightly better for the normal distribution p.d.f. than for the uniform one.

3. Thus, the optimal project's utility can be determined for the following parametrical values:

$C = 245$, $D = 93$, $R = 0.914$, $U_G = 3.64$ *(beta-distribution)*,

$C = 245$, $D = 93$, $R = 0.893$, $U_G = 3.43$ *(normal distribution)* and

$C = 245$, $D = 93$, $R = 0.889$, $U_G = 3.39$ *(uniform distribution)*.

10. Conclusions and Application Areas

The following conclusions can be drawn from the study:

1. Problems of estimating the utility of complicated and usually multilevel management systems by means of establishing and solving harmonization problems, are very urgent, especially for organization systems with a variety of quality parameters. Applications of the utility theory in

Table 2. The project's initial data

N	i	j	$c_{ij\ min}$	$c_{ij\ max}$	A_{ij}	B_{ij}
1	1	2	2	8	25	81
2	1	3	1	6	22	60
3	1	4	1	8	75	105
4	1	5	2	15	80	132

5	1	6	1	8	30	45
6	1	7	8	30	160	200
7	2	3	8	15	50	100
8	2	15	3	8	83	120
9	3	14	10	15	110	220
10	3	15	4	12	60	120
11	4	13	5	10	90	120
12	4	14	8	12	50	100
13	5	9	7	17	150	200
14	5	13	5	10	105	140
15	6	9	2	5	60	80
16	7	8	3	10	42	60
17	8	10	2	10	20	32
18	8	11	6	10	40	80
19	9	11	1	5	90	120
20	9	12	3	10	42	60
21	10	20	2	5	60	80
22	10	21	5	10	105	140
23	11	19	7	15	150	200
24	11	21	8	12	50	100
25	12	18	5	10	90	120
26	13	17	4	12	60	120
27	13	18	5	10	48	60
28	13	19	4	8	63	110
29	14	16	1	7	58	102
30	14	17	1	7	23	94
31	15	16	4	9	85	120
32	15	17	3	5	60	104
33	16	22	4	11	70	93
34	17	22	5	10	82	153
35	17	23	6	10	74	110
36	18	23	2	8	80	120
37	19	23	2	5	40	87
38	20	21	1	4	32	72
39	21	23	3	8	63	95
40	22	23	5	12	87	128

Table 3. Performance illustration of the harmonization algorithm (for a beta-distribution p.d.f.)

№ of search steps	C	D	R	*№* v *of iteration*	*Feasibility*	*Utility* $U(C,D,R)$	*Value* $U^{(v)}$ *after the* v*-th iteration*
Since values $U^{(3)}$ and $U^{(4)}$ coincide, the algorithm terminates after the fourth iteration							
0	250	95	1.000	1	Feasible	2.50	2.50
1	246	95	0.996	1	Feasible	2.90	2.90
2	242	95	0.922	1	Feasible	3.02	3.02
3	238	95	0.793	1	Feasible	2.13	3.02
4	242	93	0.861	1	Feasible	3.41	3.41
5	242	91	0.723	1	Feasible	3.03	3.41
6	244	93	0.895	2	Feasible	3.55	3.55

7	246	93	0.912	2	Feasible	3.52	3.55
8	240	93	0.814	2	Feasible	3.14	3.55
9	244	94	0.936	2	Feasible	3.46	3.55
10	244	92	0.835	2	Feasible	3.45	3.55
11	**245**	**93**	**0.914**	**3**	**Optimal**	**3.64**	**3.64**
12	243	93	0.875	3	Feasible	3.45	3.64
13	245	94	0.951	4	Feasible	3.51	3.64
14	245	92	0.855	4	Feasible	3.55	3.64

Table 4. Performance illustration of the harmonization algorithm (for a normal distribution p.d.f.)

№ of search steps	C	D	R	*№* v *of iteration*	*Feasibility*	*Utility* $U(C,D,R)$	*Value* $U^{(v)}$ *after the* v*-th iteration*
0	250	95	1.000	1	Feasible	2.50	2.50
1	246	95	0.989	1	Feasible	2.90	2.90
2	242	95	0.915	1	Feasible	2.95	2.95
3	238	95	0.782	1	Feasible	2.02	2.95
4	242	93	0.829	1	Feasible	3.09	3.09
5	242	91	0.698	1	Non-feasible	-	3.09
6	244	93	0.868	2	Feasible	3.28	3.28
7	246	93	0.885	2	Feasible	3.25	3.28
8	240	93	0.802	2	Feasible	3.02	3.28
9	244	94	0.912	2	Feasible	3.22	3.28
10	244	92	0.811	2	Feasible	3.21	3.28
11	**245**	**93**	**0.893**	**3**	**Optimal**	**3.43**	**3.43**
12	243	93	0.847	3	Feasible	3.17	3.43
13	245	94	0.921	4	Feasible	3.21	3.43
14	245	92	0.839	4	Feasible	3.39	3.43
Since values $U^{(3)}$ and $U^{(4)}$ coincide, the algorithm terminates after the fourth iteration							

Table 5. Performance illustration of the harmonization algorithm (for a uniform distribution p.d.f.)

№ of search steps	C	D	R	*№* v *of iteration*	*Feasibility*	*Utility* $U(C,D,R)$	*Value* $U^{(v)}$ *after the* v*-th iteration*
0	250	95	1.000	1	Feasible	2.50	2.50
1	246	95	0.984	1	Feasible	2.90	2.90
2	242	95	0.912	1	Feasible	2.92	2.92
3	238	95	0.765	1	Feasible	1.85	2.92
4	242	93	0.821	1	Feasible	3.01	3.01
5	242	91	0.695	1	Non-feasible	-	3.01
6	244	93	0.864	2	Feasible	3.24	3.24
7	246	93	0.882	2	Feasible	3.22	3.24
8	240	93	0.795	2	Feasible	2.95	3.24

9	244	94	0.910	2	Feasible	3.20	3.24
10	244	92	0.807	2	Feasible	3.17	3.24
11	**245**	**93**	**0.889**	**3**	**Optimal**	**3.39**	**3.39**
12	243	93	0.844	3	Feasible	3.14	3.39
13	245	94	0.918	4	Feasible	3.18	3.39
14	245	92	0.835	4	Feasible	3.35	3.39
Since values $U^{(3)}$ and $U^{(4)}$ coincide, the algorithm terminates after the fourth iteration							

recent publications are restricted to market competitive models and do not deal as yet with complicated hierarchical systems' functioning. The nowadays existing multi-attribute utility theory can be applied only to the stage preceding the product's design and determining the objectives for future market competition.

2. We suggest to implement the utility concept as a generalized system's quality estimate which takes into account several essential parameters. The latter usually define the quality of the system as a whole. We have developed a generalized harmonization problem in order to maximize the system's utility. The corresponding model is optimized by means of a two-level heuristic algorithm. At the upper level (the level of independent parameters) a relatively simple search procedure, e.g. the cyclic coordinate algorithm, has to be implemented. At the lower level partial harmonization problems to optimize the dependent parameters, have to be used. Note, that nowadays there is no formalized linkage between the system's parameters and attributes and, thus, no optimization problem can be put and solved in order to maximize the product's utility within its specific life cycle. The developed research enables implementing such a linkage, in future, on the stages of both designing and creating new products and, later on, on the stage of marketing the product.

3. For stochastic PERT-COST network projects three parameters are implemented in the model: the budget assigned to the project, the due date and the project's reliability to meet the due date on time. The harmonization model's solution is achieved by means of implementing a two-level heuristic algorithm. At the upper level a cyclic coordinate search algorithm to determine the quasi-optimal couple (budget – due date) is suggested. At the bottom level a high-speed heuristic procedure serving as a partial harmonization sub-model, is implemented: on the basis of input values (the assigned budget and the set due date) to maximize the probability of meeting the deadline on time by undertaking optimal budget reallocation among the project's activities.

4. Harmonization models can be applied directly to all kinds of PERT-COST network projects with uncertainties associated with activities' durations but without either technological risks or uncertainties on the stage of marketing the project's products. Such projects usually refer to the public service area, like constructing new hospitals, schools, stadiums, theatres, bridges and tunnels, new urban areas, factories, etc. In our opinion, those projects represent an overwhelming majority of existing projects and, thus, require good quality monitoring. For such projects we suggest to use the newly developed harmonization techniques both for estimating the project's utility and for

introducing regulating control actions at inspection points to enhance the progress of the project in the desired direction. Thus, harmonization modeling enables certain on-line control procedures for projects under random disturbances.

5. Being a regulation model, harmonization can be implemented (in a random disturbances environment) as a risk assessment tool as well. Thus, for this class of projects, harmonization, controlling and risk assessment actually meet.

References

1. Ben-Yair, A. **Harmonization models in strategic management and safety engineering,** Ph.D. Thesis, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 2004
2. Ben-Yair, A., Golenko-Ginzburg, D. and Laslo, Z. **Harmonized decision-making in managing reliability and safety,** Journal of Business Economics and Management, 6(1), 2005, pp. 43-51
3. Berny, J. **A new distribution function for risk analysis,** Journal of the Operational Research Society, 40 (12), 1989, pp. 1121-1127
4. Clark, C.E. **The PERT model for the distribution of an activity,** Operations Research, 10, 1962, pp. 405-406
5. Donaldson, W.A. **The estimation of the mean and variance of PERT activity time,** Operations Research, 13, 1965, pp. 382-385
6. Elmaghraby, S.E. **Activity Networks: Project Planning and Control by Network Models,** Wiley, New York, 1977
7. Farnum, N.R. and Stanton, L.W. **Some results concerning the estimation of beta distribution parameters in PERT,** Journal of the Operational Research Society, 38, 1987, pp. 287-290
8. Garey, U.R. and Johnson, D.C. **Computer and Intractability: A Guide to the Theory of NP-Completeness,** Freeman and Co., San-Francisco, 1979
9. Golenko, D. **Statistical Methods in Network Planning and Control,** Nauka, Moscow, 1968, (in Russian)
10. Golenko-Ginzburg, D. **On the distribution of activity time in PERT,** Journal of the Operational Research Society, 39 (8), 1988, pp. 767-771
11. Golenko-Ginzburg, D. **A two-level decision-making model for controlling stochastic projects,** International Journal of Production Economics, 32, 1993, pp. 117-127
12. Golenko-Ginzburg, D. and Gonik, A. **On-line control model for cost-simulation projects,** Journal of the Operational Research Society, 47, 1996, pp. 266-283
13. Keeney, R.L. **The art of assessing multiattribute utility functions,** Organizational Behavior and Human Performance, 19, 1977, pp. 267-310
14. Keeney, R.L. and Raiffa, H. **Decisions with Multiple Objectives: Preferences and Value Tradeoffs,** Wiley, New York, 1976
15. Keeney, R. and Winterfeldt, D. van **Eliciting probabilities from experts in complex technical problems,** IEEE Transactions on Engineering Management, 38 (3), 1991, pp. 191-201
16. Laslo, Z. and Gurevich, G. **Minimal budget for activities chain with chance constrained lead-time,** International Journal of Production Economics, 171(1), 2007, pp. 164-172
17. Laslo, Z. **Activity time-cost tradeoffs under time and cost chance constraints,** Computers & Industrial Engineering, 44(3), 2003, pp. 365-384
18. Luenberger, D.G. **Linear and Non-Linear Programming, 2nd ed.**, Addison Wesley Publishing Co., Massachusetts, 1989

19. Menipaz, E. and Ben-Yair, A. **Three-parametrical harmonization model in project management by means of simulation,** Mathematics and Computers in Simulation, 59 (5), 2002, pp. 431-436
20. Menipaz, E. and Ben-Yair, A. **Harmonization simulation model for managing several stochastic projects,** Mathematics and Computers in Simulation, 61, 2002, pp. 61-66
21. Shtub, A., Bard, J. and Globerson, Sh. **Project Management: Engineering, Technology and Implementation,** Prentice Hall International, Inc., New York, 1994
22. Taha, H.A. **Operations Research: An Introduction**, MacMillan, New York, 1987
23. Welsh, D. **Errors introduced by a PERT assumption,** Operations Research, 13, 1965, pp. 141-143
24. Williams, T.M. **Practical use of distributions in network analysis,** Journal of the Operational Research Society, 43 (3), 1992, pp. 265-270
25. Williams, T.M. **What are PERT estimates?**, Journal of the Operational Research Society, 46 (12), 1995, pp. 1498-1504

