WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods in Audit and Control

Vol. 4
No. 4
Winter
2009

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

JOURNAL OF APPLIED QUANTITATIVE METHODS

Contents

JOURNAL OF APPLIED QUANTITATIVE METHODS

Contents

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

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

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

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

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

E-mail: r_g-nosca@yahoo.com

We are happy to announce the winners of JAQM 2008 Awards.

After deliberations, the winners are:

1st Category

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

Régis BOURBONNAIS
from **Université de Paris-Dauphine**

the ***2008 JAQM Medal of Honor***

2nd Category

For the most valuable, Quantitative Methods related, paper published in JAQM, we granted to:

Alexandru ISAIC-MANIU
from **University of Economics, Bucharest, Romania**

and

Viorel Gh. VODA
from **Mathematics Institute of Romanian Academy, Bucharest, Romania**

the **2008 JAQM Best Paper Award**

3rd Category

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

Marius Emanuel POPA

from **University of Economics, Bucharest, Romania**

the ***2008 JAQM Distinction***

MULTIDIMENSIONAL MODEL FOR THE MASTER BUDGET

Daniela Ioana SANDU[1]
PhD Candidate at the University of Economics, Bucharest, Romania
Senior BI Consultant at IBM, Milan, Italy

E-mail: sdi78@yahoo.com, daniela.alexandru@it.ibm.com

Abstract: *In a dynamic business environment characterized by extreme competitiveness and the need to quickly adapt to new and changing market condition, information has became an asset. Timely and quality information are the basis for quality decisions, and only quality decisions help survive and prosper on the market. Business intelligence applications help the management take quality decisions. Business Performance Management applications steer the entire organization in the same direction, enabling the organization to translate strategies into plans, monitor execution, and provide insight to improve both financial and operational performance. A BPM implementation often combines financial with non-financial metrics that can identify the health of an enterprise from a variety of perspectives. BI and BPM applications implement multidimensional models, powerful models for data analysis and simulation. The present paper describes a multidimensional model that supports the construction of the master budget of an enterprise with simulation facilities.*

Key words: *business performance management; business intelligence; planning; budgeting*

1. Business Performance Management

Wayne Eckerson from The Data Warehousing Institute defines BPM as being *"a series of processes and applications designed to optimize the execution of business strategies"*, while Lee Geishecker, research director of the well-known research institution Gartner Inc defines BPM as being a set of *„ methodologies, metrics, processes and systems used to monitor and manage an enterprise's business performance"*.

The term *Business Performance Management – BPM* is synonym with the terms *Corporate Performance Management - CPM* , *Enterprise Performance Management - EPM* and means steering the organization in the same direction by allowing the transformation of strategies in plans, monitoring the plans execution and offering detail information on the organization evolution for improving the enterprise's operational and financial performances. A BPM implementation combines financial with non-financial metrics in order to identify the degree of health of the economic organization from a variety of perspectives.

A BPM software solution contains standard components such as:

- Planning, budgeting and forecasting: components that allows defining plans, creating budgets and forecasts;

- Financial and statutory consolidation: a component that allows realizing the consolidated balance sheet at the level of a group of companies belonging to the same holding;
- Scorecarding: a component that allows defining performance metrics and dashboards;
- Reporting and analysis;
- Business Intelligence (BI).

Figure 1. The BPM process

2. Planning, budgeting and forecasting

A *budget* is the translation of strategic plans into measurable quantities that express the expected resources required and anticipated returns over a certain period.

As defined in [8], there are several types of budgets:

- *Short term* (month to month, year to year) versus *long term* budgets (5 years);
- *Fixed* versus *rolling budgets*. A fixed budget covers a specific time frame, usually one fiscal year. At the end of the year a new budget is prepared for the following year. A rolling budget is a plan that is continually adapted so that the time frame remains stable while the actual period covered by the budget changes (for example: as each month passes, the one-year rolling budget is extended by one month so that there is always a one-year budget in place).
- *Incremental* versus *zero-based* budgeting. Incremental budgets extrapolate from historical figures: in determining the budget for the next period managers look at the previous period's budget and actuals (past figures are increased by a set percentage or by an absolute value). In zero-based budgets each new budgeting cycle starts from a zero base or from the ground up as though the budget were prepared for the first time. Zero-based budgets require managers to perform a much more in-depth analysis of each line item.

The budgeting process produces the *master budget*. The master budget brings all the pieces together incorporating the *operating budget* and the *financial budget*:

- the *operating budget* consists of budgets from each function (research and development, production, marketing, distribution, customer service) and provides the budgeted income statement;
- the *financial budget* includes capital budget, the cash budget, the budgeted balance sheet, the budgeted cash flow.

The master budget integrates operational and financial budgets and is created with an iterative process during which information flows back and forth from each element of the master budget.

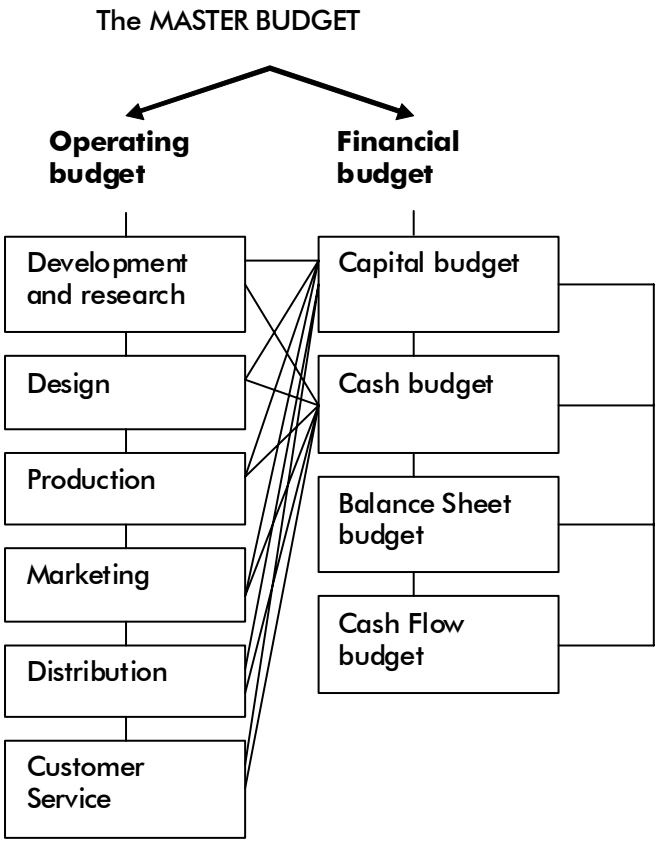


Figure 2. The Master budget components

Planning is a strategic prediction of business performance at a summary level. Plans are defined by senior managers who help the company respond to changing market conditions and opportunities. The process is frequent and must be completed quickly.

Budgeting is planning distributed to individual areas of responsibility across the business. Many more people are involved in the process and the work is done at a greater work of detail. Budgeting is a slow process, often taking weeks and performed once or twice a year.

Forecasting is a revision of the budget (sometimes in at a summarized detail) to reflect changing market conditions, strategic plan alterations, error corrections, revised assumptions in the original approved budget. Organizations typically re-forecast monthly; a handful of finance personnel take part in the process.

Table 1. Planning, budgeting and forecasting (source: [8][2])

	Frequency	**Speed**	**Detail level**	**Personnel involved**
Planning	Often	Quick	Summary	Senior management
Budgeting	Annual	Slow	Highly detailed	All departments
Forecasting	Monthly	Quick	Summary, light detailed	Finance

Budgeting performs 4 basic functions, each critical to a company in achieving its strategic objectives: planning, coordinating and communicating, monitoring progress, evaluating performance.

Planning, budgeting and forecasting are the most important management functions. Their main purpose is to enable senior managers to see the financial implications of various business scenarios. It is a continuous and rapid cycle that provides a near-real-

time response and the most likely business plan scenario becomes the target for the upcoming budget cycles.

Every year companies invest substantially to create comprehensive plans, an annual budget and several forecasts, spending heavily for specialised software, staff overtime and temporary help for data entry. Of these, budgeting it is the most difficult task. Senior managers, accountants, financial analysts, department managers spend countless hours in budget preparation, revision and consolidation. The overall result is as follows:

- organizations spend more time creating a budget than analyzing it;
- most times the budget bears little or no relation to the organization's business plan;
- after the budget it is approved no one looks at it again;
- budget holders dislike very much the tedious and lengthy process of creating, revising and submitting documents;
- budget holders usually attribute adverse variances to the finance department and favourable variable to their own performances and managerial skills. The practice is known as *slack* or *padding* and it occurs when managers believe they are going to be evaluated on their performance relative to budget. To ensure that they will achieve their budgeted figures and be rewarded, the budget revenues conservatively or exaggerate costs or do both.

Most companies use spreadsheets as main budgeting tool. Though spreadsheets are personal productivity tools, they have numerous shortcomings that prevent them from adequately manage a budgeting process of any significant size or sophistication:

- spreadsheets are two-dimensional while budgeting itself is a multidimensional process (example: budget revenue by customer, product, period, version etc);
- spreadsheets are very hard to maintain. Speed and ease in updating a budgeting model is essential for staying abreast of business change. A simple change (adding a cost centre, a department) can mean updating hundreds of spreadsheets and macros.
- Spreadsheets don't integrate well with other systems. A spreadsheet is a single-user tool. With spreadsheets is difficult to share data with other systems (ERP's , OLTP's);
- Spreadsheets models are difficult to share. A spreadsheet is a single-user tool. With spreadsheets is difficult to share data among different worksheets and workbooks. Building a spreadsheet-based solution that consolidates input from multiple users is tedious, time-consuming, very difficult to change and maintain.
- Spreadsheet models are hard to understand: chasing cell references around a spreadsheet or workbook to understand one formula is a frustrating process.

Software solutions for budgeting have to provide flexibility to accurately model the business, support multiple users, easily adapt to rapid change. Budgeting software solutions have to address the disadvantages of spreadsheets-based systems:

- should support multidimensional budgeting;
- allow fast adaptation to changing constraints, assumptions and structures;
- have data import and export functionalities;
- be easy to use for non-programmers (business users should be able to build their own models with IT department intervention);
- should allow calculations for simulations and what-if scenarios.

Leaders on the BPM market, all software providers offer planning, budgeting and forecasting solution:

- Hyperion Solutions (bought by Oracle in 2008) has Hyperion Planning;
- Cognos (bought by IBM in 2008) has Cognos 8 Planning;
- SAP has SAP SEM Intelligence (BPS);

3. Multidimensional models

A model is an abstraction of the real world. Models allow business analysts to give a form, a shape to the unknown concepts and realities. The goal of a data model is to represent in an exhaustive manner the data and information of an organization.

Legacy or OLTP (On-Line Transaction Processing) systems are the organization's operational systems: they are transaction oriented and used to managing day by day activities (input or update of orders, sales, inventory, accounting records etc). These systems use databases that implement relational models in a normalized form. A relational model is composted by:

- Entities: are objects the organization needs to analyze.
- Relationships: describe the way entities interact with each other.
- Attributes are characteristics of the entities.
- Operations such as insert, delete and update are very fast due to the normalized form of the database that allows minimum redundancy.

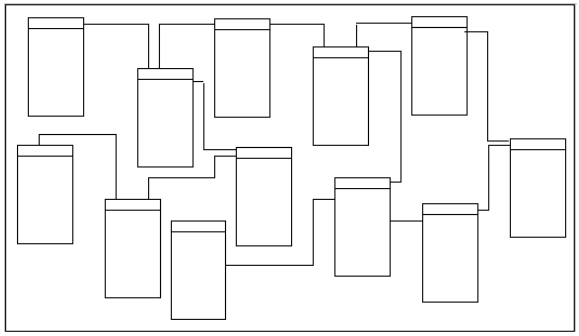


Figure 3. The relational model

Things are not as easy with data retrieval in a relational model. Retrieving the information of a query (the best sold products on all markets during the last 3 months) usually involves joining of more tables in order to find all the necessary data (in real applications joins of 10 - 20 tables are very common). The more tables involved in the query, the more data in every individual table, the more the aggregations to make (calculations like sum, average, count etc), the longer it takes to the query to retrieve the final result. The relational model is not fit for querying. The solution to this problem is a multidimensional model.

The multidimensional data model enforces simplicity by giving up to the minimum redundancy: opposite to the relational model, the multidimensional model is highly de-normalized. De-normalization and redundancy contribute to quick retrieve time because the information doesn't have to be built up from a large number of tables connected by joins.

A multidimensional model is made of two types of tables:

- A fact table: containing the measures, elements that are subject to analysis (sold quantity, price) and on which the query is build;
- Dimension tables: containing the elements on which data is to be aggregated, the analysis axis of the data (product, client, market etc).

The dimension tables (de-normalized) are linked to the fact tables with foreign key forming a star-schema. Sometimes the dimension tables are further normalized by eliminating the low-cardinality attributes in separate tables: the result is a snow-flake schema. When multiple fact tables share common dimensions we have a multi star schema.

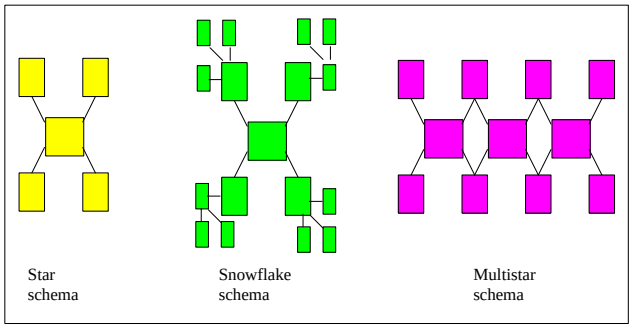


Figure 4. Multidimensional models

A star-schema forms a structure called cube. Despite its name which is a direct expression of the limited capacity of the human brain for space representation, a cube is made of n dimensions.

Dimensions are the analysis axis of data and they are made up of members organized in hierarchies based on father - child relationships. A cube contains a cell for any member combination of its dimensions.

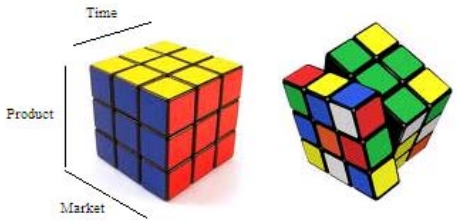


Figure 5. Rubik's cube: a 3 dimensional cube

One of the dimensions is the measures dimension: measures are the key performance indicators that the business analysts want to evaluate. To determine which of the numbers in the data might be measures, the rule to follow is: if a number makes sense when it is aggregated, then it is a measure. For example, it makes sense to aggregate daily volume to month, quarter and year. On the other hand, aggregating names or addresses would not make sense; therefore, names and addresses are not measures. Typical measures include volume, sales and cost.

The hierarchies and the dimensions allow the business analyst to slice and dice the cube according to its needs during data analysis. The cube draws its power from the fact that aggregation (total sales per quarter, per total market, per total client etc) is already calculated and the queries against the cube are very fast in response.
By multidimensional operations (drill up, drill down, pivoting, filtering) the business analyst can slice and dice according to his needs for better understanding of data.

4. Problem analysis

Multidimensional models facilitate and allow fast simulations and what-if analysis through the support they offer in implementing multiple scenarios of ‚best case' and ‚worst case' type.

The paper presents the case of a Romanian steel producing company with subsidiaries in Galaţi, Iaşi, Hunedoara and which sells steel products to national and international clients. The applications was developed used Hyperion Planning 9.3.1.

Here there are some cascading questions a steel production company could make during it's planning and forecasting process. Let's suppose the company has in plan to sell 100.000 tones of steel next year.

1) What if instead will have to sell 150.000 tones or 70.000 tones of steel (because it gains new clients or loses some of its old clients)? What's the impact on the production capacity and the workforce? In order to produce more: it would have to make new hires or buy new assembly lines and ovens? More raw materials would be needed?

2) What if the petrol price will increase? What if the price of the raw materials will increase? What if the Euro – Dollar rates will increase or decrease? What would be the impact on the final price for the client?

The analysis dimensions of the steel company are:

- Year: the years of planning and forecasting;
- Period: with months grouped in quarters (budgets and forecasts are made at a monthly level);
- Scenario: can be Budget, Forecast, Actual (an actual scenario allows comparison with real actual data and helps define the budget);
- Version: version 0, version 1, version 2, version 3 etc (for managing several versions of Budget and Forecast till the final approved ones);
- Client: a dimension containing the company's clients grouped in National clients and International clients;
- Product: steel pipes, tubes of certain length, thickness, hollow section, weight, shape, raw material composition.

The sale price of a tone of steel per product has several components:

- Base price: it is the price that represents the cost of the raw materials that participated in creating 1 tone of product;
- Extra price: it is calculated monthly. For producing stainless steel special metals, with very dynamic prices are used. These metals are nickel, chromium and molybdenum. The extra-price reflects the price of these metals on the stock market and it is influenced by the Euro-Dollar exchange rate.
- Transportation price: is the cost of transportation;
- Other price: other costs in producing steel.

Figure 6. Product and Client dimensions

The simulation needs are the following:

- Tons: what happens with revenues if the society sells an extra tons amount to a certain client, for a certain product, on a certain period or a combination of clients, products and periods?
- Base price: what happens with revenues if the base production price increases or decreases by a certain amount for a certain client, product, period or a combination of clients, products and periods?
- Extra price: what happens with revenues if the extra production price increases or decreases by a certain amount for a certain client, product, period or a combination of clients, products and periods?
- The dollar effect: how does the Euro-Dollar exchange rate affects the price components (base, extra, transportation, other) and what's the effect on the final revenues?

The selling prices are in Euros. The Euro-Dollar exchange rate affects only certain clients (external clients who operate in dollars) on all price components. For clients buying in Euro, the Euro-Dollar exchange rate affects only the extra price.

VAT is calculated only for Europe and simulations on base price and extra price are done only for products that contain nickel.

The phases for realizing plans, budgets and forecasts are:

A. Based on the historical and new contracts the company has with its clients, the commercial department performs an initial prediction for tones to sell to the clients and the selling prices. The prevision is made for tons and price components (base price, extra price, transport, others) for every combination of clients and products, for every month of the year.

B. The planning and controlling department, based on the initial prediction received as input from the commercial department and based on other parameters (the VAT percent, the Euro-Dollar rate) performs a prediction of the profit and loss account "*sold production*". Based on the payment conditions (the number of days the client pays), from the profit account "*sold production*" the balance sheet account "*payments receivables*" is calculated. The controlling department also estimates the remaining profit and loss and balance sheets accounts.

The profit and loss accounts impact on the balance sheet accounts. Part of the balance sheet accounts are calculated using automatic routines from the profit and loss accounts. For example, "*payments receivables*" can be automatically calculated form the "*sold production*" account.

C. The financial department performs simulations and what-if analysis on number of tones sold, the price components, the VAT percent, the evolution of the Euro-Dollare exchange rate. The simulations have a direct effect on the profit and loss account and the patrimonial state of the organization. Simulations are performed in different simulation versions, any simulation version will display it's own profit and loss account and patrimonial state.

5. Multidimensional model design

The prediction of the commercial activity is done at a detail level of year, scenario, version, subsidiary, product, client. The profit and loss and balance-sheets accounts are predicted at a level of detail of year, scenario, version and subsidiary. The different level of granularity for the two activities conduct us to build an application made up of two cubes:

- one cube for predicting the sold production where simulations on tones and prices are to be executed (cube Tons).
- one cube for the master budget made up of the Profit and Loss account and the Balance Sheet (cube MstBdg).

The facts for cube Tons are: tons, prices (base price, extra price, transport, other), sold production, gross revenue, net revenue, client credits.

The facts for cube MstBdg are the accounts in Profit and Loss and Balance Sheet.

The sold production and client credits calculated in cube Tons (granularity: version, subsidiary, month, client, product) feed with automatic routines the members sold production and payment receivables in the MstBdg cube (granularity: version, subsidiary, month).

Table 2. Dimensions for cubes Tons and MstBdg

Dimensions in cube *Tone*	**Dimensions in cube *MstBdg***
Measures (account) Hierarchy „parametri" Hierarchy „masuri vanzare"	Measures (account) Hierarchy „Contul de profit si pierdere" Hierarchy „Bilant contabil"
Year	Year
Scenario	Scenario
Version	Version
Period	Period
Entity (subsidiary)	Entity (subsidiary)
Client	-
Product	-

5.1. The Tons cube

It's goal is to calculate sales value derived from selling tons of different products to different clients. The calculus is performed at a very detailed level (month, product, client). The cube allows fast simulations on tons and prices.

The simulation problem exposed above can be very easy and elegantly solved by choosing the appropriate structure of the account (measures) dimension. The analyst has to design the accounts dimension so that he can fully leverage the potential of the multidimensional calculation engine. In order to satisfy the simulation needs presented above, the appropriate facts structure is presented in the figure below:

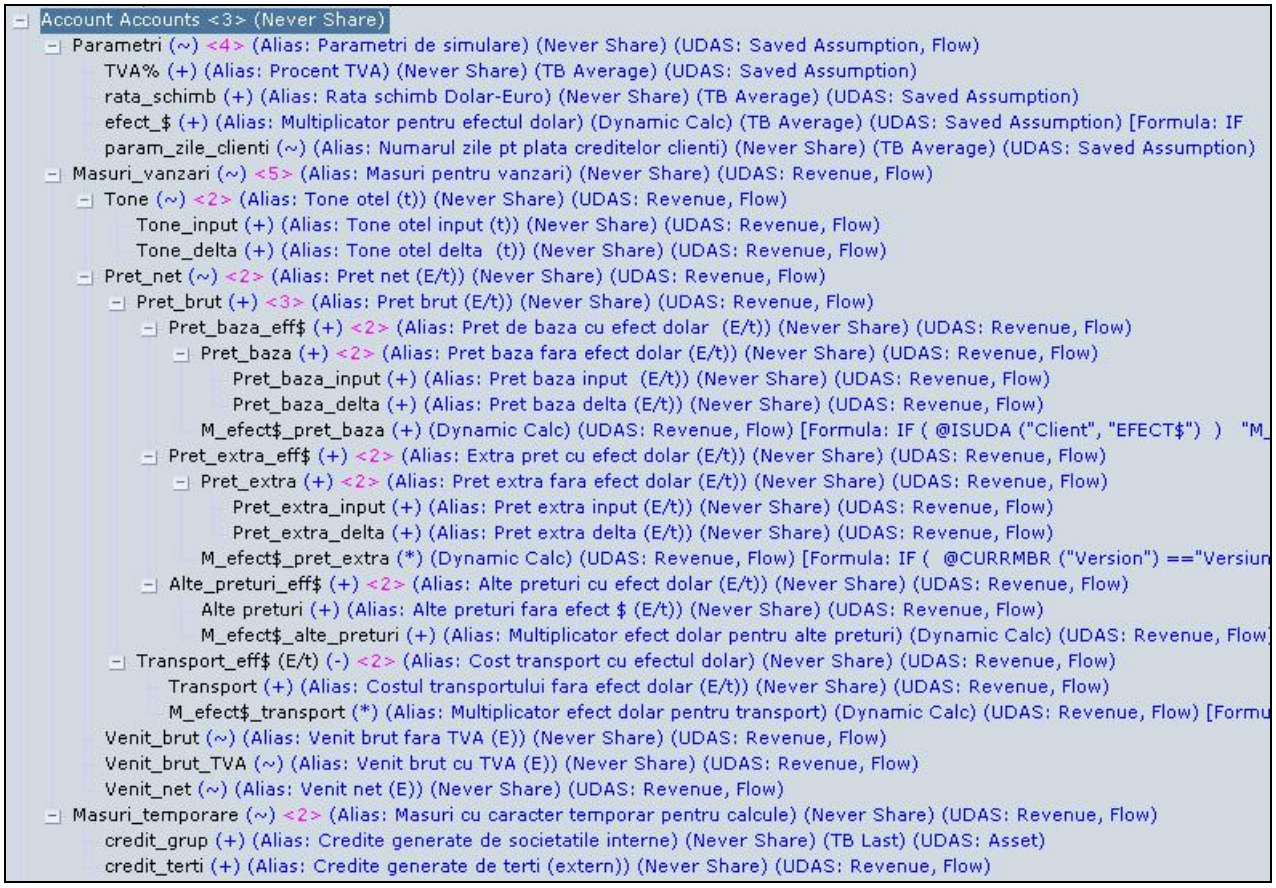


Figure 7. Accounts dimension for cube Tone

Level 0 members are members that don't have children. Data is usually inserted at level 0 member combinations.

Aggregated or upper level members are members that have children. The upper level members are automatically calculated in base of the lower level members and their consolidation operation. The consolidation operation can be:

(+) Adds the member to the result of previous calculations performed on other members (is the default operator);

(-) Multiplies the member by -1 and then adds it to the sum of previous calculations performed on other members;

(*) Multiplies the member by the result of previous calculations performed on other members;

(/) Divides the member into the result of previous calculations performed on other members;

(%) Divides the member into the sum of previous calculations performed on other members. The result is multiplied by 100 to yield a percentage value;

(~) Does not use the member in the consolidation to its parent.

For example:

(1) pret_baza_efect$ = pret_baza * M_efect$_pret_baza;

and (2) pret_baza = pret_baza_input + pret_baza_delta;

Hierarchy 'Parametri' contains:

- 'TVA%' indicates the VAT percent (is fixed for a scenario, cannot vary from one version to another);
- 'param_zile_client' indicate the client paying conditions (number of days the client pays the products he bought) ;
- 'rata_schimb': the user saves in this member the Euro-Dollar exchange rate for the current version of simulation.
- 'efect_$': is automatically calculated by the system and indicates the dollar effect multiplier.

(3) efect_$ = Euro-Dollar exchnge rate of version stasrt / Euro-Dollar exchange rate in current version.

Hierarchy 'Masuri_vanzari' contains members of type:

- input type members (tone_input, pret_baza_input, extra_pret_input, transport, alte preturi) are filled in the start version (version_0) of the BUDGET scenario with the budget data provided by the commercial. This data is the basis for simulations.
- delta type members (tone_delta, pret_baza_delta, extra_pret_delta) are used in simualtions on tones on prices (tons sold increase/decrease by certain tons, base price and extra price increase and decrease by a certain amount).
- dollar effect multiplier members indicate the effect of the Euro-Dollar exchange rate on the price component.
- Gross and net revenues.

Any final price component is calculated with the formula:

(4) price cu efect dolar = pret fara efect dolar * multiplicatorul efect dolar;

The account structure forms the net price as follows:

(5) pret net = pret brut – Transport cu efect dolar
= (pret baza cu efect dolar + pret extra cu efect dolar
+ alte preturi cu efect dolar) – Transport cu efect dolar.

Hierarchy 'Masuri_temporare' contains members:

- credit grup
- credit terti

that are used in order to calculate client credits resulted from selling products to clients. The calculation of client credits is done based on the contractual paying conditions (the number of days after delivery in which the client performs payment), conditions expressed with the member 'param_zile_client'.

For simulation, the business analyst could follow the following steps:

Step 1. Start simulation in version_0

The commercial office of the organization provides an initial version of the budget by inputting into the system:

- Tons and prices (base, extra, others, transportation) for all the combination of products, clients, geography and months;
- The VAT percent.

Step 2. Simulations in version_*n*

- 2.1. In order to keep clear the start data, data in version_0 is copied into version_*n*. Multidimensional software solutions contain data copy functionalities.
- 2.2. The business analyst inserts new exchange rates for the simulation version and the system automatically calculates the dollar effect multiplier.

Step 3. The user simulates tons and price variations in the delta type members by launching specific business rules for simulations. The business rule are parameterized calculation scripts that write down values into the delta members

Step 4. The database is calculated by launching a calc script that calculates various revenue as tons * price.

Multidimensional software contain easy to use scripting languages that allow implementation of calculations (business rules).

Step 5. Using a front-end tool to interrogate the cube, the user can investigate the result of the simulations.

Gross revenues on the simulation versions have changed due to changes of the Euro-Dollar rates, the extra-tons sold and the increase of the base price.

5.2. The Master Budget cube

Its mail goal is to offer support for realizing the predicted *Profit and Loss* and the *Balance Sheet*.

The account dimension contains two hierarchies:

- the Profit and Loss hierarchy accounts ("Contul de profit si pierdere");
- the Balance Sheet hierarchy accounts ("Bilantul contabil propriu-zis").

Figure 8. Accounts dimension for cube Master Budget

The account „Productia vanduta" (sold production) is imported from the Tons cube with an automatic routine (the corresponding fact is „Venit brut") at an aggregated level of total product and total clients.

Account „Creante comerciale interne" and „Creante comerciale externe" are imported from the Tons cube using automatic routines (the corresponding facts are „credit_grup" and „credit_terti") at an aggregated level of total product and total clients.

Very version of the MstBdg cube contains the effect of the simulations performed on the same version in the Tons cube.

6. Conclusions

Multidimensional cubes are a powerful instruments for analyzing huge volumes of data. Regardless of the software used to create multidimensional applications (Hyperion Essbase, IBM Cognos, SAP BW etc), multidimensional analysis draws its power from the following elements:

- The possibility to easily analyze the data by slicing and dicing;
- Calculations are easy to implement;
- Aggregated data is already pre-calculated (that implies a fast response time for queries).

Multidimensional models are used both in analyzing current data and as a support in planning and forecasting processes. The application designer has a major role when designing a multidimensional model for simulations. Designing a multidimensional cube is more than simple work; it can be seen as art considering that an efficiently designed account structure solves a lot of the simulation issues (especially calculations) and saves the model from future re-designing in order to adapt it to new needs and requests.

References

1. Agrawal, R., Gupta, R. and Sarawagi, S. **Modeling multidimensional databases**, Proc. of 13th Int. Conf. On Data Engineering (ICDE), IEEE Press, 1997, pp. 232-243
2. Cabibbo, L. and Torlone, R. **A logical approach to multidimensional databases**, Sixt Int. Conference on Extending Database Technology (EDBT, '98), Springer-Verlag, 1998
3. Cabibbo, L. and Torlone, R. **Quering multidimensional databases**, SEDB, 1997
4. Caraiani, C. and Dumitrana, M. **Contabilitate de gestiune & Control de gestiune**, Ed. InfoMega, Bucharest, 2005
5. Codd, E.F., Codd, S.B. and Salley, C.T. **Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate**, Associates Arbor Software, 1993
6. Feleaga, L. and Feleaga, N. **Contabilitate financiara. O abordare europeana si internationala**, editia a doua, Ed. Economica, Bucharest, 2007
7. Gyssens, M. and Lakshmanan, L.V.S. **A foundation for multi-dimensional databases**, VLDB, Athens, Greece, 1997
8. Ivanova, A. and Rachev, B. **Multidimensional models-Constructing DATA CUBE**, International Conference on Computer Systems and Techonologies – CompSysTech, 2004
9. Kerbalek, I. (coord.) **Economia intreprinderii**, Ed. Gruber, Bucharest, 2007
10. Luecke, R. **Manager's Toolkit: The 13 skills managers need to succeed**, Harvard Business School Press, Boston, Massachusetts, 2004

11. Sapia, C., Blaschka, M. Höfling, G. and Dinter, B. **Extending the E/R model for the multidimensional paradigm**, Advances in database technologies, LNCS Vol 1552, Springer-Verlag, 1999

---

[1] Daniela Ioana Sandu
Thoroughgoing studies in Informatics Systems for Management (graduated 2002) and bachelor degree in Economics Informatics (graduated 2001) at the Faculty of Economic Cybernetics, Statistics and Informatics, University of Economics, Bucharest, Daniela Ioana Sandu is currently a PhD candidate in the field of Economic Informatics at University of Economics and a Senior Business Intelligence Consultant at IBM-Italy. Her interests are in data-warehouses, business intelligence, business performance management, financial management.

[2] Codification of references:

[1]	Agrawal, R., Gupta, R. and Sarawagi, S. **Modeling multidimensional databases**, Proc. of 13th Int. Conf. On Data Engineering (ICDE), IEEE Press, 1997, pp. 232-243
[2]	Cabibbo, L. and Torlone, R. **Quering multidimensional databases**, SEDB, 1997
[3]	Cabibbo, L. and Torlone, R. **A logical approach to multidimensional databases**, Sixt Int. Conference on Extending Database Technology (EDBT, '98), Springer-Verlag, 1998
[4]	Codd, E.F., Codd, S.B. and Salley, C.T. **Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate**, Associates Arbor Software, 1993
[5]	Gyssens, M. and Lakshmanan, L.V.S. **A foundation for multi-dimensional databases**, VLDB, Athens, Greece, 1997
[6]	Ivanova, A. and Rachev, B. **Multidimensional models-Constructing DATA CUBE**, International Conference on Computer Systems and Techonologies – CompSysTech, 2004
[7]	Sapia, C., Blaschka, M. Höfling, G. and Dinter, B. **Extending the E/R model for the multidimensional paradigm**, Advances in database technologies, LNCS Vol 1552, Springer-Verlag, 1999
[8]	Luecke, R. **Manager's Toolkit: The 13 skills managers need to succeed**, Harvard Business School Press, Boston, Massachusetts, 2004
[9]	Kerbalek, I. (coord.) **Economia intreprinderii**, Ed. Gruber, Bucharest, 2007
[10]	Caraiani, C. and Dumitrana, M. **Contabilitate de gestiune & Control de gestiune**, Ed. InfoMega, Bucharest, 2005
[11]	Feleaga, L. and Feleaga, N. **Contabilitate financiara. O abordare europeana si internationala**, editia a doua, Ed. Economica, Bucharest, 2007

STANDARDS REVIEW ON MISSION OF MANAGEMENT INFORMATION SYSTEMS AUDIT

Delia BABEANU
Ph.D, Information System Supervisor, Lecturer,
Faculty of Accounting and Management Information Systems,
University of Economics, Bucharest, Romania

E-mail: delia@cig.ase.ro **Web page:** www.cig.ase.ro

Valerica MARES
Ph.D, Lecturer, Faculty of Accounting and Management Information Systems,
University of Economics, Bucharest, Romania

E-mail: maresvalerica@yahoo.com **Web page:** www.cig.ase.ro

Abstract: *The purpose of auditing is to verify that all hardware and software functions, automated processes, declared and published performance criteria are producing correct results, within the confines of system integrity, security measures and other control mechanisms, in accordance with functionality, originally envisaged, designed or modified.*

Key words: *standards for information systems audit; risks management; information security; IT governance*

Introduction

The scope of an information systems audit may be as far and wide to cover the entire lifecycle of the technology under scrutiny, including the correctness of computer calculations, or a basic test to verify compliance with a specific declared objective. The "scope" is of an audit is dependent on its declared objective, decided upon from the outset.

Audits may be initiated as a result of some concern over the management of assets. The concerned party may be a regulatory agency, an asset owner, or any stakeholder in the operation of the systems environment, including systems managers themselves. Each and every party may probably have an objective in initiating and commissioning the audit. Such an objective may be to validate the correctness of the systems performance or calculations, confirming that systems are appropriately accounted for as assets, to assess the operational integrity of a series of automated processes, to verify that confidential data is not compromised by being exposed to unauthorized persons, or it may even be a multifaceted combinations of the above mentioned aspects in addition to a wider ranging information systems issues of lesser or greater significance to an organisation, which by its very nature, may vary from one place to another. Selected various objectives of an audit will ultimately determine its scope.

The purpose of auditing is to verify if all hardware performances are used according to the software ones, at designed parameters. In order to achieve these normal working

parameters of computer networks have been defined as well as those peripheral devices.

It is important that the audit starts from the results of a previous audit of the company. The existing documents, created by a previous mission, should be analyzed, after which all the subsequent changes to the system will be verified. If these existing documents, as well as the documentation for further changes satisfy the need for information of the auditor, he will proceed to controlling the implementation of the changes.

The time period in which the audit takes place has to be well defined. Collecting samples is done using files that keep the history of the network, user rights, and hardware and software resources.

The audit of information systems is not different from other audits; it consists of the analysis of the systems referring to an activity of the company. In this sense, it is required to define information applications that represent an integrated set of programs, data and administrative procedures. Examples of such applications are: primary accounting applications, salary payment report applications, application for managing stocks, etc. The largest part of information applications are considered processes articulated around various stages like entries, processing, data storing and obtaining results (Nastase, 2007).

Standards presentation

The performed audit is based on current laws, standards and norms. One of these is standards **series 27000**. Standards that can be applied and are part of this series refer to:

The family of standards for SMSI – Information Security Management System (ISO27000 – ISO27010, http://www.iso27001security.com/html/27000.html) which covers the specifications of the system, measurements, an implementation guide, an audit guide and the management of risks.

The following are part of this category:

- **ISO 27000** – fundamental elements and vocabulary (completed at the end of 2008) which:
  - ✓ explain the terminology for all the series of standards 27000 (marketing)
  - ✓ explain basic principles and definitions that vary from one country to another
  - ✓ these principles will have an impact on other standards as COBIT (IT processes) and ITIL (Providing IT services – Service Delivery) and eliminates all confusions
- **ISO 27001** – requirements of a SMSI – Certification Process (is based on ISO 27002)
  - ✓ -certifying SMSI – published in November 2005 and operational on January 30 2006 (www.iso27001certificates.com);
  - ✓ -classification/improvement of the requirements of the PDCA process (http://27001.denialinfo.com/pdca.htm), which covers:
    - o -the scope of SMSI (figure 2), evaluating risks, selecting controls, appliance declaration, reviewing risks, SMSI internal audit, real results and measurements, plan for treating risks and controls
- **ISO 27002** – Good practice code for managing informational systems:
  - ✓ it has 11 sections which treat the protection of informational assets (it was published in April 2007)
  - ✓ -133 detailed controls (based on the process of evaluating risks and the business environment)

- ✓ -covers outsourcing purchasing and delivery services, current issues and management issues, security services at employment and during a contract of an employee, a guide for risk management and managing incidents, mobile remote or distributed communications,

- **ISO 27003** – SMSI Implementation Guide (will be available in 2009)
  - ✓ Implementing the guide that will provide support for the new requirements of the standard
    Annex B of BS7799 Standard - The second part has the following stages: overview, the responsibilities of the management, conformity with governance and rules, human resources and personnel security, managing assets, availability/continuity of business processes, managing informational incidents, access control, case studies for risk management (http://17799.standardsdirect.org/iso-17799.htm)
  - ✓ Implementing a PDCA implies identifying assets, identifying threats, evaluating and treating risks, analyzing and improving controls.
- **ISO 27004** Metrics and measurability of SMSI (at the end of 2008). The objectives of this standard are:
  - ✓ a real evaluation of SI controls and objectives
  - ✓ a real evaluation of a SMSI
  - ✓ offers indicators for management assistance
  - ✓ improving SI facilities
  - ✓ provides entries for SI audit
  - ✓ real communication at the information systems management level
  - ✓ input the process of risk management
  - ✓ output for internal comparisons and benchmarks (i.e. measuring controls and processes performance)

Figure 1. SMSI Planning stages

- **ISO 27005** SMSI risk management (end of 2008)
  - ✓ -a new risk management standard for information security
  - ✓ -risk analysis, evaluating risks from informational security (identifying assets, threats and vulnerabilities
  - ✓ -treating informational security risks (presented in figure 1)

    Annex a – goal

    Annex b – identifying and evaluating assets

    Annex c – common vulnerabilities (http://www.27001.com/catalog/7)

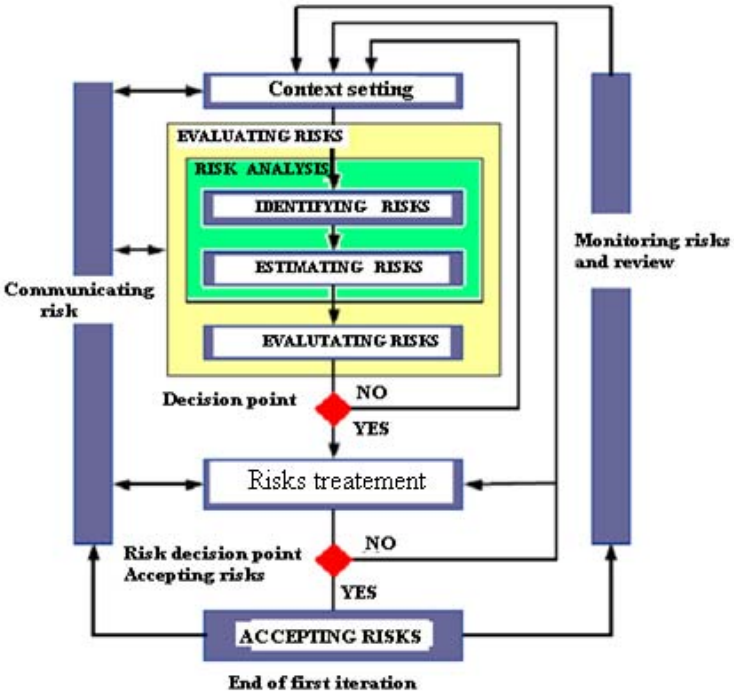


Figure 2. Risks treatment of information security
(**Source:** draft ISO 27005)

- **ISO 27006** – SMSI accreditation guide (certification contents)
  - ✓ necessary for increasing rigors and underlining the contents of certification which is required by the organization (business needs, communications and practice);
  - ✓ Operational from January 2007;
  - ✓ General requirements (impartiality guide);
  - ✓ Organizational structure applying ISO/IEC 17021;
  - ✓ Resource requirements: managerial competence, subcontracts;
  - ✓ Informational requirements – guiding certification results;
  - ✓ Process requirements – guiding SMSI audit.
- **ISO 27007** – SMSI auditing guide (from 2009)
  - ✓ Guide for auditing and SMI auditing content certification accreditation. This family of standards is represented in figure 3: Standards family applicable to a SMSI

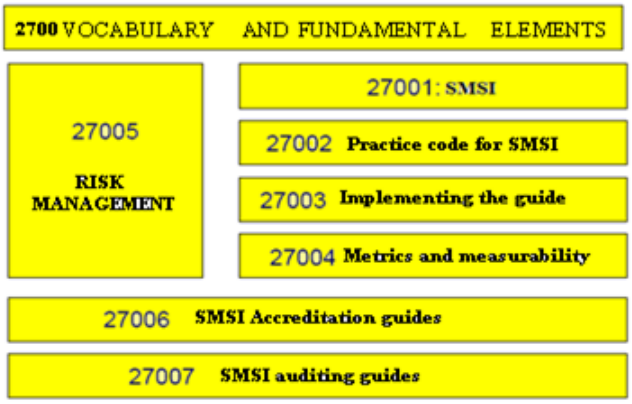


Figure 3. Standards family applicable to a SMSI

- ✓ Specific requirements for certain sectors of the economy (ISO 27011-ISO27030) – Telecom (global) ISO 27011, Health (UK) ISO 27799; Automotive (Germany; Korea; Sweden); Lottery at international level.
- ✓ Operational guide (ISO27031 – ISO27059) for which the publication date has not yet been decided. This series contains:
- ✓ -ISO 27031 ICT standard on business continuity
- ✓ -ISO 27032 cyber security
- ✓ -ISO 27033 - network security.
- ✓ -ISO 27034 - application security.

The pursued objectives can be found in the following table:

Table 1. Informational systems audit objectives

Major objectives	Implementing a good practice
	Evaluating existing or replaceable controls
	Configuring key points for information security
	Reducing frequency/impact of major incidents
Important objectives	Aligning to the internal security policy
	Integrating in the risk management program
	Identification of specific requirements for a certain activity domain
	Increasing existing investments
Other objectives	Increasing competition advantages
	Identification of requirements at the industrial branch level
	Answering a pressure by a third party
	Obtaining a minimum cost

Audit materiality, covered in S12 standard, consists of basic principles and essential procedures clearly identify, which are mandatory together with the guide for the elaboration of these procedures.

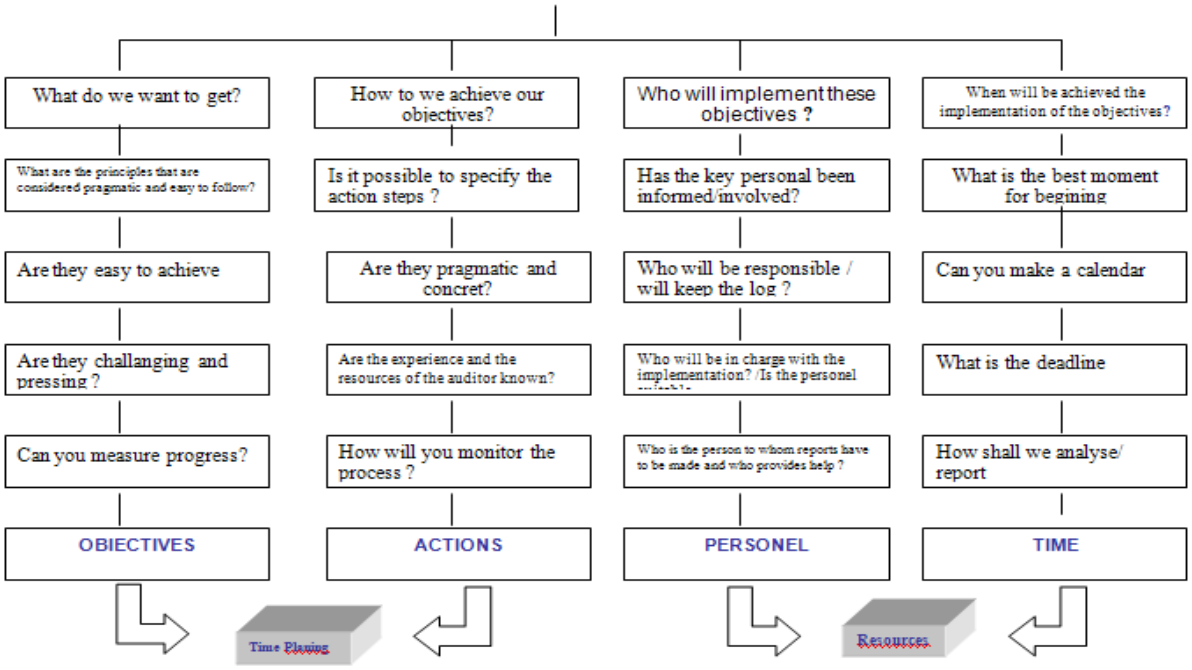


Figure 4. Strategic objectives of the audit
(Adapted for SIG after http://imm.protectiamuncii.ro and http://hwi.osha.europa.eu)

The relevance of materiality consists in the quality of information in a system which a society requires in order to publish all significant information. The materiality threshold is the one from which risks are important in evaluating a society.

This valuation can be done from a quantitative or qualitative point of view and in certain cases (an informational system permanently exposed to risks) a combination of the two methods.

The evaluation of the materiality is a judgment problem and includes considerations over the effect of organization's ability to achieve objectives in events like errors, omissions, irregularities or illegal acts which could substantially modify the results of the controls of threats in the audited area. When evaluating the materiality one has to take into account the errors accepted by the management, by the SI auditor, by the objectives assigned to the system or financial transactions processes, stored information, hardware, architecture and software, infrastructure network. Operating system, development and testing environment (www.isaca.org).

Examples of measures for evaluating the materiality:

- critical business processes supported by the system or applications (data acquiring, processing, reporting etc.)
- databases with critical information from the system or operations
- number and type of developed applications
- number of users that access the informational system
- number of managers, directors that work with classifies information according to their privileges
- critical network communication from within the system or operations
- system or operations cost (hardware, software, personnel, outsourced services, alone or in combinations)

- potential cost of errors (in terms of sales losses, lost guarantee, uncovered development costs, advertising cost required by guarantee, rectifying costs, health and safety, unnecessary production costs, etc)
- number of transactions requested over a period of time
- nature and quantity of manipulative materials
- requirements related to service contracts and costs of penalties
- Other penalties

Reporting materiality supposes determining findings, conclusions and recommendations that are to be reported. Weaknesses control should be considered materiality and reported if the absence of control causes errors in ensuring objectives controls.

Conclusion

Support information and processes, facilities, computer networks and the connection between them are the most important assets of a business. In order to manage these assets and to have business continuities, one has to implement SMSI standards in every company.

We propose transforming the components of the informational system and the information system in values and establishing a threshold for materiality based on value, computed in the respective national currency, which could be taken as the theory of materiality transformed in significance threshold, as is the case of the financial accounting audit.

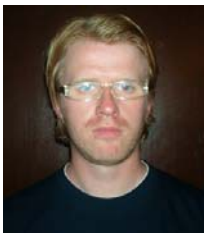
References

1. Nastase, P., Eden, A., Nastase, F., Stanciu, V., Popescu, G., Gheorghe, M., Babeanu, D., Gavrila, A. and Boldeanu, D. **Auditul si controlul sistemelor informationale,** Ed. Economica, Bucharest, 2007, pp. 17-22
2. * * * http://www.isaca.org
3. * * * http://hwi.osha.europa.eu
4. * * * http://27001.denialinfo.com/pdca.htm
5. * * * http://www.iso27001security.com/html/27000.html
6. * * * http://www.27000.org/
7. * * * http://www.27001.com/catalog/7
8. * * * http://17799.standardsdirect.org/iso-17799.htm
9. * * * http://www.iso27001certificates.com

A MODEL OF CREDIT BUREAU IN SERBIA – INSTRUMENT FOR PRESERVING STABILITY OF THE BANKING SECTOR IN CONDITIONS OF THE GLOBAL ECONOMIC CRISIS

Vladimir SIMOVIC[1]
PhD Candidate, Faculty of Organizational Sciences
University of Belgrade, Serbia

E-mail: forumvs@ptt.rs



Vojkan VASKOVIC[2]
PhD, Assistant Professor, Technical Faculty Bor
University of Belgrade, Serbia

E-mail: vaskovic@bvcom.net

Dusan POZNANOVIC[3]
MSc, Public Procurement Agency, Belgrade, Serbia

E-mail: dusan.poznanovic@belit.co.rs

Abstract: *This paper presents the characteristics of the banking system in serbia before and during the global financial crisis. The model of the credit bureau in serbia which, according to its technical characteristics and the level of business performance, represents the original solution is analyzed. Its implementation, in conjunction with other control mechanisms, has provided the stability of the banking sector in terms of crisis. Consequently, the control of liquidity in the banking sector is achieved as well as the control of the expansion of credit activities, with the maintenance of population and economy indebtedness at optimal level, which is of great importance in terms of global crisis when economic policy makers in serbia, faced with a pronounced deficit in balance of payments of the country, as one of economic policy measures aimed at improving the balance of payment position, implement the measure of controlled reduction of private demand.*

Key words: *credit bureau; financial crisis; liquidity risk; loan placement; Serbia*

1. Introduction

World economic crisis has a direct impact on the countries that dominate the international capital flows and international trade. Unlike them, the developing countries and countries undergoing the transition process are indirectly influenced by the crisis which, in the context of the impact on the financial sector, is manifested as a crisis of liquidity and difficulties in development and reform of financial institutions (Vjetrov A. et al. 2009). Serbia belongs to this group of countries.

Manifestation of the crisis of liquidity brings into question the stability of the banking system in Serbia, as the holder of the entire financial system. Our intention is to point out on the credit bureau as one of the instruments for preserving liquidity in the banking sector in Serbia as well as the stability of the same, before and during the crisis.

Information that are analyzed in this paper were collected on the basis of research conducted in the Association of Serbian Banks (ASB) under whose authority functions the first credit bureau in modern Serbian history.

The paper is organized as follows. Section 1 is the Introduction to the theme. Section 2 represents an overview of the characteristics of the banking system in Serbia in the period before and during the global financial crisis, with special emphasis on measures that the National Bank of Serbia (NBS) has undertaken with the aim of preserving the stability of the banking sector in terms of global crisis and shaken confidence of the Serbian citizens in it. Section 3 represents an overview of the functional characteristics of Serbian credit bureau model as well as its specificity which have caused it to be ranked as one of the best in the World by the World bank. Section 4 points out the importance of credit bureau in Serbia as an instrument for the preservation of liquidity in the banking sector.

2. Characteristics of the banking sector in Serbia before and during the World economic crisis

Financial system in Serbia is a network of institutions which consists of 34 banks, 22 insurance companies, 17 leasing companies and 9 voluntary pension funds. Within the financial system of Serbia banks have a very dominant role and run with 90% of total financial assets.

National Bank of Serbia carried out a radical reform of the banking sector during 2001 and 2002 which resulted in the closure of 23 insolvent banks, thus erasing almost 70% of all assets of the banking sector (Ostojic S. 2002). Reform of the banking sector led to a reduction in the number of banks in which the state is the majority owner and increased the number of banks with foreign ownership whose arrival has increased competition in the market and the efficiency of the banking sector. In today's conditions in the balance sheet amount of the banking system in Serbia, the dominant share of 75% are foreign-owned bank.

Since 2004 private sector loans in Serbia recorded dynamic growth which is a consequence of the low starting base and the fact that before 2003 private sector loans nearly were not approved at all. The main source of credit activities was the growing deposit potential, but significant funds were also provided by means of recapitalization of banks and loans from abroad. The trend of increasing dependence on loans from abroad manifested in 2005, began to drop in the second half of 2006. Consequently, in today's conditions, 7.6% of sources of funds of banks in Serbia are those loans (National bank of Serbia, 2008).

The growth of the banking sector in Serbia in 2008 was slowed as a result of restrictive monetary and prudential policy, and in the fourth quarter as a result of the global economic crisis. The period of credit expansion that lasted from 2004 ended in the last quarter of 2008. Since then credit activity achieved the minimum real growth in some months, while retail loans achieved a nominal decrease. This situation is caused by the reduction of demand for loans due to negative macroeconomic trends as well as by the reduction in credit supply due to the bad situation in terms of sources of liquidity and

minimized tendency towards risk by the banks. Retail loans recorded a growth of 20% in 2008, which is significantly below the 54% of real growth recorded in 2007. The decline in retail loans started during the 2007 thanks to the measures of the central bank issued in terms of expansion of retail loans. The fact that retail loans declined gradually is important for preserving financial stability in conditions of financial crisis. The decline of credit activity is mostly pronounced for the retail and cash loans to which the prudential measures of the NBS were targeted in 2007.

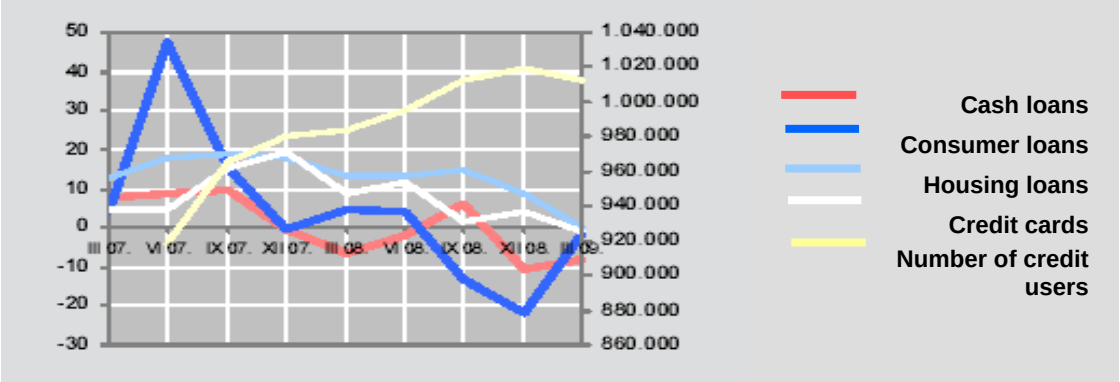


Figure 1. The growth of retail loans by purpose

National Bank of Serbia did not react only to the consequences of the crisis, but a responsible monetary policy and prudential measures in the time before the crisis acted preventively, alleviating the negative effects of the global financial crisis. In this sense, the NBS has implemented a restrictive monetary policy (high key policy rate and withdrawal of excess liquidity) prudential (comprehensive and conservative risk weights, reducing exposure to foreign currency risk, limiting the indebtedness of the population) and administrative measures (high required reserve on foreign currency savings and loans from abroad, limiting the relationship of gross retail placement and capital), and tightened control of commercial banks and also established the first private credit bureau.

In terms of the financial crisis in Serbia it turned out that the greatest vulnerability of the domestic financial system is high share of indexed loans (70%) which increased the foreign exchange and interest rate risk. The fact that the real sector (due to a low share of exports in GDP) and population (due to income in dinars) are exposed to foreign exchange risk is having the great impact on the fact that nominal depreciation leads to increase of defaults and impairs the quality of assets of the banking system.

One of the first effects of the global financial crisis in Serbia, manifested through the withdrawal of deposits from banks. Bad experience from the past, in terms of savings "trapped" in the pyramidal and some state banks, had a negative psychological impact on depositors in Serbia who widely started to withdraw deposits from banks, which have had a short-term negative effect on banks' foreign currency liquidity. In order to restore the shaken confidence of citizens in the banking sector, Serbian Government adopted a set of measures which could be systematized as follows:

- The state guarantee for savings was increased from € 3000 to € 50,000 per depositor if the bank went into bankruptcy. The decision to increase the amount of insured deposit was made on the proposal of the European Commission.

- In order to encourage savings, starting from January 1st 2009, Serbian Government temporary abolished income tax for foreign currency savings, which amounted to 20%. In 2010 this tax will be charged at a rate of 10%.
- Temporarily, until the end of the 2012, capital gains tax (20%) was abolished as well as the tax on the transfer of absolute rights (0.35%) realized through securities trading.

In addition to the Serbian Government, NBS has prepared a set of measures to mitigate the negative effects of the global crisis on the financial sector in Serbia:

- The supervision of the financial system through intensive control of daily liquidity, deposits and foreign currency reserve of banks was reinforced. A new regulatory framework that enables regular data collection on uncollectable receivables was adopted. In addition, the collection and exchange of information on the financial conditions of centrals of the banks that operate in the country was improved and control of financial accounts on a daily basis was reinforced.
- Required reserves for funds taken in overseas aren't being calculated retroactively from October 1st 2009 for bank loans from abroad (untill then the required reserve was 45%), subordinated capital from abroad (20%) and borrowing of th e financial leasing companies (20%)
- National Bank of Serbia has ordered banks to change the structure of the required reserves which are held on the account with the NBS, so that instead of the former 90% of reserves in foreign currency and 10% in dinars, now required reserves consist of 80% of reserves in foreign currency and 20% in dinars.
- Reduction of penalty interest rate for dinars from 31.75% to 23.63% and for foreign currency from 31,75% to three-month Euribor plus 10%

Thanks to the recapitalization of banks, as well as restrictive prudential and monetary policy of central bank, the banking sector in Serbia welcomed the spread of financial crisis with a high degree of resistance. Unlike banks in Europe and America, the Serbian banking sector is well prepared for external challenges. For example, depositors in Serbia are provided with a high level of protection through high required reserve in the amount of 40% for the new foreign currency savings. High capital adequacy of 28.1% (among the highest in Europe) and low dependence on bank loans from abroad, as well as a wide deposit base are mitigating the effects of the liquidity shock. Also, a third of the balance sheet sum of the banking sector are cash, deposits with NBS and securities of the central bank. The fact that local deposits are over 70% of total liabilities confirms the stable structure of sources of funds. Observed by sectors, the domestic deposit base as the primary source of financing for banks is made of population deposits (49% of total deposits). Favorable indicator from the point of the compliance of sources of financing and loans is the fact that the total retail loans are almost completely covered by term deposits of the population, and that ratio is much better than in many European countries. From the point of maturity of term deposits inconsistency between sources of financing and loans is evident - long-term loans exceed the long-term sources of financing multiple times. Thanks to the policy of required reserves on borrowing abroad and the domestic foreign currency deposits, the banki ng sector in Serbia is characterized by coverage of deposits in foreign exchange reserves amounting to 86.3% which is much more than this rate in most countries of the region in which it amounts on average 35%.

Considering the retail loans to GDP the population of Serbia is one of the least indebted in the region. In Serbia, Macedonia and Albania approved retail loans are 12-15% of GDP, while in Romania, Hungary and Bulgaria, that ratio is 20-25% of GDP. In Croatia, this percentage is up to 40%. From the point of indebtedness per capita which currently amounts to € 637, Serbia is in far more favorable position in comparison to some countries in the region in which the indebtedness per capita is up to € 3750 (Croatia). This is a direct consequent of the measures taken by NBS by which instalment of the loan can be up to 30% of monthly income and the introduction of the credit bureau. The expansion of credit activity which is not accompanied by adequate control mechanisms can jeopardize the entire banking system through the emergence of liquidity crisis. The introduction of credit bureau in Serbia aimed to enable optimal alocation of resources based on reliable information for the creditors and in conjunction with measures of NBS to ensure stability of the banking system, which is particularly important in terms of the global economic crisis.

3. Serbian credit bureau model

The model of Serbian credit bureau is the result of observed experiences of models that are more or less successfully applied in practice of a large number of countries. With certain modifications, solutions which were estimated to have been good, were accepted and relying on own resources a model, which for most of its functional characteristics represents the original solution, was developed. The ultimate result is a model of the credit bureau which was ranked by the World Bank as the as one of the best in the World. Credit Bureau in Serbia has started its operative work on 2004 as the result of the initiative of the Association of Serbian Banks with the consent of the National Bank of Serbia and the Ministry of Finance and Economics. It should be noted that the institution of the credit bureau has a long tradition in Serbia, since the Inforamtional Credit Department existed in the Kingdom of Serbs, Croats and Slovenes from 1929, but after the Second World War the whole system closed its opeations (Vaskovic V, 2007).

The basic assumption, and also the biggest advantage of the Serbian credit bureau model is the fact that the creditors (72 members of the credit bureau in Serbia) are solely responsible for the accuracy of the data shown in the credit bureau reports. This fact is ensured thanks to the unique technological process of the credit bureau in Serbia (Figure 2). The central database is located in the credit bureau, and the creditors (banks and other financial institutions) have their own part within the central database. Therefore, banks and other creditors practically rent private space within the information system of the credit bureau and are responsible for its maintenance and the accuracy of data in it.

In the banks and with the other creditors, once a day, usually before the end of business hours, procedures that draw data from their production database (on the credit activities of their clients) are initiated. As a result the documents for the exchange of data with the information system of the credit bureau are generated. Those documents are in XML format, digitally signed and encrypted. It is necessary that the data, which are to be imported by creditors in the rented private database, meet the criteria of validation. Those criteria use the logic, syntax and semantic rules for data filtering and only if all rules are satisfied the data can be stored in the database. This ensures a high degree of accuracy of data that would later appear in the credit bureau reports. Each creditor namely, the authorized person in this institution, can only access its private database, and the credit bureau can access the

private databases of the creditors based only on written consent of the client by which the report is withdrawn from the credit bureau.

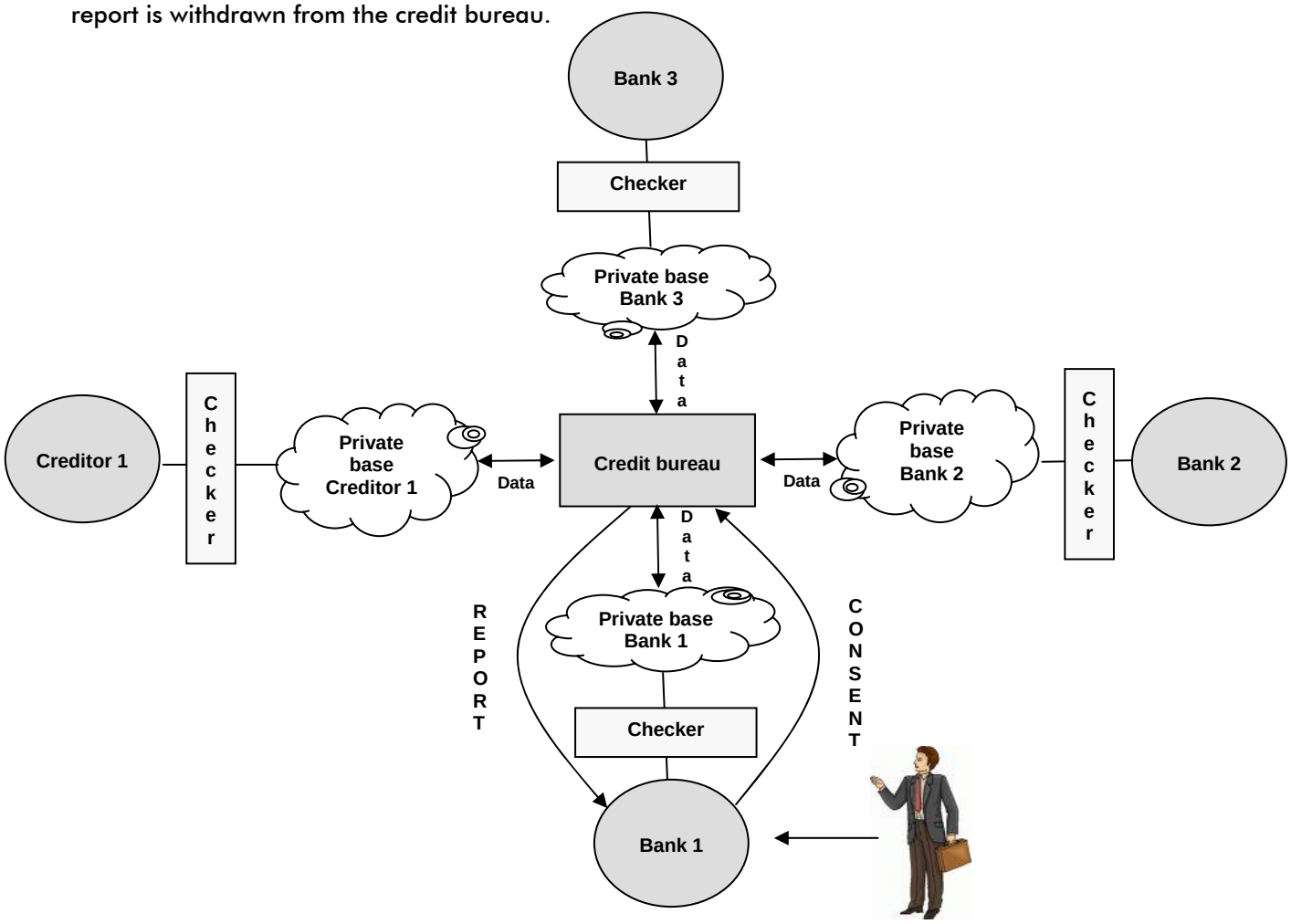


Figure 2. Technological process of credit bureau in Serbia

By the decision of the National Bank of Serbia, which is in accordance with Directive of European Parliament and The Council for Harmonization of Regulations of the Member States, it is regulated that before the loan contract is concluded, the creditor evaluates with all means available, whether the credit applicant regularly pays its obligations under previously approved loans. This means that the creditor is obliged to withdraw the report from the credit bureau. On withdrawing the credit report from credit bureau, the creditor needs to forward to the credit bureau the written consent of a natural person or legal entity for which he asks for report. Then, the data from private databases of all creditors are matched and loaded into the reporting database. The data become accessible and the information is visible in the credit report.

The good side of this model is the fact that the credit bureau can not change in any way the data stored in creditors private databases. Precisely for this reason, all the responsibility for the accuracy of data shown in credit bureau reports is fully transferred to the creditors, which from the credit bureau's point of view significantly facilitates the resolution of eventual complaints. In the case of the complaints on the accuracy of the data shown in credit bureau reports by the end users of credit lines, the request is forwarded directly to the creditor which imported the original data in the rented private space within the information system of the credit bureau. The creditor is obliged to check the

reasonableness of the complaint request and if necessary to make the changes of the data and to forward the complaint request to the credit bureau that monitors its resolution.

In current credit bureau business practice in Serbia on total number of issued reports (8,501,055) there has been 22.147 reported complaints, which is 0,26 % of the gross number of issued reports. This data clearly confirms the fact that creditors in Serbia make their credit decisions based on reliable and high quality information.

Table 1. The number of complaints in relation to the number of issued reports

YEAR	ISSUED REPORTS	NUMBER OF COMPLAINTS	PERCENTAGE
2004	65.206	0	0
2005	1.106.725	1.037	0.09
2006	1.710.999	3.389	0.20
2007	2.641.295	7.220	0.27
2008	2.976.830	10.501	0.35
SUM	8.501.055	22.147	0.26

Growth of the number of complaints is in correlation with the number of issued reports, but still is very low as a result of organizational and technological sophistication of the model of credit bureau in Serbia which by applying data validation rules and eliminating the possibility of errors during the process of data manipulation by the credit bureau provides a high degree of accuracy of data shown in credit reports.

Based on credit bureau report the creditor (bank) decides whether the applicant would be provided with the requested service. The adequate interpretation of the data shown in credit bureau reports is necessary for making the optimal decision by the creditor that minimizes the risk and simultaneously increase the number of users of its services. A model of credit scoring was developed for this purposes and currently it is used only for natural persons. This model is calculating credit score by taking the following factors into account:

1. Orderliness, and/or disorderliness in the discharge of obligations is the first and dominant factor that influences summary score (35% of total score)
- Debit rate is a second factor according to the influence on the summary score and it is 30% of total score
- Third factor is the time necessary for settling of irregularities in the discharge of obligations. The influence of this factor on the overall score is 15%.
- The number of drawn reports from the banks in the past 30 days is the fourth factor, and it influences the score in 10%.
- The length of time for the use of bank services which influences to a certain extent on the overall score in the amount of 5 % is the fifth factor.
- Sixth factor is the number of used bank services and it influences the overall score with 5 %.

One of the essential characteristics of the credit bureau model in Serbia is the fact that the counters of banks and other creditors are used as branches of credit bureau. This organization is possible thanks to the fact that for each charged fee for issuing report from the credit bureau, bank receives 40% of the total sum and the fact that banks and other creditors are responsible for the accuracy of the data shown in credit reports, so that in terms of resolving complaints, the most optimal solution is solving it in a direct contact of client and creditor. Thus, the significant operational expenses for establishing own branches on the territory of Republic of Serbia have been eliminated.

4. Credit Bureau - instrument of securing liquidity in the banking sector

The stability of the banking sector is one of the basic prerequisites of stability of the entire financial system, particularly in transitional economies in which banks have a dominant role in the financial system. The stability of the banking sector, among other things is significantly influenced by different categories of financial risks - credit risk, liquidity risk and market risks. For the purposes of this paper, special attention will be devoted to the issue of preserving the liquidity of the banking system due to the direct impact of the credit bureau on liquidity at bank level, through the control of loan placement. Maintaining liquidity of the banking system is a complex problem that involves coordinated action of central bank and individual banks.

One of the major problems in the operation of commercial banks and other creditors is maintenance of liquidity through monitoring financial discipline of bank clients. Liquidity represents the ability of the bank to have, at any time, adequate amount of funds necessary to finance the growth of assets and timely cover all the obligations that are due. Liquidity risk is the risk of emergence of negative effects on the financial result and capital of banks due to the inability of banks to meet their outstanding obligations. Liquidity risk is one of the leading financial risks in the banking sector, whose sources are often other financial risks like credit and market risks.

Problems with the liquidity of a single bank may have a significant impact on the banking sector and financial system as a whole. Current crisis points out to the great importance that liquidity of bank and the banking system has on the overall financial system and economy both on national and international level.

Liquidity in the banking system in Serbia in 2009 can be described as satisfactory. The average monthly indicator of the overall liquidity of the banking sector in March 2009 is 1.88, which can be considered a satisfactory level due to the regulatory minimum of 1.

Credit bureau has an indirect impact on the liquidity of the banking sector by reducing the credit risk which, as already mentioned, can lead to the appearance of liquidity risk. The introduction of credit bureau institution in Serbia is aimed to reduce the risk of loan placement, which in combination with other factors resulted in maintaining liquidity in the banking sector at the optimal level.

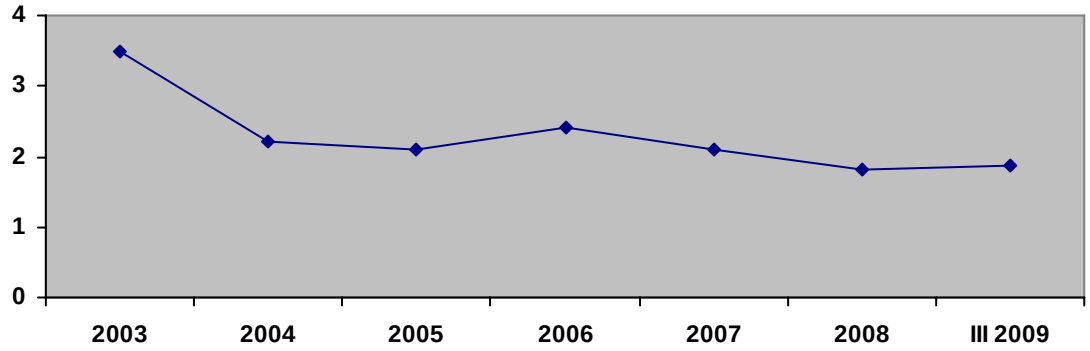


Figure 3. The average monthly indicator of the overall liquidity of the banking sector in Serbia

Data source: National bank of Serbia (Narodna banka Srbije)

The impact of the credit bureau on liquidity movement is difficult to quantify due to the fact that the liquidity is conditioned by many factors. Credit Bureau aims to provide

reliable information that would enable banks to make the adequate credit decision that allows the optimal allocation of resources and minimize the credit risk. Banks use the information on the behavior of their clients in the past, so they can predict their future behavior. The research conducted in banking sector of 34 national economies (Miller M.J, 2003) where the credit bureau operates showed that more than half of the respondents thought that the posibility of using credit information obtained from the credit bureau for making the credit decision, makes the time of loan approval shorter, lowers the costs and the default rate for more than 25%. Impact of the credit bureau on reducing the credit risk, and indirectly the risk of liquidity will depend on the quality of information which are available to the users of its services. Depending on the chosen business concept, the credit bureaus can record positive and/or negative information, about the credit activity of users of the credit lines. The research conducted by the International Finance Corporatioin (International Finance Corporation, 2006) suggests that if the credit bureau records both positive and negative information about the credit users activity, the default rate would decrease by 43% compared to the situation when making credit decisions is based solely on the negative information about credit users past behavior. Credit Bureau in Serbia keeps records in its database containing both positive and negative information on credit history of loan applicant. Due to this fact and the high quality of available information (which is confirmed by the low percentage share of complaints in the total number of issued reports) creditors in Serbia are able to realistically assess the credit risk.

For the purpose of this analysis we would point out the importance of non performing loans and the impact that the introduction of the credit bureau institution had on reduction of percentage share of non performing loans in the total amount of approved loans. According to the methodology of the IMF a loan is nonperforming when payments of interest and principal are past due by 90 days or more (Svartzman I, 2003). Non performing loans are significant due to the fact that their greater percentage share in total loans approved, leads to the reduction of the liquidity of banks, risking to jeopardize the entire banking system. The banks are pushing their efforts to make an optimal credit decisions, based on reliable information obtained from credit bureau, in order to reduce the credit risk and consequently the share of non performing loans in the total amount of approved loans. Institution of credit bureau, besides a direct impact on reducing the credit risk, indirectly affects the reduction of the percentage share of non performing loans in the total amount of approved loans, by increasing the financial discipline of the credit lines users. Figure 4 is a preview of percentage share of non performing loans in Serbia in the period 2003-2009.

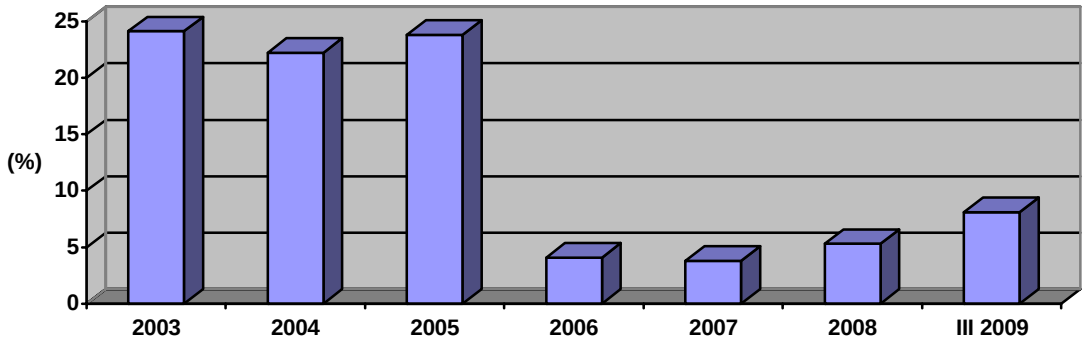


Figure 4. Non performing loans in Serbia in the period 2003-2009
Data source: National bank of Serbia (Narodna banka Srbije)

What is evident from Figure 4 is the fact that the introduction of credit bureau in Serbia had a direct impact on reduction of share of non performing loans in the total amount of approved loans from 23.8% (2005) to 4.1% (2006). Namely, the credit bureau in Serbia has started its operative work on collecting and processing data for natural persons in 2004, and the collection of data for legal enteties started in April 2006. For this reason, the reduction of share of non performing loans in the total amount of approved loans is evident only in 2006 and after, as a direct consequence of the introduction of credit bureau. In this way, the credit bureau confirmed its role as an instrument for the preservation of liquidity in the banking sector in Serbia and in the end the stability of the same, which is particularly important in terms of the World economic crisis. The results, shown in Figure 4, confirm the general expectations that the repayment ability of credit customers would decrease during recession, since the movement of non-performing loans is affected not only by the control mechanisms but also the general market trends.