APPENDIX

Nomenclature

$G(N,A)$ finite, connected, oriented activity – on – arc network of PERT-COST type;

-

$(i,j) \subset G(N,A)$ - activity leaving node i and entering node j;

t_{ij} - random time duration of activity (i,j);

c_{ij} - budget assigned to activity (i,j);

$c_{ij\ min}$ - minimal budget capable of operating activity (i,j) (pregiven);

$c_{ij\ max}$ - maximal budget required to operate activity (i,j) (pregiven); in case $c_{ij} > c_{ij\ max}$ additional value $c_{ij} - c_{ij\ max}$ is redundant;

C - budget assigned to carry out project $G(N,A)$;

D - the due date for the project $G(N,A)$;

R - the project's reliability value, i.e., its probability of meeting the deadline D on time;

$G_t^*(N,A)$ - PERT-COST graph updated at point $t > 0$;

C_t^* - the project's budget updated at point $t > 0$;

$T(G)_{c_{ij}}$ - the project's random duration on condition that budget values c_{ij} are assigned to activities (i,j);

$R(G)_{c_{ij}}$ - the project's local reliability, i.e., the probability of meeting its deadline on time on condition that values c_{ij} are assigned to $(i,j) \subset G(N,A)$, $R(G)_{c_{ij}} = Pr\{T(G)_{c_{ij}} \le D\}$;

$R(G)_{C,D} = \underset{\{c_{ij}\},\sum_{(i,j)} c_{ij}=C}{Max} R(G)_{c_{ij}}$ - the project's conditional reliability (on condition that values C and D are preset beforehand; to be calculated);

$PHM/_{C,D}$ - the partial harmonization model to optimize reliability R with independent input values C and D;

C_0 - the maximal possible budget to be assigned to project $G(N,A)$ (pregiven);

D_0 - the maximal permissible due date for the project $G(N,A)$ to be accomplished (pregiven);

R_0 - the minimal permissible reliability value for project $G(N,A)$ (pregiven);

ΔC - budget search step (pregiven);

ΔD - due date's search step (pregiven);

C_{00}, D_{00}, R_{00}, P_{00} - the best possible values of parameters C, D, R and P (pregiven);

δ_C - budget unit value (pregiven);

δ_D - time unit value (pregiven);

δ_R - reliability unit value (pregiven);

α_C - partial utility value for parameter C (pregiven);

α_D - partial utility value for parameter D (pregiven);

α_R - partial utility value for parameter R (pregiven);

$U(G) = U(C,D,R)$ - the project's utility;

$U_0 = U(C_0,D_0,R_0)$ - the project's basic utility;

$U/_{C,D}$ - conditional project's utility on the basis of unification $\left\{C,\ D,\ PHM/_{C,D}\right\}$;

$U/_{C,D}$ - conditional project's utility on the basis of unification $\left\{C,\ D,\ PHM/_{C,D},\ P_{hf}/_{C,D}\right\}$;

$CCSA\{G\}$ - the cyclic coordinate search algorithm which undertakes a search in the E^2 space of budget values C and due dates D;

$v \geq 1$ - ordinal number of a current iteration in $CCSA\{G\}$;

$CCSA^{(v)}$ - the results of the v-th current iteration in the course of carrying out $CCSA\{G\}$;

$\varepsilon > 0$ - pregiven search tolerance (accuracy) in the course of optimizing the project's utility.

---

[1] Avner Ben-Yair is lecturer in the Industrial Engineering and Management Department, Sami Shamoon College of Engineering, Israel. He was born in Moscow in 1961. He received his B.Sc. in Mechanical Engineering from the Moscow Polygraphic Institute, Russia, and his M.Sc. degree in Health and Safety Engineering and Management (Summa Cum Laude) from the Ben Gurion University of the Negev, Israel. He also received his Ph.D. degree in Industrial Engineering and Management from the Ben Gurion University of the Negev, Israel. His professional experience includes 13 years of engineering and management positions in Israeli chemical, pharmaceutical and high-tech industries. His current research interests are in economic aspects of safety, reliability and failure analysis, trade-off optimization models for organization systems, production planning, scheduling and control, cost optimization and PERT-COST models, and strategic management. He has published 40 articles in various scientific sources. His recent publications have appeared in Mathematics and Computers in Simulation, International Journal of Production Economics, Communications in Dependability and Quality Management, and Computer Modelling and New Technologies.

[2] Corresponding author

[3] Dimitri Golenko-Ginzburg is Professor-Emeritus and member of the Paul Ivanier Center on Robotics Research and Production Management, in the Industrial Engineering and Management Department of the Ben Gurion University of the Negev, Israel, as well as professor in Department of Industrial Engineering and Management, Ariel University Center in Samaria, Israel. He was born in Moscow in 1932. He received his M.A. degree in Mathematics from the

Moscow State University in 1958 and his Ph.D. degree in Applied Mathematics from the Moscow Physico-Technical Institute (Russia) in 1962. In Russia during the period 1958÷1985 he wrote 14 books and more than 200 articles. In 2007 Prof. Dimitri Golenko-Ginzburg was nominated Honorary Member of the Russian Project Management Association (SOVNET) and elected Foreign Member of the Russian Academy of Natural Sciences, the Branch of Computer Science and Cybernetics. His professional experience includes 50 years of scientific research and academic teaching in leading academic institutions in the former USSR and, after repatriating in 1985, in Israel. His current research interests are in managing and planning under uncertainty, planning and control of network projects, and industrial scheduling. During the period 1985÷2007 he published more than 170 articles. His recent publications have appeared in Mathematics and Computers in Simulation, International Journal of Production Economics, Automation and Remote Control, Communications in Dependability and Quality Management, and Computer Modelling and New Technologies.

[4] Dr. Doron Greenberg earned his doctoral degree in Economics from The University of Houston in 1992; his MBA specialized in Operation Management from the Recanati School of Management in Tel Aviv (1985); and his B.Sc. in Industrial Engineering from the Technion in Haifa (1980).
He serves as the Head of the financial branch of the
Department of Economics and Business Administration in the Ariel University Center (AUC). In 2008 he has initiated and was appointed to be the 1st director of the 1st college chapter of Global Association of Risk Professionals (GARP) in Israel which is affiliated with AUC.
Dr. Doron Greenberg published about 10 articles in various scientific source including: The Journal of Real Estate Finance and Economics, Communications in Dependability and Quality Management-An International Journal, Technological and Economic Development of Economy, Risk Management, Computer Modelling and New Technologies, Communications in Dependability and Quality Management,
He also served as a lecturer in the Recanati School for management in The Tel Aviv University, in The Negev College on behalf of Pinchas Sapir, and in the Tel Aviv College of Engineering. Dr. Greenberg gave many courses in Economics and Finance and wrote dozens articles in the local newspapers regarding public wealth and finance. His main research areas are Economics and Finance.
Dr. Greenberg serves currently as director and chair of the Investment Committee of the Technical Engineers & Technicians Professional Study fund in Israel.
Dr. Greenberg has reached a profound experience in the Israeli capital market: as a director and chair of the Investment Committee of Pension fund of agricultural in Israel, as a financial analyst for institutional investors in Israel, as chairman of Gahelet, a municipal firm which deals with savings for high education, and as a member of the board of directors of PEKAN, one of the top Israeli managing firms of mutual funds.
Dr. Greenberg devotes major part of his time and energy to the community. As such he established with other the Israeli chapter of The Transparency International Organization (TI), and serves as the CFO of the organization. He also serves as a director in several municipal firms, including: the Economic firm of Ariel and the Water & Sewage firm of Ariel.

[5]Zohar Laslo is Associate Professor and Department Chair of the Industrial Engineering and Management Department, Sami Shamoon College of Engineering, Israel. Zohar was born in 1945. He received his B.Sc. and M.Sc. in Industrial Engineering and Management from the Technion - Israel Institute of Technology, Haifa, and his Ph.D. degree in Industrial Engineering and Management (Summa Cum Laude) from the Ben Gurion University of the Negev, Israel. His professional experience includes 24 years of engineering and management positions in Israeli aeronautical, electronic, pharmaceutical and high-tech industries. His current research interests are in matrix structure performance, network design, planning scheduling and control, planning under time-cost uncertainty, quality of planning, maintainability and hazard analysis techniques, and technological education. His recent publications have appeared in Computers Modelling and New Technologies, Mathematics and Computers in Simulation, Computers and Industrial Engineering,, European Journal of Engineering Education, IEEE Transactions on Engineering Management and Communications in Dependability and Quality Management.

[6] Codification of references:

[1]	Ben-Yair, A. **Harmonization models in strategic management and safety engineering,** Ph.D. Thesis, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 2004
[2]	Ben-Yair, A., Golenko-Ginzburg, D. and Laslo, Z. **Harmonized decision-making in managing reliability and safety,** Journal of Business Economics and Management, 6(1), 2005, pp. 43-51
[3]	Berny, J. **A new distribution function for risk analysis,** Journal of the Operational Research Society, 40 (12), 1989, pp. 1121-1127
[4]	Clark, C.E. **The PERT model for the distribution of an activity,** Operations Research, 10, 1962, pp. 405-406
[5]	Donaldson, W.A. **The estimation of the mean and variance of PERT activity time,** Operations Research, 13, 1965, pp. 382-385
[6]	Elmaghraby, S.E. **Activity Networks: Project Planning and Control by Network Models,** Wiley, New York, 1977
[7]	Farnum, N.R. and Stanton, L.W. **Some results concerning the estimation of beta distribution parameters in PERT,** Journal of the Operational Research Society, 38, 1987, pp. 287-290
[8]	Garey, U.R. and Johnson, D.C. **Computer and Intractability: A Guide to the Theory of NP-Completeness,** Freeman and Co., San-Francisco, 1979

[9]	Golenko, D. **Statistical Methods in Network Planning and Control,** Nauka, Moscow, 1968, (in Russian)
[10]	Golenko-Ginzburg, D. **On the distribution of activity time in PERT,** Journal of the Operational Research Society, 39 (8), 1988, pp. 767-771
[11]	Golenko-Ginzburg, D. **A two-level decision-making model for controlling stochastic projects,** International Journal of Production Economics, 32, 1993, pp. 117-127
[12]	Golenko-Ginzburg, D. and Gonik, A. **On-line control model for cost-simulation projects,** Journal of the Operational Research Society, 47, 1996, pp. 266-283
[13]	Keeney, R.L. **The art of assessing multiattribute utility functions,** Organizational Behavior and Human Performance, 19, 1977, pp. 267-310
[14]	Keeney, R.L. and Raiffa, H. **Decisions with Multiple Objectives: Preferences and Value Tradeoffs,** Wiley, New York, 1976
[15]	Keeney, R. and Winterfeldt, D. van **Eliciting probabilities from experts in complex technical problems,** IEEE Transactions on Engineering Management, 38 (3), 1991, pp. 191-201
[16]	Laslo, Z. and Gurevich, G. **Minimal budget for activities chain with chance constrained lead-time,** International Journal of Production Economics, 171(1), 2007, pp. 164-172
[17]	Laslo, Z. **Activity time-cost tradeoffs under time and cost chance constraints,** Computers & Industrial Engineering, 44(3), 2003, pp. 365-384
[18]	Luenberger, D.G. **Linear and Non-Linear Programming, 2nd ed.**, Addison Wesley Publishing Co., Massachusetts, 1989
[19]	Menipaz, E. and Ben-Yair, A. **Three-parametrical harmonization model in project management by means of simulation,** Mathematics and Computers in Simulation, 59 (5), 2002, pp. 431-436
[20]	Menipaz, E. and Ben-Yair, A. **Harmonization simulation model for managing several stochastic projects,** Mathematics and Computers in Simulation, 61, 2002, pp. 61-66
[21]	Shtub, A., Bard, J. and Globerson, Sh. **Project Management: Engineering, Technology and Implementation,** Prentice Hall International, Inc., New York, 1994
[22]	Taha, H.A. **Operations Research: An Introduction**, MacMillan, New York, 1987
[23]	Welsh, D. **Errors introduced by a PERT assumption,** Operations Research, 13, 1965, pp. 141- 143
[24]	Williams, T.M. **Practical use of distributions in network analysis,** Journal of the Operational Research Society, 43 (3), 1992, pp. 265-270
[25]	Williams, T.M. **What are PERT estimates?**, Journal of the Operational Research Society, 46 (12), 1995, pp. 1498-1504

JOURNAL OF APPLIED QUANTITATIVE METHODS

ORTHOGONALITY MEASUREMENT FOR HOMOGENOUS PROJECTS-BASES

Ion IVAN[1]
PhD, University Professor, Department of Computer Science in Economics, University of Economics, Bucharest, Romania

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

Andrei SANDU[2]
PhD Candidate, University of Economics, Bucharest, Romania

E-mail: andreisandu@yahoo.com

Marius POPA[3]
PhD, University Lecturer, Department of Computer Science in Economics, University of Economics, Bucharest, Romania

E-mail: marius.popa@ase.ro

Abstract: *The homogenous projects-base concept is defined. Next, the necessary steps to create a homogenous projects-base are presented. A metric system is built, which then will be used for analyzing projects. The indicators which are meaningful for analyzing a homogenous projects-base are selected. The given hypothesis is experimentally verified. The projects are analyzed and measured to establish the quality level of the funds spending. Some examples of measurement are offered in this paper. The most important characteristics of the projects are identified and presented. Also, a quality characteristic framework of homogenous projects is provided by this paper. The framework is used to develop metrics that cover the properties and requirement of the projects. The implementation of the project characteristics is made within a software application, called Projects-Bases Operations Software (PBOS).*

Key words: *orthogonality analysis; quality; projects-base; homogeneity; metric*

1. Homogenous Projects-Base

The most spread definition of a project is that of a temporary endeavor undertaken in order to create a unique product, service or result. No matter their type and structure, projects can be grouped based on different criteria. A grouping can be called a program, a portfolio or simply a collection, based on the ownership of the projects contained, based on the project goals and on the project types. We a call a portfolio of projects a set of projects, not necessarily inter-dependent, which are performed by a project-oriented company at a certain point in time. If the projects are inter-dependent and strategically they are perceived as an ongoing long term correlated effort, then we are talking about a program. If the projects analyzed don't fall within the definition of portfolio or program then they can simply be referred as a collection of projects.