5. Conclusions

This paper points out the problems of the banking sector in Serbia in conditions of the global economic crisis with special emphasis on measures undertaken by regulatory authorities in order to mitigate the adverse effects of the same. The special characteristics of the Serbian credit bureau model are emphasized as well as its unique technical and technological and organizational structure that makes it an original solution in global terms. Due to the fact that banks and other creditors are solely responsible for the accuracy of data shown in credit bureau reports, high quality of data on which creditors base their credit decisions with a minimized credit risk is ensured. The unique organizational structure of the credit bureau in Serbia, which implies that the credit bureau uses the counters of banks and other creditors as own branches in order to reduce operating costs is also described in the paper.

The hypothesis that the credit bureau is one of the instruments for preserving liquidity in the banking system, and consequently for ensuring the stability of the same, which is of especially great importance in conditions of the global economic crisis, is empirically confirmed in the paper. Analyzing the impact of the credit bureau on the participation of non performing loans in the total amount of approved loans in the period from 2003 to 2009, its positive impact on reducing the liquidity risk as the basic assumption of stability of the banking sector is confirmed.

Literature

1. Miller, M..J. **Credit reporting around the globe**, Credit Reporting Systems and the International Economy, MIT Press, Cambridge, 2003
2. Ostojic S. **Osvrt na iskustva privatizacije – pouka ili ne**, Privredna izgradnja, XLV,1-2, 2002, pp.35-62
3. Svartzman I**. The Treatment of Nonperforming and Interest Accrual,** IMF, May 2003
4. Vaskovic, V. **Sistemi plaćanja u elektronskom poslovanju,** FON, Beograd, 2007
5. Vjetrov A. et al. **Svetska ekonomska kriza i posledice po Srbiju**, Institut fakulteta za ekonomiju, finansije i administraciju, Beograd, April 2009

6. * * * **Credit bureau Knowledge Guide**, International Finance Corporation, World Bank Group, Washington DC, USA, 2006

7. * * * **Report on the state of financial system**, National bank of Serbia, www.nbs.rs, September 2008

---

[1]**Vladimir SIMOVIC** has graduated on the Faculty of Economics, University of Kragujevac. He holds a Master diploma in e-Business at Faculty of Organizational Sciences, University of Belgrade. Also, he is a PhD candidate in the field of e-Business at Faculty of Organizational Sciences, University of Belgrade. His interests are in payment systems, financial management, SEPA initiative, methodology for the prevention of the credit over-indebtedness.

[2]**Vojkan VASKOVIC** enrolled at the Faculty of Organizational Sciences (computer science) and graduated in 1981. He was a post- graduate student during 1984/85. He became a Master of Science in 1989. followed by the Ph.D. in 1997. In the scope of his scientific work up to now, he has published a large number of works, a book (7) and 16 patents out of which 8 are made. He is engaged as a professor on Belgrade Business School. Also, he is engaged as an Assistant Professor on Technical faculty Bor, University of Belgrade.

[3]**Dusan POZNANOVIC** had finished Electrotechnical Faculty, University of Belgrade. Now Dusan is director of IT company "Belit d.o.o." and his specialty is managing different projects, creating new technologies in banking and financial sector. Main projects are: Credit Bureau (national centralized information system for acquisition and distribution of credit record history for individuals and companies), Creating Banking reporting system for National Bank of Serbia and Development of Public Procurement Office web portal.

JOURNAL OF APPLIED QUANTITATIVE METHODS

Software Analysis

DIFFERENT SIMULATIONS OF A BILLIARDS GAME

Lucio T. DE PAOLIS[1]
Assistant Professor, Department of Innovation Engineering,
Salento University, Lecce, Italy

E-mail: lucio.depaolis@unile.it

Marco PULIMENO[2]
BSc, ISUFI - Salento University, Lecce, Italy

E-mail: marco.pulimeno@unile.it

Giovanni ALOISIO[3]
PhD, University Professor, Department of Innovation Engineering,
Salento University, Lecce, Italy

E-mail: giovanni.aloisio@unile.it

Abstract: *Performance improvements in graphics hardware have made it possible to visualize complex virtual environments and provided opportunities to interact with these in a more realistic way. In this paper two different types of Virtual Reality applications for simulating a billiards game are presented. In one application a commercial haptic interface is used to provide a force feedback, thus rendering the interaction realistic and exciting to the user. However, there are limitations due to the use of a commercial haptic device which has not been specifically designed for this game and thus limits the workspace. Also, in the commercial device, it is not possible to use the left hand when aiming and striking the ball, as you can in a real game of billiards. In order to overcome these limitations another type of simulation has been developed using a real billiard cue; its movements are reproduced in the virtual environment using a visual marker detection system. No force feedback is provided to the player.*

In the game simulations the virtual environments have been built using the development environment XVR in the first simulator and OpenSceneGraph in the second; rigid body dynamics have been simulated utilizing the ODE and PhysX physics engines. ARToolkit was the visual marker-based detection system utilized to replicate the movements of the real cue used by the player in the virtual environment of the second simulator.

Key words: *simulations; billiards game; virtual environments*

In the field of computer entertainment new technologies have made it possible to generate new forms of human-computer interaction where some bodily feedback is provided, be it vibration or other, which is popular with players.

Haptic feedback in virtual environments makes it possible to increase the overall realism of a simulation by improving the user's experience and providing a deeper sense of being in control of the game and of participation.

In this paper two different types of Virtual Reality simulations of the billiards game are presented. The first uses a haptic device, in order to provide the user with an interactive and realistic interaction. The force feedback is provided by means of a commercial haptic interface and in this way it is possible to strike the billiard ball and to feel the contact between cue and ball. [4]

The second, in order to overcome the limitations due to the use of a commercial haptic device which has not been specifically designed for the billiards game, uses a different type of simulation which has been developed using a real billiard cue.

By means of a visual maker detection system the cue movements are replicated in the virtual environment, but no force feedback is provided to the player.

Billiards game simulations have been developed both with and without the force feedback sensation.

Gourishankar presents the HAPSTICK, a high fidelity haptic simulation of a billiards game.[5] The system incorporates a low cost interface designed and constructed for the haptic simulation of the billiards game; the device allows motion in three degrees of freedom with haptic feedback along the translation.

Takamura et al. present a billiards game simulation and the method used in this research contributes to making the game extremely realistic.[6]

Visual markers are widely used in Augmented Reality (AR) applications. Currently there are several different types of based marker tracking systems.

Zhang et al. compare several marker systems all using planar square coded visual markers. They present the evaluation results, both qualitatively and quantitatively, in terms of usability, efficiency, accuracy, and reliability.[7]

Wilczynski et al. describe internal structure and potential applications of a newly constructed system for rapid game development in augmented environments. A description of separating marker recognition and display engine of behaviour is provided.[8]

Ohshima, et al. present AR2 Hockey where two users wear see-through head mounted displays to play an AR version of the classic game of air hockey and share a physical game field, hockey sticks and a virtual puck to play in simultaneously shared physical and virtual space.[9]

The Billiards Game Simulation Based on Force Feedback

In the first type of simulation of the billiards game, in order to make the game as interactive and realistic as possible for the user, a force feedback is provided and it is possible to strike the billiard ball and to feel the contact between cue and ball. By means of a commercial haptic interface (PHANTOM Omni) a force feedback is provided, thus rendering the interaction realistic and exciting to the user.

In the game simulation it is possible to distinguish three different types of modelling: graphical, physical and haptic.

The graphical modelling consists of a set of 3D objects built using 3D Studio and imported into the XVR development environment where they are managed using the XVR scenegraph. An example of billiards with five skittles has been implemented.

Since in the real game it is possible to use your left hand when aiming and striking the ball, in the play modality it is possible to fix the cue movement in the desired direction in order to allow a more careful aim and a more stable interaction in the virtual environment. In addition it is possible to choose the force amplification with which the ball is hit.

Each object of the scenegraph is modelled from the physical point of view defining the geometry, the mass, the inertia, the stiffness and the contact friction with another one. The ODE library is used to carry out the physical modelling definition and to define the dynamics for simulating the billiards game.

Regarding the haptic modelling of the objects that are present in the virtual scene, the utilization of the OpenHaptics library makes it possible to exercise control at a lower level of the haptic interface. The cue is modelled as a rigid body and, in the play modality, its position and orientation are linked, using a spring-damper system, to the position and orientation of the haptic interface stylus.

Figure 1 shows the interactions with the virtual environment using a haptic interface.

Figure 1. Haptic interaction

The limitations of the simulation are due to the use of a commercial haptic device which has not been specifically designed for the billiards game. Because of the limited workspace of the haptic device used, it is not possible to perform some shots, which, in the real game, require wide movements in order to be carried out. In addition, it is not possible to use your left hand in order to stabilize the cue and to obtain a more precise stroke, as would happen in a real game of billiards. For this reason some modifications have been introduced in the simulation; in particular it is possible to fix the chosen direction of the cue during the strike and also to decide on the force amplification with which to hit the billiard ball.

The Billiards Game Simulation Based on Marker Detection

In the second simulator of the billiards game, the player is not provided with a force feedback because a real cue is used instead of a haptic interface.

By means of a marker detection system the movements of the real cue are replicated onto the virtual one and this is able to interact with the other objects on the virtual billiards table.

In this way, although players cannot feel the contact with the virtual ball, they can carry out all the game procedures with a real cue and, as in the real game, they can use their left hand in order to stabilize the cue and to obtain a more precise stroke.

Figure 2 shows a game phase using the developed billiards game simulator.

Figure 2. A billiards game phase

Regarding the construction of the virtual environment, the same models utilized in the first simulator have been imported in OpenSceneGraph, the 3D graphics toolkit used in this simulation.

OpenSceneGraph is an open source high performance and cross platform 3D graphics toolkit written in Standard C++ and OpenGL; it is used in many flight simulators, games and virtual reality visualization systems. It includes a wide range of features among which there is a complete scene graph, support for a wide range of image and 3D model formats.[10]

OpenSceneGraph is more compatible with ARToolkit, the software utilized to manage the interactions in the virtual environment, and for this reason it has been chosen over XVR.

To implement the dynamics of the rigid bodies that make up the virtual game environment PhysX was the preferred choice.

The NVIDIA PhysX SDK is a physics engine used in a wide variety of console games and game engines.[11] Like ODE it allows rigid body dynamics simulation and collision detection; in addition it offers a wide range of other features such as simulation of deformable objects, advanced character control, articulated vehicle dynamics, cloth and clothing authoring and playback, advanced fluid simulation.

PhysX is free for non-commercial and commercial use on PC platforms, but it is not open source like ODE.

The visual marker-based detection system which was utilized in order to replicate the movements of the real cue used by the player in the virtual environment is ARToolkit.

ARToolkit is a software library for building Augmented Reality applications and uses square markers each carrying a unique pattern which is a planar bitmap enclosed by a black border.[12]

Pattern recognition proceeds in two stages: recognition of the pattern boundaries and correlation of the interior pattern with the patterns stored in a database.

These markers are observed by a single camera and the tracking software uses computer vision techniques to calculate the marker position and orientation from the captured image.

Markers can be used as a tangible interface to handle virtual artefacts or as user interface elements. Tracking is impeded whenever the marker to be tracked is not fully and clearly visible within the camera image; chances of full visibility can be improved by using several markers fixed to a rigid object.

The offsets between the markers must be well-known and there must be some components in the application which are able to calculate the final position of the object from the valid tracking input.

The accuracy of tracking depends on many parameters in the processing chain: the quality of the camera images, calibration of the camera, lighting, size and visibility of the reference marker, the size of the marker to be tracked. If only one of these factors is not optimally set, the results of tracking may be inaccurate or even unusable.

In the developed simulation of the billiards game it is possible, by means of a webcam, to detect a marker grid used to define the position of the reference system with respect which the movements of the real cue are calculated; these movements are detected by means of a marker placed on the cue.

Figure 3 shows the interactions with the virtual environment using a marker-based detection system.

Figure 3. The marker-based detection system

The use of a second marker on the cue was not considered because it would have had to be placed in the visual field of the camera and thus close to the other one. This solution is not feasible because the second marker would impede the cue movement during the stroke.

The movements of the real cue are replicated on the virtual one that is modelled as a physical body provided with mass of its own. The force applied to the ball is calculated by a physic engine and is based on the speed of the real cue during the stroke.

The utilization of a physic engine as PhysX permits the modelling of the physical proprieties of the virtual objects and hence defines their dynamic behaviour by means of masses, frictions, etc.

Without the utilization of a haptic device, the force feedback due to the contact between cue and other objects of the billiard table is lost, but the use of a real billiard cue overcomes the limitations produced by the use of a commercial haptic interface which is not specific to the billiards game.

A marker-based detection system was preferred to another type of tracking system, such as an optical tracker, because it provides a solution which is both cheap and simple to build.

Evaluation Test

This marker-based simulator, based on a marker detection system, allows the player to handle a real billiard cue and thus to carry out all the strokes permitted in the real game, but no force feedback is provided to the player.

In order to validate the simulator, some tests have been carried out in order to check if the system is also able to detect the rapid strokes normally made in the real game.

To evaluate the performances of the tracking method based on marker detection, a test application has been developed able to store the following positions of the billiard cue detected by means of the tracking system during the stroke. This application makes it possible to draw the trajectory obtained from the following positions detected by the tracking system and to compare it with the linear path of the real cue during a stroke.

In this way it is possible to evaluate the ability of the system to detect the cue positions in the cases of slow and fast strokes and to estimate the validity of the chosen method. The data are also stored for future processing.

In the test phase ARToolKit Plus was chosen for use and just one marker to define the position of the reference system with respect to the movements of the real cue; in this way it was possible to achieve a higher degree of accuracy in marker detection and a reduction in processing time. These improvements could easily be transferred to the application.

The carried out tests highlight that the detection system is able to correctly register the billiard cue trajectory in the case of slow strokes; however, when a rapid stroke occurs, the number of detected cue positions decreases and the real trajectory departs slightly from the ideal one. Figures 4 and 5 show the following positions of the tip cue detected by the tracking system in the case of a slow stroke and the difference between the ideal (purple line) and real (yellow line) trajectories.

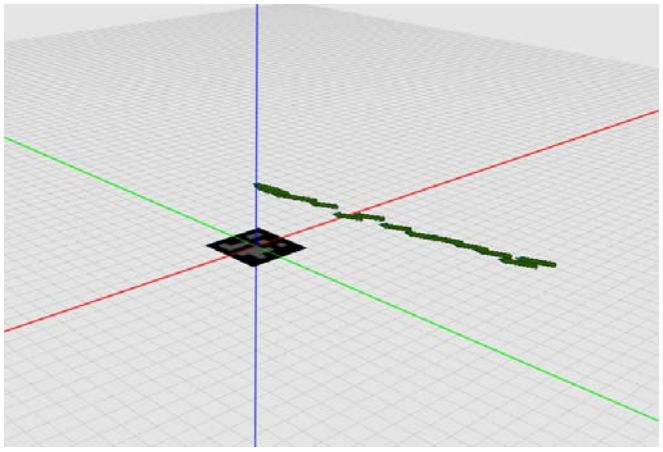


Figure 4. Following positions of the tip cue in the case of a slow stroke

Figure 5. Difference between the virtual and real trajectories in the case of a slow stroke

Figures 6 and 7 show the following positions of the tip cue detected by the tracking system in the case of a rapid stroke and the difference between the ideal (purple line) and real (yellow line) trajectories.

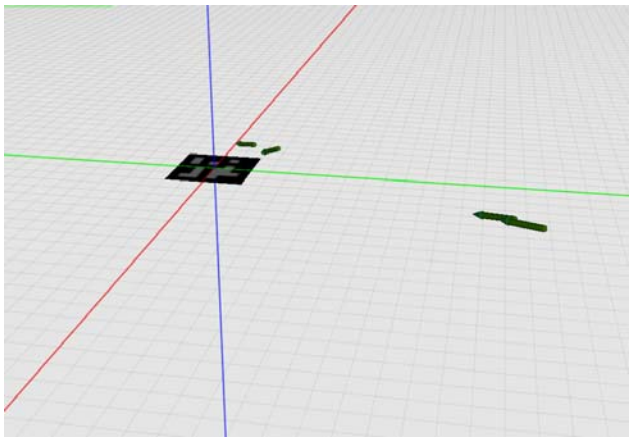


Figure 6. Following positions of the tip cue in the case of a fast stroke

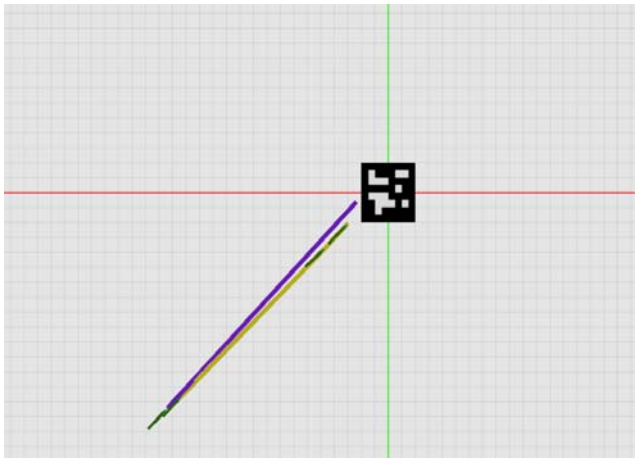


Figure 7. Difference between the virtual and real trajectories in the case of a fast stroke

Future Work

The analysis carried out to obtain a first validation of the marker-based billiards game simulation is qualitative and the test results highlight that the method used to detect the cue movements is not optimal.

It is probable that the use of a webcam provided with higher frame rate and resolution and a more appropriate lighting of the game area would be useful in obtaining better results.

An improvement in the simulation could be obtained by changing the modality of the stroke and splitting it into two different phases; in the first phase of the stroke only the movement of the billiard cue would be detected whereas in the second one the previously acquired data would be processed in order to obtain the correct force to apply to the ball.

In this way it would be possible to detect and to correct the errors due to the tracking system, but it is necessary to verify if the delay due to this processing remains enough short during the simulation.

In addition, a quantitative analysis could be obtained by means of a comparison with measurements obtained using a more accurate tracking system, such as an optical tracker, where the margin of error is well known.

Bibliography

1. De Paolis, L., Pulimeno, M. and Aloisio, G. **The Simulation of a Billiard Game Using a Haptic Interface,** in Roberts D. J. (ed.) "Proceedings of the 11th IEEE International Symposium on Distributed Simulation and Real Time Applications", Los Alamitos, CA: IEEE Computer Society, 2007, pp. 64-67
2. Gourishankar, V., Srimathveeravalli, G. and Kesavadas, T. **HapStick: A High Fidelity Haptic Simulation for Billiards,** in "Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, World Haptics 2007", Los Alamitos, CA: IEEE Computer Society, 2007, pp. 494-500
3. Ohshima, T. *et al.* **AR²Hockey: A Case Study of Collaborative Augmented Reality,** in "Proceedings of the 1998 Virtual Reality Annual International Symposium", Los Alamitos, CA: IEEE Computer Society, 1998, pp. 268-275
4. Takamura, Y. *et al.* **A Virtual Billiard Game with Visual, Auditory and Haptic Sensation,** in Pan, Z. (ed.) "Technologies for E-learning and Digital Entertainment", Heidelberg, Springer Berlin, 2006, pp. 700-705

5. Wilczyński, Ł. and Marasek, K. **System for Creating Games in Augmented Environments,** in "Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering", Los Alamitos, CA: IEEE Computer Society, 2007, pp. 926-931
6. Zhang, X. Fronz, S. and Navab, N. **Visual Marker Detection and Decoding in AR Systems: A Comparative Study,** in "Proceedings of the 2002 International Symposium on Mixed and Augmented Reality", Los Alamitos, CA: IEEE Computer Society, 2002, pp. 97-106
7. * * * **ARToolKit Home Page,** http://www.hitl.washington.edu/artoolkit/
8. * * * **Nvidia PhysX Main Page,** http://www.nvidia.com/object/nvidia_physx.html
9. * * * **OpenSceneGraph Home Page,** http://www.openscenegraph.org/projects/osg

---

[1]**Lucio Tommaso De Paolis** is an Assistant Professor of Information Processing Systems at the Department of Innovation Engineering of the University of Salento (Italy).
He received a Degree in Electronic Engineering from the University of Pisa (Italy) and since 1994, first at the Scuola Superiore S.Anna of Pisa and then at the University of Salento, his research interest has concerned the study of the interactions in the virtual environments and the development of Virtual Reality and Augmented Reality applications in medicine and surgery.
He is a member of the Society for Medical Innovation and Technology (SMIT), member of the Southern Partnership for Advanced Computational Infrastructure Consortium (SPACI Consortium) and member of the Italian Movement of Modelling and Simulation (MIMOS).
He teaches Computer Science at the Faculty of Sciences of the Salento University.

[2]**Marco Pulimeno** is contract worker at the Salento University where he obtained a Bachelor Degree in Computer Science Engineering. His main interests include Augmented and Virtual Reality technologies applied to gaming and medicine.

[3]**Giovanni Aloisio** is Full Professor of Information Processing Systems at the Engineering Faculty of the University of Lecce. His research interests are in the area of High Performance Computing, Distributed and Grid Computing and are carried out at the Department of Innovation Engineering of the University of Lecce. He has been a co-founder of the European Grid Forum (Egrid) which then merged into the Global Grid Forum (GGF). He has founded SPACI (Southern Partnership for Advanced Computational Infrastructures), a consortium on ICT and grid computing among the University of Lecce, the University of Calabria and HP Italia. The Consortium is a follow-up of the SPACI project funded by the Italian Ministry of Education, University and Technological Research, to pursue excellence in the field of Computational Science and Engineering. He is the author of more than 100 papers in refereed journals on parallel & grid computing.

[4] De Paolis, L., Pulimeno, M. and Aloisio, G. **The Simulation of a Billiard Game Using a Haptic Interface,** in Roberts D. J. (ed.) "Proceedings of the 11th IEEE International Symposium on Distributed Simulation and Real Time Applications", Los Alamitos, CA: IEEE Computer Society, 2007, pp. 64-67

[5]Gourishankar, V., Srimathveeravalli, G. and Kesavadas, T. **HapStick: A High Fidelity Haptic Simulation for Billiards,** in "Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, World Haptics 2007", Los Alamitos, CA: IEEE Computer Society, 2007, pp. 494-500

[6]Takamura, Y. *et al*. **A Virtual Billiard Game with Visual, Auditory and Haptic Sensation,** in Pan, Z. (ed.) "Technologies for E-learning and Digital Entertainment", Heidelberg, Springer Berlin, 2006, pp. 700-705

[7]Zhang, X. Fronz, S. and Navab, N. **Visual Marker Detection and Decoding in AR Systems: A Comparative Study,** in "Proceedings of the 2002 International Symposium on Mixed and Augmented Reality", Los Alamitos, CA: IEEE Computer Society, 2002, pp. 97-106

[8]Wilczyński, Ł. and Marasek, K. **System for Creating Games in Augmented Environments,** in "Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering", Los Alamitos, CA: IEEE Computer Society, 2007, pp. 926-931

[9]Ohshima, T. *et al*. **AR²Hockey: A Case Study of Collaborative Augmented Reality,** in "Proceedings of the 1998 Virtual Reality Annual International Symposium", Los Alamitos, CA: IEEE Computer Society, 1998, pp. 268-275

[10]* * * **OpenSceneGraph Home Page,** http://www.openscenegraph.org/projects/osg

[11]* * * **Nvidia PhysX Main Page,** http://www.nvidia.com/object/nvidia_physx.html

[12]* * * **ARToolKit Home Page,** http://www.hitl.washington.edu/artoolkit/

DATABASE SECURITY - ATTACKS AND CONTROL METHODS

Emil BURTESCU[1]

PhD, Associate Professor, Department of Accounting and Management Informatics, University of Pitesti, Pitesti, Romania

E-mail: emil.burtescu@yahoo.com, emil.burtescu@upit.ro

Abstract: *Ensuring the security of databases is a complex issue for companies. The more complex the databases are the more complex the security measures that are to be applied are. Network and Internet connections to databases may complicate things even further. Also, each and every additional internal user that would be added to user base can create further serious security problems. This pupose of this paper is to highlight and identify the main methods and facets of attack on a database, as well as ways to deflect attacks, through focusing on the delicate issue of data inference.*

Key words: *attack; control; impact; inference; security*

Like all tangible assets that have to be protected by a company, valuable information stored in its computer system database is probably the most precious of assets of the company that must be protected.

Safety measures must be an integral part of any database, right from the start, at the inception and design phase. Modern approaches employed to assure the security of databases address security and protection defenses at all levels: physical, network, host, applications and data.

It goes without saying that the first of such measures has to be applied starting at the physical level and to then progress right through, reaching the data level at the other end. Initially, companies have had a rather simplistic approach, mainly due to primitive and rudimentary nature of earlier attacks, as well as the simple nature and construction of the then prevalent networks with very limited complexity if any, and did therefore focus on assuring security at the physical level. That then involved basic measures such as limiting access to locations that only authorized personnel may have access to data.

More recently, due to the rapidly changing and increased size as well as complexity and expansion of company information systems, AAA type measures began to be used (Authentication, Autorisation, Access).

Currently the necesary security measures are far more complex. These are meant to stop the highly sophisticated attacks from external attackers, and especially, from those who may very well have access to the company's internal network.

1. Classical attacks

The focus of attacks on the company's databases are motivated by the following factors:

- Databases are the mass of information which the company works with;
- Databases can reveal private data by processing public data.

Database security is relative in the next situations:

- Theft and fraud;
- Loss of confidentiality/privacy;
- Loss of privacy;
- Loss of integrity;
- Loss of availability.

The hazards which make these things happen are due in large amount to deliberate human action. Natural type hazards or random events have an impact only on data integrity and availability.

To ensure a minimum security of the databases the following requirements must be satisfied:

- Physical integrity of databases;
- Logical integrity of databases;
- The integrity of each element which composes the database;
- Access control;
- User identification;
- Availability.

The physical and **logical integrity** of databases will require the focus of efforts for protecting the physical integrity of databases, especially the recordings against destruction. The easiest way to do that is represented by regular backups.

The integrity of each element forming the database will assume that the value of each field may be written or changed only by authorized users and only if there are correct values.

The access control is being done taking into consideration the restrictions of the database administrator. DBMS will apply the security policy of the database administrator (DBA).

This must meet the following requirements:

- **Server security**. Server security involves limiting access to data stored on the server. It is the most important option that has to be taken in consideration and planned carefully.
- **Connections to the database**. Using the ODBC will have to be followed by checking that each connection corresponds to a single user who has access to data.
- **Access control table**. The access control table is the most common form of securinga database. An appropriate use of the table access control involves a close collaboration between the administrator and the base developer.
- **Restriction tables**. Restriction tables will include lists of unsure subjects who could open set off sessios.

Secure IP addresses. Some servers may be configured to receive only queries from hosts that are in a list. Oracle servers allow blocking queries that are not related to the database.

Cancellation of the Server Account. The ability to suspend an account when guessing the password is tried after a predefined number of attempts (usually 3).

Special tools. Special programs such as Real Secure by ISS which will alert in case of intrusion attempts. Oracle has an additional set of authentication methods: Kerberos security; Virtual private databases; Role-based security; Grant-execute security; Authentication servers; Port access security.

User identification will allow at any time to be known who does anything in the system. All the operations performed by users will be stored and will form a history of access. Checking the history of all hits is sometimes hard and requires a considerable workload.

Availability will allow the required data to be available for an authorized user.

2. Attacks specific to the databases

Unlike other types of data, databases may be subject to unilateral actions, in which an unclassified user has access legitimately to public information but on which he is able to infer classified information. These types of actions are called inference.

After such actions two situations are distinguished which lead to the disclosure of secret data from public data: data aggregation and association.

Two cases of inference which often appear in databases: **data aggregation** and **data association**.

Data aggregation problem arises whenever a set of information is classified at a higher level than individual levels of involved data.

Example: Military field - Individual location of vessels is unclassified, but the overall information about the location of the entirefleet is secret. Commercial - Total sales reports of different branches of the company can be seen as less confidential than the global reports of the company.

Data association problem arises whenever two values taken together are classified at a higher level than the one of each value.

Example: The list containing the names of all employees and a list containing salaries are unclassified, and a combined list containing the names and the salaries of employees is considered classified.

A first step in countering these types of attacks is the protection of sensitive data-data that must not be made public. It is considered as being sensitive data facts that are inherently sensitive, from a sensitive source, are declared sensitive, come from a recording or an attribute which is sensitive or are not sensitive in relation with other sensitive data.

Applying one or more methods of attack, and in combination with a weak protection of databases, several sensitive data types may be displayed:

Accurate data. When the database does not implement any protection mechanism, the extracted data is exatcly the exepcted ones. Queries are simple and obvious.

Bound data. In this situation an attacker can determine the range of values which the searched value can have.

Existing data. Data are classified but which can be emphasized that the existence by a process of inserting data protection mechanisms, operation refused by the protection mechanisms of the database because the data already exist.

Negative data. After some seemingly innocent queries sensitive data can be displayed. A query will be able to display data whose existence is not known, these being sensitive.

Probable data. Their existence is highlighted by complex attacks.

The success of attacks on databases relies heavily on the skills and training of the attacker and less on the automation mechanisms of attack. They use pretty much their

knowledge and statistical tools, and because in this these attacks are also called statistical attacks or statistical inference attacks.

An attacker, after he passed all levels of protection and reached the database, he will try progresively a series of attacks: **direct**, **indirect** and by **tracking**.

Direct attacks are obvious attacks and are successful only if the database does not implement any protection mechanism. The displayed results will be the ones required and expected. If this attack fails then the attacker moves to the next.

Indirect attacks are attacks that are executed when it is desired the extraction of other data than tose that are displyed. Combinations of queries are used some of them having the purpose to cheat thesecurity mechanisms.

The tracking attack is applied to the databases that have impemented a supression mechanism for the claims that have dominant results. This type of attakc is used against databases that have short answers to queries. Attacks are based on the principle to which if a direct query has as result a small number of answers, the denial of the main claim will result in zero. If the answer to a complex claim is not displayed due to the supression mechanism for claims with dominant result, then the database will be queried with a set of claims and the answer to these claims will be studied, following that from these sensitive data to be extracted. In literature, this type of attack is called Linear System Vulnerability.

3. The risk

Focusing efforts to ensure database security must be done considering firstly the impact the loss of data has on the business. The final purpose must bear in mind the assurance of confidentiality, integrity, availability and data non repudiation. If the first three objectives are already classic, the last one, the non repudiation, it is necesarry in electronic transactions for confirming authenticity.

A quantitative approach of the risk is preferable than a qualitative approach because it offers a more tangible value of the situation. Even so we will still work with subjective data, estimated after an evaluation process.

If in the case of hardware loss it will be easier to estimate the loss using the cost of replacing the component, in case of a data loss the operation is far more complex. In this case we will discuss about costs for recovery. For a quantitative approach we will start from the formula for calculating the risk:

Risk = Impact x Probability

To estimate the impact we have to ask ourselves if: data can be rebuilt or restored; how long does it take to rebuild data; it is because of a deliberate action or because of accidental causes; the lsot data have special character (military, secret service, confidential).

$\text{Impact}_a = \sum_{i=1}^{n} \text{Impact}_i$ where: Impact_a – Total impact for asset **a**; i – impact zone, (1 to 4, **confidentiality**, **integrity**, **avalability** and **nonrepudiation**).

Calculus of impact value is exemplified in next table.

Table 1. Impact value calculus

Impact zone	Impact value (USD)	Observations
Confidentiality	5 000	Loss due to the data theft. Very difficult evaluation. It implies the cooperation of several departments.
Integrity	1 000	Loss due to data spoilage. It involves costs for checking data integrity. Complex operations.
Avalability	800	Loss due to the unavailability of data. It involves costs for restoration and availability of data. Complex operations and under pressure.
Nonrepudiation	100	Loss due to orders denial.
Total	6 900	

The probability to happen an incident on the databases must be estimated by the analysis team for the security risk. The probability of generating natural events that can disrupt the smooth functioning of databases may be provided by the organizations in the field. Another category is represented by the threats specific to every company and which are linked to the human factor. Once the risk management is mature the estimation of probability of some events to occur would be more precise. Creating diagrams of evolution of the risk would help the company to concentrate its efforts one the areas that are the most affected and enhance methods of control on these locations.

4. Control methods

The classical methods of ensuring database security, the partition of database, cards, encryption, etc. Are able mostly to accomplish their tasks. Yet these are not sufficient.

Supposing we have already implemented the security mechanisms which permit us to know who the user is, then the only thing left is that we have to see what the user does. Using appropriate mechanisms for logging and auditig operations we will be able to see what every user has done and in case of incident the user to be held responsible.

Once this is made, we can go to the phase of attack control. This will involve the implementation of a mechanism that will not permit displaying sensitive data.

The options that can be chosen for such mechanism are the following:

Suppressing the applications with sensitive results. The requests for access for database elements that have as result displaying sensitive results are rejected without any response.

Results approximation. The results of request will be approximated in such way that the attacker will not be able to determine the exact values. In the case of such request the system will be able to display results close to the real ones.

Limiting the results of a request that reveals sensitive data. Limiting the result of a request which reveals sensitive data can be done in the case in which this is 1 (one).

Combining results. Combining the results from several request will create even a greater confusion for the attacker.

All these can be embedded in a monitor type mechanism which will implement de security policy of the company. Access to data will be done according to the user's classification and data classification. There is sometimes confusion between the user and the end-user. An end-user must have access only to one or more programs that run

applications. A user is defined as being that person who has access to a computer system. The end-user is actually an operator and so it should stay.

Persons who work in security field or are on the other side of the barricade agree that a security ensurance system must resist 3-5 days to fulfill its purpose.

Other security controls to ensure the security of databases include control elements that are not based on the computer. Here we include **policies**, **agreements** and **other administrative control elements** different than the ones who sustain control elements based on the computer. From this category we have:

- Security policy and emergency situations plan;
- Staff control;
- Placing the equipment in safe conditions;
- Escrow agreements;
- Mainentance agreements;
- The physical control of access.

5. Conclusions

Database security presents features that must be seriously taken into account.

The first option, for a secure database is represented by its optimal protection.

Ensuring database security must be done from outside ton inside, this involving ensuring security starting from the physical level and ending with the data level (physical, network, host, applications and data).

Databases are a favourite target for attackers because of the data these are containing and also because of their volume. Datawarehouse is the ultimate goal.

Efforts to ensure database security are considerably higher than for the other types of data. It is easier to implement an access list for a great number of files than an access list for the elements of a database.

Database security mechanisms should not irritate their users.

References

1. Burtescu, E. **Problems of Inference in Databases** in „Education, Research & Business Technologies". The Proceeding of the 9th International Conference on Economic Informatics, INFOREC Printing House, Bucharest, 2009
2. Burtescu, E. **Databse security**, in „Knowledge Management. Projects, Systems and Technologies", International Conference „Knowledge Management. Projects, Systems and Technologies", Bucharest, November 9-10, 2006, INFOREC Printing House, Bucharest, 2006
3. Hsiao, S.B. and Stemp, R. **Computer Security**, course, CS 4601, Monterey, California, 1995
4. McCarthy, L. **IT Security: Risking the Corporation**, Prentice Hall PTR, 2003
5. Proctor, P.E. and Byrnes, F.C. **The Secured Enterprise**, Prentice Hall PTR, 2002
6. * * * **Security Complete**, Second Edition, SYBEX Inc., 2002

[1]Emil BURTESCU has graduated the Polytechnics University of Bucharest, Faculty of Aerospace Engineering. He holds a PhD diploma in Economic Cybernetics and Statistics at Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest Academy of Economic Studies.

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

A MATHEMATICAL MODEL FOR TUMOR VOLUME EVALUATION USING TWO-DIMENSIONS[1]

John P. FELDMAN
MSc, Department of Nuclear Engineering, Ben-Gurion,
University of the Negev, Beer-Sheva, Israel

E-mail:
Ron GOLDWASSER
MSc, Department of Nuclear Engineering, Ben-Gurion,
University of the Negev, Beer-Sheva, Israel

E-mail: ron.goldwasser@gmail.com
Shlomo MARK[2]
PhD, Senior Lecturer, Negev Monte Carlo Research Center (NMCRC) and
Department of Software Engineering, Shamoon College of Engineering, Beer Sheva, Israel

E-mail: MarkS@sce.ac.il
Jeremy SCHWARTZ[3]
Research Assistant, Negev Monte Carlo Research Center (NMCRC) and
Department of Software Engineering, Shamoon College of Engineering, Beer Sheva, Israel

E-mail: jrschwar@gmail.com
Itzhak ORION[4]
PhD, Senior Lecturer, Department of Nuclear Engineering, Ben-Gurion,
University of the Negev, Beer-Sheva, Israel

E-mail: iorion@bgu.ac.il

Abstract: *Many recent papers present different ways to show the volume of a tumor from a few two-dimensional images. The three-dimensional fundamental shape of tumors is assumed to be a hemi-ellipsoid as presented in different studies. The three measurements were essential for tumor volume calculations: length, width, and height. Tumor volume measurements task is a very intensive routine is cancer research. Recent papers present how to reconstruct the 3-D tumor from a set of 2-D images, this in order to find the tumor volume. In this paper we report on a new approach to calculating the volumes based on measurements of two dimensions, length and width, after having identified a statistical constant that replaced the need of measuring the tumor height. Results: Our new method was examined on a subcutaneously implanted tumor placed on a mouse's thigh. The width, length, and height of tumors were measured, in four groups of BALB/c mice, using a digital caliper. The tumor dimensions were periodically measured for several weeks. It was shown that this new model can assist in tumor volume measurements using digital images, and in CT scan tumor size assessments.*

Key words: *tumor volume; tumor reconstruction; tumor imaging; hemi-ellipsoid; mice model*

1. Introduction

When using medical imaging that involve radiation such as computed tomography, it is important to minimize the patient exposure. The high exposure is delivered due to the need of a number of images to reconstruct the tumor shape, and dimensions (Junzhou, 2006).

The three-dimensional fundamental shape of tumors is assumed to be a hemi-ellipsoid in different studies. Researches that use the assumption of a hemi-ellipsoid tumor shape were published regarding breast cancer (Wapnir, 2001), prostate cancer (Sosna, 2003) (Egorov, 2006) cervical cancer (Mayr, 2002), glioma cancer (Schmidt, 2004), and others (James, 1999).

A typical ellipsoid volume is giving by:

$$V = \frac{\pi}{6}(length)\cdot(width)\cdot(height) \qquad (1)$$

This study aims to assist in tumor volume measurements by developing a new model that reduces the essential number of dimensions for the volume, and therefore reduces the number of images needed.

The new mathematical model for tumor volume measurement was investigated using a mice model, which is described in the methods section.

2. Methods and Materials

The mice model

In order to examine a new method of tumor volume measurements a subcutaneously implanted tumor was placed on a mouse's thigh. The width, length, and height of tumors were measured, in four groups of BALB/c mice, using a digital caliper. In cancer treatments, determination of the growth rate of tumor volume as a function of time is a standard method of determining the efficiency of a particular treatment.

Since changes in the growth rate reflect the efficiency or inefficiency of a treatment, extreme precision in the measurements is critical. The KHJJ tumor line was derived from a primary mammary tumor arising in a BALB/c mouse, after implantation of a hyperplastic alveolar nodule (Rockwell, 1972). Four groups of mice—26 individuals altogether—were tested in the experiments. All the mice were of the same BALB/c type and of similar size (28 ± 1.4 SD gram average weight). There were two separate groups for each gender (see Table 1). The mice received a tumor transplant on the thigh. The mice were kept and treated according to the Ben-Gurion University of the Negev guidelines for treating animals in scientific experiments.

Table 1: Members of each group of mice and the period of the measurements

Group	Research period	Number of mice in group	Gender
1	63 days	8	male
2	39 days	2	male
3	36 days	5	female
4	43 days	11	female

Preparation of the tumor segments

The KHJJ tumor segments were prepared from a KHJJ tumor about 200 mm^3 in size. This tumor was taken from a mouse that had been sacrificed moments before dissection. The tumor had to be without signs of necrosis and with smooth margins. The tumor was separated properly from any remaining healthy tissue and washed three times with PBS. Then the tumor was cut through the middle to ensure that there was no necrotic tissue inside. The segments were prepared by cutting the tumor into 1 mm^3 sections for transplantation. These segments were kept moist by dripping PBS on them until implantation. This procedure was carried out in sterile conditions under a sterile hood.

Tumor transplantation

Each mouse, before the transplant, was anesthetized with a low dose of 80 ml/kg Ketamine and 8 ml/kg Xylazine anesthetics for about 40 minutes. The tumor segments were prepared for dissection a few moments before the process.

The mouse's right leg was shaved, the thigh skin was pulled up with forceps, and then a subcutaneous slit was made along the skin. A 1 mm^3 tumor segment was inserted using a trocar into the small pocket under the skin on the thigh. A small, flat stick covered with antibiotic ointment was pressed on the skin slit while the trocar was removed.

Tumor volume measurements

Once the tumor became palpable 5–10 days after transplant, its size was measured using a digital caliper. Only one person measured all the tumors in the experiments to prevent observation differences, since it was found that measurements by more than one person can lead to different results. The tumor was measured every other day, within four hours before or after the previous measurement. The tumor was measured between the skin surface layers. The length and width were measured with an accuracy of 0.01 mm using a digital caliber. The length was measured along the imaginary longitude of the leg; the width was measured in the direction of the latitude. The height was measured between the leg surface layer and the upper skin of the tumor. The caliper was placed perpendicular to the tumor so that the height could be measured properly. The tumor was measured from a volume of about 50 mm^3 until it had grown up to 1500 mm^3. Each volume measurement was repeated three times for verification.

Measuring tumor volume is problematic mainly because of inaccuracy in the measurement of tumor height, which contributes the largest error to volume results. The difficult part in achieving precise measurements is determining where to position the caliper in order to measure height. It is clear that a new approach eliminating the need to measure the height, while still providing a precise assessment of the tumor volume, could be very helpful. Tumor length and width can be measured very accurately because they can be observed directly. The 0.01 mm instrumental error associated with the caliper can lead to a 0.5%–1.0% error in a volume of 50 mm^3. The intensive repetitions in this study indicate that the total error for length and width is about 0.1 mm each, leading to a 3% error in the resulting volume. The tumor height error is very large when measuring with a caliper, around 0.5 mm, and can add a 7% error to the volume value. The total estimated volume error is around 10% (Dethlefsen 1968).

In order to minimize the volume error, we suggest an approach that relies only on length and width measurements to calculate volume. Such a calculation may a priori reduce

the uncertainty regarding the volume if a geometrical dependence of volume on length and width can be established:

$$Volume = f(width, length) \tag{2}$$

3. Results and Discussion

The dimensions of the tumor in the mouse model were measured only from the time it reached a volume of 50 mm^3 until it had grown to 1500 mm^3.

Different tumor growth rates were observed for different mice. Figure 1 shows examples from two mice groups—a male group and a female group. The graphs demonstrate a change in tumor dimensions over time.

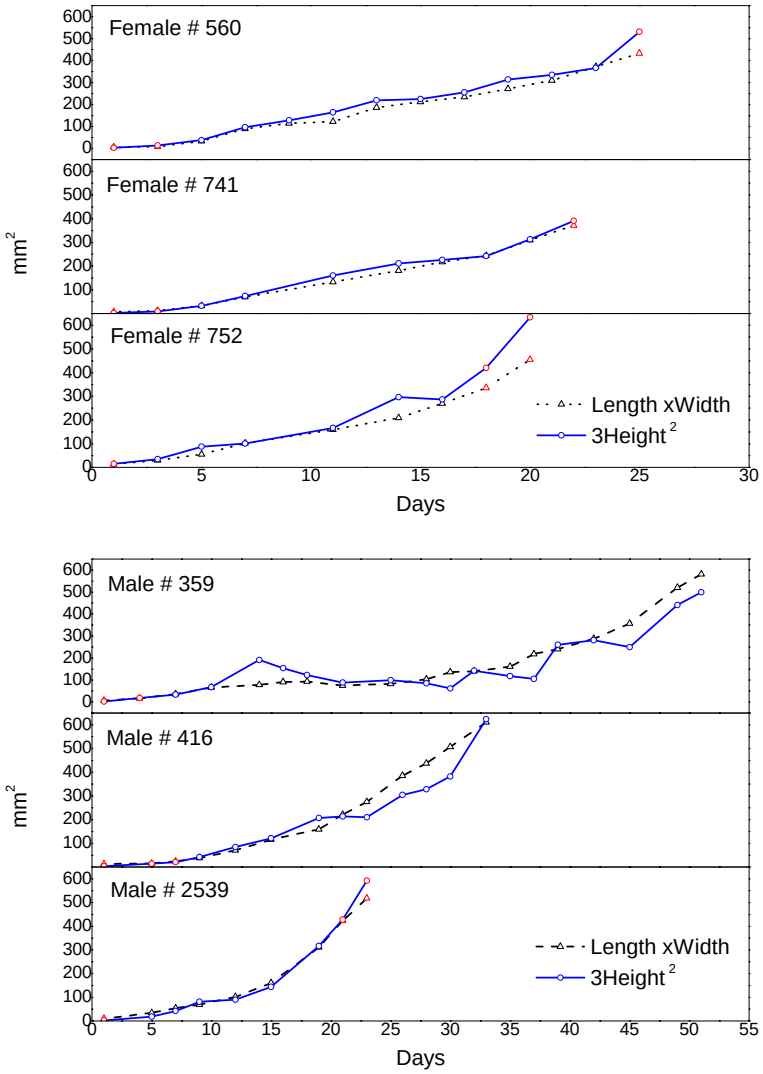


Figure 1. (Length × width) and (3 × height^2) over time. The red points represent data pertaining to volumes that were found to be irrelevant: (a) three female mice; (b) three male mice

We compared the average product of length and width with the square of height. A rough fit was discovered between the two values, with the square of the height multiplied by a factor of 3 being approximately equal to the length multiplied by the width:

$$3H^2 \cong LW \tag{3}$$

The examples in Figure 2 agree with Equation 3 for most of the points that represent different volumes. In addition, it should be noted that the different tumor growth rates did not affect the fit. A linear fit of all the results, shown in Figure 2, resulted in the following equation:

$$H = 1.63\sqrt{LW} \tag{4}$$

The correlation represented by Equation 4 was found to be high, with a linear correlation coefficient of R = 0.919. For a normal distributed data set, the likelihood estimator can be obtained by a least squares analysis (Alfassi, 2005). Therefore this correlation coefficient shows best the validity of Equation 4.

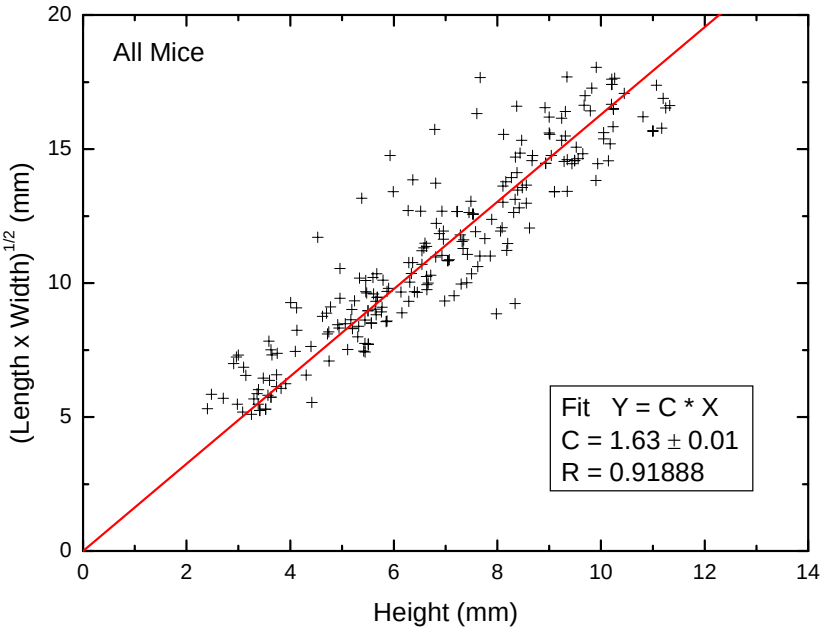


Figure 2. The linear fit of the square root of length × width corresponded to the height in all the measurements of mice.

In order to determine accurately the relationship between H^2 and LW, these values were analyzed separately for each gender, with the following results:

$$\text{Females: } f = \frac{H}{\sqrt{LW}} = 1.58 \pm 0.01 \tag{5}$$

$$\text{Males: } f = \frac{H}{\sqrt{LW}} = 1.69 \pm 0.03$$

A new formula for calculating tumor volume without the use of tumor height was obtained from the analysis of the measurements:

$$V = \frac{\pi}{6} f \left(length \cdot width\right)^{\frac{3}{2}} \tag{6}$$

The new formula is based on some symmetry assumptions that inherent in the classical volume formula, as the classical volume formula is a simple expression of an ellipsoid volume. A comparison of the new volume calculation with the classical calculation based on three dimensions, seen in Figure 3, shows no apparent difference in volume values. The total mouse mass growth-rate usually differs between males and females. The difference in the tumor growth-rate can be explained by the distinction in tumor growth-rate between genders.

The total volume error can be reduced using this new method in tumors placed subcutaneously to around 4%, compared to the 10% error that was obtained in the old-fashioned volume measurements. The error bars in Figure 3 are larger for the classical volume values and smaller for the new volume results.

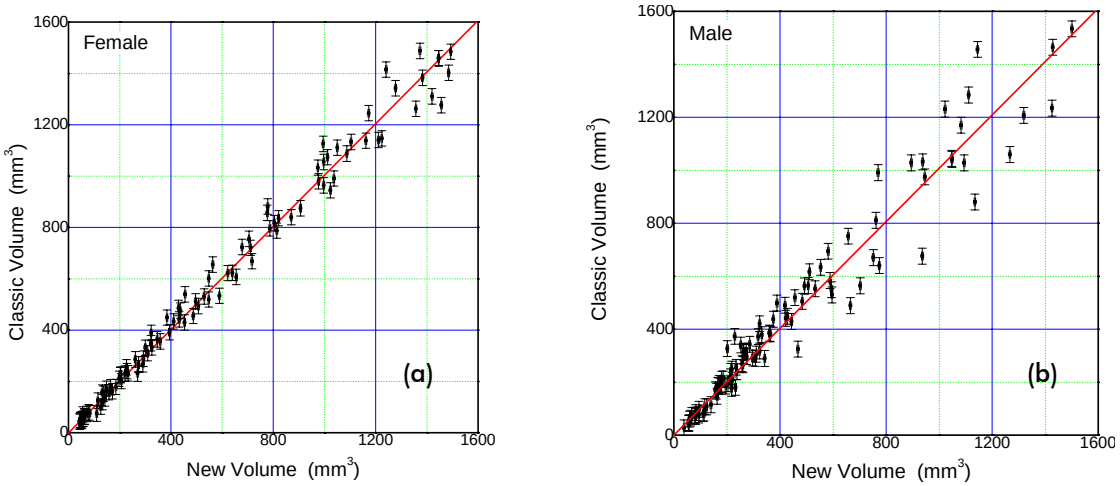


Figure 3. A comparison of the volume calculated according to the classical three-dimensional formula with the volume calculated according to the new method. The fit shows a good match between the results of the two methods: (a) in female mice; (b) in male mice.

4. Conclusions

The new method for tumor volume calculations was studied using the mice model described above. This model showed improved error estimation for the tumor volume. The old method results were plotted against the new method results, shown in Figure 3. The linear fit line for this graph obtained a slope of 1.00, proving the consistency of the new method. The new method of calculating volume should reduce the error in the volume measurements because it depends on visual measurements that can be accurately obtained. Since the nominal errors for length and width are similar, using the new method, presented in Equation 4, should limit the height error to about 1.5 times the error in length. This new method, not only reduces the error of the tumor volume, but also reduces the number of parameters needed to be measured down to two (length and width) from three (including the height). The new method can be helpful in several cases where a digital photo of a tumor can be taken, and may shorten the time needed for handling mice in the lab (see Appendix).

This work presents a test-case study of the new mathematical method offered for a tumor placed subcutaneously. These findings should encourage future studies towards reducing the amount of medical imaging scans needed for internal tumor volume reconstruction.

References

1. Alfassi, Z.B., Boger, Z., and Ronen, Y., **Statistical treatment of analytical data,** Boca Raton (FL), CRC Press, 2005
2. Dethlefsen, L.A., Prewitt, J.M.S. and Mendelsohn, M.L., **Analysis of tumor growth curves**, J Natl Cancer Inst., 40, 1968, pp. 389-405
3. Egorov, V., Ayrapetyan, S. and Sarvazyan A.P. **Prostate Mechanical Imaging: 3-D Image Composition and Feature Calculations**, IEEE Transactions on Medical Imaging, Vol. 25, No. 10, October 2006
4. James. K., Eisenhauer, E., Christian, M., Terenziani, M., Vena, D., Muldal, A. and Therasse, P. **Measuring Response in Solid Tumors: Unidimensional Versus Bidimensional Measurement**, J Natl Cancer Inst, Vol. 91, No. 6, March 17, 1999
5. Junzhou, H., Xiaolei, H., Metaxas, D. and Banerjee, D. **3D tumor shape reconstruction from 2D bioluminescence images**, 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 6-9 April 2006, Biomedical Imaging, Macro to Nano, 2006
6. Mayr, N.A., Taoka, T., Yuh, W.T.C., Denning, L.M., Zhen, W.K., Paulino, A.C. *et. al*. **Method and timing of tumor volume measurement for outcome prediction in cervical cancer using magnetic resonance imaging**, Int. J. Radiation Oncology Biol. Phys., Vol. 52, No. 1, 2002, pp. 14–22
7. Rockwell, S.C., Kallman, R.F. and Fajardo, L.F., **Characteristics of a serially transplanted mouse mammary tumor and its tissue-culture-adapted derivative,** J Natl Cancer Inst, 49, 1972, pp. 735-49
8. Schmidt, K.F., Ziu, M., Schmidt, N.O., Vaghasia1, P., Cargioli, T.G., Doshi, S. *et. al*. **Volume reconstruction techniques improve the correlation between histological and in vivo tumor volume measurements in mouse models of human gliomas**, Journal of Neuro-Oncology, 68, 2004, pp. 207–215
9. Sosna, J., Rofsky, N.M., Gaston, S.M., DeWolf, W.C. and Lenkinski, R.E. **Determinations of prostate volume at 3-tesla using an external phased array coil Comparison to pathologic specimens,** Academic Radiology Vol. 10, Issue 8 , 2003, pp. 846-853
10. Wapnir, I.L., Barnard, N., Wartenberg, D. and Greco, R.S. **The inverse relationship between microvessel counts and tumor volume in breast cancer**, Breast J., 7(3), May-Jun, 2001, pp. 184-8

Appendix

We developed a Windows-utility-program to insert a digital photograph of a tumor (shown in Figure 4). This program could be used to measure tumor length and width, and the tumor volume could be calculated using the new method presented in this paper. It would also be possible to use the classical method of calculating volume, if desired, once the tumor height has been determined. The caliper is not necessary to measure the tumor volume if one uses this program with digital photographs of the tumor. The program can be downloaded from http://www.bgu.ac.il/~iorion/.

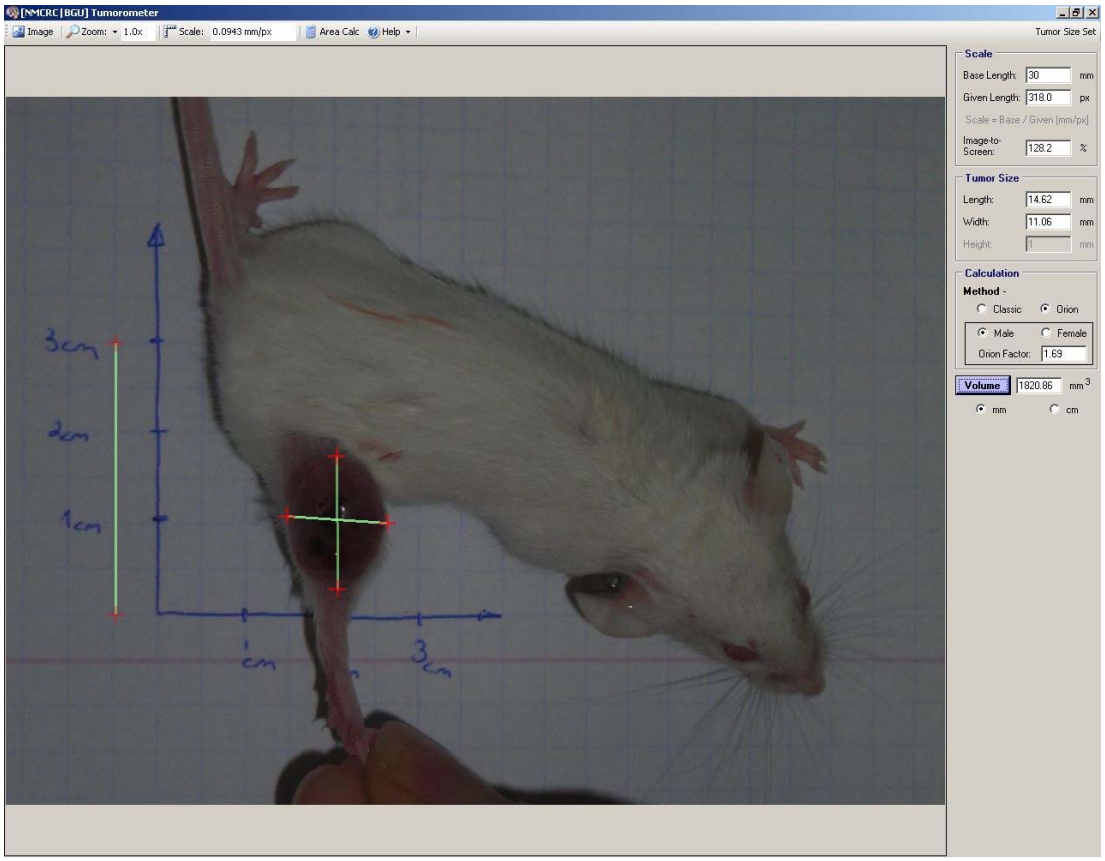


Figure 4. An example of tumor volume measurements using the Windows utility program

---

[1] **Acknowledgments:** We thank Prof. Brenda Laster of the Department of Nuclear Engineering, Ben-Gurion University of the Negev, for allowing us to use her research labs.

[2] **Shlomo Mark** is a senior lecturer at SCE Shamoon college of engineering. He earned his Ph.D. in nuclear engineering and an M.Sc. in **Biomedical Engineering and in Managing and Safety Engineering** . He works in the Department of Software Engineering. and he is the Head of the NMCRC - Negev Monte Carlo Research Center, Shamoon College of Engineering. His main research interests are Scientific programming, Computational modeling for Physical, environmental and medical applications, Monte Carlo Simulations, Develop, upgrade, and improved Monte Carlo based codes, by using software engineering mythologies.

[3]**Jeremy Schwartz** is a software developer with a broad range of programming experience from research, consulting, and academic settings. He graduated in 2006 from North Carolina State University (USA) with undergraduate degrees in Chemical Engineering and Computer Science. Mr. Schwartz specializes in writing highly functional, well-documented and efficient code for both online and offline environments. Having written training materials, developed data analysis tools, and rebuilt several websites from the ground up, he understands the iterative process by which user requirements are transformed into deliverables.

[4]**Itzhak Orion** is a senior lecturer at Ben-Gurion University. He earned a Ph.D. in nuclear engineering and an M.Sc. in nuclear physics. He works in the Department of Nuclear Engineering. He visited several institutes around the world as a collaborating scientist. He serve as an advisor to the Israeli atomic energy research centers in the subject of radiation simulations. His main research interests are radiation, Monte Carlo simulations, computational modeling and applications for medical physics.

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

A FORECASTING MODEL WITH CONSISTENT ADJUSTMENTS FOR ANTICIPATED FUTURE VARIATIONS

Chin-Lien WANG[1]

PhD Candidate, Department of Industrial Engineering and Enterprise information, Tunghai University, Department of Business Administration, Ling Tung University, Taichung, Taiwan

E-mail: alvinw3@gmail.com

Li-Chih WANG[2]

PhD, University Professor, Department of Industrial Engineering and Enterprise information, Tunghai University, Taichung, Taiwan

E-mail: wanglc@ie.thu.edu.tw

Abstract: *Due to the limitation of most statistical forecasting models ignoring contextual information, judgmental adjustment is a widespread practice in business. However, judgmental adjustment still suffers with many kinds of biases and inconsistency inherent in subjective judgment. Our approach uses an adjustment mechanism concerning only with critical cue factors evaluated with genetic algorithm to alleviate problems caused by collinearity and insignificant sporadic variables usually arising in least square type estimators, and to derive more realistic parameter estimation. In case there are anticipated variations in the forecasting horizon and can't be handled by the model alone, this adjusting mechanism, formulated in a set of equations, can be used to assess mixed effect of cue factors consistently and effectively without subjective judgment involved. Empirical results reveal that this adjustment mechanism could significantly reduce MAPE of forecasts across seasons with improvement mainly coming from large size adjustments.*

Key words: *judgmental adjustment; seasonal index realignment; genetic algorithm; calendar effect*

1. Introduction

Owing to the limitations of statistical forecasting methods generating forecasts solely based on historical data [1-3][3], or not including critical explanatory variables, as pointed out in [4], judgmental adjustments taking advantage of contextual information or non time series information [5] become a widespread practice in business to improve forecasting accuracy [5, 6-8], especially for model-based forecast in variable environments. As [9] put it, the benefits should be greatest where series are subject to high noise and/or where the signal is relatively complex. Blattberg and Hoch [10] also argue that forecasters

can use econometric models effectively only if they have a built-in adjustment mechanism to capture the changing environment. Also see [4, 11, 12].

Most researchers in judgmental adjustment agree that, if the contextual information used is reliable, the performance of judgmental adjustments will be better than that of judgmental adjustments using unreliable one [5, 13, 14]. Hence, the eliciting of contextual information from experts using quantitative techniques such as Delphi and cross impact analysis [15], the screening [4], and classification such as the use of 5 structured types of causal forces proposed by [16], as well as processing, such as [17, 18], among many others advocating the use of decomposition method, of relevant contextual information, all are important aspects of judgmental adjustment.

However, judgmental adjustments still have all kinds of bias like cognitive bias, double-counting bias, political bias, and so on inherent in judgmental forecasting and still have the issue of inconsistency [19–21].

The objective of this study focuses on proposing an adjustment mechanism capable of handling and reflecting, without subjective judgement, detailed changes anticipated in the forecasting horizon but could not be handled by the regression model alone, which incorporating all the critical variables estimated with GA in such a way as to be more realistically conformed with the real world. Thus, this adjustment mechanism, a natural extension of the model, consisting of seasonal index realignment and proportional adjustment in a set of equations, is able to make appropriate adjustments consistently and effectively improving the forecasting accuracy of initial forecasts of the model compared favorably to other alternative like Box-Jenkins ARIMA [22-23] with adjustment.

The remainder of this paper is organized as follows. Section 2 describes our forecasting model in detail, including formulation of a regression model, a brief introduction to the feature and process of GA in model fitting, subsequent model checking, and the process of e-composition, as well as the adjusting mechanism consisting of seasonal index adjustment and proportional adjustment, as well as a combination of the former two. Section 3 portrays the background and design of our empirical research. Section 4 depicts the empirical results of model fitting and model checking, and a comparative analysis of forecasting results of various adjustment methods assessed with MAPE on per item basis, percentage of correct direction adjustment, and IMP on per adjustment basis, from the perspective of adjustment size and lead time, as well as an illustration with two graphical examples, and discussions. Finally, a conclusion is drawn in section 5.

2. The forecasting model

2.1. Formulation of a regression model

The first objective in our forecasting model involves decomposing the promotional sales of products of a company into simple components easy to handle. Eq. (1) of our regression model is motivated by Dick R. Wittink *et al.*'s analytical models in [24-26]. The model can be formulated as

$$S_{it} = \lambda_{it}\left(P_{it} / \widehat{P}_i\right)^{\theta_{it}} \prod_{l=1}^{n} \mu_{l\,it}^{\;D_{l\,it}} \prod_{r=1}^{o} w_{rit}^{\;H_{rit}}\, \ell^{\varepsilon_{it}} \qquad , \forall\ t \in Q \qquad (1)$$

Where, *i* denotes an item number, $i = 1,2,3....,I$; *t* denotes specific number of period referenced, $1 \le t \le T$. *T* is the total number of normal periods. While *I* is the total number of items involved.

Q denotes the set of referenced periods.

Z denotes the set of periods to be forecasted.

S_{it} is the total unit sales of the item *i* in period *t* under a retailer, for weekly sales, *t* actually represents a certain week in the referenced periods.

λ_{it} denotes the normal unit sales (base sale) of the item *i* in period *t* without any promotion under a retailer.

$\widehat{P}_i$ is the list price of item *i*.

P_{it} is the discount price of item *i* during period *t* under a retailer.

θ_{it} denotes the coefficient of price elasticity of item *i* during period *t* under a retailer.

D denotes an indicator parameter(or dummy variable) of non-price promotion mix.

$D_{l\,it}$ is the *l*-th component of a vector of *n* indicator parameters of non-price promotion mix ($D_{1\,it}$, $D_{2\,it}$, …, $D_{n\,it}$) of item *i* in period *t*. $D_{l\,it} = 1$ denotes a promotion mix of type *l* arises, the default value of $D_{l\,it} = 0$.

$\mu_{l\,it}$ denotes the non-price promotion effect parameter of type *l* non-price promotion mix ($D_{l\,it}$), a combination of certain non-price promotion activities, of item *i* during normal period *t* under a retailer.

H denotes an indicator parameter of holiday.

H_{rit} is the *r*-th component of a vector of *o* indicator parameters of holiday (H_{1it}, H_{2it}, … , H_{oit}) to indicate whether there is any holiday(s) in a certain period *t* or not. $H_{rit} = 1$ denotes a holiday of type *r* arises, the default value of $H_{rit} = 0$.

w_{rit} denotes holiday effect parameter of holiday type *r* in period *t* of item *i*.

ε_{it} denotes the residual error.

Besides, additional notations listed below may be helpful in the following sections.

φ denotes the weekend effect, which is derived via GA based on data in mixed periods.

$d(t_1)$ denotes the length of sub-period t_1 of *t*, $t \in Z$. $0 \le d(t_1) \le 7$.

$d(t_2)$ denotes the length of sub-period t_2 of *t*, $t \in Z$. $0 \le d(t_2) \le 7$. $d(t_1)+d(t_2)= d(t)$, because in a week there are at most two different kinds of promotion mixes held.

δ denotes the duration of the weekend.

δ_{t1} denotes the duration of weekend covered by sub-period t_1.

Take natural logarithm in both sides of Eq. (1), we get the following:

$$\ln S_{it} = \ln \lambda_{it} + \theta_{it} \ln\left(P_{it} / \widehat{P}_i\right) + \sum_{l=1}^{n} D_{lit} \ln \mu_{lit} + \sum_{r=1}^{o} H_{rit} \ln \omega_{rit} + \varepsilon_{it}\ ,\ \forall t \in Q \qquad (2)$$

Thus, a nonlinear model like Eq. (1) is transformed to a linear regression model [27-28], which is the underlying model to conduct model fitting and model checking in this study.

2.2. Model fitting--parameter estimation with GA

To take into account of all the influential and sporadic cue factors in various sub-periods of the training period, the number of variables may amount to such a quantity that conventional parameters estimation method like ordinary least square, maximum likelihood method, and so on may become incompetent, due to the issue of collinearity [29-30], insignificant parameters, [4, 31] or small sample size. Therefore, in this study we use a customized genetic algorithm (GA) which could estimate parameters effectively and efficiently [32].

2.2.1. Features and procedures of GA in this study

GA simulates Darwin's biological evolution through stochastic crossover and mutation by selecting encoded individuals (solutions) in the population with higher fitness via a fitness function to generate population of individuals (reproduction) more fitted to the environment (better solutions) from generation to generation [33-35].

The initial population is randomly created in the encoded form of a binary matrix, there are *pop* rows, each row of binary string in the matrix is an individual (solution) which encompasses β chromosomes, each chromosome, representing a parameter, is composed of γ genes, while each gene is represented by a binary code.

Each individual is evaluated by the fitness function, check Eq. (3), in each generation, the best α % ($1 \leqq \alpha \leqq 6$) of the population are kept as elites to the next generation, the remaining of the population are created by a randomly selected pairs of individuals conducting a multi-point crossover [36], n + o + 2 points in total, for each one of each pair to reproduce offspring, forming a random recombination of individuals' ingredients of genes, to search new solution space and possibly better solution.

After that, a one-bit mutation is performed [37], with a view to creating new pieces of gene originally not possessed by members of the population, through randomly selected genes within each individual, this occasional random change in genes could open the door to new possibilities of better solutions. Afterwards, each encoded individual in the population is decoded back to a string of real numbers of parameters, and each individual is evaluated by the fitness function…, the iterative process goes on and on until a termination condition is met.

In this study, parameters like crossover probability (P_c) and mutation probability (P_m) of GA are designed to vary with the number of generations processed or others, such as the minimum level of moving average percent of improvement (MAPI) in fitness function value within certain number of generations, to keep proper diversity of the population, while retaining the convergence capability, to circumvent getting stuck too early at local solutions in its search process and derive satisfying results [38-39].

In estimating parameters of complicated multivariate nonlinear models, GA is generally considered to be better than other alternatives such as nonlinear least square,

maximum likelihood estimation, and so on, due to its parallel search capability [40-41], even based on small size dataset, it is capable of deriving satisfying results.

The fitness function of GA may be formulated as

$$FV_i = MAPE_i = (\sum_{t=1}^{T} \left| \ln S_{it} - \ln \tilde{S}_{it} \right| / \ln S_{it}) / T \qquad , \forall t \in Q \tag{3}$$

Where the term $\left| \ln S_{it} - \ln \tilde{S}_{it} \right|$ is the absolute value of difference between natural logarithm of the actual sales volume ($\ln S_{it}$) of the *i*-th item and natural logarithm of the estimated sales volume ($\ln \tilde{S}_{it}$) of the same item in period *t*. *T* denotes the number of normal periods. The objective of GA is to find a solution with the minimal $MAPE_i$. The smallest $MAPE_i$ found is updated once a smaller one is found in the solution search process. After model fitting, every effect parameter in Eq. (2) is derived in real value.

2.3. Model checking

In this section, a regression diagnostics focused on normality and independence is performed to see if critical assumptions of linear regression are violated, based on Eq. (2). If these assumptions are severely violated, particularly if collinearity arises among predictor variables, bias may be a serious issue in model fitting or even in model specification.

Normality test is conducted through One-Sample Kolmogorov-Smirnov test [42], and Q-Q plot [43]. Independence test in this study consists of two parts, namely, multi-collinearity test and autocorrelation test. The former is performed via condition index, whereas the latter is performed via ACF checking [44].

2.4 The re-composition of effect parameters

As the cycle length of CPG industry is about *52* weeks long, let *t′* = *t* + 52, denoting the corresponding week to be forecasted in a new year. A modified naïve sales forecasting method considering cycle length to forecast unit sales of item *i* of period *t′* in a new year, see Williams (1987), according to sales data of week *t* in the referenced year, would be

$$\hat{S}_{it'} = \eta_i \pi_{it} \ln\left(P_{it'} / \hat{P}_i\right)^{\theta_{it'}} \prod_{l=1}^{n} \mu_{it'}{}^{D_{lit'}} \prod_{r=1}^{o} \omega_{rit}{}^{H_{rit}}, \qquad t' = t + 52, \forall t' \in Z. \tag{4}$$

Where, η_i denotes the average normal sale of item *i* across referenced periods. π_{it} denotes the seasonal index of item *i* in period *t, and Z* denotes the set of periods to be forecasted.

So far, all the parameters in Eq. (4) are already derived via GA. Let $e_{1it'}$ denotes the price effect multiplier of item *i* in forecasting period *t′*, $e_{2\ it'}$ and $e_{3\ it'}$ denote the effect multiplier of a non-price promotion mix and a specific holiday effect, respectively. In each group of indicator parameters at most one condition will arise in each period. We get

$$\hat{S}_{it'} = \eta_i \pi_{it} e_{1it} e_{2it} e_{3it} , \quad t' = t + 52, \forall t' \in Z. \tag{5}$$

In its re-composed form, Eq. (5) can be used to forecast weekly unit sales. Actually, parameters estimated through GA, based on observations in the training periods, can be recombined as Eq. (5) in responding to expected promotional campaigns in the forecasting horizon specified in the promotion proposals to perform out of sample forecasting without any adjustment in the following empirical study.

2.5. The adjusting mechanism of this study

The mechanism of this study stresses that adjustments of forecasting are based on the anticipated changes of the context of promotions and holidays in the forecasting periods, which can not be handled by the regression model alone, in this study a set of equations are formulated to do this job, they are natural extension of the model. The objective is to improve the performance of the model, making our final forecasts more closely reflect these changes in prospect.

2.5.1. Seasonal index adjustment (SIA)

Based on domain knowledge, sales volume of the last week or average sales volume of the last few weeks (adjusted with calendar effect) in the reference periods is a better predictor to sales of the first few weeks in the forecasting periods next year than sales of the same weeks in the referenced periods in Taiwan, check Figures. 1-2. Thus, the corresponding formula can be modified from Eq. (5) and formulated as

$$\hat{S}_{it'} = \eta_i \pi_{it(t=\Omega)} e_{1it'} e_{2it'} e_{3it'} e_{4it'(t'=t+52-ws)} \quad , \forall t' \in Z \tag{6}$$

Eq. (6) is an example of modified Naïve method used for multiple-step out-of-sample forecasting considering cycle length which is a year. In which, $\eta_i \pi_{it(t=\Omega)}$ stands for normal (base) sales of the last week(s) in the training periods and is the most recent related data available. While $e_{4it'}$ stands for pre LNY (Lunar New Year) effect arises annually in a period of around 4 weeks right before LNY. In this period, sales volumes are usually much higher than usual even without any promotion. Since pre LNY effect has not been incorporated as a variable into our regression model for parsimonious purpose, it will dominate the seasonal indices in corresponding periods, hence, it's quite intuitive to use these indices as proxy variable of pre LNY effect, denoted as $e_{4it'}$. Because there is usually a time shift of the timing of LNY from year to year, to forecast unit sales of weeks after LNY in next year, the week number referenced corresponding to the week in the forecasting horizon has to be adjusted.

Let *LNY(t')* denote the week number of LNY in the year to be forecasted, *LNY(t)* denote the week number of LNY in the referenced year, then, let *ws = LNY(t') – LNY(t)* represents the number of weeks shifted between two different years as the week of LNY is concerned. If *ws* > 0, it means the sequence of week of *LNY(t')* in the forecasting year will be *ws* weeks later than that of *LNY(t)* in the referenced year; on the other hand, if *ws* < 0, it means the sequence of week of LNY(t') in the forecasting year will be *ws* weeks earlier than

that of *LNY(t)* in the referenced year. Therefore, the most right term in Eq. (6) $e_{4it'}$ will be replaced by $\pi_{i(t-ws)}$. Thus Eq. (6) will become

$$\hat{S}_{it'} = \eta_i \pi_{it(t=\Omega)} e_{1it'} e_{2it'} e_{3it'} \pi_{i(t-ws)} \quad , \forall t' \in Z \tag{7}$$

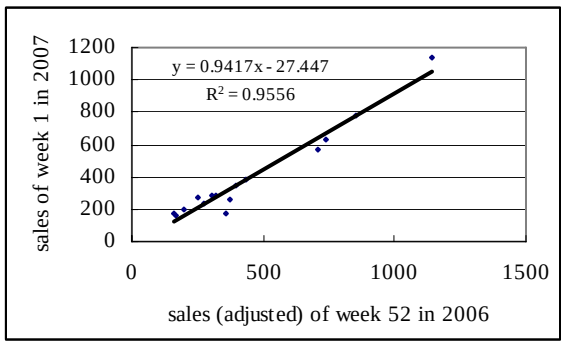


Figure 1. Regression of week 52's sales (2006) to sales of week 1 in 2007

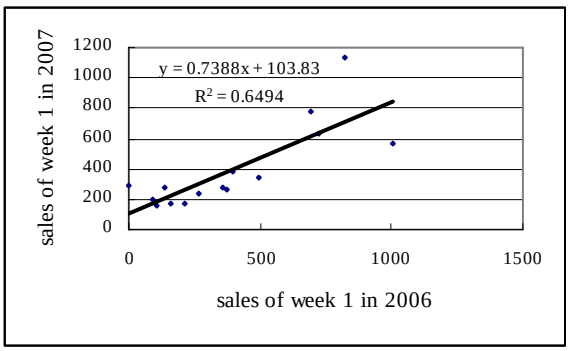


Figure 2. Regression of week 1's sales (2006) to sales of week 1 in 2007

We also find that sales of the same weekly order as that in a different year after LNY are more aligned than weekly sales of the same ordinary sequential order between different calendar years, check Figures 3-4, in which R^2 from sales of weeks after LNY in 2007 regressed against those of weeks after LNY of 2006 is 0.8, much better than the ordinary week n corresponding to the same week n, n = 1,2,..,5, regression between different years, which only has a R^2 of 0.628, please check Figure 4.