According to the Romanian Explicatory Dictionary (DEX), homogeneity is "the attribute of an object or a group or a physical-chemical system of having the same characteristics allover".

From a statistical point of view, homogeneity of data means:

- collecting data from homogenous statistical entities;
- fall within the same definitions and methodologies of calculation in relation to the scope of time and space;
- description of evolution of status within a timeframe which is not subject to major modifications of the analyzed process;
- refers variables by using the same measure of unit – this is most commonly used when evaluating economical indicators in real or comparable prices.

Empirically speaking, if we have a collectivity identified by a characteristics which falls within the interval (a, b) and which has an average X^{MED}, it is said that the elements of that collectivity form a homogenous collectivity if $a = X^{MED} - 2.5\% * X^{MED}$ and $b = X^{MED} + 2.5\% * X^{MED}$.

The above result is the output of analyzing the input of 300 specialists facing the challenge of determining the needed heights of people from a team, so that the team would be considered homogenous. By analyzing the data sets there was obtained the lower limit, the upper limit and the average height. Statistical analysis showed the results are stable and representative. The length of the interval equals 5 % of the average height value.

Coming back to projects-bases, homogeneity is built following those steps:

- a program with clear objectives is defined;
- guidelines for drawing the projects are presented;
- the guidelines used for assessing the projects are presented;
- the templates to be used for detailing the project and structuring the budget are given;
- there is a given key words list;
- there are a given certain thresholds for texts and different thresholds for budget figures.

There are projects-bases for different industries and different purposes within those industries, like projects-bases for building hospitals, research for environment protection, putting up education centers around the country and so forth. They are all influenced and diverse due to the experience of the people who propose the projects, due to the proposed

objective, due to the resources used for implementation, due to techniques used and so forth.

The main goal is that those projects increase homogeneity so that the projects have a better chance of getting to implementation, of delivering quality as well as being considered a success.

On existing projects-bases there have to be analyzed the projects and to be identified and diminished the causes that lead to lack of homogeneity.

2. Milestones for Building Homogenous Projects-Base

In order to build strong project proposals, there must be put in place a strategy for deploying the processes of creating, evaluating, financing, implementing and auditing the projects, so that the projects will address a high level of quality:

- define the financing purpose – the objective must be clearly stated and must targeted to be achieved only by skilled professionals;
- define the eligibility criteria – there must be strict criteria in place, including years of experience, relevance of previous projects, clearly defined contribution;
- creating a guide for creating project proposals – clear guidelines using precise terminology and references which can only be accessible and understood by true professionals;
- creating the evaluation booklet and a full process for handling the evaluation – put in place clear evaluation criteria and detail on every step of the evaluation and even focus on appeals that may arise after the evaluation ends;
- put in place clear quality thresholds as related to management, documentation, guidelines – projects that do not comply with at least one of the thresholds are automatically disqualified;
- the scoring has to be representative enough for the overall quality of each project, as for small differences between projects to be sufficient to distinguish between a selected and a non-selected project;
- during the lifecycle of the project there has to be taken into account not to be major differences between what has been planned and what has been delivered;
- auditing has to establish that what has been delivered is in line with expectations;
- nevertheless – the satisfaction degree of end users for all homogenous applications has to be at the same level.

The main target beneath the homogeneity concept is that of identifying different ways of reaching the highest quality level possible.

3. Set of Indicators to Be Used for Analyzing Quality of Projects

Projects should all encompass the following quality characteristics: gradualness, complexity, orthogonality, correctness, completeness.

Gradualness is a quality characteristic that refers to the degree by which the project has been addressed following a waterfall like approach. Therefore, it is of interest the degree by which the project title is connected with the project abstract, the degree to which the project abstract is connected with the introduction, and so forth. An indicator, having

values in the range 0 to 1 will show the degree of logical coherence of the project flow, 0 showing a total lack of gradualness, whereas 1 will show a gradualness which is fully implemented.

Project *complexity* is the universal measuring unit for all the projects. Complexity represents a characteristic, which is common to all the projects. Projects differentiate through the complexity degree. Some projects are simple. Others are complex. The resources asked by a project vary very much, depending on its complexity. Thus, a project could be realized in a few weeks or it can last more months; a project may need from one team with few people to teams with hundreds.

Orthogonality means determining the degree to which two text entities are different, considering their presentation means and content. Two projects will be considered orthogonal if their texts only have in common the domain specific technical terms, whereas the others words differ in terms of frequency and positioning within the sentences. It is more important that projects texts are orthogonal and not completely different, because very different texts could mean that the two projects do not belong to the same domain.

Correctness as applicable to projects means that the project text is accepted as being in accordance with the base requirements for the domain that is being financed. Correctness envisages to the naming of processes, technologies and operations, the usage of appropriate concepts, models and the usage of proper variables. Therefore, by respecting the industry standards and code of professional ethics all that previously existed has to be correctly cited and referenced in the bibliography. Correctness also refers to the logic behind activities planning, resource consumption and estimating. If project valuation procedure is transparent, correctness means that auto valuation differs very little from the official valuation.

Completeness is the project quality characteristic that shows the degree by which, when turning project plan into action, the desired project outcomes are obtained. The level of completeness is to be considered one if supplementary costs are not present. The more the completeness level is closer to zero, the more it is clear that the project responsible for project delivery has not taken into account some activities, has badly estimated durations or has not properly managed the deliverables dependencies.

The indicators to be used on testing homogenous projects, depicted in figure 1, need to fulfill a series of conditions and must possess some properties as follows:

- sensitivity – the slightest variations of exogenous variables will produce slight variations for the indicator itself;
- non-catastrophic – there will not be scenarios when, for small variations of exogenous variables, we will have large variations of the indicator;
- non-compensatory – for different levels of independent exogenous variables there are obtained different values of the indicator;
- representativeness –there is defined a clear correspondence between intervals of possible value of the indicator and qualitative levels of the process;
- efficacy – the level of correlation between the effort of reaching a certain level for exogenous variables and the effects generated by the decisions taken upon the level of the indicators.

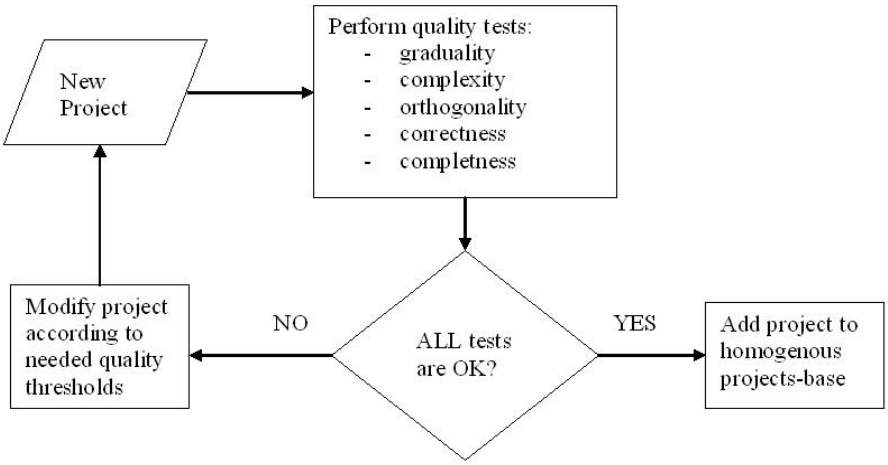


Figure 1. Quality characteristics of homogenous projects

In (Romulus Arhire, 2000), it is demonstrated that an aggregated indicator cannot have three characteristics at the same time: sensitivity, stability and non-compensation. For avoiding the appearance of the compensatory effect, which is specific to aggregated indicators when making a hierarchy based on several complexity characteristics, it is used the grouping technique called *cluster*.

4. Homogenous Projects-Bases Metrics

Homogenous projects-bases have several things in common, like:

- the texts do not differ significantly in terms of length;
- common vocabularies;
- frequency level differences;
- close complexity levels;
- common citations.

The system of indicators specific to homogenous projects-bases includes components that allow differentiations even within very narrow intervals. Therefore, there are taken into account the indicators presented in the previous chapter.

Every indicator is defined, the influence factors are presented and it is demonstrated that at very small variations of the involved factors, there are obtained large variations of the indicators, therefore making those indicators representative for homogenous projects. Those indicators allow the grouping of homogenous projects in project classes, therefore having an even higher degree of homogeneity and the quality will be closer to the desired maximum level.

Homogenous projects involved in obtaining financing are built following clear guidance, which comprises of:

- vocabulary, with key words;
- structure based on chapters;
- structure of tables to be used for presenting the numerical information;
- calculation algorithms;
- maximum lengths of texts;
- verification keys;
- restrictions regarding the thresholds for eligible spending.

If the project proposal is done in an assisted approach, the requestor is not allowed to move to the next step if he has not satisfied the guidelines imposed by the financer. If we have a threshold of maximum 500k EUR of financing for a project and the applicant is requesting 600k, the request will not be processed and if the requester insists for more than 5 times without taking into account the imposed limitations, then the application will reject the input and the project data will be removed completely.

Quality of homogenous projects possesses a characteristic called *gradualness*.

The project text is composed of C1- title, C2 – abstract, C3 – objectives, sub objectives, …, Cn – audit. The project proposal owner creates first C1. Starting from C1, he creates C2, then C3, by following a water flow approach. The indicator has to determine the degree to which C1 is included in C2, C2 is included in C3 and so on. The indicator will tend to be zero if the vocabularies are almost disjunctive and will tend to be 1 as long as most of the words of C1 are found in C2 vocabulary, most of the words of C2 are found in C3 vocabulary and so on.

$$GRD = \frac{\sum_{i=1}^{M-1} \frac{(card(Ci \cap Ci+1))}{card(Ci)}}{M-1}$$

where:

- Ci – set of distinct words from chapter i;
- M – maximum number of words;
- card(*arg*) – total number of words from a given collection.

One approach to measuring *complexity* is that of taking into account the variety of existing resource types and their quantities.

$$C = \sum_{i=1}^{m} q_i * \log_2 q_i$$

where:

- C – project complexity;
- m – total number of distinct resources used by the project;
- q_i – quantity of the resource with index i.

A very simple implementation of such an indicator for *orthogonality* would be, (Andrei Sandu, 2008):

$$ORTO = \frac{NW_{both}}{NW_{\max}}$$

where:

- NR_{both} – total number of words, which are part of the project domain standard terminology, which are found in both texts analyzed;
- NR_{max} – the highest of the total number of domain specific words of the two projects.

Therefore, for this indicator a value of 1 means the two projects are cloned one after the other, whereas a value of 0 means that the two projects are orthogonal.