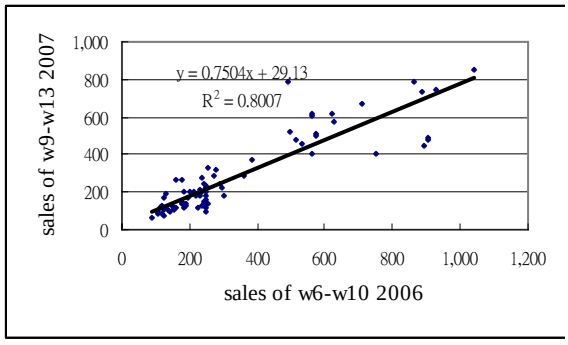


Figure 3. Regression of sales of w6-w10 2006 to that of w9-w13 in 2007

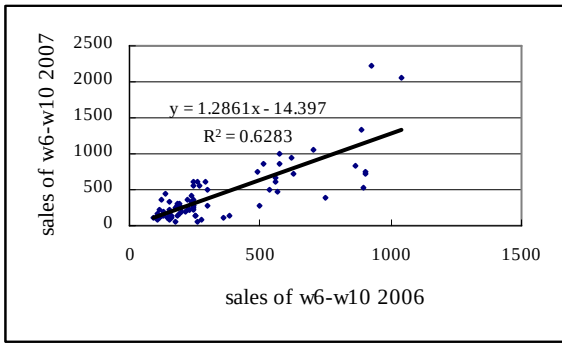


Figure 4. Regression of sales of w6-w10 2006 to that of w6-w10 in 2007

Based on the finding mentioned above, to forecast the sales of weeks after the week of LNY in a new year, denoted as $\hat{S}_{it'}, t' > LNY(t')$, Eq. (8) can be of use, which is modified from Eq. (7) :

$$\hat{S}_{it'} = \eta_i \pi_{i(t-ws)} e_{1it'} e_{2it'} e_{3it'}, \quad t' > LNY(t') \quad , \forall t' \in Z \tag{8}$$

2.5.2. Proportional adjustment (PA)

Quite often, in the week of LNY, there is a small part of pre LNY present prior to the eve of LNY, or the last week of pre LNY is mixed with a small part of LNY in the referenced period, but the condition of the corresponding week in forecasting horizon is different, in these cases, to get a proper estimation of these effect multipliers of calendar effect in the forecasting periods, we must get them restored to regular ones (a whole week only covered by purely one kind of holiday related effect like pre LNY effect or holiday effect of LNY) first, and then proceed to calculate the changed mixed effect in the forecasting period.

The adjusting equations are used to calculate the mixed effect of pre LNY and LNY present in the same week:

$$e_{3it} = \frac{(d(t_1) - d(\delta_{t_1}) + d(\delta_{t_1})\varphi)e_{4it_1} + (d(t_2) - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_1}))\varphi)e^*_{3it_2}}{d(t) + d(\delta_t)(\varphi - 1)}, \forall t \in (Q \cup R) \quad (9)$$

In Eq. (9), the part $(d(t_1) - d(\delta_{t_1}) + d(\delta_{t_1})\varphi)e_{4it_1}$ represents the sum of the effect of the last week in pre LNY in the duration of weekdays covered by sub-period t_1 and the effect of the last week in pre LNY in the duration of weekend covered by sub-period t_1 times the weekend effect. While the term $(d(t_2) - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_1}))\varphi)e^*_{3it_2}$ stands for the sum of the effect of regular LNY in the duration of weekdays covered with sub-period t_2 in the referenced period and the effect of regular LNY in the duration of weekend covered by sub-period t_2 in the referenced period times the weekend effect. Every parameter in Eq. (9) except e^*_{3it} is known, so e^*_{3it} can be obtained. Note that the daily effect of each normal weekday in a week is assumed to be 1. Eq. (9) actually is a daily effect weighted average formula of mixed weekly effect of pre LNY and LNY present in the same week.

It follows that the mixed effect in the forecasting period ($e_{3it_{1'}}$) could be computed through the following formula:

$$e_{3it'} = \frac{(d(t_1') - d(\delta_{t_{1'}}) + d(\delta_{t_{1'}})\varphi)e^*_{3it_{1'}} + (d(t_2') - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_{1'}}))\varphi)e_{rit_{2'}}}{d(t) + d(\delta_t)(\varphi - 1)}, \forall t' \in Z \quad (10)$$

Where, $e_{rit2'}$ may be effect of the last week in pre LNY or just base effect equal to 1. In Eq. (10), the part $(d(t_1') - d(\delta_{t_{1'}}) + d(\delta_{t_{1'}})\varphi)e^*_{3it_{1'}}$ stands for the sum of the effect of the regular LNY in the duration of sub-period t_1' in the weekdays in the forecasting period and the effect of the regular LNY in the duration of sub-period t_1' in the weekend times the weekend effect. While the part of $(d(t_2') - d(\delta_t) + d(\delta_{t_{1'}}) + (d(\delta_t) - d(\delta_{t_{1'}}))\varphi)e_{rit_{2'}}$ represents the sum of the effect of the last week in pre LNY in the duration of sub-period t_2' in the weekdays in the forecasting period and the effect of the last week in pre LNY in the duration of sub-period t_2' in the weekend times the weekend effect.

In the same token, regular effect of the last week in pre LNY ($e^*_{4it1'}$) in the referenced periods can be derived via Eq.(11). Then, the mixed effect of $e_{4it1'}$ could be obtained with Eq. (12):

$$e_{4it} = \frac{(d(t_1) - d(\delta_{t_1}) + d(\delta_{t_1})\varphi)e^*_{4it_1} + (d(t_2) - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_1}))\varphi)e_{3it_2}}{d(t) + d(\delta_t)(\varphi - 1)}, \forall t \in (Q \cup R) \quad (11)$$

$$e_{4it'} = \frac{(d(t_1') - d(\delta_{t_1'}) + d(\delta_{t_1'})\varphi)e^*_{4it_1'} + (d(t_2') - d(\delta_{t'}) + d(\delta_{t_1'}) + (d(\delta_{t'}) - d(\delta_{t_1'}))\varphi)e_{3it_2'}}{d(t) + d(\delta_{t'})(\varphi - 1)}, \forall t' \in Z \quad (12)$$

2.5.3. Total adjustment (TA)

As the combination of both SIA and PA, TA is the most comprehensive adjustment in this study.

3. Empirical Study

3.1. The background of empirical study

This study has a focus on the adjustment of model-based forecast of weekly unit sales of several series of CPG products, manufactured by Company A, under retailer B. Company A is a leading manufacturer specialized in dehumidifier and deodorizer products in Taiwan. While retailer B is an international outlet of DIY products.

A sales data set of *10* items in 2007, aggregated from retailer B's outlets, coupled with price promotion, non-price promotion data, as well as promotion proposals, which were set up in 2007, of the first 4 months in 2008, are used to conduct our empirical study. The training dataset covers two periods, the first period covers the whole year of 2007, the dataset of this period can be denoted as sample A, forecasting horizon is the first 6 weeks of 2008. The second period ranges from the beginning of 2007 to the 10th week of 2008, the training data of this period can be denoted as sample B, forecasting horizon ranges from 11th week to 16th week of 2008.

The underlying equation used in model fitting was Eq. (2), all the effect parameters in Eq. (2) were estimated through GA with objective function set as Eq. (3) and constraints set realistically from contextual knowledge, such as price elasticity parameter to be in the range of [0, -8], non-price promotion effect multiplier to be in the range of [1, 5], holiday effect multiplier to be in the range of [1, 2]. Besides, the number of types of non-price promotion mixes n in Eq. (2) was set to be 7, therefore, there are about 7 types of different combination of promotion activities across seasons, and the number of holiday type o was set to be 4, which means there are about 4 different types (according to duration of holiday) of holidays each year, to reflect the actual business settings. GA programs were run with Matlab 7.1. Effect parameters estimated with GA and mixed effect parameters reassessed in the mixed periods on both sample A and sample B were recomposed according to the expected variations of promotions in the promotion proposal as initial forecasts without any adjustment. The multiple-step out-of-sample forecasts with ARIMA were run with SPSS 13. In busy season, because of intensive promotion campaigns, ARIMA tends to underestimate unit sales, the rule of adjustment for forecasts of ARIMA can be formulated as

ARIMA ad = forecast of ARIMA * 1.2 (13)

However, in time of off season, ARIMA tends to overestimate unit sales, the rule of adjustment can be formulated as

ARIMA ad = forecast of ARIMA * 0.8 (14)

Thus, the performance of various adjustment methods in this study can be compared with their counterpart of ARIMA.

3.2. The design of empirical study

In order to take both the busy season and off season into account to have a proper assessment of the performance of different adjustment methods, the forecasting horizon is designed to consist of two periods of equal duration, the first period includes the first 6 weeks of 2008 which covers the busiest season, ie, the LNY season in Taiwan, and can be denoted as busy season, while the second one starts from the 11th week and ends at the16th week of 2008, which is one of the off seasons in the same year and can be denoted as off season.

As the forecasting target is concerned, 10 items of products were selected to conduct our empirical research. The relevant prices and promotion activities can be found in promotion proposals which actually are the source of anticipated variations in promotions. Another source of anticipated variations in calendar effects is the calendar.

To properly evaluate the performance of various methods in adjusting original forecasts of the model, which can be denoted as NA, made by regression model, SIA was performed first to adjust NA, followed by PA to adjust the same NA. Then, the combination of SIA and PA were used to adjust NA, denoted as TA. The busy season was the first forecasting horizon, and off season was the second forecasting horizon. In addition, for the purpose of comparative reference, forecasts with Box and Jenkins ARIMA were derived, adjustments of forecasts from ARIMA were derived with Eq. 15 in busy season, and Eq. 16 in off season, respectively.

4. Empirical results

4.1. The results of model fitting

The estimated error in terms of MAPE in general is below 3%, except item 10. Most parameters derived are consistent with our expectations, such as the effect parameters of μ_1 to μ_3 are increasingly bigger from 2.391 to 2.992 for sample A and from 2.382 to 2.848 for sample B, respectively, because more effort and expenditure are made for the bigger number type of promotion therein, and μ_5 is larger than μ_4 because non-price promotion type 5 employs direct mail in addition to what type 4 has.

4.2. The results of model checking

The normality test, consisting of one-sample Kolmogorov-Smirnov test and Q-Q plot, in which, this model passed the test with data from sample A or sample B without problem based on standardized error term ε_{it} in Eq. (2) and natural logarithm of predicted weekly unit sales denoted as $\ln \tilde{S}_{it}$ in Eq. (3). However, in independence test, the measures of condition index and results of ACF showed complex but interesting consequences in both samples, in which 5 out of 10 items have autocorrelation problems for sample B. As for sample A, there are 2 items have the same kind of problems, however. As collinearity is concerned, according to [44], if the condition index is above 10 and below 30, there may

have a minor problem of collinearity, if condition index is above 30 and below 100, there may have moderate to severe collinearity issue. In our empirical study, 4 out of 10 items may have moderate to severe collinearity problems, 5 items may have severe collinearity problem for sample B, the condition looks similar for sample A, check Table A1 in Appendix A. If model parameters are estimated by OLS, it is quite possible to have serious bias issues.

4.3. Comparing and analyzing results from various kinds of forecasting adjustment methods

The performance of weekly sales forecasting adjustment from various methods in terms of MAPE can be displayed in Table 1 and Table 2. Each cell with negative adjustment performance is in bold face. Among these adjustment methods, without taking advantage of any adjustment, the MAPE of sales forecasting with the regression model, that is NA, in average, is 17.96% and 37.06% in busy season and off season, respectively.

If forecasts are adjusted with SIA (seasonal index adjustment), the average performance across items in terms of MAPE is 19.51% for busy season, which seems a little worse than NA, check Table 1. However, for off season, the average figure of SIA is 24.77%, a significant improvement of 33.16% of initial MAPE, check Table 2. There are 5 items improved out of 10 because of SIA for busy season, while there are 7 items get improved due to SIA's contribution for off season. The relatively poor performance of SIA in busy season may be attributed to the already good performance of NA compared to that of ARIMA without adjustment.

If adjustment is conducted with PA (proportional adjustment of mixed effect in mixed periods) in busy season, we see an improvement from 17.96% to 14.98%, a 16.59% improvement in average. The number of items with negative results is reduced to 3 also, check Table 1. However, for off season, it's a different story for PA, with MAPE just slightly reduced from 37.06% to 34.37%, a small improvement of 7.26% overall in average. Nevertheless, 5 out of 7 adjustments performed gets improvement in MAPE, besides, item 3, 8, and 10 didn't perform any PA adjustment, therefore, their MAPE are the same as that of NA, and almost only 1 out of 6 weeks needs to get adjustment with PA in off season, their contribution to the improvement of MAPE is therefore trivial, check table 2.

Table 1. Comparison of the accuracy of various forecast adjustment methods in busy season 2008

MAPE of various forecast adjustment methods						
item	**NA**	**SIA**	**PA**	**TA**	**ARIMA***	**ARIMA ad**
1	22.65%	18.18%	16.30%	11.87%	39.00%	**60.14%**
2	9.00%	**16.53%**	7.55%	**15.12%**	10.84%	**24.41%**
3	28.91%	28.67%	17.48%	20.91%	52.29%	42.75%
4	24.93%	21.90%	15.76%	18.66%	41.09%	29.31%
5	9.88%	**14.15%**	**11.40%**	**10.20%**	16.02%	**20.01%**
6	21.30%	19.43%	**21.82%**	18.75%	37.46%	30.14%
7	13.80%	**23.77%**	13.17%	**25.62%**	21.19%	**37.57%**
8	8.95%	**18.11%**	7.81%	**15.65%**	27.79%	20.32%
9	23.14%	17.34%	**24.10%**	15.97%	39.04%	26.85%
10	16.99%	**17.01%**	14.45%	16.05%	23.23%	19.35%
AVG	17.96%	**19.51%**	**14.98%**	16.88%	30.80%	**31.09%**

Note: forecasting with ARIMA (1, 1, 1)

If TA (total adjustment) is performed, since it combines both SIA and PA, for busy season, average MAPE reduced from 17.96% to 16.88%, the reduction of MAPE amounts to an average of 6.01% in improvement. For off season, the improvement is even more significant, the MAPE of NA improved from an average of 37.06% to an average of 22.78%, about 38.51% improvement over MAPE of initial forecasts. Note that if both SIA and PA have positive contribution in improving forecast accuracy, TA can be very effective, check item 1 and item 4 in Table 1, also item 1, 4, 5, 6, and 7 in Table 2.

If the adjustment of ARIMA is of concern, 6 out of 10 items get improved in busy season, check Table 1, but average percent of improvement from adjustment is a negative -0.942%, overall performance of ARIMA adjustment seems to be worse than that of PA and TA, however, it still is better than SIA in busy season.

Table 2. Comparison of the accuracy of various forecast adjustment methods in off season 2008

MAPE of various forecast adjustment methods						
item	**NA**	**SIA**	**PA**	**TA**	**ARIMA**	**ARIMA ad**
1	33.03%	20.93%	23.27%	12.12%	46.95%	18.95%
2	12.77%	**33.43%**	**16.58%**	**36.78%**	21.23%	20.80%
3	36.16%	12.37%	36.16%	12.37%	40.07%	**52.05%**
4	39.28%	23.41%	38.52%	24.32%	38.86%	17.87%
5	90.76%	46.36%	85.31%	41.77%	29.58%	23.08%
6	59.41%	8.83%	56.82%	10.43%	15.94%	**26.55%**
7	37.50%	32.12%	25.04%	19.38%	32.02%	28.37%
8	24.50%	19.55%	24.50%	19.55%	16.56%	8.38%
9	17.09%	**19.03%**	**17.36%**	**19.44%**	11.51%	**11.93%**
10	20.13%	**31.63%**	20.13%	**31.63%**	34.87%	**47.90%**
AVG	37.06%	24.77%	34.37%	**22.78%**	28.76%	**25.59%**

In off season, the adjustment of ARIMA displays a performance obviously better than its counterpart in busy season, the number of items get improved in MAPE remains the same as that in busy season. However, overall percentage of improvement from adjustment amounts to 11.02%, a performance much better than its counterpart in busy season.

4.3.1. Analysis of various adjustment methods from the perspective of adjustment size

In this subsection, the percentage of correct direction in adjustments over total adjustments is used to measure the performance of various adjustment methods. Whether the direction is correct or not depends on the comparison between initial forecast and the actual sale, if initial forecast is under-forecast, the correct direction of adjustment should be adjusted upwards, regardless of adjustment size. On the other hand, if initial forecast is over-forecast, the correct direction of adjustment should be adjusted downwards, regardless of adjustment size. However, if the initial forecast is within the range of [actual unit sales - 3%*actual unit sales, actual unit sales + 3%*actual unit sales], any subsequent adjustment with result less than or equal to initial over-forecast or any adjustment with result more than or equal to the under-forecast is perceived as adjustment in the correct direction.

In this study, any adjustment with result within less than 10% range of the initial forecast, regardless of adjustment direction, is regarded as small adjustment, otherwise, it's a large adjustment. In Table 3, all the small adjustment, in both busy season and off season,

has the ratio of adjustment with correct direction regardless of adjustment method, to be less than 60%.

On the other hand, large size adjustments seem to have a much more consistent and better performance in average than that of small ones, with the average correct ratio at least over 70% in off season except ARIMA adjustment, particularly, in busy season, an average ratio of correct direction around 86% is recorded for methods proposed in this study, whereas for adjustment of ARIMA, the percentage is 73.33%, still is not bad. This result is not surprising, in the literature, there are considerable similar evidences (Fildes and Goodwin, 2007; Syntetos et al., 2009).

Among three adjustment methods proposed in this study, SIA seems to have the best performance in terms of the ratio of correct direction adjustment in both small and large adjustments in off season, the overall ratio of correct direction adjustment on the basis of per adjustment is about 80%, but it does not necessarily mean that SIA provides the most positive contribution to improvement of forecast accuracy, because in Table 4, whether the adjustment is over-adjusted or not is not taken into account. Table 4 may offer some remedies in this regard. However, in busy season, PA seems to be the winner, with an overall ratio of correct direction adjustment close to 80%, whether it offers the most positive contribution to forecast accuracy improvement or not, still have to be crosschecked with other criteria like IMP in Table 4 to have an adequate assessment.

A measure called IMP, which can be used to evaluate the adjustment improvement, may be formulated as

$$IMP = APE_{ini} - APE_{ad} \quad (15)$$

Where, APE denotes absolute percentage error, APE_{ini} denotes APE of initial forecast, while APE_{ad} denotes APE after adjustment.

In Table 4, with the only exception of SIA applied in busy season, large adjustments consistently and significantly outperform small adjustments in terms of IMP, regardless of the adjustment method. The only exception is SIA which implies that most large size SIA adjustments with correct direction in Table 3 are actually over-adjusted. Note that in busy season, all three adjustment methods using small adjustment, the average IMP are all negative, among them, more than half of small adjustments made by SIA and TA are in correct direction, this means that there are serious issue of over-adjustment in the small size adjustments of these two adjustment methods.

Table 3. Comparing the performance concerning direction of adjustment of various adjust

	busy season 2008				off season 2008			
AD method	ratio of small ad	% of correct direction in small ad	ratio of large ad	% of correct direction in large ad	ratio of small ad	% of correct direction in small ad	ratio of large ad	% of correct direction in large ad
SIA	30/60	53.33%	30/60	83.33%	14/55	57.14%	41/55	**87.80%**
PA	5/20	40.00%	15/20	**93.33%**	1/8	0.00%	7/8	71.43%
TA	25/60	56.00%	35/60	82.86%	12/60	50.00%	48/60	79.17%
ARIMA ad	0/60	--	60/60	73.33%	0/60	--	60/60	58.33%

Table 4. Comparing IMP of various forecast adjustment methods in either small or large djustments

	busy season 2008		off season 2008	
AD method	avg IMP from small adjustments	avg IMP from large adjustments	avg IMP from small adjustments	avg IMP from large adjustments
SIA	-0.21%	**-5.63%**	0.66%	10.8%
PA	-3.86%	**14.37%**	-1.7%	**37.3%**
TA	-0.395%	4.44%	1.37%	28.8%
ARIMA ad	--	-1.036%	--	2.89%

As large size adjustments are concerned, in Table 3, even though PA is not the best performer in terms of percent of correct direction adjustment, however, in Table 4, it does have the best performance in terms of IMP, this implies that the number of over-adjustments of PA is the smallest. In busy season, PA still is the best performer in terms of IMP, it is the best one even from the viewpoint of percent of correct direction adjustment, crosscheck with Table 1, obviously, it provides the most consistent and the largest contribution to the improvement of forecast accuracy in busy season among various methods. However, in off season, because of the relatively less frequency of PA adjustments made, even though it still offers the best performance in terms of IMP, check Table 4, its overall contribution to improvement of forecast accuracy in terms of MAPE is not impressive.

On the other hand, the performance of TA in terms of IMP in off season though is not the best among these methods, due to its highest ratio of large size adjustment, check Table 3, TA still provides the most positive contribution to improvement of forecast accuracy in terms of MAPE, check Table 2.

4.3.2. Analysis of various adjustment methods from the perspective of lead time

In Figure 6, each forecasting horizon is divided into two parts, namely, the first 3 weeks and the second 3 weeks in both busy season and off season. Obviously, in busy season, the performance of various adjustment methods is relatively more stable than that of its counterpart in off season. Among different adjustment methods, PA seems to have the best performance in terms of IMP across different seasons, TA ranked second, and SIA still is the worst performer. Note that PA is only conducted in the second half of the forecasting horizon, check Figure 5, unlike other methods, it appears as a point in each season.

The performance of adjustment of ARIMA looks not so stable in different lead time of busy season, however, in off season, its performance is parallel to others in shape but obviously inferior to others.

4.4. Illustration of typical adjustment of different methods with two examples

To further present the detailed results of each adjustment method compared to initial forecast and Box and Jenkins ARIMA's forecasts and their simple adjustments as a yardstick, two graphs are drawn, see Figures 6-7. Note that to expose the detailed results of adjustments more clearly, all data from week 11 to week 50 of 2007 are cut off. The forecasting horizon is the first 6 weeks of 2008 (the busy season) in Figure 6.

Note that the original forecasts of the regression model are expressed in purple square dots in purple line, after adjustment of SIA, which are expressed in yellow triangle dots with yellow line, in weeks 3-4, the direction of adjustment are not correct, therefore, SIA

actually made obvious negative improvements. There is no adjustment made in the first 4 weeks for PA, However, in weeks 5-6, due to excellent adjustments of PA, which are expressed in empty blue diamonds with bold blue line, the forecast points are moved much closer to the actual sales, check TA in weeks 5-6 in Figure 6. Also note that down in the bottom are points formed by ARIMA, after 20% upward adjustment, these points are moved closer to the reality.

To illustrate the adjustment performance of different adjustment methods and initial forecasts as well as forecasts and adjustment of ARIMA in a bigger scope, the same item is used in Figure 7, in which, initial forecasts of the model are expressed in purple dots which are not close to the reality, after adjustments of SIA, check yellow triangle dots in weeks 11-16. In weeks 15-16, owing to the adjustment of PA, the forecasts are moved much closer to the reality. Because of the relative poor performance of NA in the first 4 weeks in Figure 7, and PA doesn't make any adjustment in this period, its overall performance is unlike its performance in Figure 6. TA, due to its combination of SIA and PA, particularly SIA, which has a good performance and therefore push overall TA closer to the reality.

The points of ARIMA stays in relative high positions from week 11 to week 16 and are also forming the very one most far away from the reality, after a 20% downward adjustment, they are moved much closer to the actual sales in average, check the bold green line with empty circles in Figure 7.

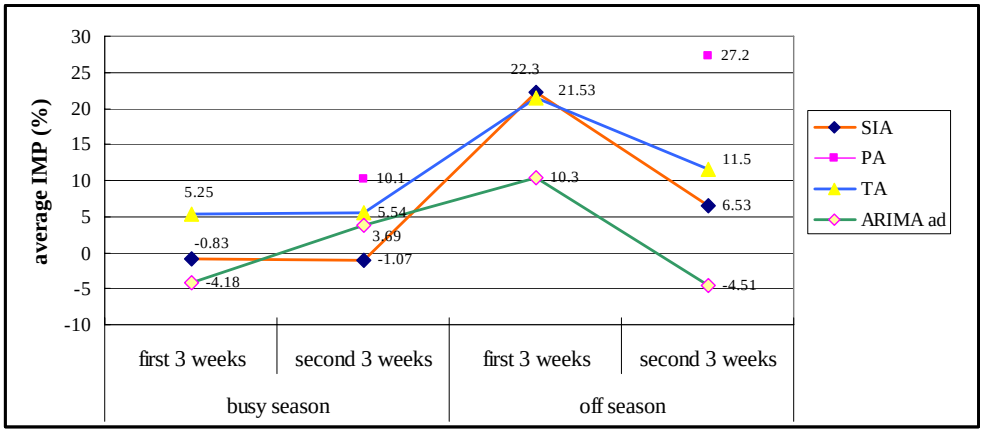


Figure 5. Comparison of average IMP of various adjustment methods on different lead times

4.5. Discussions

From the above analysis and explanations, on a per adjustment basis, PA offers the most improvement in terms of both MAPE and IMP in busy season, check Table 1 & 4, it also has the highest percentage in correct direction adjustment in busy season. In off season, since it is rarely used (the mixed effect condition is rare in comparison), its total contribution is not impressive.

TA, on the other hand, is more comprehensive in off season, and provides the most contribution in improving MAPE, even though in terms of percentage of correct direction adjustment, it is not the best performer. Crosscheck Table 3 and Table 4, it is easy to see that, the correct direction is the prerequisite for an adjustment of any kind to improve forecasting accuracy, but due to the issue of over-adjustment, many correct direction adjustments still have negative contributions to forecasting accuracy.

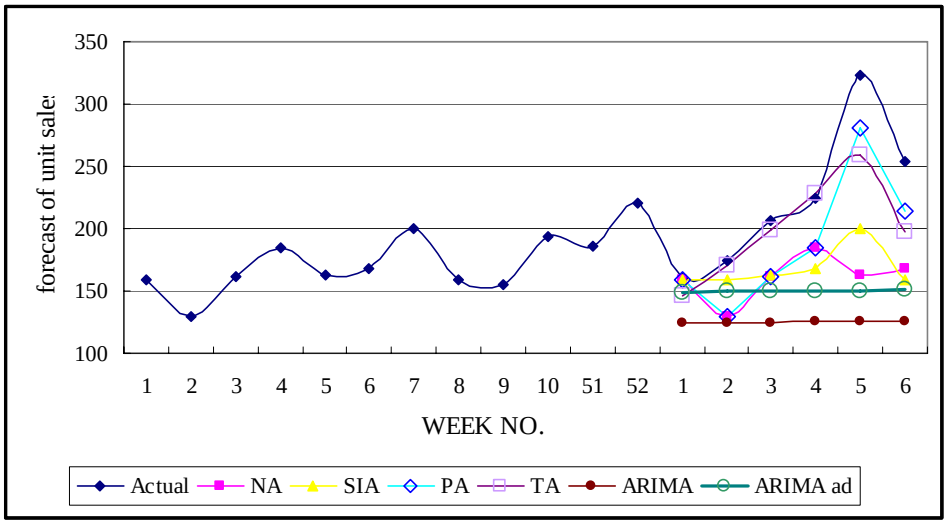


Figure 6. Comparison of adjustment performance in busy season 2008 with different adjustment methods on item 4.

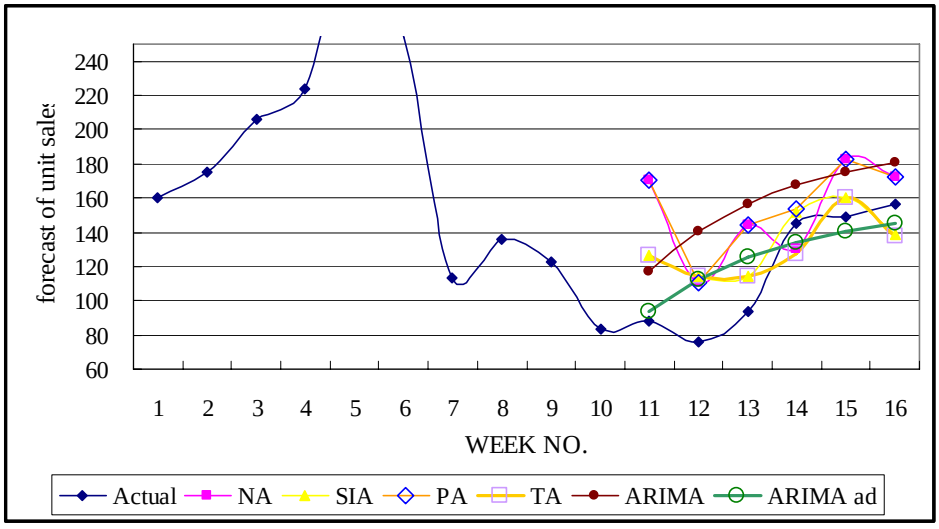


Figure 7. Comparison of adjustment performance in off season 2008 with different adjustment methods on item 4.

As adjustment of ARIMA is concerned, its performance in terms of MAPE is negative in average in busy season, nevertheless, in off season, the average MAPE after adjustment is a not bad 25.59%, a performance quite close to that of TA in the same season. Therefore, if efficiency is an important issue in forecasting, ARIMA with adjustment is a good alternative in off season. Otherwise, TA is the best tool of adjustment in off season. In busy season, PA is the best choice in forecast adjustment, due to its significantly much better performance.

Besides, with the only exception of SIA in busy season, if the performance is measured in terms of both percentage of correct direction adjustment and IMP in average, large size adjustment has a significant advantage over small size adjustment, regardless of season. Our model assumes that there will be no big difference between actual promotion activities and those specified in promotion proposals in forecast horizon, if this is not true,

there will be larger MAPE incurred in the original forecasts and various types of forecast adjustments. The relatively more accurate performance of original forecasts in busy season than in off season may due to the fact that promotion and calendar effects are so strong that they dominate unit sales in busy season, while in off season, these effects are much less obvious and much less frequent as in busy season, other factors like seasonal index realignment, competitors' actions and so on may have critical impacts on unit sales therein.

5. Conclusions

The forecasting adjustment mechanism proposed in this study only concerning with realignment of seasonal indices and the anticipated mixed effect of certain variables, such as the multiplier of the effect of promotion mix, and the multiplier of holiday effect, already incorporated in the regression model and assessed with GA which is more flexible and is capable of deriving more realistic coefficient of variables than most other conventional alternatives. Therefore, adjustment mechanism proposed in this study is a necessary and natural extension of the regression model. And in the process of forecast adjustment, subjective judgement based on contextual information is minimized.

Among three adjustment methods embedded in the adjustment mechanism of this study, SIA focuses on the realignment of seasonal index in forecasting horizon in a different year than the year of referenced periods, PA provides the necessary reassessment of mixed effect in mixed periods and is capable of offering the most contribution to the improvement of forecasting accuracy on per adjustment basis and is also the best performer in average in busy season. However, in off season, since both SIA and PA provide positive contribution in improving forecasting accuracy, and TA combines the above two adjustment methods, it offers the most comprehensive and most reliable adjustment in off season of this study, even though it's not necessarily the best performer in improving initial model-based forecast on a per adjustment basis in average.

In off season, ARIMA with adjustment, which just moves down the original forecast by 20%, also provides very good forecast accuracy close to that of TA in average, besides, ARIMA is embedded in most statistical software package and is very handy, in case efficiency is an important requirement in forecasting, ARIMA with adjustment intuitively is a good alternative.

References

1. Armstrong, S. **Forecasting by extrapolation: Conclusion from 25 years of research**, Interfaces 14, 6, 1984, pp. 52-66
2. Armstrong, S. **Long-range forecasting: from crystal ball to computer**, New York; Wiley, 1985
3. Armstrong, S., Collopy, F. and Thomas, Y. J. **Decomposition by causal forces: a procedure for forecasting complex time series**, International Journal of Forecasting, 21, 2005, pp. 25-36
4. Belsley, D. A., Kuh, E., and Welsch, R. E. **Regression Diagnostics. Identifying Influential Data and Sources of Collinearity**, John Wiley & Sons, New York, 1980
5. Belsley, D.A. **Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise**, Journal of Econometrics, *20*, 1982, pp. 211-253

6. Berilant, J., de wet, T., and Goegebeur, Y. **A goodness-of-fit statistic for Pareto-type behavior**. Journal of Computational and Applied Mathematics, 186, *1*, 2005, pp. 99-116
7. Blattberg. R. C. and Hoch, S. J. **Database models and managerial intuition! 50% model + 50% manager**, Management Science, 36. 8, 1990, pp. 887-899
8. Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. **Time series analysis: Forecasting and control (3rd ed.)**, Englewood Cliffs, NJ7 Prentice Hall, 1994
9. Bunn, D. W. and Salo, A. A. **Adjustment of forecasts with model consistent expectations**, International Journal of Forecasting, 12, 1996, pp. 163-170
10. Carroll. R. J. and Ruppert. D. **Transformation and Weighting in Regression**, Chapman and Hall, New York, NY, USA, 1998, pp. 115-160
11. Dalrymple, D. J. **Sales forecasting practices, results from a United States survey**, International Journal of Forecasting, 3, 1987, pp. 379-391
12. De Gooijer, J. G. and Hyndman, R. J. **25 years of time series forecasting**, International Journal of Forecasting, 22, 2006, pp. 443- 473
13. De Jong, K. A. and Spears, W. M. **A formal analysis of the role of multi-point crossover in genetic algorithms**, Annals of Mathematics and Artificial Intelligence 5(1), 1992, pp. 1–26
14. Edmundson, R. H., **Decomposition: A strategy for judgmental forecasting**, Journal of Forecasting, 19, pp. 305-314, IEEE Transactions on Evolutionary Computation, 3(2), 1990, pp. 124-141
15. Eiben, A. E., Hinterding, R. and Michalewicz, Z. **Parameter control in evolutionary algorithms**, IEEE Transactions on Evolutionary Computation, 3(2), 1999, pp. 124-141
16. Fildes, R. and Goodwin,P. **Good and bad judgement in forecasting: Lessons from four companies**, Foresight, Fall, 2007, pp. 5-10
17. Foekens, E. W., Leeßang, P. S. H. and Wittink, D. R.. **Varying parameter models to accommodate dynamic promotion effects**, Journal of Econometrics, 89, 1999, pp. 249-268
18. Franses, P. H. and McAleer, M. **Testing for unit roots and non-linear transformations**, Journal of Times Series Analysis, *19*, 1998, pp. 147-164
19. Goldberg, D. E. **Genetic Algorithms, in Search, Optimization & Machine Learning**, Addison-Wesley, Boston, MA, USA, 1989
20. Goldberg, D. E. **Simple genetic algorithms and the minimal deceptive problem**, in: Davis, L. (ed.) "Genetic Algorithms and Simulated Annealing", Hyperion Books, New York, NY, USA, Morgan Kaufmann, San Francisco, CA, USA, 1987, pp. 74-88
21. Goodwin, P. and Fildes, R. **Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy?** Journal of Behavioral Decision Making, 12, 1999, pp. 37-53
22. Heerde, H. J. Van, Leeflang, P. S. H. and Wittink, D. R. **Flexible decomposition of price promotion effects using store-level scanner data**, 2002b
23. Heerde, H. J. Van, Leeflang, P. S. H. and Wittink, D. R. **How promotions work: Scan*Pro-based evolutionary model building**, Schmalenbach Business Review, 54, 2002a, pp. 198-220
24. Hendry, D. F. **Econometric modeling**, Lecture notes for the PhD course in econometric modelling and economic forecasting, Department of Economics, University of Oslo, 2000
25. Holland, J. H. **Adaptation in natural and artificial systems**, University of Michigan Press, Ann Arbor, MI, USA. (extended new edition, MIT Press, Cambridge, MA, USA, 1992), 1975
26. Klassen, R. D., and Flores, B. E. **Forecasting practices of Canadian firms: Survey results and comparisons**, International Journal of Production Economics, 70, 2001, pp. 163-174

27. Kleinmutz, B. **Why we still use our heads instead of formulas: towards an integrative approach**, Psychologieal Bulletin, 107, 1990, pp. 296-310
28. Lilliefors, H. W. **On the Kolmogorov-Smirnov test for normality with mean and variance unknown**. Journal of the American Statistical Association, 64, 1967, pp. 399-402
29. Lim, J. S. and O'Connor, M. **Judgmental forecasting with time series and causal information**, International Journal of Forecasting, 12, 1996, pp. 139-153
30. Liu, Z., Zhou, J., and Lai, S. **New adaptive genetic algorithm based on ranking**, Proceedings of the Second International Conference on Machine Learning and Cybernetics, 2003
31. Makridakis, S. and Maclachlan, D. **Judgmental biases in applied decision making situations**, Working Paper, INSEAD, Fountainbleau, France, 1984
32. Mathews, B. P. and Diamantopoulos, A. **Managerial intervention in forecasting: An empirical investigation of forecast manipulation**, International Journal of Research in Marketing, 3, 1986, pp. 3-10
33. Nikolopoulos, K., Fildes, R., Goodwin, P. and Lawrence, M. **On the accuracy of judgemental interventionson forecasting support systems**, Working paper 2005/022, Lancaster University Management School, 2005
34. Pham, D. T. and Karaboga, D. **Genetic algorithms with variable mutation rates: Application to fuzzy logic controller design**, Proceedings of the I MECH E Part I. Journal of Systems & Control Engineering, 211(2), 1997, pp.157–167
35. Rawlings, J. O., Pantula, S. G. and Dickey, D. A. **Applied regression Analysis—A Research Tool**, SpringerVerlag, New York INC. New York, NY, USA, 1998, pp. 371-372
36. Remus, W., O'Connor, M. and Griggs, K. **Does reliable information improve the accuracy of judgmental forecasts?** International Journal of Forecasting, 11, 1995, pp. 285-293
37. Salo, A. A. and Bunn, D. W. **Decomposition in the assessment of judgmental probability forecasts**, Technological forecasting and Social Change, 149, 1995, pp. 13-25
38. Sanders, N. and Ritzman, L. P. **The need for contextual and technical knowledge in judgmental forecasting**, Journal of Behavioral Decision Making, 5, 1992, pp. 39-52
39. Sanders, N. R. and Manrodt, K. B. **Forecasting practices in US corporations: Survey results**, Interfaces 24, 1994, pp. 92-100
40. Scapolo, F. and Miles, I. **Eliciting experts' knowledge: A comparison of two methods**, Technological Forecasting & Social Change, 73, 2006, pp. 679-704
41. Schaffer, J. D., Caruana, R. A., Eshelman, L. J., and Das, R. **A study of control parameters affecting online performance of genetic algorithms for function optimization**, In Proceedings of the Third International Conference on Genetic Algorithms, 1989, pp. 51–60
42. Skalak, D. B. **Prototype and feature selection by sampling and random mutation hill climbing algorithms**, In "Proceedings of the Eleventh International Conference on Machine Learning", New Brunswick, NJ. Morgan Kaufmann, 1994, pp. 293-301
43. Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R. and Goodwin, P. **The effects of integreating management judgement into intermittent demand forecast**s, International Journal of Production Economics, 118, 2009, pp. 72-81
44. Tversky, A. and Kahneman, D. **Judgement under uncertainty: Heuristics and biases**, Science, 185, 1974, pp. 1124-1131
45. Wang, C. L., and Wang, L. C. **A GA-based sales forecasting model incorporating promotion factors**, Working Paper., Department of Industrial Engineering and Enterprise information, Tunghai University, Taiwan, ROC, 2009
46. Williams, T. M. **Adaptive Holt-Winters forecasting**, Journal of Operational Research Society, 38 (6), 1987, pp. 553-560

Appendix A.