Correctness envisages things like using preexisting knowledge, formulas without altering the initial contents, putting a certain professional order in the operations undertaken, in applying procedures and processes as to demonstrate that the project proposal owner is experienced within his competency.

In case of partial valuation, (Nicolae Enescu, 2008), it is needed to determine the weight of correct criteria from the whole set of criteria.

The relative correctness indicator, RCI, is computed as follows:

$$RCI = \frac{\sum_{i=1}^{n} R_{ia}}{\sum_{i=1}^{n} R_{iMax}} * w$$

where:

- R_{ia} – actual result of valuation, based on criterion i;
- R_{iMax} – maximum result of valuation, based on criterion i;
- w – weight of correct criteria from the whole set of criteria.

There are three matrixes to be considered as relevant in project *completeness* valuation:

- A-C matrix – shows all resources used for performing each activity;
- A-TM matrix – shows the allocation of project team members on activities
- A-A matrix – shows the dependencies between the activities.

One indicator of completeness, CMPL, has the following form:

$$CMPL = \frac{3 - \frac{\Delta_1 + \Delta_2 + \Delta_3}{n}}{3}$$

where:

- n – number of project planned activities;
- Δ_1 – number of activities which have at least one difference between the planed and the executed A-C matrix;
- Δ_2 – similar to Δ_1, except it applies for A-TM matrix;
- Δ_3 – similar to Δ_1, except it applies for A-A matrix.

According to the guide for project proposals, there is imposed a certain structure and a project will be considered incomplete if any of the following is true:

- missing chapters;
- missing activities;
- missing links between activities;
- missing formulas.

Therefore, a simple indicator stating:

(total expected item – missing items) / total expected items

will show the level of completeness for any of the above stated criteria. A value of 1 will be interpreted as complete project, whereas any value < 1 will be considered as incomplete and the project will be considered for rejection or additional information will be asked for.

The implementation of the above-mentioned indicators is done within PBOS application, in dedicated procedures named accordingly – *computeGraduality, computeComplexity, computeOrthogonality, computesCorrectnes, computeCompleteness*.

5. Experimental Results

It is considered the objective of evaluating the projects sent by students at a master in project management and grading them accordingly. Around 100 distinct projects have been received. They have all been sent via e-mail for being evaluated. The format of sending consists in both a Word document with the whole project description as well as a collection of text documents that are in fact the chapters of each project.

The sample of projects is considered representative for the purpose of our study of homogeneity because all students who submitted the project proposals have taken part at the same classroom where they where explained the purpose of the projects, the structure, information formatting and the marking criteria. Also, they are all using the same terminology in terms of quality management and project management, as presented during the classroom hours.

The imposed project structure is as follows: title, abstract, introduction, presented problem, proposed solution, results, analysis, conclusions, bibliography and annexes.

The reason for sending distinct text files resides in the fact that they are all used as input in the PBOS application for performing the operations for testing homogeneity and nevertheless for grading the projects of the students.

After having input all the data in the application there was tested the orthogonality of each project text in relation with the text of all other projects in the projects-base. The result is as depicted in table 1.

Table 1. Computation of orthogonality each 2 by 2 projects

Project ID	Project Name	Project ID	Project Name	Orthogonality
1	GISG	2	OCMC	0.76
1	GISG	3	SDDE	0.95
1	GISG	4	DIII	0.88
...				
42	MIC	1	GISG	0.92
42	MIC	2	OCMC	0.87
42	MIC	3	SDDE	0.70
...				
104	TVD	101	OIIL	0.83
104	TVD	102	MMAR	0.77
104	TVD	103	SVIC	0.74

Next, for each project there has been computed an average orthogonality and a completeness value as well, by taking into account if all chapters have been sent for the projects. Based on the average orthogonality and the completeness indicator, there is establishes a ranking, as average between the previously mentioned two indicators, as depicted in table 2.

Table 2. Ranking based on avg. orthogonality & completeness

Project ID	Avg. Orthogonality	Completeness	Ranking
1	0.93	1	0.9650
68	0.85	1	0.9250
76	0.83	1	0.9150
...			
98	0.75	0.9	0.8250
15	0.64	1	0.8200
30	0.63	1	0.8150
...			
56	0.62	0.8	0.7100
94	0.59	0.8	0.6950
86	0.65	0.7	0.6750

The results are now converted into marks. Therefore, we obtain a summary that looks as depicted in table 3.

Table 3. Summary of marks

Mark	Percentages
7	8.42 %
8	57.89 %
9	32.63 %
10	1.05 %

The table above is a clear indication that the efforts to create a homogenous project-base have succeeded, in the sense that the level of quality for the projects comprised is very similar. In order to minimize the impact of projects that obtain a lower mark, for example mark 7, and therefore decrease the level of homogeneity a stricter filter has to be put in place when accepting projects. One should reject projects which have an average orthogonality < 0.7 or a completeness indicator < 0.8.

6. Conclusions

The main goal for homogenous projects-bases is that of wisely spending the total funding capacity. Therefore working with homogenous projects-bases is similar to an active attempt of moving quality towards the upper bounds in all involved stages.

The financing programs become clear if:

- they are defined completely and they contain procedures and clear documentation;
- the teams who made the project proposals are trained in drafting high quality proposal and are helped in implementing the projects afterwards;
- there is a process of creating a hierarchy of project offers, so that the risk of having differences between the project proposal – costs, durations, fulfillment degree and so forth – and the effective project roll-of is minimal.

In project management, the efficiency of financing is given by:

- full utilization of resources;

- not over passing the planned financing threshold;
- keeping any supplementary spending within reasonable limits, which do not affect the overall estimated level.

In order to assure a homogenous projects-base creation, different indicators, like gradualness, complexity, orthogonality, correctness and completeness can be applied to projects that are enrolled for obtaining financing. The resulted projects-base allows this way the extraction of information on the long run in relation to teams, projects that have been proposed and implemented, as well as qualifications obtained for those projects.

One simple way of increasing the homogeneity of a projects-base is by modifying its acceptance levels for the metrics used on projects proposals. Like in the given example, an average orthogonality < 0.7 or a completeness indicator < 0.8 will increase the chances of having only good projects in the projects-base – to be marked 8, 9 or even 10.

References

1. Arhire, R. **Complexity Valuation of Systems of Programs,** PhD thesis, ASE, Bucharest, 2000
2. Bodea, C. **Knowledge management into modern university**, ASE Printing House, Bucharest, 2007
3. Enescu, N. **Models for Evaluating the Correctitude of Distributed IT Systems,** PhD thesis, ASE, Bucharest, 2008
4. Gareis, R. **Happy Projects!, 2nd Edition,** ASE Printing House, Bucharest, 2006
5. Ivan, I. and Sandu, A. **Projects-bases Management Software,** in "The Proceedings of 20th IPMA World Congress on Project Management", vol. 2, Shanghai, China, October 15 - 17, 2006, pp. 801-805
6. Sandu, A. **Indicators for Measuring the Quality of Projects-base Components,** in "4th International Conference for Applied Statistics", NIS Printing House, Bucharest, 20-22 November, CD format
7. * * * www.projectstoday.com; India's largest database on new projects.

[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 Academy of Economic Studies, 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 Computer Science in Economics at Faculty of Cybernetics, Statistics and Economic Informatics from the Academy of Economic Studies. 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, conferences and congresses at numerous universities from Greece, Ireland, Germany, France, Italy, Sweden, Norway, United States, Holland, Australia, China 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. Also, he was expert assessor in many research programmes like RELANSIN, INFOSOC, CALIST and CEEX. He holds many awards and diplomas in research activity.

His main interest fields are: software metrics, optimization of informatics applications, developments and assessment of the text entities, efficiency implementation analysis of the ethical codes in informatics field, software quality management, data quality management and so forth.

List of Main Publications (2006 – 2009)

Ion IVAN, Marius POPA and Paul POCATILU (coordinators) – **Structuri de date,** ASE Printing House, Bucharest, 2008, vol. I Tipologii de structuri de date; vol. II Managementul structurilor de date

Ion IVAN and Catalin BOJA **Practica optimizarii aplicatiilor informatice,** ASE Printing House, Bucharest, 2007

Ion IVAN, Catalin BOJA and Cristian CIUREA **Metricile sistemelor colaborative,**: ASE Printing House, Bucharest, 2007

Ion IVAN, Gheorghe NOSCA, Sergiu CAPISIZU and Marius POPA **Managementul calitatii aplicatiilor informatice,** Bucharest: ASE Printing House, 2006

Ion IVAN, Traian BADICA and Marius POPA **Procese de agregare software,** Studii si Cercetari de Calcul Economic si Cibernetica Economica, vol. 42, no. 1, 2008, pp. 69 – 85

Ion IVAN, Eugen DUMITRASCU and Marius POPA **Evaluating the Effects of the Optimization on the Quality of Distributed Applications,** Economic Computation and Economic Cybernetics Studies and Research, vol. 40, no. 3-4, 2006, pp. 73 – 85

Ion IVAN, Cătălin BOJA, Marius VOCHIN, Iulian NITESCU, Cristian TOMA and Marius POPA **Using Genetic Algorithms in Software Optimization,** Proceedings of the 6th WSEAS International Conference on Telecommunications and Informatics, Dallas, Texas, USA, March 22 – 24, 2007, pp. 36 – 41

Ion IVAN, Marius POPA, Catalin BOJA, Cristian TOMA and Dragos ANASTASIU **Software for Structured Text Entities Dependency Graph Building,** Proceedings of the 2007 WSEAS International Conference on Computer Engineering and Applications, Gold Coast, Queensland, Australia, January 17 – 19, 2007, pp. 224 – 230

Ion IVAN, Cristian TOMA, Marius POPA and Catalin BOJA **Secure Architecture for the Digital Rights Management of the M-Content,** Proceedings of the 5th WSEAS International Conference on Information Security and Privacy, Venice, Italy, November 20 – 22, 2006, pp. 196 – 201

[2]Andrei SANDU has graduated the Faculty of Cybernetics, Statistics and Economic Informatics in 2005. He holds a Masters degree in Project Management. He held positions as project manager, technical team leader and software developer. Currently, he is portfolio project manager in banking field. Also, he is PhD student enrolled in a PhD programme on Computer Science in Economics.

His key skills refer to the following: project management, business management skills, banking skills, coaching team members, technical team leadership and IT&C specialist in business oriented software development and maintenance.

List of Main Publications (2006 – 2009)

Andrei SANDU **Knowledge Based Valuation of Projects-bases,** The Proceedings of Ninth International Conference on Informatics in Economy, Education, Research & Business Technologies, Bucharest, Romania, May 7-8, 2009, Bucharest: ASE Printing House, pp. 507-512

Andrei SANDU **Indicators for Measuring the Quality of Projects-base Components,** 4th International Conference for Applied Statistics, Bucharest, Romania, November 20-22, 2008, Bucharest: NIS Printing House

Andrei SANDU **Reingineria bazelor de proiecte,** International Symposium of Young Researchers 2008 (6th Edition), Kishinev, Moldavia Rep., April 18-19, 2008, Kishinev: ASEM Printing House, pp. 343-344

Ion IVAN and Andrei SANDU **Projects Hierarchy Based on Duration and Complexity,** 22nd IPMA World Congress Proceedings - "Project Management to Run", Rome, Italy, November 9-11, 2008, CD format

Ion IVAN and Andrei SANDU **Collection Analysis Metrics Applicable to Projects-bases,** The 3rd International Conference "Knowledge Management – Projects, Systems and Technologies", Bucharest, Romania, October 23-25, 2008

Andrei SANDU **Security of Projects-bases,** Proceedings of 8th International Conference on Informatics in Economy, INFORMATICS IN KNOWLEDGE SOCIETY, Bucharest, Romania, May 17-18, 2007, Bucharest: ASE Printing House, pp. 974-983

Ion IVAN and Andrei SANDU **Projects-bases Management Software,** The Proceedings of 20th IPMA World Congress on Project Management, Shanghai, China, October 15 - 17, 2006, vol. 2, pp. 801-805

[3]Marius POPA has graduated the Faculty of Cybernetics, Statistics and Economic Informatics in 2002. He holds a PhD diploma in Economic Cybernetics and Statistics. He joined the staff of Academy of Economic Studies as teaching assistant in 2002. He has been university lecturer since 2006. Currently, he is university lecturer in Economic Informatics field and branches within Department of Computer Science in Economics at faculty of Cybernetics, Statistics and Economic Informatics from Academy of Economic Studies.

He is the author and co-author of 6 books and over 100 articles in journals and proceedings of national and international conferences, symposiums, workshops in the fields of data quality, software quality, informatics security, collaborative information systems, IT project management, software engineering. From 2009, he is a member of the editorial team for the Informatica Economica Journal. Between 2003 and 2008 he was a member of the editorial team for the journal Economic Computation and Economic Cybernetics Studies and Research.

He was involved as project manager or research team member in research projects on following topics: virtual intelligent manufacturing processes, developing and testing an automated system of risk analysis, diagnosis and decision to support the medical act, system of quality assessment services generated by mobile applications in electronic business, system of quality assessment for on-line public services for citizens and businesses, system of indicators for evaluating IT project management, collaborative informatics systems in the global economy, methodology of applications development for managing the IT project portfolios, evaluation system of the entities based on text, models to estimate the cost of e-business applications, model base for software quality management, designing and implementing the virtual enterprise and platform for estimating costs of testing object oriented software prototypes.

Currently, he is involved as project manager in a research project on topic Implementation of the Quantitative Methods in Distributed Informatics System Audit and as research team member in techniques for classification and recognition with applications in identification documents similarity.

He is certified as project manager IPMA Level D, Certified Project Management Associate. Also, he holds more awards and diplomas for his research activity.

His interest fields are: software engineering, informatics security, software development, project management, informatics audit, data and software quality.

He is member of Research Center of Excellence (ECO-INFOSOC), Association for Promoting the Higher Education in Computer Science in Economics (INFOREC) and Association for Development through Science and Education (ADSE).

List of Main Publications (2005 - 2009)

Ion IVAN, Marius POPA and Paul POCATILU (coordinators) **Structuri de date,** Bucharest: ASE Printing House, 2008, vol. I Tipologii de structuri de date; vol. II Managementul structurilor de date

Marius POPA **Evaluarea calitatii entitatilor text – Teorie si practica,** Bucharest: ASE Printing House, 2005

Marius POPA **Detection of the Security Vulnerabilities in Web Applications,** Informatica Economica, vol. 13, no. 1(49), 2009, pp. 127 – 136

Cristian TOMA, Marius POPA, Catalin BOJA and Miruna VASILACHE **Secure Electronic Cards in Public Services,** Informatica Economica, vol. 12, no. 2(46), 2008, pp. 80 – 85

Cristian TOMA, Marius POPA and Catalin BOJA **Smart Card Based Solution for Non-Repudiation in GSM WAP Applications,** WSEAS Transactions on Computers, vol. 7, Issue 5, May 2008, pp. 453 – 462

Ion IVAN, Dragos ANASTASIU, Catalin BOJA, Marius POPA and Cristian TOMA **Structures Text Entities Dependecy Graph Building - Theory and Practice,** WSEAS Transactions on Computers, vol. 6, Issue 5, May 2007, pp. 835 – 842

Ion IVAN, Cristian TOMA, Marius POPA and Catalin BOJA **Secure Platform for Digital Rights Management Distribution,** WSEAS Transactions on Computers, vol. 6, Issue 3, March 2007, pp. 478 – 485

SECURITY MANAGEMENT IN A MULTIMEDIA SYSTEM

Emanuil REDNIC
PhD candidate, Economic Informatics Department
University of Economics, Bucharest, Romania

E-mail: emanuil.rednic@gmail.com

Andrei TOMA[1]
PhD candidate, Economic Informatics Department
University of Economics, Bucharest, Romania

E-mail: hypothetical.andrei@gmail.com

Abstract: *In database security, the issue of providing a level of security for multimedia information is getting more and more known. For the moment the security of multimedia information is done through the security of the database itself, in the same way, for all classic and multimedia records. So what is the reason for the creation of a security management system and a set of security rules for this type of information? The reason is the fast progress of multimedia information usage in enterprise activities. More and more distributed activities are based on processing multimedia information in real time, which is why the implementation of such a security system is important.*

Key words: *database security; distributed activities; multimedia information; relational databases; multimedia processing; watermarking; metadata; image resources multimedia filters*

1. Introduction

For more than two decades, a new issue appeared in the database technologies: how to store, manage, manipulate, archive multimedia information. At the beginning the information was not stored in the database, but only a logical reference of the physic location from the hardware, and of course all the others characteristics of the multimedia data was saved in relational tables, like: height, RGB, resolution, format. In this way a pre metadata system for managing this kind of information was created.

Furthermore, due to increased number of operations over these data and the necessity of reducing the access, processing and manipulation and other time costs, brought about the storing of the information in the database. A lot of opposing opinions appeared,

saying that databases were not designed for storing these multimedia types of data, because sustaining these data would unjustifiably load the database, due to non character based information. Why were these objections raised? The answer comes from the limited hardware resources available at the time. Fortunately the progress of database technology was followed by the progress of hardware and now we can argue that the limitation of hardware resources is no more a problematic issue in this matter.

Although the hardware limitation problem was solved, another minus was, is and will be discussed in this field: the limited resource which is time. Most database producers tried to create a system of metadata information for multimedia data in order to obtain increased flexibility, speed and scalability. One of the top producers in the field of multimedia databases is Oracle.

But how did Oracle start with multimedia database management? Oracle designed a new type of data, called long raw; then they went further by creating the BLOB's, which were used to store the multimedia information directly in the database, so that it could be managed using the same tools and could participate in the same DML operations as other types : text, numeric, time. Even with this, BLOB's were not able to manage the metadata information, so a new technology had to be introduced within the Oracle Databases, known as Oracle InterMedia.

As mentioned above, Oracle InterMedia is not a specific standalone Oracle Product, it is included in Oracle Database System, to put it in Oracle's own terms, "Oracle InterMedia Enables Databases to understand the real nature of images".

Oracle InterMedia is built on the database kernel and operates as a privileged component of the database. The advantages of using Oracle InterMedia to store images are, as follows:

- Both the descriptions of an image and the image itself can be stored using industry standard formats;
- InterMedia's objects model and methods make application programming simple in the development phase;
- InterMedia's applications maintenance become much easier as well;
- The metadata information and indexes are now created automatically by the InterMedia system;
- InterMedia eliminated the necessity of parsing the information about the image;

2. Characteristics of Oracle InterMedia

One of the most important types which was introduced by Oracle InterMedia in order to manage the image data is OrdSys OrdImage. Its design can be seen below in *Figure 1. OrdSYS. OrdImage Data Design* .

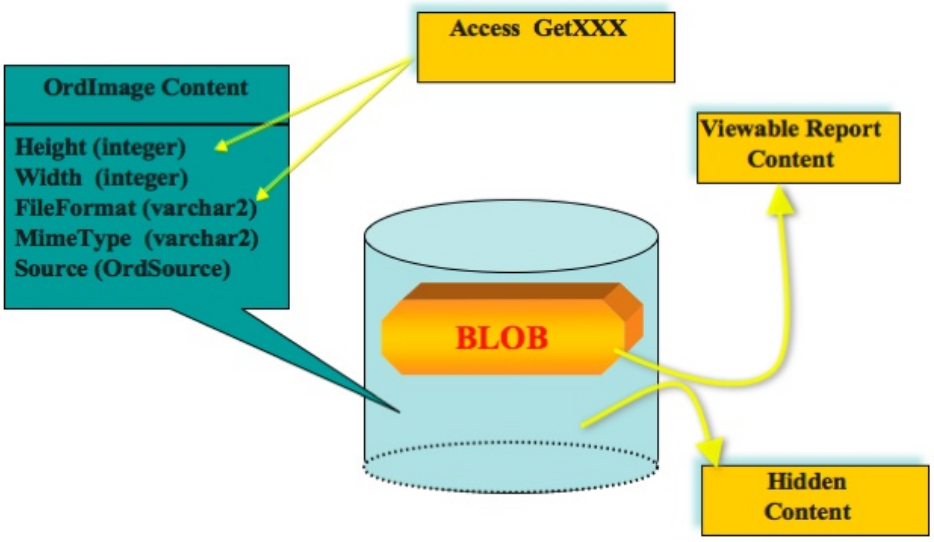


Figure 1. OrdSYS.OrdImage Data Design

As it is a Java based object, besides the Get Methods, OrdImage has as well a set of Set Methods and in all multimedia processing these GET/SET methods appear in pairs, as shown in the below sample PL/SQL procedure:

contrast.prc

```
create or replace procedure contrast(p_cod test_inter.cod%type) is
Image ORDSYS.ORDImage;
begin
SELECT photo INTO Image FROM test_inter t where t.cod=p_cod FOR UPDATE;
Image.process('contrast = 50');
Image.setProperties;
UPDATE test_inter SET photo = Image WHERE cod=p_cod;
COMMIT;
end contrast;
/
```

Figure 2. The content of "Contrast" procedure

Most of the important characteristics of Oracle InterMedia are:

- Searching and indexing/archiving of images records will be most useful if their metadata can be searched. Searching the large images can be efficient only if indexes are available to support the search. In a common way, index searching have been accomplished by complex algorithms that parses image metadata and load it into a series of indexed tables. Oracle InterMedia extract metadata from an image into an XML document, which can be stored in a single XMLType column, in the same table that contains the image column. Indexing this text column offers the robust search capabilities which lead to a faster DML;
- Flexibility: Just after the images are stored in the database, these can be manipulated like any other relational data. Set of images can be updated, deleted, copied to another table(s), by using simple SQL queries or as well PL-SQL code;
- Image manipulation: Once an image is stored in the Oracle Database as an

InterMedia object, it can be manipulated easily: the image format can be changed, RGB (red green blue collor pallete), image scaling, image cropping, image resizing, or image rotation/inversion;

- Space management: Even if the hardware limitation is not anymore problem, still how much space these data is using. Reducing the storage is a requirement for performance feature of data access as well for backup and recovery solutions. Within Oracle Database, this management includes: compressing images; changing the format, will decrease the size of the image, like changing a bitmap image to a jpeg one; remove and check the unused space; resizing and shrinking data files.

It has to be mentioned that all the features of the database itself can be applied to the InterMedia as well, like encryption, auditing access, backup and recovery, data replication, archiving. One of the most important operations in managing high amounts of multimedia data is the creation of data warehouses: features like materialized views and summary management offer a very high speed in information analysis and retrieval. Data warehouses users and applications can search for patterns in order to create/associate and retrieve summary information, which depends on the filter applied to images data, which is used in the search operations.

Oracle InterMedia provides as well, for application development, the freedom to create specific types of operations as the entire metadata system is based on the XML system, a known standard in the n-tier based applications.

3. Multimedia processing

Multimedia processing has always been based on the pixel mapping matrix. This can be achieved by sequential access, which means that each pixel is saved with its characteristics: RGB, lightness and color intensity. The image is first divided in groups of pixels, then the splitting process goes deeper to the pixel level.

Another method used more frequently because of the way graphic information is stored and as well as for the shorter processing time. This is the method used by OCR - Optical Character Recognition Software. A section of the image resource is much easier to process than the entire resource itself. Types of sections that are used by OCR software are presented in *Figure 3. Graphic Sections Used by OCR* .

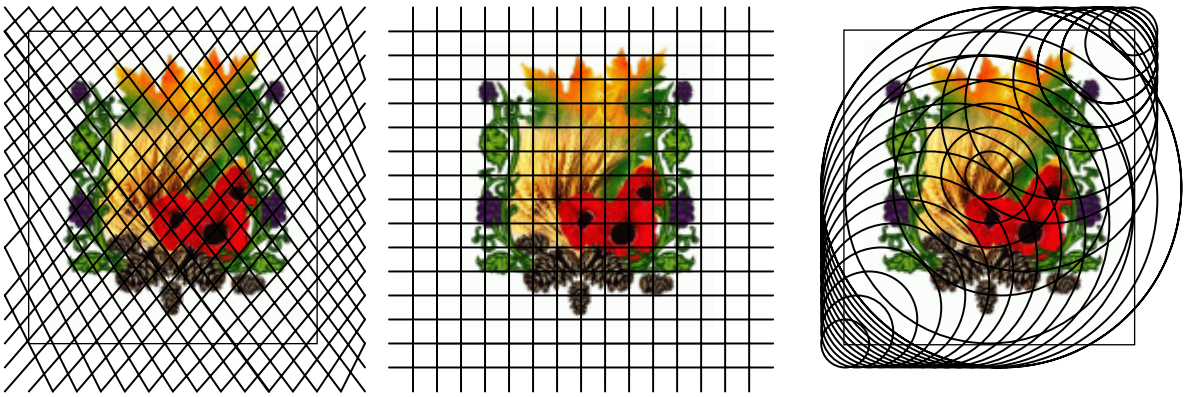


Figure 3. Graphic Sections Used by OCR

Other graphics processing operations are related to altering the color palette of the image resource, the orientation, translation, resizing:

- Transforming the image to grayscale is presented in *Figure 4. Image GrayScale Process*:

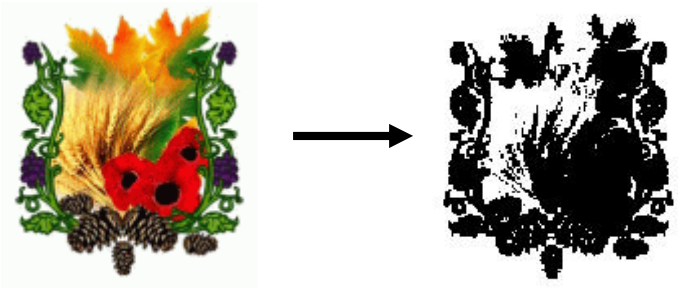


Figure 4. Image Grayscale Process

As a main characteristic OCR handles the recognition of an image by using either predefined patterns or custom ones, which are always based on fragments cropped from the image resource itself.