Table A1. Results of model checking

	Sample A			Sample B		
item	Normality	CI[a] (mean of max)	ACF	Normality	CI (mean of max)	ACF
1	O[b]	105.341	X[c]	O	93.677	X
2	O	118.517	X	O	6.358	X
3	O	116.247	O	O	82.962	O
4	O	51.337	O	O	78.099	O
5	O	67.519	O	O	170.917	X
6	O	21.686	O	O	113.379	O
7	O	77.332	O	O	91.224	X
8	O	40.015	O	O	118.214	O
9	O	195.611	O	O	227.116	O
10	O	374.608	O	O	191.684	X

Note: a. CI denotes condition index.
b. O denotes a success to pass the test.
c. X denotes a failure to pass the test.

[1]**Chin-Lien Wang,** is a PhD candidate of Industrial Engineering & Enterprise Information at Tunghai University in Taiwan. He got his MBA from Michigan State University in 1985, and got his MS in computer science from DePaul University in 1989. At present, he teaches and does research at the Department of Business Administration, Ling Tung University in Taiwan. His current research interest includes business modeling, search methods, and time series forecasting.
Department of Industrial Engineering and Enterprise information, Tunghai University, 181 Section 3, Taichung Harbor Rd. Taichung, 40704 Taiwan, ROC
Department of Business Administration, Ling Tung University, 1 Ling Tung Rd. Taichung,
Taiwan, ROC

[2]**Dr. Li-Chih Wang** is a Professor and Chairman of Industrial Engineering and Enterprise Information at Tunghai University, Taiwan. He earned his B.S. degree in Industrial Engineering from Tunghai University, Taiwan and M.S. and Ph.D. degrees in Industrial and Systems Engineering from The Ohio State University, U.S.A. His teaching covers logistics and supply chain management, systems analysis, manufacturing planning and control system, and ERP-II at the levels of undergraduate, master's, and doctorate.
Dr. Wang is currently working on supply chain management (SCM) and advanced planning & scheduling (APS) system development. He has conducted extensive research leading to the current academic and industrial projects (e.g., memory module, electronic assembly, TFT-LCD). Dr. Wang has published three SCM/APS related books which are widely referenced in Taiwan and numerous scholarly papers and reports in these areas. His research has been supported by both government agencies (e.g., National Science Council, Ministry of Education, Ministry of Economy) and industry (e.g., Kingston, Macronix, EverTek, Farcent).
Dr. Li-Chih Wang has consulted for a number of high tech industries and served as Chief Advisor for developing Digichain's supply chain planning (SCP) and APS software/system which has been implemented in a number of high tech industries. He also actively promotes SCM concept and system in government, research institute, academy and industry for the last 15 years.

[3] Codification of references:

[1]	Armstrong, S. **Forecasting by extrapolation: Conclusion from 25 years of research**, Interfaces 14, 6, 1984, pp. 52-66
[2]	Dalrymple, D. J. **Sales forecasting practices, results from a United States survey**, International Journal of Forecasting, 3, 1987, pp. 379-391
[3]	Kleinmutz, B. **Why we still use our heads instead of formulas: towards an integrative approach**, Psychologieal Bulletin, 107, 1990, pp. 296-310
[4]	Bunn, D. W. and Salo, A. A. **Adjustment of forecasts with model consistent expectations**, International Journal of Forecasting, 12, 1996, pp. 163-170
[5]	Sanders, N. and Ritzman, L. P. **The need for contextual and technical knowledge in judgmental forecasting**, Journal of Behavioral Decision Making, 5, 1992, pp. 39-52
[6]	Makridakis, S. and Maclachlan, D. **Judgmental biases in applied decision making situations**, Working Paper, INSEAD, Fountainbleau, France, 1984

[7]	Mathews, B. P. and Diamantopoulos, A. **Managerial intervention in forecasting: An empirical investigation of forecast manipulation**, International Journal of Research in Marketing, 3, 1986, pp. 3-10
[8]	Klassen, R. D., and Flores, B. E. **Forecasting practices of Canadian firms: Survey results and comparisons**, International Journal of Production Economics, 70, 2001, pp. 163-174
[9]	Goodwin, P. and Fildes, R. **Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy?** Journal of Behavioral Decision Making, 12, 1999, pp. 37-53
[10]	Blattberg. R. C. and Hoch, S. J. **Database models and managerial intuition! 50% model + 50% manager**, Management Science, 36. 8, 1990, pp. 887-899
[11]	Nikolopoulos, K., Fildes, R., Goodwin, P. and Lawrence, M. **On the accuracy of judgemental interventionson forecasting support systems**, Working paper 2005/022, Lancaster University Management School, 2005
[12]	Fildes, R. and Goodwin,P. **Good and bad judgement in forecasting: Lessons from four companies**, Foresight, Fall, 2007, pp. 5-10
[13]	Remus, W., O'Connor, M. and Griggs, K. **Does reliable information improve the accuracy of judgmental forecasts?** International Journal of Forecasting, 11, 1995, pp. 285-293
[14]	Lim, J. S. and O'Connor, M. **Judgmental forecasting with time series and causal information**, International Journal of Forecasting, 12, 1996, pp. 139-153
[15]	Scapolo, F. and Miles, I. **Eliciting experts' knowledge: A comparison of two methods**, Technological Forecasting & Social Change, 73, 2006, pp. 679-704
[16]	Armstrong, S., Collopy, F. and Thomas, Y. J. **Decomposition by causal forces: a procedure for forecasting complex time series**, International Journal of Forecasting, 21, 2005, pp. 25-36
[17]	Edmundson, R. H., **Decomposition: A strategy for judgmental forecasting**, Journal of Forecasting, 19, pp. 305-314, IEEE Transactions on Evolutionary Computation, 3(2), 1990, pp. 124-141
[18]	Salo, A. A. and Bunn, D. W. **Decomposition in the assessment of judgmental probability forecasts**, Technological forecasting and Social Change, 149, 1995, pp. 13-25
[19]	Tversky, A. and Kahneman, D. **Judgement under uncertainty: Heuristics and biases**, Science, 185, 1974, pp. 1124-1131
[20]	Armstrong, S. **Long-range forecasting: from crystal ball to computer**, New York; Wiley, 1985
[21]	Sanders, N. R. and Manrodt, K. B. **Forecasting practices in US corporations: Survey results**, Interfaces 24, 1994, pp. 92-100
[22]	Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. **Time series analysis: Forecasting and control (3rd ed.)**, Englewood Cliffs, NJ7 Prentice Hall, 1994
[23]	De Gooijer, J. G. and Hyndman, R. J. **25 years of time series forecasting**, International Journal of Forecasting, 22, 2006, pp. 443- 473
[24]	Foekens, E. W., Leeßang, P. S. H. and Wittink, D. R.. **Varying parameter models to accommodate dynamic promotion effects**, Journal of Econometrics, 89, 1999, pp. 249-268
[25]	Heerde, H. J. Van, Leeflang, P. S. H. and Wittink, D. R. **How promotions work: Scan*Pro-based evolutionary model building**, Schmalenbach Business Review, 54, 2002a, pp. 198-220
[26]	Heerde, H. J. Van, Leeflang, P. S. H. and Wittink, D. R. **Flexible decomposition of price promotion effects using store-level scanner data**, 2002b
[27]	Carroll. R. J. and Ruppert. D. **Transformation and Weighting in Regression**, Chapman and Hall, New York, NY, USA, 1998, pp. 115-160
[28]	Franses, P. H. and McAleer, M. **Testing for unit roots and non-linear transformations**, Journal of Times Series Analysis, *19*, 1998, pp. 147-164
[29]	Belsley, D.A. **Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise**, Journal of Econometrics, *20*, 1982, pp. 211-253
[30]	Belsley, D. A., Kuh, E., and Welsch, R. E. **Regression Diagnostics. Identifying Influential Data and Sources of Collinearity**, John Wiley & Sons, New York, 1980
[31]	Hendry, D. F. **Econometric modeling**, Lecture notes for the PhD course in econometric modelling and economic forecasting, Department of Economics, University of Oslo, 2000
[32]	Wang, C. L., and Wang, L. C. **A GA-based sales forecasting model incorporating promotion factors**, Working Paper., Department of Industrial Engineering and Enterprise information, Tunghai University, Taiwan, ROC, 2009
[33]	Holland, J. H. **Adaptation in natural and artificial system**s, University of Michigan Press, Ann Arbor, MI, USA. (extended new edition, MIT Press, Cambridge, MA, USA, 1992), 1975
[34]	Goldberg, D. E. **Simple genetic algorithms and the minimal deceptive problem**, in: Davis, L. (ed.) "Genetic Algorithms and Simulated Annealing", Hyperion Books, New York, NY, USA, Morgan Kaufmann, San Francisco, CA, USA, 1987, pp. 74-88
[35]	Goldberg, D. E. **Genetic Algorithms, in Search, Optimization & Machine Learning**, Addison-Wesley, Boston, MA, USA, 1989

[36]	De Jong, K. A. and Spears, W. M. **A formal analysis of the role of multi-point crossover in genetic algorithms**, Annals of Mathematics and Artificial Intelligence 5(1), 1992, pp. 1–26
[37]	Skalak, D. B. **Prototype and feature selection by sampling and random mutation hill climbing algorithms**, In "Proceedings of the Eleventh International Conference on Machine Learning", New Brunswick, NJ. Morgan Kaufmann, 1994, pp. 293-301
[38]	Liu, Z., Zhou, J., and Lai, S. **New adaptive genetic algorithm based on ranking**, Proceedings of the Second International Conference on Machine Learning and Cybernetics, 2003
[39]	Pham, D. T. and Karaboga, D. **Genetic algorithms with variable mutation rates: Application to fuzzy logic controller design**, Proceedings of the I MECH E Part I. Journal of Systems & Control Engineering, 211(2), 1997, pp.157–167
[40]	Schaffer, J. D., Caruana, R. A., Eshelman, L. J., and Das, R. **A study of control parameters affecting online performance of genetic algorithms for function optimization**, In Proceedings of the Third International Conference on Genetic Algorithms, 1989, pp. 51–60
[41]	Eiben, A. E., Hinterding, R. and Michalewicz, Z. **Parameter control in evolutionary algorithms**, IEEE Transactions on Evolutionary Computation, 3(2), 1999, pp. 124-141
[42]	Lilliefors, H. W. **On the Kolmogorov-Smirnov test for normality with mean and variance unknown**. Journal of the American Statistical Association, *64*, 1967, pp. 399-402
[43]	Berilant, J., de wet, T., and Goegebeur, Y. **A goodness-of-fit statistic for Pareto-type behavior**. Journal of Computational and Applied Mathematics, 186, *1*, 2005, pp. 99-116
[44]	Rawlings, J. O., Pantula, S. G. and Dickey, D. A. **Applied regression Analysis—A Research Tool**, SpringerVerlag, New York INC. New York, NY, USA, 1998, pp. 371-372
[45]	Williams, T. M. **Adaptive Holt-Winters forecasting**, Journal of Operational Research Society, 38 (6), 1987, pp. 553-560
[46]	Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R. and Goodwin, P. **The effects of integreating management judgement into intermittent demand forecast**s, International Journal of Production Economics, 118, 2009, pp. 72-81

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

INDUCTION OF MEAN OUTPUT PREDICTION TREES FROM CONTINUOUS TEMPORAL METEOROLOGICAL DATA[1]

Dima ALBERG[2]
PhD Candidate, Department of Information Systems Engineering,
Ben-Gurion University of the Negev, Beer-Sheva, Israel

E-mail: alberg@bgu.ac.il

Mark LAST[3]
PhD, Associate Professor, Department of Information Systems Engineering,
Ben-Gurion University of the Negev, Beer-Sheva, Israel

E-mail: mlast@bgu.ac.il

Avner BEN-YAIR[4,5]
PhD, Department of Industrial Engineering and Management,
Sami Shamoon College of Engineering, Beer-Sheva, Israel

E-mail: avnerb@sce.ac.il

Abstract: *In this paper, we present a novel method for fast data-driven construction of regression trees from temporal datasets including continuous data streams. The proposed Mean Output Prediction Tree (MOPT) algorithm transforms continuous temporal data into two statistical moments according to a user-specified time resolution and builds a regression tree for estimating the prediction interval of the output (dependent) variable. Results on two benchmark data sets show that the MOPT algorithm produces more accurate and easily interpretable prediction models than other state-of-the-art regression tree methods.*

Key words: *temporal prediction; inductive learning; time resolution; regression trees; split criteria; multivariate statistics; multivariate time series*

1. Introduction

The time dimension is one of the most important attributes of massive continuous temporal data sets or continuous data streams [13][6], where data arrives and has to be processed on a continuous basis. Sources of such data include real-time monitoring devices as: meteorological stations, traffic control systems, financial markets, etc. If we extract a portion of data arrived over a finite time period and store it in a persistent database, it becomes a temporal dataset. Generally, the time dimension is represented in a temporal

dataset as a calendar variable which has an agglomerative structure consisting of several time granules [22]: for example, 60 seconds represent 1 minute, 60 minutes represent one hour, etc. The correct choice of the pre-processing time granularity very often predetermines the accuracy and interpretability of data stream mining algorithms.

We are interested in prediction of temporal continuous variables, since they are abundant in most data streams mentioned above. However, the existing prediction methods (such as Regression Tree Models [3, 6, 18-19, 21, 24-25]) do not take the time dimension and time granularity into account, since they were developed for mainly static (time-invariant) databases. In this work, we present a new prediction algorithm capable to induce an accurate and interpretable model for estimating the values of a given continuous output variable in a massive temporal data set or a continuous data stream.

2. Problem Statement and Prior Work

In many data streams, the data is available as time-continuous statistical moments (mean, variance, etc.) calculated over pre-defined measurement cycles rather than raw values sampled at discrete points in time. Examples of such data streams include: meteorological data, financial data, factory control systems, sensor networks, etc. For example, a meteorological station might be continuously storing mean and variance estimation for a large number of meteorological attributes at predefined time intervals (e.g., every 10 minutes). A prediction model that is built using multiple statistical moments instead of discretely sampled exact values is likely to have a lower update cost, since as long as an attribute value remains within the prediction interval, fewer updates to the model will be required.

However, supervised predictive data mining models like regression models (GLM, MARS[10]) and Regression Trees (M5 [21], M5' [25], CART [3], GUIDE [19], RETIS [18], MAUVE [24], (M)SMOTI [6]) widely used at present for prediction of continuous target variables do not utilize multiple statistical moments of input and target attributes. Time-series prediction models (ARIMA, ARCH [9], and GARCH [2]), which carry out simultaneous prediction of continuous target variables represented by statistical moments are frequently non stable on significant volumes of non-stationary data and require labor-consuming reassessment at uncertain time intervals [1]. Another difficulty is to produce an interpretable set of prediction rules for such cases. Indeed, even supposing that it would be possible to build an accurate regression tree or a set of logical rules using the time dependent input attributes, the resulting model is likely to be very intricate and essentially impossible to interpret [12].

Interval prediction is an important part of the forecasting process aimed at enhancing the limited accuracy of point estimation. An interval forecast usually consists of an upper and a lower limit between which the future value is expected to lie with a prescribed probability. The limits are sometimes called forecast limits [26] or prediction bounds [5], whereas the interval is sometimes called a confidence interval [12] or a forecast region [16]. We prefer the more widely-used term prediction interval, as used by Chatfield [7] and Harvey [14], both because it is more descriptive and because the term confidence interval is usually applied to interval estimates for fixed but unknown parameters. In our case, a prediction interval is an interval estimate for an (unknown) future value of the output (dependent) variable. Since a future value can be regarded as a random variable at the time

the forecast is made, a prediction interval involves a different sort of probability statement from that implied by a confidence interval. The above considerations cause a need of a model which can process the incoming data in real time and on the basis of the received results to make interval prediction of target time dependent continuous variables with a given level of statistical confidence.

2. The Proposed Data Stream Mining Methodology

2.1. The Model Induction Algorithm Overview

The proposed algorithm is aimed at inducing a prediction model for a case where every input (predictive) temporal variable is continuous and in a given sliding window k, the two statistics of mean and variance are calculated for each measurement cycle. The output (predicted) variable is the mean value of a continuous temporal variable calculated for the future sliding window $k+\Delta$.

As inputs the algorithm receives the time resolution interval j, the two first statistical moments of each temporally continuous input variable $\overline{X}$ with user predefined lag Δ history and temporally continuous output variable Y, as well as the significance level α. In the conventional regression tree algorithm, the objective is to build an inductive predictor assuming the following functional form:

$$Y = \{\hat{y}_{jk}\} = f\{x_{jk-\Delta}\} \tag{1}$$

where the predicted target variable in a sliding window k is represented as a function of input numerical variables in $k-\Delta$ sliding window, where Δ is a user-specified prediction lag parameter.

The proposed algorithm will build an inductive predictor of the following form:

$$Y = \left\{\hat{y}_{jk}, w\left(T\left(\Theta_{Y_{jk}}\right)\right)\right\} = f\left\{\bar{x}_{jk-\Delta}, \hat{s}^2_{x_{jk-\Delta}}\right\}, \tag{2}$$

where

$$T\left(\Theta_{Y_{jk}}\right) = w\left(\Theta_{Y_{jk}} - \overline{\Theta}_{Y_j}\right)^T \overline{\Psi}(\Theta)\left(\Theta_{Y_{jk}} - \overline{\Theta}_{Y_l}\right),\ \Theta_{Y_{jk}} \in \left\{\bar{y}_{jk}, \hat{s}^2_{y_{jk}}\right\}. \tag{3}$$

Where for time resolution j in sliding window k, $\hat{y}_{jk}$ is the predicted value of temporally continuous output variable Y, $\Theta_{Y_{jk}}$ is the mixture mean variance parameter for output variable Y, $w\left(T\left(\Theta_{Y_{jk}}\right)\right)$ is the mixture density estimation weight for the output variable Y and $\bar{x}_{jk-\Delta}$, $\hat{s}^2_{x_{jk-\Delta}}$ are mean and standard deviation estimators of a temporally continuous input variable X for time resolution j in sliding window $k-\Delta$. Finally, in (3) for time resolution j in the sliding window k, the joint variable $T\left(\Theta_{Y_{jk}}\right)$ is the mean-

variance estimator based on the two first statistical moments of the output variable Y, $\overline{\Theta}_{Y_j}$ is the vector of the means of the parameter $\Theta_{Y_{jk}}$ and $\overline{\Psi}(\Theta)$ represents the within-group normalized covariance matrix for parameter $\Theta_{Y_{jk}}$, which estimates normalized mean variance covariance values of the predicted output variable Y. The confidence interval of the joint variable $T(\Theta_{Y_{jk}})$ can be approximated using the F distribution as follows:

$$UB, LB\left(T\left(\Theta_{Y_{jk}}\right)\right) = \frac{2(\ell_j - 1)(\ell_j + 1)}{\ell_j(\ell_j - 2)} \cdot F\left((\alpha/2),(1-\alpha/2),2,\ell_j - 2\right), \tag{4}$$

where ℓ_j is the number of measurement cycles in the sliding window k for time resolution j and α is the user-specified significance level (default value is 0.05).

When the variance is independent of the mean value of the variable, the values of the joint variable $T(\Theta_{Y_{jk}})$ are expected to lie inside the confidence interval (see (4)) implying that the interaction between mean and dispersion variables does not add information about the behavior of the corresponding input variable. However, when some values of the joint variable are found outside the boundaries UB and LB (see (4)), it can be said that the interaction between mean and dispersion variables adds further information about the input variable $\overline{X}$. These outliers provide sufficient information for the output variable prediction and therefore we can consider only the outliers when evaluating the candidate split points of an input variable. In case when no outliers are found, the algorithm checks the possibility to switch to the higher time resolution and if the higher resolution represents the initial (raw) time resolution, the algorithm proceeds as the regular RETIS [18] algorithm.

The impurity merit (5.1) and (5.2) contains two parts, left and right, whereas the major objective is to find the optimal split point $\overline{X}$, which minimizes the expression in (6):

$$Var\left(T(\Theta)^L\right) = p^L \sum_{i=1}^{N_L} w_X^L \left(T(\Theta)_i^L - T\left(\overline{\Theta}\right)^L\right)^2 \tag{5.1}$$

$$\left(Var\left(T(\Theta)^R\right) = p^R \sum_{i=1}^{N_R} w_X^R \left(T(\Theta)_i^R - T\left(\overline{\Theta}\right)^R\right)^2\right) \tag{5.2}$$

$$X^* = \arg\min_{T(\Theta)}\left(Var\left(T(\Theta)^L\right) + Var\left(T(\Theta)^R\right)\right) \tag{6}$$

Here $T(\Theta)^L$ and $T(\Theta)^R$ are the left (right) joint mean variance estimator values of target variable Y, p^L and p^R are the relative number of N_L (N_R) cases that are assigned to the left (right) child, while w_X^L (w_X^R) is the left (right) mixture density estimation weight for the target variable Y.

Thus, the best split at a node is defined as the split, which minimizes the weighted variance of the joint mean variance estimator.

2. 2 Performance Metrics

A validation dataset for time resolution j and sliding window w_j is made up of $k \in \{1,...,N\}$ instances (sliding windows), each mapping an input vector $(x_1,...,x_A)^k$ to a given target y_k. The error is given by: $e_k = \hat{y}_k - y_k$, where $\hat{y}_k$ represents the predicted value for the k-th input pattern (2). The overall performance is computed by a global metric, namely the Mean Absolute Error (MAE) and Root Mean Squared (RMSE). However, the RMSE is more sensitive to high volatility errors than MAE. In order to compare the accuracy of trees from different domains the % of Explained Variability (EV) defined as:

$$EV(T) = \frac{(SSE(Mean) - SSE(T))}{SSE(Mean)} \cdot 100\% \quad (7)$$

Here $Mean$ is a majority rule predictor, which always predicts the mean value of the training set, $SSE(Mean)$ and $SSE(T)$ are sum of square errors from the mean value and the value predicted by the evaluated regression tree (T) model, respectively. Another possibility to compare regression tree models is the Cost Complexity Measure (CCM) defined as:

$$CCM(T) = RMSE(T) + \alpha \cdot TS(T). \quad (8)$$

Here $RMSE(T)$ is the estimated error cost of regression tree T, $TS(T)$ is the number of terminal nodes in the tree, and α is the user defined non-negative cost complexity parameter adopted from [9], where it is shown that for a given complexity parameter, there is a unique smallest subtree of the saturated tree that minimizes the cost-complexity measure, which actually quantifies the tradeoff between the size of the tree and how well the tree fits the data.

3. Experimental Results

The performance of the MOPT algorithm proposed in the Section 2.1 was evaluated on ElNino data set from the UCI Machine Learning Repository [23]. The selected data set consist from numerical attribute types and belong to the multivariate spatio temporal regression domain. Finally, the performance of the complete Mean Output Prediction Tree (MOPT) algorithm was evaluated on the second data set, which represents a multivariate continuous data stream collected at a meteorological station in Israel during a period of about 8 years.

The algorithm performance is compared to four state-of-the-art prediction algorithms implemented by the Java API of WEKA [25]: M5P Tree [25] (Bagging M5P tree), M5-Rules [21] (Bagging M5-Rules), RepTree (Bagging RepTree) and by our implementation of the RETIS [18] algorithm (RETIS-M). The main difference between RETIS[18] and RETIS-M algorithm concludes in more fast splitting criterion implementation.

3.1. El Nino Data Set

The El Nino/Southern Oscillation (ENSO) cycle of 1982-1983, the strongest of the century, created many problems throughout the world. The El Nino dataset consists of the following attributes: buoy, date, latitude, longitude, zonal winds (west<0, east>0), meridian winds (south<0, north>0), relative humidity, air temperature and sea surface temperature. Data was taken from the buoys from as early as 1980 for some locations publicly available UCI Machine Learning Repository [23]. Other data that was taken in various locations are rainfall, solar radiation, current levels, and subsurface temperatures. The experimental data is represented using a single (daily) time resolution and it consists of 178,080 data instances. Important to note, that all data readings were taken at the same time of day and the target (predicted) variable is the subsurface temperature.

Finally in order to evaluate the predictive performance, the set of all examples was split into learning and testing examples sets in proportion 70:30.

The results in Table 1 show that under RMSE and Explained Variability criterions the MOPT and the RETIS-M algorithms are more accurate than other proposed algorithms in terms of t-test pair-wise difference. We have denoted by * the cases where the p value of the difference between MOPT and other algorithms is smaller than or equal to 5%. The MOPT algorithm outperforms significantly the other algorithms in the terms of cost complexity measure. Finally, we will to consider that our proposed MOPT Tree is more interpretable than RETIS-M tree in terms of Tree Size measure (7 vs. 23).

Table 1. El Nino data set learners comparison

Learner	RMSE	TS	CCM	EV
B-M5 Rules	0.84*	7	1.01*	0.46*
B-M5P Tree	0.83*	10	1.07*	0.47*
B-REPTree	1.57*	5	1.69*	NA
M5 Rules	0.86*	7	1.03*	0.45*
M5P Tree	0.84*	8	1.03*	0.46*
MOPT	**0.60**	7	0.77	0.62
REPTree	1.57*	3	1.64*	NA
RETIS-M	**0.63**	23	1.18*	**0.60**

3.2. Israel Meteorology Data Set

In this experiment we used the data collected at a meteorological station in Israel during a period of about 8 years (from 01/07/1997 to 31/08/2005). Spatio- temporal meteorological attributes (such as pressure, temperature, solar radiation, horizontal wind: direction, speed, gust speed, gust time, and vertical wind: down-up and up direction) are measured constantly in time and saved every 10 minutes in the form of mean and variance. The selected data set exceeds 1,500,000 records. The total number of temporal and meteorological attributes collected at the three stations is 22. Our first experiment was run on the summer months (JUN, JUL and AUG) only. The experimental data was represented using 5 time resolutions (10, 30, 60, 90 and 120 Minutes). The algorithms were run for 11:00-12:00 and 23:00 – 24:00 hours prediction.

The aim of this experiment was to compare the different state-of-the-art algorithms for different time resolutions in order to be able to predict wind directions for short time range (now-casting) up to 8 hours sliding window ahead. We have shown that the most state-of-the-art algorithms gave the same or poorer quality of results and less interpretable

trees as the proposed Mean Output Prediction Tree (MOPT) algorithm. The results also pinpoint the fact that sometimes there was no need to use very high time resolution, but only lower time resolutions, since the statistical measures checked (i.e. RMSE, Cost Complexity Measure and Percentage of Explained Variability) were similar.

Tables 2, 3 compare between the proposed MOPT algorithm and four state-of-the-art algorithms: Modified RETIS (RETIS-M), M5P and REPTree in five time resolution scales in terms of Cost Complexity and Explained Variability measures. For more effective evaluation of the MOPT algorithm we performed short term prediction of 11:00 and 23:00 hour. The sliding window size and the prediction lag were set to 8 hours and 3 hours, respectively. Thus for predicting 11 hour wind direction we collected data from 00:00 to 8:00 and for predicting 23 hour wind direction we collected data from 12:00 to 20:00. In each prediction case, the main issue is to predict wind direction 3 hours ahead therefore fast robust and accurate prediction algorithm producing a compact model is needed. For each time resolution, we preprocessed the raw data and calculated the first two statistical moments for each attribute in every measurement cycle. The MOPT algorithm refers to each input attribute as a 2-dimensional array (two moments × number of instances) and determines the split point with the aid of two moments target variable impurity and the variance of the input variable. As in the previous experiments, the differences are considered statistically significant when the p-value of the t-pair-wise test statistic is smaller than or equal to 5% which signed by *.

Table 2. MOPT and state-of-the-arts models cost complexity Measure (CCM) cross resolutions results for 11:00 hour prediction

TR	MOPT	RETIS-M	M5P	REPTree
10	116.61	227.28*	117.18	124.23
30	117.20	245.48*	149.48*	132.07
60	120.58	251.13*	172.78*	143.31
90	120.32	236.58*	171.03*	139.11
120	114.61	240.73*	188.88*	148.70*

In 10 minutes resolution the M5P slightly outperforms the proposed MOPT model and significantly better than other models. In other resolutions the MOPT model significantly better than other state-of-arts models. This result pinpoint to the fact that adding second moments to split criterion improves quality of prediction for higher time resolution.

Table 3. MOPT and state-of-the-arts models cost complexity measure (CCM) cross resolutions results for 23:00 hour prediction

TR	MOPT	RETIS-M	M5P	REPTree
10	43.95	60.04*	55.00	53.01
30	43.27	59.84*	60.53*	59.74*
60	45.45	62.28*	53.38	60.72*
90	44.55	59.76*	57.76	74.45*
120	44.53	61.82*	53.60	60.00*

By comparison to state-of-the-art algorithms, the MOPT algorithm demonstrates more stable prediction accuracy with a more compact tree size in 23:00 hour prediction (for example in 10 minutes resolution the size of MOPT tree is 502 versus 2947 of M5P). In this

case, the regression tree pruning procedure may significantly reduce the final size of the tree, but this procedure is out of scope in the proposed MOPT approach because our main purpose is to build accurate and compact tree with minimal access to the sliding window training data.

Table 4. MOPT and state-of-the-arts models explained variability (%EV) cross resolutions results for 11:00 hour prediction

TR	MOPT	RETIS-M	M5P	REPTree
10	45.24%	32.46%	60.18%	59.96%
30	45.35%	25.35%	41.99%	50.26%
60	44.59%	25.78%	29.73%	43.43%
90	44.32%	30.45%	29.69%	53.98%
120	48.65%	32.39%	23.09%	37.34%

Table 5. MOPT and state-of-the-arts models explained variability (%EV) cross resolutions results for 23:00 hour prediction

TR	MOPT	RETIS-M	M5P	REPTree
10	29.3%	-6.1%	22.7%	8.7%
30	30.7%	-5.6%	-10.4%	12.1%
60	25.6%	-14.7%	8.5%	12.7%
90	28.5%	-4.8%	-13.3%	-43.9%
120	28.3%	-13.3%	12.9%	16.4%

The final stage of this experiment presented in the Tables 4-5 demonstrates the comparison of Percentage of Explained Variability EV between five defined models and time resolutions for 11:00 and 23:00 Hours respectively. The cells with negative explained variability percent indicate the fact that the induced model is poorer (less accurate) than a simple majority rule mean model. For example, RETIS-M, M5P and REPTree models have not contributed to the explained variability of the 23:00 hour prediction. Important to emphasize, that three state-of-the-art algorithms did not scale well to the low time resolution of 120 minutes.

4. Conclusions

In this work, we have presented the two moments (mean-variance) Mean Output Prediction Tree algorithm (MOPT), which is able to predict large amounts of massive temporal data sets. The proposed algorithm differs from the state-of-the-art regression algorithms in the splitting of each input and output feature to two moments according to the input time resolution and it can also identify the most appropriate prediction time resolution that minimizes the prediction error and builds more compact interval based regression tree.

The two conducted experiments indicate that the proposed algorithm produces more accurate and compact models by comparison to the modern state-of-the-art regression tree algorithms.

5. References

1. Alberg, D., Shalit, H. and Yosef, R. **Estimating stock market volatility using asymmetric GARCH models**, Applied Financial Economics, vol. 18(15), 2008, pp. 1201-1208
2. Bollerslev, T.A. **Conditionally heteroskedastic time series model for speculative prices and rates of return**, Review of Economics and Statistics, vol. 69, 1987, pp. 542–547
3. Breiman, L. and Friedman, J. **Estimating Optimal Transformations for Multiple Regression and Correlation**, J. Amer. Statist. Assoc., vol. 80, 1985, p. 580
4. Breiman, L., Friedman, J., Olshen, R. and Stone, C. **Classification and Regression Trees**, Wadsworth Int. Group, Belmont California USA, 1984
5. Brockwell, P. and Davis, R. **Time Series: Theory and Methods**, 2nd ed., New York: Springer-Verlag, 1991
6. Ceci, M., Appice, A. and Malerba, D. **Comparing simplification methods for model trees with regression and splitting nodes**, Foundations of Intelligent Systems, 14th International Symposium, vol. 2871 of LNAI, 2003, pp. 49–56
7. Chatfield, C. **The Analysis of Time Series**, 5th ed., London: Chapman and Hall, 1995, pp. 431-441
8. Cortez, F. and Morais, R.D. **A data mining approach to predict forest fires using meteorological data**, in Neves, J., Santos, M.F. and Machado, J. (eds.) "New Trends in Artificial Intelligence", Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, 2007, pp. 512-523
9. Engle, R. **Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation**, Econometrica, vol. 50, 1982, pp. 987-1007
10. Friedman, J. **Multivariate adaptative regression splines**, Annals of Statistics, 1991, pp. 1-19
11. Gama, J., Rocha, R., Medas, P., Wah, B. and Wang, J. **Accurate decision trees for mining high-speed data streams**, KDD 2003, 2003 , pp. 523-528
12. Granger, C. and Newbold, P. **Forecasting Economic Time Series**, 2nd ed., Academic Press, New-York, 1986
13. Han, J., Cai, D. and Chen, Y. **Multi dimensional analysis of data streams using stream cubes in data streams models and algorithms**, ed. Aggarwal, C., 2007, pp. 103-125
14. Harvey, A. **Forecasting Structural Time Series Models and the Kalman Filter**, C U P, Cambridge, 1989
15. Hulten, G., Spencer, L. and Domingos, P. **Mining time-changing data stream**, Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001, pp. 97-106
16. Hyndman, R. **Highest-density forecast regions for non-linear and non-normal time series models**, Journal of Forecasting, vol. 14, 1996
17. Kamber, M. and Han, J. **Data Mining: Concepts And Techniques**, 2nd ed., The Morgan Kaufmann Series in Data Management Systems, 2007
18. Karalic, A. **Linear regression in regression tree leaves**, in "Proceedings of International School for Synthesis of Expert", vol. 10(3), 1992, pp. 151-163
19. Loh, W.Y. **Regression tree models for designed experiments**, 2nd Lehmann Symposium, Knowledge, 2006, pp. 210-228
20. Ma, J., Zeng, D. and Chen, H. **Spatial-temporal cross-correlation analysis: a new measure and a case study in infectious disease**, in Mehrotra, S. *et. al.* (eds) Informatics, ISI, LNCS 3975, 2006, pp. 542-547
21. Quinlan, J. **Learning with continuous classes**, in "Proceedings of the 5th Australian Joint Conference on Artificial Intelligence", Singapore World Scientific, 1992
22. Spranger, S. and Bry, F. **Temporal data modeling and reasoning for information statistical view of boosting**, The Annals of Statistic, vol. 38(2), 2006, pp. 337-374

23. Vens, C. and Blockeel, H. **A simple regression based heuristic for learning model trees**, Intelligent Data Analysis, vol. 10(3), 2006, pp. 215-236
24. Wang, Y. and Witten, I. **Inducing of model trees for predicting continuous classes**, in "Proceedings of the 9th European Conference on Machine Learning", Springer-Verlag, 1997, pp. 128-137
25. Wei, S., Dong, Y. and Pei, J.-B. **Time Series Analysis**, Redwood City Cal Addison-Wesley, 1990
26. * * * **UCI, Machine Learning Repository**, http://archive/ics/uci edu/ml/index html

[1]**Acknowledgment:** This study was partially supported under a research contract from the Israel Ministry of Defense.

[2]**Dima Alberg** is currently Ph.D. candidate in the Department of Information Systems Engineering, Ben-Gurion University of the Negev under the supervision of Prof. Mark Last. He is also Engineer in the Industrial Engineering and Management Department of SCE - Shamoon College of Engineering. Dima was born in Vilnius, and repatriated to Israel in 1996. He received his B.A and M.A. in Economics and Computer Science from Ben-Gurion University of the Negev. His current research interests are in time series data mining, data streams segmentation and computer simulation. His recent publications include the Journal of Business Economics and Management, Applied Financial Economics, and Communications in Dependability and Quality Management.

[3]Mark Last is currently Associate Professor at the Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel and Head of the Software Engineering Program. Prior to that, he was a Visiting Research Scholar at the US National Institute for Applied Computational Intelligence, Visiting Assistant Professor at the Department of Computer Science and Engineering, University of South Florida, USA, Senior Consultant in Industrial Engineering and Computing, and Head of the Production Control Department at AVX Israel. Mark obtained his Ph.D. degree from Tel-Aviv University, Israel in 2000. He has published over 140 papers and chapters in scientific journals, books, and refereed conferences. He is a co-author of two monographs and a co-editor of seven edited volumes. His main research interests are focused on data mining, cross-lingual text mining, software testing, and security informatics.
Prof. Last is a Senior Member of the IEEE Computer Society and a Professional Member of the Association for Computing Machinery (ACM). He currently serves as Associate Editor of IEEE Transactions on Systems, Man, and Cybernetics, where he has received the Best Associate Editor Award for 2006, and Pattern Analysis and Applications (PAA). Prof. Last is very active in organizing cooperative international scientific activities. He has co-chaired four international conferences and workshops on data mining and web intelligence.

[4] **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.