- Multimedia Mirroring is presented in *Figure 5. Mirroring Process*

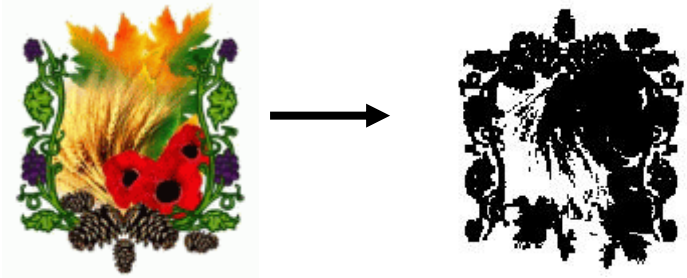


Figure 5. Mirroring Process

- Negative image: is creating by resetting all the pixels, so as each value is calculated by the complementary to the value of black color, like is shown in *Figure 6. Image Negative Process*

Figure 6. Image Negative Process

4. Multimedia Security System

The issue of multimedia security resources is a subject that has to be taken into account by multimedia application developers administrators as well by the users. This can be achieved both on the database level and on the GUI level.

On the database level, security can be achieved by the following:

- On the user level:
  - With the database mechanism of managing users: by granting/restricting access, creating roles, privileges at DDL operations, or/and at DML operations;
  - With LDAP mechanisms by integrating LDAP users into the database. In order to achieve this, Oracle Database comes with the facility of an integrating technology called Enterprise User Security.

- On the data level:
  - Using watermarking as a means to generate modified images;
  - Embedding links on metadata level.

Watermarking is the process of embedding information in content. When watermarking is done by digital means we refer to digital watermarking.
Watermarking classification is achieved by the level of visibility, so as:

- Visible watermarking, where the watermark is visible to the user when the image is read (it can be read through the same means as the image);
- Invisible watermarking, where the watermark is invisible to the user (it cannot be read through regular image reading).

Invisible watermarking can have one of two purposes:

- To transmit the information to the user by indirect means, which assumes that the image reading software is used in conjunction with other software which has the purpose to read the additional information;
- To function as a preconstituted means of proving ownership of the multimedia content in the event that some user might decide to infringe on the rights of the owner (this is usable in a court of law by the means of a technical expertise).

A good example of visible watermarking is superimposing the actual signature of the author or the name of the rights owner. It is advantageous to do this at the moment of content delivery without altering the original content.

Invisible watermarking can be used because of the intention of the rights owner (as a means of controlling infringement or as a means of transmitting additional information) or it can have a technical justification.

For example, the developer can opt to store the information in such a way that, while it is not viewable when the image itself is viewed, it can be read by other tools. This

can be justified by performance constraints or by standardization (to avoid difficulties created by inhomogeneous formats)

Another aspect to be taken into account is that when the information is not intended to be read by the user in any way, we are in fact talking about a steganographic signature.

A steganographic signature is inserted into an object in order to be able to prove ownership in case of infringement as it has no impact on the legal use of the content.

Watermarking implementation can be achieved by either public or private means. In the case of public watermarking, when each image is stored in the database, a copy of it is automatically created in another table. On the backup table, each image will have applied to it a distinguishing watermark. How a watermark should look like: either the name of the database user posted on a specific corner, transversal or any other combination of the text position, or a predefined word like : demo or specimen. It's recommended to create a parallel relation between the word used for watermark and role that the user has in the database. This allows for the use of a set of predefined words can be used: superuser, dba, sysadmin, orcladmin, like is is detailed in *Figure 7. Database watermark*

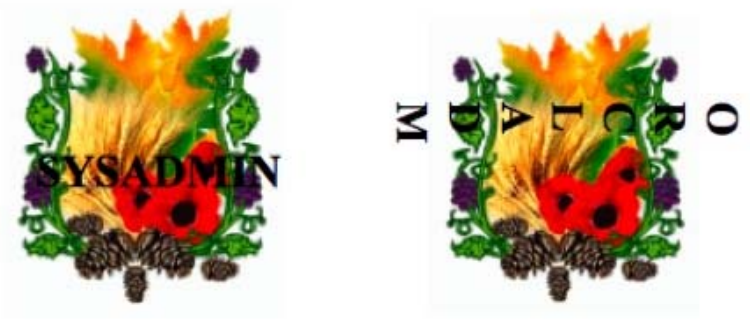


Figure 7. Database watermark

Watermarking can be either public, made automatically by adding the predefined watermark word or the custom one as it is shown in *Figure 7. Database Watermark*, or private. Private watermark can be either visible by the common user or not, this depends of the DML privileges he has on that specific image data. A sample of these custom DML privileges are shown below:

In the same way watermark can be either private or public, through the information that is added to the picture, the watermark can be either visible by the graphic views or human eyes or none. This leads to the conclusion that private watermarks are obtained by altering graphic information at pixel level and involve a small number of pixels so as the change cannot be detected by common graphic tools.

Embedded links on metadata level are a different method of including CMI in digital content is to include a link to the information instead of the information itself. Including a link to the information might lead to proprietary systems where the link only has significance to specific software used so the solution is to include some sort of standardized link such as an object identifier.

A significant argument for the use of embedded links is that the size of the information an owner wants to include in the content can be very large (for example, the owner can have an interest to tie to any type of content the terms of use associated to it) so it might not be practical to include it in the content itself.

Another possible use stems from the fact that while the overhead might not be significant for content of significant size, a standardized system might be preferable.

Implementation of embedded links on metadata level can be done through metadata stored in the database. Just as the graphic information is already provisioned in the watermarking implementation, the metadata information is required to be provisioned in order to provide the source information for embedded links. Why is necessary to create provisioning tables for the metadata information? As a standard for security of database information XML, metadata information created by Oracle InterMedia should not be altered directly, so that the information generated by the system will remain the same as it was at the time the image was stored in the database, so called T0. After the metadata information was provisioned it can be used to create a specific repository for all images that are managed in the database. Also, since the information is XML organized, it can be easily displayed in a web page based form, which can be used afterwards for the embedded links.

Embedded links can be either created dynamically or statically; the method used depends on the security standards and requirements that are mandatory in the implementation of the multimedia database application. The types of the embedded links are either web link based, which are the classically URL used for accessing a web page using the HTTP protocol, or either a document identifier, which can be either the classical path for accessing the document on the disc, or custom ones, by adding in the name of the document a specific coding: for the name of the document, the person who altered it, the version/date of the image resource, or either a random code (an alphanumerical representation for example).

5. Graphics related metrics in multimedia security systems

In order to create metrics systems for a multimedia security system, it is mandatory to first establish the entities involved in it. The main entity involved is human, followed by the graphics resources. Also, a main factor involved for creating the graphics metrics are the letters of the Latin alphabet, as described in *Formula 1*, for the case of the watermarking process:

$$GPW = (TU * FL(U,L) + TUX) * NC \qquad (1)$$

GPW – graphical process index for watermarking
L – total letters of Latin alphabet
TU – total number of users involved in the system
TUX – total number of users integrated from other directory systems
NC – number of colors used by watermark resource

FL – is a function defined on the users set, and the total letters of the alphabet, and it shows how many letters are necessary to write a user identifier

Example

For *U* - user is Scottyy , *FL(Scotty,L) = 4*, the letter *T* repeats twice in the user identifier.

We can consider *L* a constant number of letters. In this case, the *FL* function is not influenced by *L*, but for example, when applying *FL* for a non English/American user set from a Directory Service, letters like "Y", "Q", will not appear in users' identifiers. For example, in the case of a Romanian common/traditional base of names, we will not have those to letters included. Considering the above, it is mandatory to define FL as depending on L.

In Directory Services management, TU, presented in *Formula 2*, depends as well on TO (total operations made in the LDAP system), as it is described in the article "*Identity Management in University System*".

$$TU(TO) = TO + \Phi * TO \qquad (2)$$

- Total operations *TO*
- Total connections *TCN*
- Total authentication failures TF
- Total binds *TB*
- Total unbinds *TU*
- Total searches *TS*
- Total compares *TCC*
- Total modifications *TM*
- Total modifications of user distinguishing name *TMD*
- Total additions *TA*
- Total deletions *TD*

Φ is an indicator that can measure the instability of the LDAP system, due to high amounts of managed users, and takes values in the interval [0, 1]

where:

$$TO = TCN + TCN + TF + TB + TU + TS + TCC + TM + TMD + TA + TD \qquad (3)$$
$$TBV = TB + TU \qquad (4)$$
$$TR = TCN + TF \qquad (5)$$
$$TMO = TS + TCC + TM + TMD + TA + TD \qquad (6)$$
$$TO = TBV + TR + TMO \qquad (7)$$

For the embedded links, the graphics metrics is presented in *Formula 8*.

This depends, by the following:

$$GPEL = GPELD + GPELS + \alpha \qquad (8)$$

GPEL - graphical process index for embedded links
GPELD - graphical process index for dynamically embedded links
GPELS - graphical process index for statically embedded links

α influences *GPEL* an index that represents the number of embedded links not used anymore in system, used very rarely, less than 3 times in a month, or which are invalidated and not removed yet from the system.

The correlation between *GPELD* and *GPELS*, is presented in *Formula 9*, made by Ω – correlation index, so as

$$\Omega = GPELD/GPELS \qquad (9)$$

Ω can be either lesser or greater than 1 depending on the implementation of the multimedia security system.

In conclusion, the proposed multimedia/graphics metrics system can be extended, depending, of course on the level of complexity of the implemented system. In the present article, the definition of the graphics metrics system started from the multimedia security features involved in a real system: watermarking and embedded links. However, the metrics can be extended with RGB metrics, InterMedia transactional metrics. The last ones will be discussed/presented in a future article.

6. Conclusions

In an enterprise organization the necessity of archiving the documents in the databases, not only the textual information, but as a scanned copy as well appears more and more frequently. The necessity of using multimedia databases in document management has thus increased as well. Solutions for the archiving of business documents are taken into consideration as well multimedia databases as a way to implement them.

Why use multimedia databases? First of all because of the scalability advantages stemming from storing large amounts of multimedia data, secondly because of the flexibility of accessing the information, short time for either DDL and DML operations, easily achieved import/export of the information, either by using the facility of the relational databases or using databases files.

Multimedia databases can be used as well in various fields such as medical, cadastral, shipping, mailing, geographical, geodesic, transportation activities and the list goes on; as a common point for these activities are that these involve processing, in a significant percentage, multimedia information and to a lesser extent alphanumerical ones like in the case of financial/banking activities.

Besides the advantages, the multimedia databases also have some limitations, such as those relating to searching the information using a multimedia filter (a filter is an amount of pixels, even random, even consequently, from an image or a specific division or

subdivision of the image). For the moment the only way to search for specific multimedia information is by interrogating the metadata information of the multimedia information. This feature of interrogating the databases records by a multimedia filter would increase the security of multimedia information, for example through the possibility of a user to create a password based on specific pixels from an image; should the picture be altered, the password would be altered as well; in order to change the password the RGB information of a pixel would be changed. This topic is however outside the bounds of the present article and will be discussed in a future paper.

Bibliography

1. Dhananjavan, L. **Oracle Identity And Management – all in one**, ORACLE, USA, 2007
2. Mauro, J. **Oracle interMedia Managing Multimedia Content**, ORACLE, USA, 2008
3. Mavria, S. **Oracle interMedia Feature Overview,** ORACLE, USA, 2005
4. Saha, S. **Oracle Application Server 10g Administration I,II,** ORACLE, USA, 2004
5. Saha, S. **Oracle Application Server 10gr2 Administration I,II,** ORACLE, USA, 2004
6. Velicanu, M. and Rednic, E. **Identity Management in University System**, Revista Informatica Economica, no. 2, 2008, pp. 71-74

[1] Andrei TOMA has graduated the Faculty of Cybernetics, Statistics and Economic Informatics, Economic Informatics specialization, within Academy of Economic Studies Bucharest in 2005 and the Faculty of Law within the University of Bucharest in 2005. In present, he is taking his LLM in International and European Law at the University of Amsterdam and follows doctoral research at the Academy of Economic Studies. He is currently interested in IT Law as well as Computer Science issues and seeks an interdisciplinary approach to legal issues related to Computer Science.

JOURNAL OF APPLIED QUANTITATIVE METHODS

THE ROLE OF AMENITIES AND QUALITY OF LIFE IN AGRICULTURE PRODUCTIVITY RURAL ECONOMIC GROWTH A CASE STUDY OF SINDH

F.M. SHAIKH
Assistant Professor, Department of Agricultural Economics,
SZABAC-Dokri-Larkana, Sindh, Pakistan

E-mail: faizshaikh@hotmail.com

M. SALEEM RAHPOTO
Assistant Professor, Department of Economics,
SALU-Khairpur, Sindh, Pakistan

E-mail: rahpotosaleem@hotmail.com

Abstract: *The present study investigates the role of amenities facilities on quality of life in Rural Sindh and how it reflects the agriculture production in Rural Sindh. A structural model of regional economic growth is estimated using data from 200 rural households by using simple random technique, for this study a structured questionnaire designed to get the reliable results from the chosen sample. Five districts Larkana, Shikarpur, Jacobabad, Khairpur and Sukkur of Rural Sindh where sample has been conducted. It was revealed that education as well amenities facilities has positively associated with the agricultural production. It was observed that rural areas where facilities like, road, education, health and amenities facilities were there they have impact on the Agriculture production and enjoying the highest level of yield compare with the traditional farming activities. In areas where Farmers were educated and they were using all the modern implements and their usage. Results suggested that amenity characteristics can be organized into consistent and meaningful empirical measures that move beyond ad hoe description of amenities. In addition to insights into the influence of local characteristics ranging from tax burdens between amenities, quality of life, and local economic performance exits in various villages of rural Sindh. It was further revealed that only 20% of the villages of all five districts where quality of drinking water was available.*

Key words: *Economic; Growth; Amenities; Development; Quality; Rural*

Introduction

Rural Economic structure in many parts of Pakistan has undergone significant changes over past one decade Traditional good producing sectors such as agriculture and manufacturing are giving way to service producing sectors such as those that support natural resources based leisure activities. Numerous studies have been documented in western countries but not a single study was conducted in Pakistan. The quality of life plays an increasingly important role in community economic development and growth Duller and

Duller, Halstead and duller. In detail review of literature Gotten suggest that the arguments for using amentias attributes s an economic growth tool appear toy be very powerful. Even through such strategy fails to provide jobs in rural areas concern has been expressed those how ever such changes may leads to the high yield performance in agriculture sector in case of Rural Singh. While shift to market based (e.g., extractive and manufacturing activities) to non market based (recreation and retirement) activities are well documented in the rural economic growth literature. (Walker and Duller) The impact in the shift on the structure of regional economies and the well being of rural residents is not well understood the entire market structural changes in Pakistan. Our farmers are mostly illiterate and middleman are getting 10-20% of their income when they are selling their products a well as crop to the different markets, except few owners (progressive farmers) who knows the tactics of the market. This research investigates the nature and extends of economic structural changes in the regional markets in rural areas and how farmer is taking the right decision. The main objectives of this research are (1) the construction of family of consistent measures of mentis and quality of life and (2) the determination of role of amenities and quality of life attributes in production performance and economic growth in rural Sindh.

The current SR model view addressed the different issues of rural house hold and their consumption pattern and production system traditional as well progressive approaches are applied. In general we assume that both households and firms are free to make decision about the production derived both from the consumption of the market good and non market amenities. Progressive farmers are maximized the profit by utilizing modern inputs mechanized operations.

Precisely we construct two central hypotheses in this research.

H_0: Amentias has positive role in agricultural production and economic growth
H_1: Amentias has no impact in agricultural production and economic growth

According to above hypotheses our goal is to examine formally and rigorously the level and degree of this hypothesized relationship as it relates to amentias.
Based upon SR model

$$1)\ P^* = f\ (E^* I^* \backslash\ \Omega^P)$$

$$2)\ E^* = g\ (P^* I^* \backslash\ \Omega^E)$$

$$3)\ I^* = g\ (P^* E^* \backslash\ \Omega^I)$$

Where P, E and I are equilibrium levels of population, production, profit and income level and Ω^P and Ω^E and Ω^I are the set of variables describing initial conditions and other historical information contained in the letter set of information are measure of amentias attributes. Relying on the equilibrium conditions laid out above, a simple linear representation of those conditions can be expressed as

$$4)\ P^*\alpha_{0P} + E^* + \beta_{2P} I^* + SD_{IP}\ \Omega^P$$

$$5)\ E^*\alpha_{0E} + E^* + \beta_{2E} I^* + SD_{IE}\ \Omega^E$$

6) $I^*\alpha_{0I}+E^*+\beta_{2I}I^*+\Sigma\delta_{II}\,\Omega^I$

More over production response, income level likely to adjust to their equilibrium levels with substantial lags (i.e. Initial conditions) Partial adjustment equations to the equilibrium levels are

7) $Pt=Pt\text{-}1+\Lambda_{p}\,(P^*\text{-}Pt_{\text{-}1})$

8) $Et=Et\text{-}1+\Lambda_{E}\,(E^*\text{-}Et_{\text{-}1})$

9) $It=I_{t\text{-}1}+\Lambda_{p}\,(I^*\text{-}I_{t\text{-}1})$

After slight rearrangement of terms this yields

10) $\Delta P=P_t\text{-}P_t\text{-}1=\Lambda_P\,(P^*\text{-}P_t\text{-}1)$

11) $\Delta E=E_t\text{-}E_t\text{-}1=\Lambda_P\,(E^*\text{-}E_{T\text{-}}1)$

12) $\Delta I=I_t\text{-}I_t\text{-}1=\Lambda_P\,(I^*\text{-}I_t\text{-}1)$

Measures of Amenity and Quality of Life Attributes

In current research we propose five broad based indices of amenity and quality of life attributes climate, land, water, winter recreation and developed recreational infrastructure. We capture regions climate conditions such as temperature level of recreational activities, agricultural farming system progressive as well traditional mode of cultivation significantly historical and cultural dimensions. In the set of land variables we want to capture the regions land resources such as the percentage of the acres. The set of available timely water for irrigation as well pure drinking water facilities, roads and other infrastructure facilities.

Data Set and Methodology

A structural model of regional economic growth is estimated and data were collected from 200 rural households by using simple random technique, for this study a structured questionnaire has been designed to get the reliable results from the chosen sample. Five districts Larkana, Shikarpur, Jacobabad, Khairpur and Sukkur of Rural Sindh where sample has been conducted.

Empirical Specification

In the design of the $\Omega^{P,}$ Ω^{E} and ΩI vectors, we follow the logic proposed by Wagner and Duller we hypothesize that there are four broad classifications of factors influence regional economic growth, market, labor, government and amenity attributes. The data drew from the Primary sources as well s secondary sources.

Market

In this category, we attempt to capture the factors that influence the demand side of the regional market. Generally these factors are designed to describe the region's market size and consumption ability. We used five Variables

1. Population of the Area
2. Percentage of the population who are engaged in Farming
3. Percentage of population who are educated.
4. Percentage of the household who direct sell to the market
5. Percentage of the farmers who indirect sell their crop yield in the market.

Progressive Farmer

In this category intend to use the modern implements and input for the better production, and also have knowledge about the markets. The variable that measure human capital flow and income generated from the crop significant to capture the influences of this side of the market on the regional economic growth. The variables we use to capture the progressive farmers are chacteristics are

1. Percentage of the Progressive farmers with high School diploma.
2. Knowledge towards agriculture inputs

Government

The local government's finance is fundamental to the growth potential of the region. On the other hand high taxes on agriculture sector are generally deemed to be determinable to the local economic growth. Yet on the other hand government revenue finances the local infrastructure and public serves that may attract household firms to establish in the region, then eventually simulate the local economy. We used two variables to represent the local Government finance.

1) Agricultural tax
2) Total government spending on Agriculture sector

Amenities

The role of amenity attributes is hypothesized to be a central factor in the economic growth of the rural areas of Sindh through its impact on quality of life within the region. As described above, these five sets of amenity variables are (1) climate (2) recreational infrastructure (3) land (4) Water (5) Consumption pattern

Empirical Results

For this analysis there are two distinct sets of empirical results first, the principal component analysis used to construct the measures of amenity attributes and second the reduced from results of our model SR(Shaikh, Rahpoto) of regional economic growth.

Amenity Measurement

The results from the principal components analysis for the five board measures of amenity attributes are report in Tables 1-4 (see the appendix) the final measure accounts for 33% of the variation of six separate input variables Only February and July temperature do not play an important role in final measure. Higher values of climate measure tend to be associated with southern regions like Jacobabad and Shikarpur. The developed recreational infrastructure measure is intended to capture the role of amentias that tend to be more artificial or man made (table-2). Fourteen separate variables are used to construct this particular amentias attribute measure. Individual measures that that determine the final amenity measure include number of park and recreational departments with in the country the number of cricket grounds and other facilities of sports. The number of play grounds recreational centers is significantly contributed in the economic growth, due to the relatively large number of variables introduced in this measure, coupled with large number of variables not loading into the final principal component measure only 18% of the cumulative variance is explained. The progressive farmer measure intended to describe the nature and tertian and land resources and input used for high growth in productivity. Given these results districts if Upper Sindh would tend to be enjoying the high yield compare with lower Sindh. Again due to the large number of variables not loading into the final principal component measure only 20% of the cumulative variance is explained. The availability of the water for irrigated land captures the water resources available in Pakistan and especially in Sindh. The final principal component measure used for this analysis tends to b emphasis value add businesses associated with water resources and Agriculture sector is depending the water availability and time constraints are also important in the yield of the different crops available in door step-in rural Sindh. The measure as defined by principal components analysis appear to be identifying those districts that tend to have reasonably high levels recreational development combine with amenity based as proposed those area of that have sonly higher level of raw amentias. The final measure of amenity attributes used in the analysis capture recreational activities and opportunities where farmers have enjoy the facilities.

Growth Model

As predicted by theory, Initial conditions play and important role in detraining overall growth levels (table3) .The negative and statistically significant coefficient on initial levels for each respective equation all else constant reinforces prior results of the rural renaissance. Countries like Pakistan that had higher levels of population, employment and precipitate income at the beginning of the period (2005) tided to experience lower tats of overall growth and development. Higher initial levees of population appear to lead to higher employment growth while higher level in per capital income. Given theses results the scone hypotheses laid out above appear to hold true and there appear to be patterns of convergence. In Table-4 where there is big gap in the production performance of progressive farmers compare with the traditional farmers. If you just look at the difference 20-30 percent.

Other interesting results including the strong negative relationship between population percentage and economic growth. A higher percent of the population they are

illiterate and 70% are engaged in Agriculture where only 3% are progressive farmers. The income results coupled with the initial income condition results above, might be due the lower level of income in these areas to begin with. Of the five amenities attributes measures all five appear to play a significant role in the regional economic growth.. Based on the simple reduced from the results some strong pattern between amenities and economic development and growth become apparent. Climate appears to strongly influence growth levels in population, have no role employment growth, and have weak influence on per capita income growth. Given that the regions with high climate amenity scores tend to be retirement destination areas, this result seems to make intuitive sense. Similarly, countries with higher levels of water amenities as measured by our simple principal component index also tend to be associated with higher levels of population and income growth, but water amenities do not appear to influence job growth directly. These results might b capturing retirement migration, the growth in recreational demand for natural resources amenities, or higher end residential areas within the commuting shed of urban areas.

Developed recreational infrastructure is strongly associated with population employment, and income growth rates. Recall that the index is driven by availability of parks, play grounds and other recreational facilities, it was revealed that where facilities and amenities facilities were found those villages farmers were progressive and they have innovative strategies of production distribution and consumption of their yield and market structure where they sell their crop yield.