[5] Corresponding author

[6] Codification of references:

[1]	Alberg, D., Shalit, H. and Yosef, R. **Estimating stock market volatility using asymmetric GARCH models**, Applied Financial Economics, vol. 18(15), 2008, pp. 1201-1208
[2]	Bollerslev, T.A. **Conditionally heteroskedastic time series model for speculative prices and rates of return**, Review of Economics and Statistics, vol. 69, 1987, pp. 542–547
[3]	Breiman, L. and Friedman, J. **Estimating Optimal Transformations for Multiple Regression and Correlation**, J. Amer. Statist. Assoc., vol. 80, 1985, p. 580
[4]	Breiman, L., Friedman, J., Olshen, R. and Stone, C. **Classification and Regression Trees**, Wadsworth Int. Group, Belmont California USA, 1984
[5]	Brockwell, P. and Davis, R. **Time Series: Theory and Methods**, 2nd ed., New York: Springer-Verlag, 1991
[6]	Ceci, M., Appice, A. and Malerba, D. **Comparing simplification methods for model trees with regression and splitting nodes**, Foundations of Intelligent Systems, 14th International Symposium, vol. 2871 of LNAI, 2003, pp. 49–56

[7]	Chatfield, C. **The Analysis of Time Series**, 5th ed., London: Chapman and Hall, 1995, pp. 431-441
[8]	Cortez, F. and Morais, R.D. **A data mining approach to predict forest fires using meteorological data**, in Neves, J., Santos, M.F. and Machado, J. (eds.) "New Trends in Artificial Intelligence", Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, 2007, pp. 512-523
[9]	Engle, R. **Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation**, Econometrica, vol. 50, 1982, pp. 987-1007
[10]	Friedman, J. **Multivariate adaptative regression splines**, Annals of Statistics, 1991, pp. 1-19
[11]	Gama, J., Rocha, R., Medas, P., Wah, B. and Wang, J. **Accurate decision trees for mining high-speed data streams**, KDD 2003, 2003 , pp. 523-528
[12]	Granger, C. and Newbold, P. **Forecasting Economic Time Series**, 2nd ed., Academic Press, New-York, 1986
[13]	Han, J., Cai, D. and Chen, Y. **Multi dimensional analysis of data streams using stream cubes in data streams models and algorithms**, ed. Aggarwal, C., 2007, pp. 103-125
[14]	Harvey, A. **Forecasting Structural Time Series Models and the Kalman Filter**, C U P, Cambridge, 1989
[15]	Hulten, G., Spencer, L. and Domingos, P. **Mining time-changing data stream**, Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001, pp. 97-106
[16]	Hyndman, R. **Highest-density forecast regions for non-linear and non-normal time series models**, Journal of Forecasting, vol. 14, 1996
[17]	Kamber, M. and Han, J. **Data Mining: Concepts And Techniques**, 2nd ed., The Morgan Kaufmann Series in Data Management Systems, 2007
[18]	Karalic, A. **Linear regression in regression tree leaves**, in "Proceedings of International School for Synthesis of Expert", vol. 10(3), 1992, pp. 151-163
[19]	Loh, W.Y. **Regression tree models for designed experiments**, 2nd Lehmann Symposium, Knowledge, 2006, pp. 210-228
[20]	Ma, J., Zeng, D. and Chen, H. **Spatial-temporal cross-correlation analysis: a new measure and a case study in infectious disease**, in Mehrotra, S. *et. al.* (eds) Informatics, ISI, LNCS 3975, 2006, pp. 542-547
[21]	Quinlan, J. **Learning with continuous classes**, in "Proceedings of the 5th Australian Joint Conference on Artificial Intelligence", Singapore World Scientific, 1992
[22]	Spranger, S. and Bry, F. **Temporal data modeling and reasoning for information statistical view of boosting**, The Annals of Statistic, vol. 38(2), 2006, pp. 337-374
[23]	* * * **UCI, Machine Learning Repository**, http://archive/ics/uci edu/ml/index html
[24]	Vens, C. and Blockeel, H. **A simple regression based heuristic for learning model trees**, Intelligent Data Analysis, vol. 10(3), 2006, pp. 215-236
[25]	Wang, Y. and Witten, I. **Inducing of model trees for predicting continuous classes**, in "Proceedings of the 9th European Conference on Machine Learning", Springer-Verlag, 1997, pp. 128-137
[26]	Wei, S., Dong, Y. and Pei, J.-B. **Time Series Analysis**, Redwood City Cal Addison-Wesley, 1990

THE IMPACT OF PACING MODE AND THE ANATOMICAL POSITION OF PACING LEAD ON THE INCIDENCE OF HEART FAILURE[1]

Oana STANCU (DINA)[2]
PhD, Associate Professor
University of Medicine and Pharmacy "Carol Davila", Bucharest, Romania

E-mail: dina.oana@yahoo.com

Iulia TEODORESCU (GABOR)[3]
Cardiologist MD, PhD, Researcher
Clinical Hospital "Sfantul Ioan", Bucharest, Department of Internal Medicine and Cardiology

E-mail: iulia.gabor@ttconsult.ro

Catalina Liliana ANDREI
PhD, Associate Professor
University of Medicine and Pharmacy"Carol Davila", Bucharest, Romania

E-mail: ccatalina97@yahoo.com

Emanuel RADU[4]
Cardiologist MD, Researcher
Clinical Hospital "Sfantul Ioan", Bucharest, Department of Angiography

E-mail: esr_manu@yahoo.com

Octavian ZARA[5]
Cardiologist MD, Researcher
Clinical Hospital "Sfantul Ioan", Bucharest, Department of Angiography

E-mail: octavzod@gmail.com

Abstract: *In Romania in the last decade, pacing is playing an increasingly important role in the management of cardiac disease. If, at first, attention of the cardiologists and researchers was focusing on the electrical rather than functional effects of pacing, the fact that pacing the RV may initially improve cardiac function but may induce heart failure over time, has led to a change in direction.*

This study evaluates comparative the clinical outcome as well the incidence and predictors of heart failure in 38 patients with VVIR pacing, VDDR and DDDR pacing implanted in "Sf. Ioan" Hospital, Bucharest, over a period of 2 years. We also intended to evaluate the long-term effects of alternative right ventricular pacing sites on LVEF.

Key words: *right ventricular pacing; pacemaker syndrome; RV pacing sites; VVIR; VDDR; DDDR pacing modes*

1. Introduction

VVI pacing alone has shown a higher risk of sudden death compared with non-paced patients with a similar degree of heart failure [1][6]. A retrospective study of long-term follow-up between VVI and DDD pacing showed that DDD enhances survival compared with VVI in patients with heart failure and AV block [2]. Nielsen et al. demonstrated that VVI pacing for sinus node disease was associated with the development of congestive heart failure over a 5 year follow-up period.

On the other site the expectation that the hemodynamic benefits of atrioventricular synchrony would lead to a reduction in cardiac mortality, a reduced risk of heart failure, and a better quality of life were not proven by all the clinical trials. The MOST study which followed for three years the cardiovascular mortality and morbidity in patients with DDDR cardio stimulation toward patients with VVIR cardio stimulation showed no statistical differences between the two groups. In exchange, concerning the heart failure episodes and the quality of life, the study proved the superiority of the DDDR stimulation.

During ventricular pacing the asynchronous ventricular activation may lead to abnormal regional myocardial blood flow and metabolic disturbancies which can reduce systolic and diastolic left ventricular (LV) function[3,5]. These functional abnormalities seam to have enhanced effects over time. Some studies have shown that long-term right ventricular apical (RVA) pacing induces abnormal histologic changes and asymmetrical LV hypertrophy and thinning [6,7,9]

The choice of pacing site in the right ventricle is an other important issue. No recommendation could be made so far concerning the location of the right ventricular pacing site.

The right ventricular apex, does not seem to lead to best haemodynamic results, although it is easy accessible and best for electrode stability. [10,11,12]. While acute haemodynamic studies find that outflow tract or dual-site pacing are best for haemodinamic reasons, most of the controlled studies with permanent pacing found no significant difference to right ventricular apical pacing

In a previous clinical study, on 547 patients we have shown that the relation between VVIR pacing and the development of the pacemaker syndrome is likely to be complex. Age, comorbidity and haemodinamic status before pacing are factors that influence the appearance of the pacemaker syndrome. The patient group over 85 years had a higher incidence of worsening heart failure than the other age groups. The patients with EF> 40% before pacing had a better outcome than those with impaired left ventricular systolic function.

The data of our previous study has shown that VVIR pacing may not induce directly heart failure but may increase the risk of developping atrial fibrillation, an important precipitant of heart failure.

In animal studies, pacing at the right ventricular outflow tract (RVOT) has been shown to decrease the asynchrony of activation so that it seems to ameliorate the reduction in LV function and prevent the wall motion abnormalities and the impair of the LV function.

2. Methods

The study, included 38 patients, men and women, which needed permanent pacing, hospitalized in the Internal Medicine And Cardiology Department of "Sfantul Ioan" Hospital, over a two and a half years period, between January 2007 and June 2009. Patients who refused to sign the written consent and those with serious (severe) coagulation disorders, chronic patients with dialysis or with cancer in terminal stages were excluded.

The patients were admitted for pacemaker implantation due to following diseases:

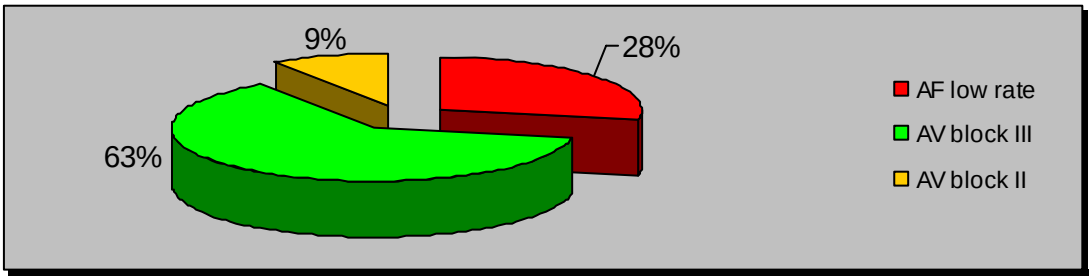


Figure 1. Patients' history

The follow up after the implant was planned at 1 month, 3 month, 12 month and 48 month.

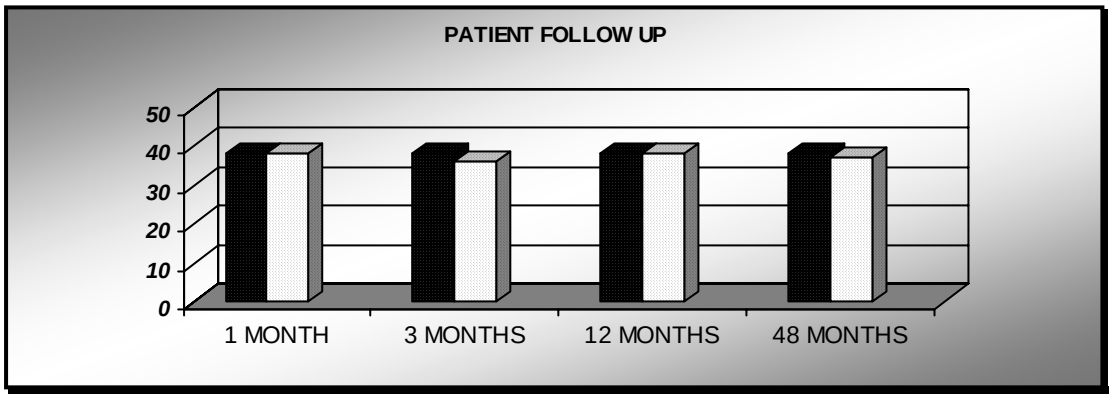


Figure 2. Patient follow up

We selected the pacing type, following a simple alghoritm that regarded the aethiology and the anatomical and functional status of the atria (see figure 3).

All patients underwent implantation of a single chamber or a dual-chamber pacemaker using one active fixation atrial lead and one active fixation bipolar ventricular lead.

The ventricular pacing lead was inserted into the RV through subclavian vein puncture. Under fluoroscopic guidance, the ventricular lead was positioned in the right ventricular apex or in the RVOT (by advancing the lead through the tricuspid valve and then withdrawing it and positioning the tip against the interventricular septum and verifying the position using multiple fluoroscopic views).

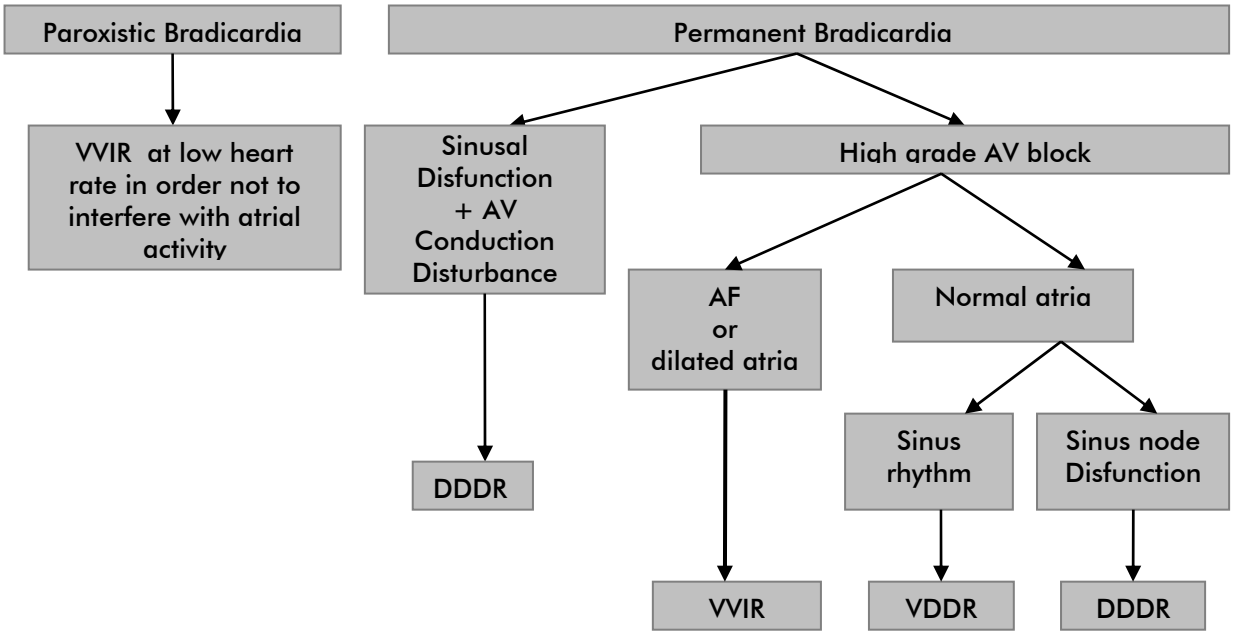


Figure 3. Pacing's type selection algorithm

The ventricular pacing leads were positioned at a stable position to obtain a satisfactory pacing threshold value (mean 1.1 ± 0.2 V) at a pulse width of 0.5 ms and R-wave sensing value (mean 12.3 ± 0.1 mV). Atrial leads were positioned to the right atrial lateral wall. After implantation, optimization of the AV delay was performed by using pulsed Doppler echocardiography of the transmitral blood flow.

Patients were evaluated before implant by a complete clinical examination. Cardiac risk factors, cardiac and associated non cardiac pathology were identified and concomitant medication was recorded.

For a proper evaluation of heart failure a special attention was given to include the patients in different NYHA classes according with their symptoms. The symptom screening, prior to the clinical examination and echocardiogram was made by the physician by asking the same questions in order to evaluate symptoms of heart failure.

The real effort capacity was estimated by standard 6 minutes walking test.

Before and after the implant , the end systolic and end diastolic volumes of the left ventricle and the ejection fraction (Simpson method in two and four chambers incidence) were measured.

Echocardiographic measurements were made in M mode and two-dimensional echocardiography (2DE). Measurements of left ventricular end-diastolic volume (LVEDV), left ventricular end-systolic volume (LVESV), and EF were obtained using the software installed on the ultrasound equipment, with LVEDV measurements at the time of mitral valve closure and LVESV measured on the image with the smallest LV cavity. The papillary muscles were excluded from the volumes. Biplane Simpson's rule volumes were obtained from the apical four- and two-chamber views.

M mode parameters were measured according to the American Society of Cardiology.

The severity of MR was appreciated from Doppler color-flow in the conventional parasternal long-axis and apical four-chamber images. Mitral regurgitation was

characterized as: *mild* (jet area/left atrial area <10%), *moderate* (jet area/left atrial area 10% to 20%), *moderately severe* (jet area/left atrial area 20% to 45%), and *severe* (jet area/left atrial area >45%).

3. Results

As the result of the pre-procedural evaluation, the patients were included into three groups with different pacing mode:

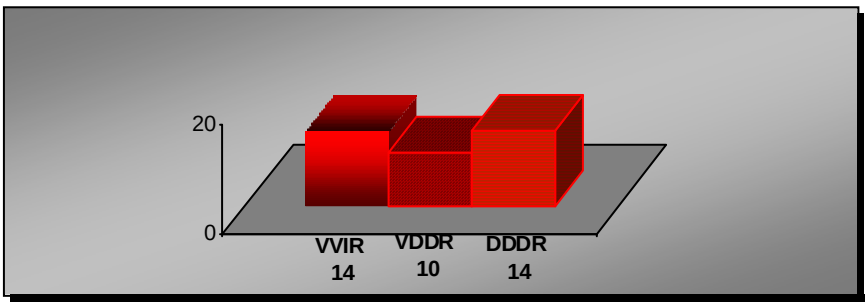


Figure 4. The distribution of the patients by the pacing mode

The incidence of heart failure at screening, at 3 months and at 12 months was the following:

Table 1. The incidence of heart failure

Pacing mode	Number of patients	Number with heart failure Screening	Number with Heart failure At disclosure	Number with heart failure 12 month
VVIR	14	2	1	5
VDDR	10	3	0	2
DDDR	14	3	0	1
Total	38	8	1	8

At pre-discharge echocardiography, there was no significant difference from baseline values in LVEF, cardiac output, as measured by Doppler echocardiography, or transmitral A- and E-wave ratio in all the three groups.

Significant changes between the three pacing modes were found at the 3 months and 12 months follow up, as shown in the table:

Table 2. Changes found at follow up checks

	VVIR	VDDR	DDDR	p value
NYHA class				
▪ Baseline	2.8 ± 0.3	2.5 ± 0.2	1.3 ± 0.2	< 0.01
▪ 12 months	3.2 ± 0.3	2.4 ± 0.3	3.1 ± 0.3	< 0.05
QRS (ms)				
▪ Baseline	136 ± 22	130 ± 26	129 ± 27	< 0.01
▪ 12 months	147 ± 26	132 ± 27	125 ± 18	< 0.01
LVEF (%)				
▪ Baseline	35 ± 6	31 ± 7	33 ± 6	< 0.01
▪ 12 months	38 ± 5	30 ± 6	30 ± 7	< 0.05
Severe MR				
▪ Baseline	3	2	3	NS
▪ 12 months	4	2	2	NS

Mitral regurgitation improved by at least one grade in 8 of 14 (57,14%) patients with severe regurgitation.

The mean QRS duration was significantly longer in VVIR paced patients than in VDDR or DDDR pacing.

We subdivided the patients in two different groups, those with right ventricular apex pacing (RVA) and those with right ventricular outflow tract (RVOT) pacing.

The differences between the 2 subgroups are seen in the following table:

Table 3. Subgroups' particularities

	Right ventricular apex pacing	RVOT/RS pacing	p Value
Baseline echocardiography			
LVEF	54 ± 11%	55 ±12%	**0.72**
E/A ratio	0.9 ± 0.3	1.1 ± 0.2	**0.30**
Optimal atrioventricular interval(ms)	140 ± 39	146 ± 36	**0.47**
Discharge echocardiography			
LVEF	55 ±2%	58±11%	**0.36**
E/A ratio	1.0 ± 0.3%	1.1 ± 0.3%	**0.25**
Optimal atrioventricular interval (ms)	141 ± 38	145 ± 38	**0.49**
6 months echocardiography			
LVEF	47 ±3%	56 ±12%	**0.38**
E/A ratio	0.9 ± 0.3%	1.0 ± 0.3%	**0.28**
Optimal atrioventricular interval (ms)	140 ± 38	146 ± 38	**0.49**

4. Discussion

Although VVIR pacing is effective in preventing symptomatic bradyarrhythmias, it has been demonstrated to be associated with a significant negative inotropic effect and with an increased rate of congestive heart failure. This pacing modus has led also to a greater increase in the grade of mitral regurgitation than the bicameral pacing modes.

RV apical pacing frequently produces an LBBB pattern with alteration in myocardium depolarization and contraction which results in a reduction of the LVEF, an prolongation of conduction intervals. Our study reveals that these changes are more important over longer periods of time than immediately after pacing.

RVOT or RS pacing can improve cardiac performance, over that obtained with RV apical pacing, despite the presence of AV synchrony. This improvement appears in every pacing mode, but is more efficient in dual chamber pacing.

5. Conclusions

Even if for over 4 decades of cardiac pacing, the right ventricular apex (RVA) has been the main site for right ventricular lead placement. during RVA pacing, larger QRS duration, impairment of LV diastolic function with significant reduction in global LV function is present.

Pacing at the RVOT is associated with more synchronous ventricular activation with a narrower QRS duration and with lower incidence of deterioration in global LV systolic and diastolic function.

Over long-term follow-up the clinical benefit seems to be greater.

RVOT pacing for routine pacemaker implantation might be the answer for preventing and treating congestive heart failure in paced patients.

Further studies may be necessary in order to compare the benefits of RVOT pacing compared with classical RVA pacing in patients with risk for heart failure.

References

1. Saxon, L.A., Stevenson, W.G., Middlekauff, H.R. and Stevenson L.W. **Increased risk of progressive haemodynamic deterioration in advanced heart failure patients requiring permanent pacemakers,** Am Heart J, 125, 1993, pp. 1306–1310
2. Alpert, M.A., Curtis, J.J., Sanfelippo, J.F. *et al*. **Comparative survival after permanent ventricular and dual chamber pacing for patients with chronic high degree atrioventricular block with and without preexisting congestive heart failure,** J Am Coll Cardiol, 7, 1986, pp. 925–932
3. Rosenqvist, M., Isaaz, K., Botvinick, E.H., *et al*. **Relative importance of activation sequence compared to atrioventricular synchrony in left ventricular function,** Am J Cardiol, 67, 1991, pp. 148–156
4. Nielsen, J.C., Bottcher, M., Nielsen, T.T., Pedersen, A.K. and Andersen, H.R. **Regional myocardial blood flow in patients with sick sinus syndrome randomized to long-term single chamber atrial or dual chamber pacing—effect of pacing mode and rate,** J Am Coll Cardiol, 35, 2000, pp. 1453–1561
5. Skalidis, E.I., Kochiadakis, G.E., Koukouraki, S.I. *et al*. **Myocardial perfusion in patients with permanent ventricular pacing and normal coronary arteries,** J Am Coll Cardiol, 37, 2001, pp. 124–129
6. Prinzen, F.W., Cheriex, E.C., Delhaas, T. *et al*. **Asymmetric thickness of the left ventricular wall resulting from asynchronous electric activation: a study in dogs with ventricular pacing and in patients with left bundle branch block,** Am Heart J, 130, 1995, pp. 1045-1053
7. Karpawich, P.P., Justice, C.D., Cavitt, D.L., Chang, C.H., Lee, M.A., Dae, M.W., Langberg, J.J. *et al*. **Effects of long-term right ventricular apical pacing on left ventricular perfusion, innervation, function and histology,** J Am Coll Cardiol, 24, 1994, pp. 225–232
8. Karpawich, P.P., Justice, C.D., Cavitt, D.L. and Chang, C.H. **Developmental sequelae of fixed-rate ventricular pacing in the immature canine heart:an electrophysiologic, hemodynamic, and histopathologic evaluation,** Am Heart J, 119, 1990, pp. 1077–1083
9. Van Oosterhout, M.F., Prinzen, F.W., Arts, T. *et al*. **Asynchronous electrical activation induces asymmetrical hypertrophy of the left ventricular wall,** Circulation, 98, 1998, pp. 588–595
10. Thambo, J.B., Bordachar, P., Garrrigue, S. *et al*. **Detrimental ventricular remodeling in patients with congenital complete heart block and chronic right ventricular apical pacing,** Circulation, 110, 2005, pp. 3766–3772
11. Adomian, G.E. and Beazell, J. **Myofibrillar disarray produced in normal hearts by chronic electrical pacing,** Am Heart J, 112, 1986, pp. 79–83
12. Nahlawi, M., Waligora, M., Spies, S.M. *et al*. **Left ventricular function during and after right ventricular pacing,** J Am Coll Cardiol, 4, 2004, pp. 1883–1888

13. Boucher, C.A., Pohost, G.M., Okada, R.D. *et al.* **Effect of ventricular pacing on left ventricular function assessed by radionuclide angiography,** Am Heart J, 106, 1983, pp. 1105–1111
14. Tse, H.-F., Yu, C., Wong, K.-K. *et al.* **Functional abnormalities in patients with permanent right ventricular pacing,** J Am Coll Cardiol, 40, 2002, pp. 1451-1458

---

[1]**Acknowledgments:**
This study was supported entirely by the Romanian Ministry of Education and Research through **T**he National University Research Council (CNCSIS) which we gratefully acknowledge.

[2] **Dr. Oana Stancu**, MD-Internal Medicine and Cardiology, PhD, Lecturer
- Graduated at University of Medicine and Pharmacy"Carol Davila", Bucharest, 1990;
- Assistant professor - Internal Medicine at University of Medicine and Pharmacy"Carol Davila"-1995
- PhD- 2002;
- Specialist - Cardiology 2004
- since 2005 - lecturer at University of Medicine and Pharmacy"Carol Davila", Bucharest;
- since 1995 - MD at the Saint John Clinical and Emergency Hospital, Department of Internal Medicine and Cardiology
Training in Pacing and Electrophysiology :

- 1996- training in cardiac pacemakers implants VVI,VDD,DDD-Military Hospital-Bucharest, in Cardiac Surgery Department;
- 1998,1999-Training in pacing and electrophysiology, in Austria, Wien, Wilheminenspital;
- 2000-training in pacing and electrophysiology in Austria, Wien, Allgemeines Krankenhaus;

2001-Training in pacing and electrophysiology, Biotronik Company, Berlin, Germany.

[3] Dr.**Teodorescu Iulia**- Cardiologist MD, PhD, researcher;
- graduated at University of Medicine and Pharmacy "Carol Davila", Bucharest, 1994;
- working by contest in Saint John Clinical and Emergency Hospital, Bucharest from 1995 in Internal Medicine Department;
- researcher from 1995;
- PhD from 2002;
- cardiac echography classes, pacing and electrophysiology preoccupations;

- 1995-training in periphery angiography - National Institute of Cardiology Bucharest -Catheters and Angiography Department
- 1996-training in cardiac pacemakers implants VVI, VDD, DDD - Military Hospital - Bucharest, in Cardiac Surgery Department;
- 1998, 1999-Traing in pacing and electrophysiology offered by BIOTRONIK, in Austria, Wien, Wilheminenspital; pacing and electrophysiology preoccupations;
- 2000-Training in pacing and electrophysiology offered by BIOTRONIK, in Austria, Wien, Allegemeines Krankenhaus;
- 2001-training in pacing and electrophysiology offered by BIOTRONIK, Germany, Berlin

[4] **Emanuel Radu**, MD, Researcher, Saint John Clinical and Emergency Hospital, Bucharest
- Graduated at University of Medicine and Pharmacy, Bucharest "Carol Davila";
- Working by contest in Saint John Clinical and Emergency Hospital, Bucharest, from 2000, Angiography Deparment;
- Specialist in Cardiology since 2008

[5] **Zara Octavian Dumitru**, MD, Researcher,resident in training in Interventional Cardiology,
- Graduated at „Vest University"-Vasile Goldis-1998;
- Working by contest in Saint John Clinical and Emergency Hospital, Bucharest, from 2000, Angiography Deparment;
- resident in training in Cardiology-2002

[6] Codification of references:

[1]	Saxon, L.A., Stevenson, W.G., Middlekauff, H.R. and Stevenson L.W. **Increased risk of progressive haemodynamic deterioration in advanced heart failure patients requiring permanent pacemakers,** Am Heart J, 125, 1993, pp. 1306–1310
[2]	Alpert, M.A., Curtis, J.J., Sanfelippo, J.F. *et al.* **Comparative survival after permanent ventricular and dual chamber pacing for patients with chronic high degree atrioventricular block with and without preexisting congestive heart failure,** J Am Coll Cardiol, 7, 1986, pp. 925–932

[3]	Rosenqvist, M., Isaaz, K., Botvinick, E.H., *et al.* **Relative importance of activation sequence compared to atrioventricular synchrony in left ventricular function,** Am J Cardiol, 67, 1991, pp. 148–156
[4]	Nielsen, J.C., Bottcher, M., Nielsen, T.T., Pedersen, A.K. and Andersen, H.R. **Regional myocardial blood flow in patients with sick sinus syndrome randomized to long-term single chamber atrial or dual chamber pacing—effect of pacing mode and rate,** J Am Coll Cardiol, 35, 2000, pp. 1453–1561
[5]	Skalidis, E.I., Kochiadakis, G.E., Koukouraki, S.I. *et al.* **Myocardial perfusion in patients with permanent ventricular pacing and normal coronary arteries,** J Am Coll Cardiol, 37, 2001, pp. 124–129
[6]	Prinzen, F.W., Cheriex, E.C., Delhaas, T. *et al.* **Asymmetric thickness of the left ventricular wall resulting from asynchronous electric activation: a study in dogs with ventricular pacing and in patients with left bundle branch block,** Am Heart J, 130, 1995, pp. 1045-1053
[7]	Karpawich, P.P., Justice, C.D., Cavitt, D.L., Chang, C.H., Lee, M.A., Dae, M.W., Langberg, J.J. *et al.* **Effects of long-term right ventricular apical pacing on left ventricular perfusion, innervation, function and histology,** J Am Coll Cardiol, 24, 1994, pp. 225–232
[8]	Karpawich, P.P., Justice, C.D., Cavitt, D.L. and Chang, C.H. **Developmental sequelae of fixed-rate ventricular pacing in the immature canine heart:an electrophysiologic, hemodynamic, and histopathologic evaluation,** Am Heart J, 119, 1990, pp. 1077–1083
[9]	Van Oosterhout, M.F., Prinzen, F.W., Arts, T. *et al.* **Asynchronous electrical activation induces asymmetrical hypertrophy of the left ventricular wall,** Circulation, 98, 1998, pp. 588–595
[10]	Thambo, J.B., Bordachar, P., Garrrigue, S. *et al.* **Detrimental ventricular remodeling in patients with congenital complete heart block and chronic right ventricular apical pacing,** Circulation, 110, 2005, pp. 3766–3772
[11]	Adomian, G.E. and Beazell, J. **Myofibrillar disarray produced in normal hearts by chronic electrical pacing,** Am Heart J, 112, 1986, pp. 79–83
[12]	Nahlawi, M., Waligora, M., Spies, S.M. *et al.* **Left ventricular function during and after right ventricular pacing,** J Am Coll Cardiol, 4, 2004, pp. 1883–1888
[13]	Boucher, C.A., Pohost, G.M., Okada, R.D. *et al.* **Effect of ventricular pacing on left ventricular function assessed by radionuclide angiography,** Am Heart J, 106, 1983, pp. 1105–1111
[14]	Tse, H.-F., Yu, C., Wong, K.-K. *et al.* **Functional abnormalities in patients with permanent right ventricular pacing,** J Am Coll Cardiol, 40, 2002, pp. 1451-1458

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

DIFFERENT APPROACHES USING THE NORMAL AND THE EXPONENTIAL DISTRIBUTION IN THE EVALUATION OF THE CUSTOMER SATISFACTION

Antonio LUCADAMO[1]
PhD, Assistant Researcher, Tedass,
University of Sannio, Benevento, Italy

E-mail: alucadam@unina.it

Giovanni PORTOSO[2]
PhD, Associated Professor, SEMEQ Department, Faculty of Economics,
University of Eastern Piedmont "A. Avogadro", Novara, Italy

E-mail: portoso@eco.unipmn.it

Abstract: *The Customer Satisfaction is generally evaluated using the data collected with questionnaires. The data are organized on an ordinal scale and, for this reason, it's convenient to transform them in pseudo-interval support. The psychometric methods used for this transformation generally hypothesize that the latent variable has a normal distribution. Sometimes, particularly when the frequencies are concentrated on the left extreme or on the right extreme of the distribution, this assumption brings to preposterous results. In these cases the use of other types of distribution, as, for example the exponential distribution, is preferable. In this paper we show how the results of a survey can change using the normal distribution, the exponential distribution or the two distributions alternatively. We use, in fact, the results coming from the different transformations, to apply a multilevel model.*

Key words: *customer satisfaction; normal distribution; exponential distribution; multilevel models*

1. Introduction

One of the problem of the Customer Satisfaction is the quantification that converts on a metric scale the judgements about services or products. A simple technique is the so-called "direct quantification": this technique hypothesizes that the modalities of a qualitative character are at the same distance, but this hypothesis is not respected in many situations (Marbach, 1974). For this reason it is preferable to use an alternative technique, the "indirect quantification", that consists in assigning real numbers to the categories of the qualitative variable. In this type of quantification the numbers are not equidistant but they depend on a latent variable. Different measurement techniques have been developed during the years (Thurstone, 1925, Guilford, 1936, Torgenson, 1958) based on the hypothesis that the model

is normally distributed. This assumption can be realistic in a psychometric field, but it is not always valid in the Customer Satisfaction, especially if the judgements are all extremely positive or extremely negative. More recent techniques have been proposed, based for example on the use of logit and probit models, on structural equation models and so on. In next section we introduce the psychometric quantification, underlining the problems that can arise in some situations; then we show how the use of another kind of quantification can solve these pitfalls and, in the following paragraphs we propose the use of a combined technique, showing the results that we obtain on real data.

2. The psychometric quantification

In the psychometric quantification, the modalities x_i $(i = 1,2,\ldots,r)$ of a qualitative variable X, are associated to the values of a quantitative latent variable Z, normally distributed. Let $F(i)$ be the cumulative relative frequency, corresponding to x_i and let $\Phi^{-1}[F(i)]$ the inverse of the cumulative distribution function, the quantile z_i associated to x_i can be expressed as $z_i = \Phi^{-1}[F(i)]$. To obtain the new scores, we simply calculate the expected values $E(Z_i)$ over all the X variables in the data-set. The assumption of the normal distribution, when the frequencies are prevalently on the left extreme or on the right extreme of the distribution, leads to strange results. In fact the scores will be negative if the modalities are almost on positive side and vice-versa (the results in Table 1 can help to understand the situation) (Portoso, 2003a).

Table 1. Quantification with the normal distribution of the judgements given on two different services

Judgements	Frequencies of the first service	Frequencies of the second service
Very negative	350	10
Negative	80	20
Indifferent	40	40
Positive	20	80
Very positive	10	350
Totals	500	500
Expected quantile	0.0729	-0.0729

It is easy to see that the first service has many negative judgements, so the frequencies are prevalently on the left side of the distribution, but the expected quantile has a positive value. For the second service there is instead the inverse situation, in fact the frequencies are on the right side, but expected value of the quantile is negative.

This incongruity leads to use a distribution that could better express, in a numerical way, the categorical variables characterized by this particular structure. The exponential distribution seems to be the right solution.

3. The exponential quantification

In this section we show how to determine a quantification based on the negative and on the positive exponential distribution. Before introducing the new procedure, it's necessary to describe briefly the two cited distributions.

3.1. The negative exponential distribution

Let consider

$$\begin{cases} \psi(z) = \exp(-z) \quad if \quad (0 \le z \le \infty) \\ \psi(z) = 0 \qquad otherwise \end{cases} \tag{1}$$

where Z is a quantitative variables.

It can be assumed as the relative density function, in fact:

$$\int_0^{\infty} \psi(z)dz = \int_0^{\infty} \exp(-z)dz = 1 \tag{2}$$

The mean and the variance are defined as follow:

$$E(Z) = \int_0^{\infty} z\psi(z)dz = \int_0^{\infty} z\exp(-z)dz = [-z\exp(-z)]_0^{\infty} - \int_0^{\infty} -\exp(-z)dz = 1 \tag{3}$$

$$Var(Z) = \int_0^{\infty} z^2 \exp(-z)dz - [E(Z)]^2 = 2 - 1^2 = 1 \tag{4}$$

The variable can be then standardized in the following way:

$$S = (Z - 1) = Z - 1 \tag{5}$$

with

$$\begin{cases} f(s) = \exp(-s-1) \quad if\,(-1 \le s \le \infty) \\ f(s) = 0 \qquad\qquad otherwise \end{cases} \tag{6}$$

This is a relative frequency density function, in fact:

$$\int_{-1}^{+\infty} \exp(-s-1) = 1 \tag{7}$$

The cumulative distribution function is:

$$\begin{cases} \Psi(s) = \int_{-1}^{s} \exp(-t-1)dt = 1 - \exp(-s-1) \quad if \quad (-1 \le s \le \infty) \\ \Psi(s) = 0 \quad otherwise \end{cases} \tag{8}$$

3.2. The positive exponential distribution

Let consider

$$\begin{cases} \psi(y) = \exp(y) \quad if \quad (-\infty \le y \le 0) \\ \psi(y) = 0 \qquad otherwise \end{cases} \tag{9}$$

that can be assumed as the relative density function, in fact:

$$\int_{-\infty}^{0} \psi(y)dy = \int_{-\infty}^{0} \exp(y)dy = 1 \tag{10}$$

The mean and the variance are defined as follow:

$$E(Y) = \int_{-\infty}^{0} y\psi(y)dy = \int_{-\infty}^{0} y\exp(y)dy = [y\exp(y)]_{-\infty}^{0} - \int_{-\infty}^{0} \exp(y)dy = -1 \tag{11}$$

$$Var(Y) = \int_{-\infty}^{0} y^2 \exp(y)dy - [E(Y)]^2 = 2 - (-1)^2 = 1 \tag{12}$$

The variable can be then standardized in the following way:

$$P = (Y - (-1)) = Y + 1 \tag{13}$$

The cumulative distribution function of P is :

$$\begin{cases} \Psi(p) = \int_{-\infty}^{p} \exp(t-1)dt = \exp(p-1) \quad if \quad (-\infty \le p \le 1) \\ \Psi(p) = 1 \quad otherwise \end{cases} \tag{14}$$

3.3. The quantification

To build the scores, both for the negative exponential distribution and for the positive one, it's necessary to consider the relative frequencies $f(i)$ and the cumulative relative ones $F(i)$.