The final amenity measures, attributes that support winter recreational activities is positively related to growth rate in population, employment , income, and yield performance. Demand for the winter recreational activities such that Cricket, Mallah, Kabadi and football, have been and expected to continue growing in the rural areas of Sindh. In the last in general conclusion that all statistically significant amenity attributes are positively related to the economic growth in rural Sindh.

Conclusions and Policy Implications

As the demand for natural resource amenity attribute increases many rural areas of Pakistan are in position to capitalize economically on available resources endowments While the rural growth and Natural resource management literature are acknowledged the important of amenity attributes in economic performance knowledge of Market and progressive farmers are enjoying the high yield per acre compare with the traditional farmers in Rural Sindh and progressive farmers are more conscious about the amenities facilities in their villages has and At the same time, at lest a part of the skills that citizens of developing countries like Pakistan possess have been acquired in universities and training institutions in developed countries. Education, health and other necessities were available for the villages where progressive farmers are conducting there. Types similarly our amenity measures did not account for any spatial affects across country any advice in to those rural areas that may be said to be amenity poor. The policy implication appears to be simple and straight forward . Rural areas endowed with key natural resources amenities can manage those resources to capture growth more effectively. Given expected levels of growth in the demand f recreational uses of these resources, the future growth and development potential of many rural areas may be additionally tied to range of tourism activities.

References

1. Andrews's **FM comparative studies of life Quality Comments on the Current State of the Art and some Issues for Future Research**, The quality of Life comparative Studies Szlai and FM Andrews ed Newbury Park CA SAGE, 1980
2. Angrist, J., Krueger, A. **Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments**, Journal of Economic Perspectives, vol. 15, no.4, 2001, pp.66-85
3. Barkley D. Henry, M.S. and Bao, S. **The Role of local School Quality and Rural employment and Population Growth**, Rev Regnlstud 28, summer 1998, pp. 81-102
4. Bartik, T., Who, J. **Benefit for the state and the local economic development policies?** Kalamazoo MI: W.E , Upjohn Institute, 1991
5. Beale, C.L., Johnson, K.M. **Tthe identification of the recreational countries in non metropolitan areas of the USA' Population Resources**, Policy. Rev. 17, 1998, pp. 37-53
6. Dissart, J.C., Deller, S.C. **Quality of life in planning literature,** J Planning lit. August,15, 2000, pp. 135-16
7. Dorf R.J., Emerson, M.J. **Determinants of Manufacturing Plant Location for Non Metropolitan Communities in the West North Central Region of the US,** Journal Regl. Sci. 18, April 1978, pp.109-120
8. Duffy N.E. **The Determinants of State Manufacturing Growth Rates: A Tow Digits Level Analysis**, Journal Regl. Sci 34, May 1994, pp. 137 –162
9. Duffy N.E., Dena K.T. **The Effect of Federal wilderness on county growth in their inter mountain Western Untied States,** Journal Regl. Sci. 38, February 1998, pp. 109-36
10. Duller S.C., Tsia, T.S. **Examinations of the Wage Curve: A Research Note,** J. Regl. Anal Policy 28, no. 2, pp.3-54
11. Ellerman, D. **Policy Research on Migration and Development**, World Bank Policy Research, Working Paper No WP-t13, Sussex University, CANADA, access, 2003
12. English D., Marcouiller, B.K. and Cordell, H.K. **Linking Local Amenities with Rural tourism Incidence: Estimates and effects Soc**., Natur. Res., 13, 2000, pp.185-202
13. Fulton, J.A., Fuguitt, G.L.V. and Gibson, R.M. **Break Polistan Migration Streams**, Rural Sociology, 62 Fall, 1997, pp. 363-384
14. Gottlieb, P.D. **Amenities as an economic Development Tool: is There Enough Evidence?**, Econ Develop, quart, August, 1994, pp. 207-85
15. Graves, P. **A life Cycle Empirical Analysis of Migration and Climate**, Journal of Urban Economics, 6, 1979, pp. 135-47
16. Greenwood, M.J. **Humana Migrations Theory Modes and Empirical Studies**, Journal Regl. Sci., 25 November, 1985, pp. 521-544
17. Hildebrandt, N., McKenzie, D. **The Effects of Migration on Child Health in Mexico,** World Bank Policy Research Working Paper 3945, 2005
18. Mansuri, G. **Temporary Migration and Rural Development**, mimeo, World Bank, 2006
19. Nord, M., Cromartic, J.B. **Migration: the increasing importance of rural natural Amenities,** Choices, Third quarter, 1997, pp. 31-32
20. Porell, F.W. **Inter metropolitan Migration and quality of life,** Journal of Rel. Sci., 22, May, 1982, pp.137-158
21. Rudzitis, G. **Amenities increasing draw people to the rural west,** Develop. Perpect 14, August, 1999, pp. 23-28
22. Stagier, D., Stock J. **Instrumental variables regression with weak instruments**, Econometrics 65(3), 1997, pp.5 57-86

23. Walzer N., Deller, S.C. **Rural issues and Trends: Role of Visioning Program Community visioning Programs,** Practices and Experiences N. Walzer ed. New York, Praeger, 1996
24. Walzer, N., Deller, S.C. **Rural issues and trends: role of visionary programs**. Community visionary Programs; Practices and Experiences, N. Walzer, ed. New York, Praeger, 1996
25. Wanger, J.E., Deller, S.C. **Measuring the effects of economic Diversity on growth and stability**, Land Economics, 74,Nov,1998, pp. 541-556
26. Yang, D. **International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks**, NBER Working Paper 12325, 2006
27. * * * **Economic Survey of Pakistan** various issues from 2000-2008
28. * * * **The International Migration Agenda and the World Bank, Managing Risks Enhancing Benefits**, World Bank at http:/www.worldbank.org/migration, 2006

Appendix

Table 1. Principal Component Eigenvectors:

Climate Variables	Eigenvectors
Climate Average Temperature 0.5016	
Average annual precipitation	0.5387
January Temperature	0.5160
January sunny days	0.0391
July Temperature	0.0747
July humidity	0.04300
Cumulative variance explained	46.17

Table 2. Principal Component Eigenvectors: Developed Recreational

zonal departments	0.2368
Urban facilities Variables	**Eigenvectors**
#Tour operations and sightseeing tour operations	0.4884
#Playground and re carrion centers	0.0287
#Private and Public swimming pools	0.0785
#organized camps	0.5550
#Tourist attractions and historical places	0.32739
#Amusement places	0.3559
#fairgrounds	0.6534
#Local and country park	0.0135
#Private and public golf courses	0.0313
#ISTEA funded greenway trails	0.4908
Estimated of acres of urban and built up and from 1995 national	0.3300
Resources Inventory (NRI)	0.6680
Cumulative variance explained	16.69%

Table.3. Estimation on Reduced Form

Intercept-	52.174	74.102	152.007
	(8.511)	(5.730)	(18.949)
Population in 2005	0.0001	0.000.5	-0.0001
	(0.245)	4.00	

Table.4. Progressive farmers comparison to Traditional Farmers in Rural Sindh

No. Acres 10-acres	**Yield of Progressive Farmers**	**Yield of the Traditional Farmers**
Crop Rice	78/acre	45/acre
Wheat	45/Acre	35/Acre
Cotton	35/Acre	20/acre

Survey-2008

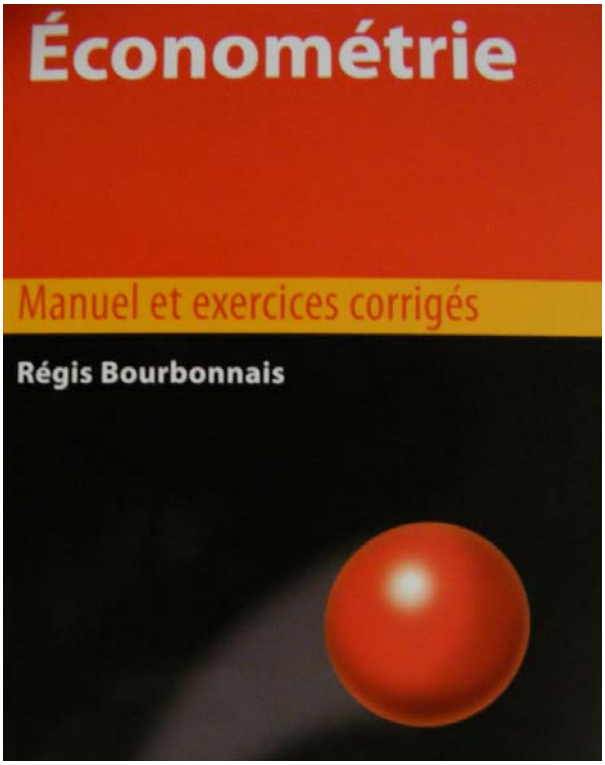
JOURNAL OF APPLIED QUANTITATIVE METHODS

Bogdan Vasile ILEANU
PhD, University Assistant, Department of Statistics and Econometrics
University of Economics, Bucharest, Romania

E-mail: bogdan.ileanu@csie.ase.ro; ileanub@yahoo.com
Web page: http://www.bodo.ase.ro

Key words: Économétrie; Régis Bourbonnais; exercices; time series

Book Review on ECONOMETRICS. COURSE AND SOLVED EXERCISES ("ÉCONOMÉTRIE. MANUEL ET EXERCICES CORRIGES" 7e édition), by Régis BOURBONNAIS, Maison d'Edition DUNOD, Paris, 2009



The book (edited in French) is built on the previous editions. This, the seventh edition, adds a new chapter to the previous –one, "Introduction to econometric Panel- data analysis

This new edition book is structured in 13 chapters in which there are combined in a well-balanced measure the theoretical and practical econometrics.

In the **first chapter** there are briefly presented the concepts, ideas, definitions, role, advantages and disadvantages of econometrics.

After the general overview of econometrics approach the **second chapter** presents in a detailed but simply way the simple regression model. Studying examples, brief mathematical demonstration combined with economical application the reader can obtain solid information about parameters estimations, effects and tests of the regression model's hypotheses, about the validity and quality of the regression model and in the end of chapter we may find the description of one goal of the econometric approach: the forecast.

The **third chapter** extends the econometric analysis to the multiple regression. In the eight subchapters there are presented general problems of the multiple regression such as: parameters estimation, statistical tests, analysis of variance and forecasting, and also

some particular discussion such as: the influence and role of binary variable, the stability and model specification tests. The chapters end with a set of exercises which helps the reader to complete and consolidate his knowledge about econometric extended application.

In the **chapters four and five** the author describes in an extended way the violation of the simple and multiple regression hypotheses. In the fourth chapter there is a focus on the partial correlation and multiple correlation. There is a special approach of the multicollinearity process by treating presence/detection, effects and possibility of correction → optimal model selection. In the next chapter the focus is changed to autocorrelation, heteroskedascity and variables errors. These problems are treated in sense of detection, effects and possibility of avoiding/eliminations or at least of reduction of the negative effects.

A short chapter about non-linears models is the **sixth chapter.** The chapters has a general overview about parameters estimation with particular analyses about exceptions given by non-linear approach. Here we can meet the terms of exponential and polynomial regression function and also the diffusion models.

In the **chapter seven** there are presented the LAG models and their particularities regarding specification, estimation, hypotheses testing and prediction.

Times series in a detailed description are presented in the **chapters nine, ten and eleven**. The description starts with an introduction in time series. The reader will be familiar with concepts like white noise, unit root, stationary process, random walk, non-stationary process, autoregressive (AR), moving averages (MA), ARMA, ARIMA, SARIMA models and Box-Jenkins methodology for model identification . The models VAR, ARMAX and Granger or Sims causality with their particularities are described in the tenth chapter. Concepts tests and model estimation about cointegrated series can be found in the eleventh chapter.

An introduction in econometric approach with qualitative variables is described in the **twelfth chapter**. Here there are presented the particularities of the models which include qualitative variables with more specified cases for the Logit, Probit or Tobit models.

In the **last chapter, the thirteenth** can be found an introduction in panel data econometric analysis. A short presentation of model specifications, homogeneity tests and parameters estimation is also described in a global form. All the examples are made with specific programs used in econometric analysis like RATS, EXCEL, Eviews, etc.

The book ends with a case study and a collection of exercises.

The combination between mathematic theories, economic implication, case studies along with general or particular exercises recommend the book to a large area of readers from many domains.