In this way we can define the following quantity (empirical distribution of cumulative frequencies):

$$G(i) = F(i-1) + f(i)/2 \quad i = 1, 2, \ldots, r \tag{15}$$

If we consider the negative exponential distribution, we can compare formula (6) and (15) and we obtain the standardized quantile:

$$s_i = -1 - \ln[1 - G(i)] \tag{16}$$

The same procedure can be applied for the positive distribution, using formula (13) and (15); in this case we will obtain the standardized quantile in the following way:

$$p_i = 1 + \ln[G(i)] \tag{17}$$

To verify the importance, in particular situations, of using the exponential distribution, instead than the normal distribution, we can consider the value in table 2 in which there are the absolute frequencies about the judgments given to 8 different services by 500 judges.

Table 2. Absolute frequencies of the judgements given to 8 different services

	Services							
Judgments	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Very negative	496	470	20	180	20	10	2	1
Negative	1	16	50	50	120	20	4	0
Indifferent	1	8	360	40	220	40	6	0
Positive	1	4	60	50	120	350	96	0
Very positive	1	2	10	180	20	80	392	499
Total	500	500	500	500	500	500	500	500

In this table we can note that the services A and B received many negative judgments, the services F, G and H many positive judgments and the services C, D and E had a quasi-symmetric distribution.

This table is important to understand what happens when we apply the different kinds of quantification.

The results are shown in table 3.

Table 3. Quantiles associated to the relative cumulate frequencies centred on every judgment category in the hypothesis of exponential and normal distribution

Services		JUDGMENTS					Expected quantile
		Very negative	Negative	Indifferent	Positive	Very positive	
A	Exp neg.	-0.315	3.962	4.298	4.809	5.908	-0.274
	Norm	-0.010	2.457	2.576	2.748	3.090	0.012
B	Exp neg.	-0.385	2.124	2.912	3.828	5.215	-0.177
	Norm	-0.075	1.705	2.054	2.409	2.878	0.047
C	Exp neg.	-2.912	-1.408	0.307	0.917	0.990	0.093
	Norm	-2.054	-1.341	0	1.405	2.326	-0.001
D	Exp neg.	-0.802	-0.472	-0.307	-0.108	0.714	-0.114
	Norm	-0.915	-0.228	0	0.228	0.915	0
E	Exp neg.	-0.980	-0.826	-0.307	0.833	2.912	-0.056
	Norm	-2.054	-0.994	0	0.994	2.054	0
F	Exp pos.	-3.605	-2.219	-1.303	0.287	0.917	0.082
	Norm	-2.326	-1.751	-1.282	-0.025	1.405	-0.012
G	Exp pos.	-5.215	-3.828	-3.017	-1.120	0.502	0.091
	Norm	-2.878	-2.409	-2.097	-1.175	0.274	-0.067
H	Exp pos.	-5.908	-5.215	-5.215	-5.215	0.309	0.296
	Norm	-3.090	-2.878	-2.878	-2.878	0.003	-0.004
						General mean	0.002

Here we can see that, in a Customer Satisfaction analysis, when the frequencies are very high for judgments extremely positive or extremely negative, the use of the normal distribution is not an appropriate way to effectuate the quantification. Using the exponential distribution leads to better results, in fact we can see that for the services A and B that presented value extremely negative, we have that the expected value of the quantile is

negative if we use the negative exponential distribution, while using the normal quantification it will be positive. For the services C, D and E there are no substantial differences between the use of the normal or of the exponential distribution, but the first one seems to be preferable; for this services in fact we had a symmetric distribution. For services G and H instead, the calculation of the expected quantile shows that the use of positive exponential distribution leads to positive values, while using the normal distribution we will have negative values.

Of course the observation of the expected quantile can not be the only instrument to decide if considering the normal distribution or the exponential distribution as latent variable, but we need an indicator that could help in the choice. A possible solution is given in Portoso (2003b) that introduces an useful index to decide which kind of distribution is better to apply in the different situations.

3.4. The EN index

The EN index is an indicator that assumes values between -1 and +1. The value -1 is assumed when all the frequencies are associated to the first modality (in this case we have maximum negative concentration), while when there is maximum positive concentration the value assumed by the index will be +1. If the frequencies are balanced in a symmetric way then the EN index will be equal to 0. The index has the following formulation:

$$EN = \sum_{i=1}^{r/2} (f_{r-i+1} - f_i)(r - 2i + 1)/(r - 1) \qquad (18)$$

where *r* is the number of modalities and if they are odd the value r/2 is round off to the smaller integer while f_i are, as already stated, the relative frequencies associated to the modality *i*, f_{r-i+1} are the frequencies associated to the opposite modality and $r - 2i + 1$ is the difference between the position of the two opposite modalities.

An alternative formulation of the index can be the following:

$$EN = 1 - 2\sum_{i=1}^{r-1} F(i)/(r - 1) \qquad (19)$$

that presents some similarities with the Gini index and where $F(i)$ have already been defined as cumulative relative frequency and r is the number of modalities of the qualitative variables. If the value of the index *EN* is close to 0, the use of normal distribution doesn't generate any problems, but if the absolute value of this index grows then the use of exponential distribution can lead to better results. The problem is to define a threshold to decide which distribution is better to apply. Portoso, with empirical attempts, showed that a value of the EN bigger than 0.2 in absolute value, indicates that the use of the exponential distribution is preferable to the normal one. In the following sections we first introduce briefly the multilevel models and then we verify what happens to the results of an analysis using the different kinds of quantification.

4. The multilevel models

Multilevel models suppose that in a hierarchical structure, the upper levels can influence the lower ones (Snijders, Bosker 1999). The basic model is the so called empty model defined as follows:

$$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij} \tag{20}$$

In this formula there is a dependent variable Y_{ij} given by the sum of a general mean (γ_{00}), a random group effect (U_{0j}) and a random individual effect (R_{ij}). In this way the variability is divided in two parts: in fact, in this model it's assumed that the random variables U_{0j} and R_{ij} are mutually independent, normally distributed with zero mean and variances equal to τ^2 and σ^2. The total variance is then the sum of the two variances and we can compute the intra-class correlation coefficient:

$$\rho = \tau^2 / (\tau^2 + \sigma^2) \tag{21}$$

If this coefficient is significant, it is possible to effectuate a Multilevel Analysis (Hox, 2002). A first model is the Random Intercept Model that can be defined as follows:

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + R_{ij} \quad \text{with} \quad \beta_{0j} = \gamma_{00} + U_{0j} \tag{22}$$

In the equation (20) if we consider the j subscript for the coefficient β_1 we will have the Random Slopes Model. In this case too we can see that there is a fix effect (γ_{10}) and a random ones (U_{1j}).

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + R_{ij} \text{ with } \beta_{0j} = \gamma_{00} + U_{0j} \text{ and } \beta_{1j} = \gamma_{10} + U_{1j} \tag{23}$$

5. A case study

The application concerns a survey about Patient Satisfaction. The patients answered to 30 items relative to the services received during the staying in the hospital. They gave a score between 1 and 7 and furthermore they furnished information about the gender, the age and the education. To apply a multilevel model we need a variable relative to the second level and this is the experience of the head physician of the different wards. The aim is to verify if the Customer Satisfaction (CS) depends on the different variables and, above all, if the different quantifications leads to dissimilar results. For this reason we adopt the Normal Quantification, the Exponential Quantification and a Mixed Quantification (Normal or Exponential). The first two quantifications have already been illustrated, instead the third one is based on the use of the *EN* index. We use in fact the normal quantification for the items that have the *EN* index lower than a fixed threshold and the exponential distribution for the items with *EN* larger than the threshold. Furthermore, to compute the new scores, we used a geometric mean for the exponential quantifications because of lower sensitivity to

extreme values and an arithmetic mean for the normal distribution. In Table 4 we show the number of the items transformed using the two distributions, according to the different thresholds, arbitrarily assumed and considering that it is possible a larger series of values.

Table 4. Number of the two distributions used according to the different thresholds

Threshold	0	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1
N° of exponential	30	29	26	25	22	17	9	1	0
N° of normal	0	1	4	5	8	13	21	29	30

For all the criterions we then compute the overall CS as the sum of the new value that every individual has for the 30 items. In the building of the model, the only significant variable for the individual level is the age and the model that we adopt is a Random Intercept Model, so we can write:

$$CS_{ij} = \beta_{0j} + \beta_1 Age_{ij} + R_{ij} \quad \text{with} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}(Exp_j) + U_{0j} \qquad (24)$$

The results that we obtain are reported in Figure 1.

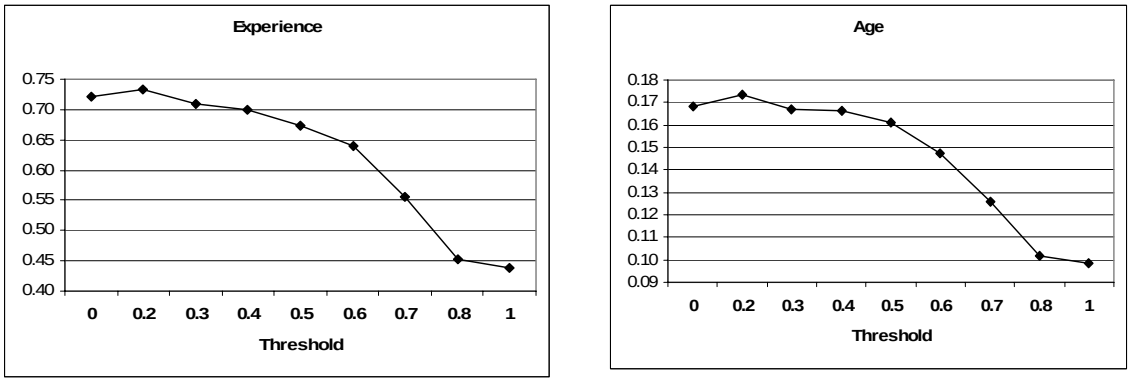


Figure 1. Coefficients of the two explicative variables with the different thresholds

We can note that the coefficients relative to the experience of the head physician (γ_{01}) and to the age of the patients (β_1) are both positive, so the CS is higher for older patients and for people that were nursed in department with expert doctors. Furthermore they increase if we consider an EN threshold that goes from 0 to 0.2 and then they both decrease considerably when the threshold is higher than 0.2, reaching a minimum by using only the normal as latent variable. Moreover they are all statistically significant and there are no substantial differences in the values of the t-ratio. The value of the *EN* = 0.2 is the value that Portoso (2003b) indicated as critical for the choice between exponential and normal distribution.

6. Considerations and perspectives

In the Customer Satisfaction or in the evaluation of other services there is a quantification problem that can not be solved using the direct quantification, because it doesn't answer to the reality. The use of the indirect quantification, with the assumption of a continuous latent distribution, is in this case preferable, but the choice can not be always the normal distribution. Using its standardization, the exponential distribution, negative or

positive, has been assumed as an alternative to the normal when this one is not appropriate. The exponential distribution assures results that are more consistent with the shape of the empirical distribution and, furthermore, it guarantees distances between the modalities more adherent to the psychological continuum with which the judgments are expressed. The problem about the choice of the right distribution was discussed in an empirical way in a previous work; in this paper, the results that we obtain, introducing also a second step of the analysis, confirm the idea of using the exponential distribution instead than the normal one when the EN index is higher than 0.2 or smaller than -0.2. Obviously these are only results that comes from a restricted number of analyses and the definition of the threshold for the choice between normal and exponential distribution must be studied deeply. Furthermore some other indexes could be proposed and not only the use of normal and exponential distribution must be taken into account; our proposal is in fact to consider, in next works, the use of other distributions too.

References

1. Guilford, J.P. **Psychometrics Methods,** McGraw-Hill, New York, 1936, pp. 255-258
2. Hox, J.J. **Multilevel Analysis. Techniques and Applications,** Lawrence Erlbaum Associates, 2002
3. Marbach, G. **Sulla presunta equidistanza degli intervalli nelle scale di valutazione,** Metron, Vol. XXXII: 1-4, 1974
4. Portoso, G. **La quantificazione determinata indiretta nella customer satisfaction: un approccio basato sull'uso alternativo della normale e dell'esponenziale,** Quaderni di dipartimento SEMeQ, 53, 2003a
5. Portoso, G. **Un indicatore di addensamento codale di frequenze per variabili categoriche ordinali basate su giudizi,** Quaderni di dipartimento SEMeQ, 66, 2003b
6. Snijders, T. A. B. and Bosker R. J. **Multilevel Analysis. An introduction to basic and advanced multilevel modelling,** SAGE Publications, 1999
7. Thurstone, L. L. **A method of scaling psycological and educational tests,** J. Educ. Psychol., 16, 1925, pp. 443-451
8. Torgenson, W.S. **Theory and Methods of Scaling,** Wiley, New York, 1958

[1] **Antonio Lucadamo**, Philsophy Doctor in Statistics with the thesis "Multidimensional Analysis for the definition of the choice set in Discrete Choice Models"- department of Mathematics and Statistics, University of Neaples "Federico II". From 2007-2009 he had a post-doc research fellowship in Statistics about the Customer Satisfaction – department SEMEQ, University of Eastern Piedmont, Novara. Currently he has a post-doc research fellowship about the use of multivariate statistical methods in the spectrometry –Tedass, University of Sannio, Benevento. His research field are the Discrete Choice Models (Logit and Probit), the Multidimensional Analysis of Data (Principal Component Analysis, Cluster Analysis), the Multivariate Analysis for the definition of the sample size in clinical trials, the Rasch Analysis, the Quantification Methods and the Multilevel Regression for the evaluation of Customer Satisfaction.

[2] Graduated in Economics from the University of Bari in 1965 by a vote of 110 cum laude. Assistant Professor since 1971. In 1982 he became Associate Professor at the Faculty of Economics of Bari. He then taught at the Faculty of Economics in Turin. In the nineties, with 4 other colleagues, he formed the Faculty of Economics of Novara, where he currently teaches. Declared fit to full professor, he is not called by the Board of Faculty of Economics of Novara for lack of quorum. He has lectured in Anthropometry, Statistics, Statistical Methodology, Statistical Quality Control, Multivariate Statistical Analysis, Statistical Sampling, Theory of Probability and Mathematical Statistics, Market Research. His publications cover the methodological aspects and applications developed in the field of demographics and the financial sector, anthropometry, statistical analysis of the results of gambling, use of latent variables in the customer satisfaction, normalization of indices of association.

THE IMPACT OF FINANCIAL CRISIS ON THE QUALITY OF LIFE

Lindita ROVA

MSc, Lecturer, Economic Department,
"Eqrem Cabej" University, Gjirokastra, Albania

E-mail: lindita_rova@yahoo.com

Romeo MANO

PhD Candidate, Department of Mathematics, Physics and Informatics,
"Eqrem Cabej" University, Gjirokastra, Albania

E-mail: manoromeo2002@yahoo.com

Abstract: *The quality of life is a relatively new concept, which is continually changing and for which there is not yet a wholly satisfactory definition. The quality of life involves human, social-economic and health characteristics. The manifold nature of the quality of life led to the development of various patterns for measuring it. The quality of life is determined by subjective and objective indices and this allows us to have a clearer overall picture of it.*

The aim of this research paper is to measure the standard of living based on six main levels determined by the European Foundation for the Improvement of Living and Working Conditions-Eurofound: 1. Employment; 2. Total income; 3. Family & Home; 4. Social life & Community involvement; 5. Medical care & health security insurance and 6. Knowledge & education

A very interesting part of the paper is the research into the impact of economic crisis on the quality of life. The study is based on a questionnaire which is filled by 200 persons living in Gjirokastra and Saranda. The analysis of the results from the questionnaire is carried out employing the logistic regress method.

Key words: *quality of life; economic crisis; welfare; objective indices; subjective indices; logistic regression method; Albania*

1. What is the concept of quality of life; a historical overview

> *"People are the real wealth of a nation. Thus, the main real aim of development is to provide the basis for an environment where people could be able to lead a long, healthy and creative live. Human development brings about an increase in the number of possible alternatives offered to the individual, of which the most important are to live longer and healthier and to get used to enjoying a respectably high standard of living. Other alternatives include political freedom, human rights and self-respect."*
>
> UNDP 1990

The quality of life is a relatively modern concept, which has undergone significant changes, and for which, there does not exist a universally excepted definition. The quality of

live involves human, social, economic and health characteristics (Lamau, 1992; Treasury Board of Canada Secretariat, 2000).

Historically speaking the quality of life as a concept appeared first in the 50s referring to a high living standard in the new consumer society just created. The way in which it was measured was the accumulation of wealth (money, new cars, houses, furniture, modern electrical appliances etc). Later, in the 60s, the concept extended to include elements such as education, health care, economic and technological development as well as the protection of the Free World (Fallowfield, 1990). During the 70s, times in which, the development schemes failed to improve the living standards of the poor strata of the society, there was a heated long-running debate over the real goals of the economic development as well as the political goals supposed to be achieved by the developed countries. Thus, the concept of basic human needs as the fundamental estimated parameter of development schemes was brought to light (Nagamine, 1981:1).

This perception, highlighting the human basic needs and reduction of poverty did not except economic growth as the only goal of economic development. On the contrary a special interest was shown in the quality of life of the poor strata of the society and the distribution of income. Meeting the basic needs has often been identified with those needs providing the bases for a living standard in accordance with human dignity. Put otherwise, this means that different generally accepted patterns should be studied considering food, clothing, housing, health, education, social safety, working conditions and human freedom (Moon, 1991:5).

In the 60s and 70s the concept of wellbeing (as a subjective perception) was extended to include happiness or life enjoyment as a new dimension in asserting the quality of life. For many researchers it was considered to be the most important (Argyle, 1996) and later, fronted by Amartia Sen, several new elements were added to it, to determine human development. In Sen's research (1993) individual ability was introduced as a concept making for the first time the connection between the quality of live and people's ability to be involved in important activity mutually beneficial of both individual and society. In the 90s, by putting these theories together one could talk about human development, which in fact is a multi dimensional concept.

The fundamental idea of human development is that wellbeing is a crucial parameter of development and that the individual is the bases of every level of this development.

The quality of life as a concept involves several subjective and objective elements and every study on the quality of life should consider all these elements. The subjective space refers to the wellbeing and pleasures one gets from the environment in which one lives, whereas the objective space refers to the pleasures of the individual related to social political requirements; material wellbeing, social status, good medical conditions, etc (Noll, 1998).

Studies on the quality of life based only on objective indices cannot be complete. For this reason, a serious study should analyze objective and subjective indices both from the qualitative and quantitative point of view. According to European Foundation for the Improvement of Living and Working Conditions (Eurofound) the quality of life is a clear indication related to economic wellbeing in social context and has 6 main areas of study:

1. Employment
2. Total income
3. Family & Home
4. Social life & Community involvement
5. Medical care & health security insurance
6. Knowledge & education

2. Objective indices of the quality of life assessment.

This study emphasizes some of the objective parameters measuring wellbeing as a component of the quality of life and will stop to analyze mainly its subjective elements. Among the objective elements we should highlight the parameters universally accepted as important in measuring economic growth for example:

1. Gross Domestic Product
2. Gross Domestic Product per capita
3. Investments as percentage of Gross Domestic Product
4. Average salary
5. Average income per capita after tax
6. Minimal pay (salary)

Employment & unemployment indicators:

1. Population (Active – Non active) according to group age & Gender
2. Employment
3. Unemployment
4. Full-time, Part- time employment
5. Employment according to economic fields or occupation etc.

Parameters for measuring material goods & services:

1. Number of cars
2. Number of schools- primary & secondary schools
3. Number of hospitals, clinics etc.
4. Illiteracy rate
5. Number of pupils & students
6. Number of doctors
7. Number of computers etc.

Indicators for public expenses:

1. Public expenses as percentage of GDP
2. Public expenses for health care
3. Private expenses for health care
4. Expenses for old age pensions as percentage of GDP
5. Expenses for employment programs as percentage of GDP
6. Expenses for social welfare as percentage of GDP etc.

The combination of all these indicators can provide a clear picture over the level of wellbeing and development of a country; it can help even to compare the countries among themselves. But the quality of life comprises elements other than the above mentioned ones. The most delicate and, therefore, the most important factor in the process of its evaluation is undoubtedly the implication of subjective elements in it.

3. The subjective indices of quality of life assessment

The main goal of this study is precisely the assessment of the subjective elements and their impact on determining the quality of life. A questionnaire was compiled for this reason comprising several categories which takes place in this determination. Specifically in the questionnaire is asked the evaluation of the interviewee for:

- Material wellbeing (income, purchasing power, housing)
- Wealth

- Political stabilization and security
- Family life
- Employment and work safety
- The connection between employment and personal life
- Culture and sports
- Social quality evaluation

The survey is conducted in the Southern Region of Albania including three districts: Saranda, Gjirokastra and Tepelene. Initially we thought to carry out the survey based on 250 questionnaires. The questionnaire contains 38 questions and 221 persons were questioned out of 250. The survey area was divided in several groups in order to have a complete cover of the area according to the formula: 30% in rural areas, 30% at random choice (random sample), 10% in private business, 10% in family environments, 10% employed in the private sector, 10% employed in the public sector. The questionnaire was distributed in the whole region according in the following way: 60% in Gjirokastra, 25% in Saranda and 15% in Tepelena according to the number of inhabitants in the three districts respectively. The questionnaire contains six questions related to the resent economic and financial crisis, in order to study its link with the quality of life. Data processing was carried out based on logistic regress method and here are some of the results.

a) Assessment of quality of life elements

The first part of the survey deals with material welfare as one the main elements of determining the quality of life. Material welfare can be determined in two ways or according to two principal view points, from the objective one (income level and their growth rate, individual purchasing power etc) as well as the subjective one (the satisfaction he gains from possessing material goods). Put otherwise, two individuals with the same purchasing power ability judge their situation differently according to their own requirements. The outcome of the survey is interesting: 47% of the interviewees feel relatively rich, 14% feel rich, 28% relatively poor, 5% very rich and 6% very poor (figure 1).

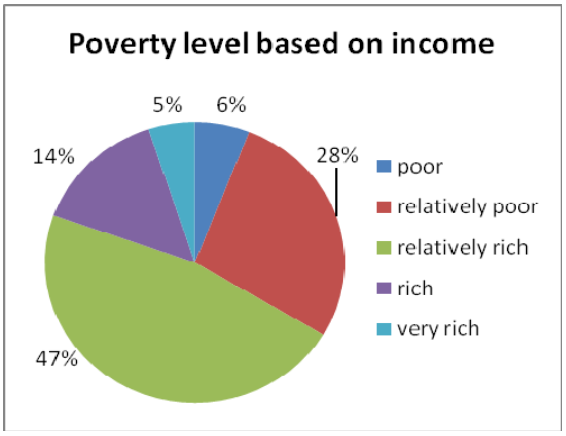


Figure 1. The poverty level based on incomes

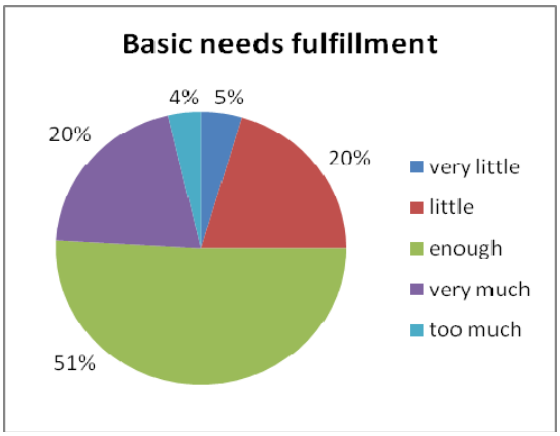


Figure 2. The basic needs fulfillment

The rate of meeting their basic needs divides the survey area into two parts: 51% of them meet their basic needs adequately, 20% little and the same percentage much, 4% very little and 5% meet their basic need too much. We think that the answers to the above mentioned questions have a considerable margin of insincerity. That's because the Albanians are too proud to accept their level of poverty (graph nr 2).

Concerning employment the survey reveals that 60% of the interviewees were employed, 7% unemployed, 5% retired (graph nr 3)

Concerning medical care in the region 30% of the interviewees asses it as adequate, 18% as good, 17% as not good at all, 7% as very good (graph nr 4). Social stability and safety were assessed to be as follows: 35% of relatively high level, 33% relatively low level, 14% very low and only 3% judged it to be very high (graph nr 5)

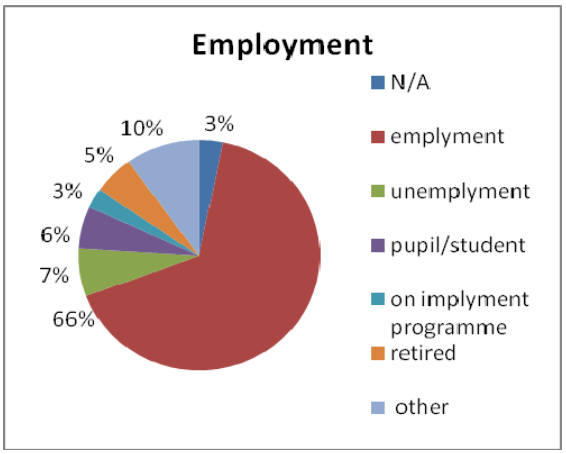


Figure 3. Employment

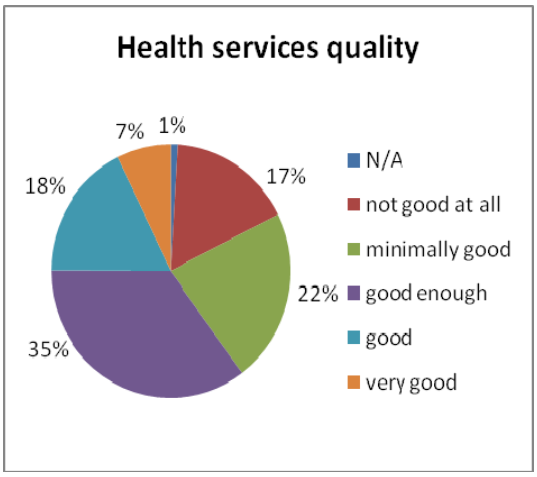


Figure 4. Health services quality

In graph number 6 represents the results of five questions concerning cultural life and sports (question nr 13), the quality of social services such as education etc offered in the region (question number 14), assessment on the environment related to greenery and pollution (question 15), every day facilities such as roads, traffic, public transport system (question 16) and as a syntheses of all the above mentioned questioned there is question number 17: how would you assess the quality of your life.

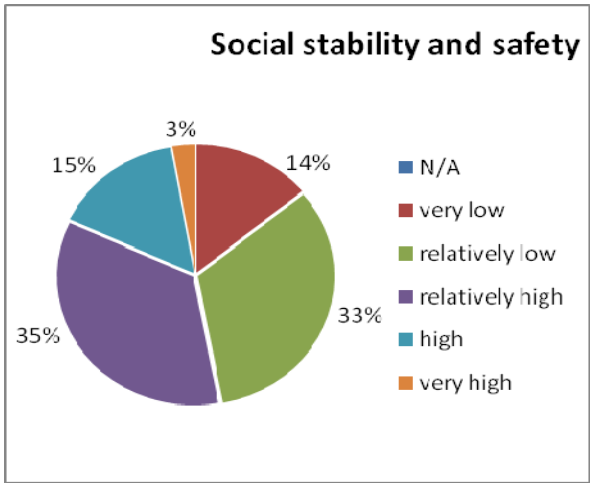


Figure 5. Social stability and safety

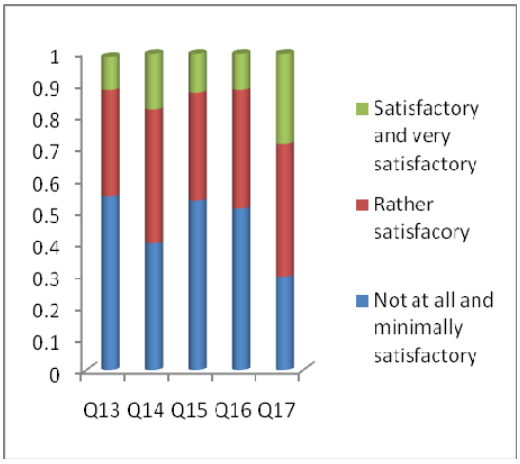


Figure 6. The results for questions 13, 14, 15, 16 and 17

Again the outcome of the survey is interesting: more than 50% think that the cultural and sport life (Q13), the environment their live (Q 15) and the everyday facilities (Q 16) offer them little or too little satisfaction. The quality of the public services (Q 14) is thought to be little or not at all satisfactory from 39% of the interviewees, averagely satisfactory from 40% and 17% of them feel satisfied and very satisfied. As a synthesis of all the above questions the interviewee is asked to express their feelings about their quality of life (Q 17) and here we have contradictory results. 30% of them are minimally or not at all satisfactory, 42% are rather satisfactory and the rest, which means 28% feels from satisfactory to very satisfactory (graph nr 7).

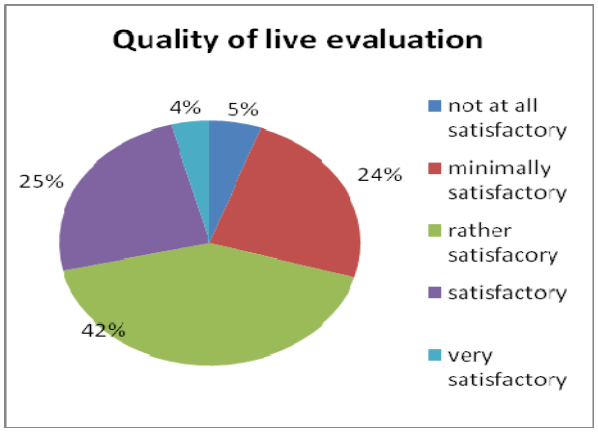


Figure 7. Quality of life evaluation

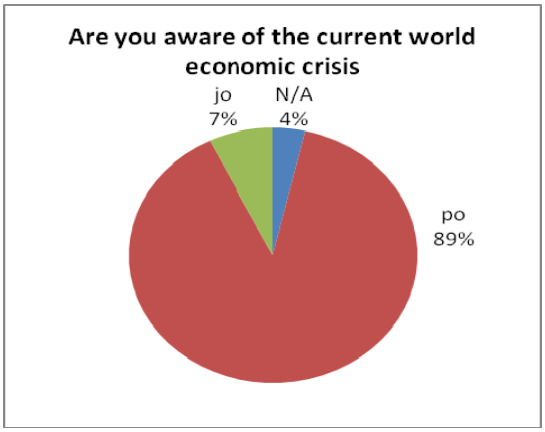


Figure 8. The world crisis information

b) The relation between the economic crisis and quality of life

The part of the survey which tends to show the assessment of the impact of the economic crisis concerning the quality of life starts with the question: are you aware of the resent world economic crisis? The survey reveals that 89% of the interviewees answered positively, 7% of them answered negatively and 4% didn't give an answer (graph nr 8)

Logistic regress applied to understand the link between question number 18 (Are you aware of the current world economic crisis) as a reactive variable (the Y of the logistic regress) and question number 21 (What has happened with the general price level) as a predictive variable (the X of the logistic regress) shows an odd ratio of a very high level (1.8099). This goes to show that according to the interviewees' price increase from a base level to another lever with one unit increases 1.8 times the possibility of crisis sensitivity on the consumers part (table nr 1 logistic regress)

The link between question number 18 as a reactive variable and question 20 as a predictive variable reveals an odd ratio lower than the unit (0.8867) which means that in mean time there are no positive chances for an increase in the family income. On the contrary it is expected to be in decline due to the impact of the economic crisis on the real income.

Table 1. Logistic regression between Q18 and Q19-23

Logistic regression	
Dependent Y	Y
Method	Enter
Sample size	207
Cases with Y=0	15 (7.25%)
Cases with Y=1	192 (92.75%)

Overall Model Fit

Null model -2 Log Likelihood	107.62585
Full model -2 Log Likelihood	99.65450
Chi-square	7.9714
DF	5
Significance level	P = 0.1578

Coefficients and Standard Errors

Variable	Coefficient	Std.Error	P
X19	0.0505	0.4062	0.9010
X20	-0.1203	0.2834	0.6712
X21	0.5933	0.4530	0.1904
X22	0.0630	0.4496	0.8886
X23	-0.9422	0.4775	0.0485
Constant	3.4529		

Odds Ratios and 95% Confidence Intervals

Variable	Odds Ratio	95% CI
X19	1.0518	0.4745 to 2.3317
X20	0.8867	0.5088 to 1.5451
X21	1.8099	0.7448 to 4.3982
X22	1.0650	0.4412 to 2.5710
X23	0.3898	0.1529 to 0.9938

Classification table

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	0	15	0.00 %
Y = 1	0	192	100.00 %
Percent of cases correctly classified			92.75 %

The impact of the economic crisis on the interviewees' income shows the following: 22.6% think that their income is the same in rapport to the crisis, 45.7% have a slight decrease in their income, 27.1% have a considerable decrease and 4% have an increase in their income. From the logistic regression analyses it appears that the difference in the income level has no impact on crisis sensitivity because the odd ratio is near the unit.

The method of logistic regress was applied in our study in order to find the link between question number 18 (Are you aware of the current world economic crisis) and the questions 30 (Where do you live; in rural or urban area), question 32 (The type of housing; owned or rented) and question 33(The educational level). All the three variables seem to have had a considerable impact on the interviewees' answers on the recent crisis. The most considerable impact seems to be noticed on variable 32 (The type of housing) which means that those families that live in rented accommodation, thus having a higher living cost, tend to be more liable to recent crisis impact. Such a family feels the economic effect of the crisis three times more than the families that live in their own accommodation (odd ratio 3.1373),

Table 2. Logistic regression between Q18 and Q30, 32, 33

Logistic regression	
Dependent Y	Y
Method	Enter

Sample size	207
Cases with Y=0	16 (7.73%)
Cases with Y=1	191 (92.27%)

Overall Model Fit

Null model -2 Log Likelihood	112.65429
Full model -2 Log Likelihood	108.74713
Chi-square	3.9072
DF	3
Significance level	P = 0.2717

Coefficients and Standard Errors

Variable	Coefficient	Std.Error	P
X33	0.5772	0.3647	0.1134
X30	0.6319	0.6381	0.3220
X32	1.1434	1.0707	0.2856
Constant	-1.4615		

Odds Ratios and 95% Confidence Intervals

Variable	Odds Ratio	95% CI
X33	1.7811	0.8715 to 3.6399
X30	1.8813	0.5386 to 6.5713
X32	3.1373	0.3847 to 25.5820

Classification table

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	0	16	0.00 %
Y = 1	0	191	100.00 %
Percent of cases correctly classified			92.27 %

Of no little importance seems to be the difference in sensitivity to the crisis in the rural areas compared to the urban ones. According to the study the possibility that families living in the rural area feels the crisis is two times more than the families living in the urban ones (odd ratio 1.8813). As well as this, the link between the education level and crisis sensitivity due to a better understanding of things, shows that the increase in the education level with one unit increases the crisis sensibility by 1.7811 times.

If we add the political impact on the above mentioned regress (question number 24: Are you personally interested in politics?) it is obvious that crisis sensitivity remains the same whether one lives in the rural or urban area. The same can also be said for the education level, but things are different when it comes to the kind of accommodation people live in, case in which the crisis sensitivity appears to increase considerably (from 3.1373 to 4.0951). This means that the political impact of economic crisis on the families with rented accommodation is more significant.

Table 3. Logistic regression between Q18 and Q30, 32, 33 and Q24

Logistic regression

Dependent Y	Y
Method	Enter
Sample size	207
Cases with Y=0	16 (7.73%)
Cases with Y=1	191 (92.27%)

Overall Model Fit

Null model -2 Log Likelihood	112.65429
Full model -2 Log Likelihood	95.69744
Chi-square	16.9568
DF	4
Significance level	P = 0.0020

Coefficients and Standard Errors

Variable	Coefficient	Std.Error	P
X33	0.4737	0.3882	0.2224
X30	0.5468	0.6664	0.4119
X32	1.4098	1.1208	0.2085
X24	-0.9250	0.2869	0.0013
Constant	1.7539		

Odds Ratios and 95% Confidence Intervals

Variable	Odds Ratio	95% CI
X33	1.6059	0.7503 to 3.4371
X30	1.7278	0.4680 to 6.3790
X32	4.0951	0.4552 to 36.8408
X24	0.3965	0.2260 to 0.6958

Classification table

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	0	16	0.00 %
Y = 1	0	191	100.00 %
Percent of cases correctly classified			92.27 %

On the other hand the level of the interviewees' interest in politics, which normally expresses their level of trust in it and the alternatives offered by it, has a major impact on their perception of the crisis. As the answers to question number 24 (Are you personally interested in politics?) show a decrease in the interest in politics (odd ratio has a much lower level than the unit, 0.3965) which means that the lower the interest in politics the lower the sensitivity towards the crisis becomes.

As a conclusion we can say that politics and momentary political events such as elections, political instability, and frequent rotation in office and frequent changes in government increase the level of the sensitivity towards the economic crisis. The answer to the question: how you forecast the crisis in the months to come 54.7% of the interviewees' answered "I don't know", 20.4% of them think the situation is going to improve and 23.9% think that it is going to deteriorate.

4. Conclusions

The above survey shows the following results:

1. In the quality of live assessment a very important part is played by subjective indicators along side with the objective ones. Therefore it is necessary that they are taken into consideration in its assessment. Of great importance in this context are the methods applied to measure them. Certainly every method should be based on surveys where the individuals express their subjective assessment concerning the integral components of the quality of life.
2. These indices are relative and of national character related to culture, tradition and the other peculiarities of each nation.
3. The quality of health care in the region is inadequate because 75% of the interviewees value it from satisfactory to not good at all.
4. Cultural life and sports in the region is poor, this because 88% of the interviewees give a negative answer to it, offering them little or no satisfaction at all.
5. Public services are beginning to improve. Consequently, there is a good attitude to them. Therefore 55% of the interviewees assess themselves to be rather satisfied, satisfied and very satisfied. Nevertheless, if it is thought that the requirements for the public sector will be on the increase, the assessment on their quality will probably change.
6. An increase in the amount of information and sensibility has been of great help to the citizens to know and assess the economic crisis of the recent months. The number of those interviewed with knowledge and perception on the crisis and its link with income level, purchasing power level and the general price level is high. All these factors increase the stress level and have a negative impact on the quality of life.
7. Interest shown in politics, which indirectly measures the trust and commitment level of broad masses in it, was low. 60% of the interviewees are little or not at all interested in politics.
8. In this research paper it is not possible to state out if the quality of life has been on the increase or otherwise because of the lack of the basis for comparison. This study will provide the basis for a further research to compare the Albanian quality of live in the future.

5. Recommendation

1. During the process of assessment of the quality of life, it is indispensable that decision making authorities in cooperation with the civil society should take into account objective indicators as well as subjective ones.
2. It is high time that human development in Albania was considered a priority as it is nowadays in all developed countries. Besides the classical concept (economic growth) it should be extended to include such new concepts as social, cultural, educational development, increase security and qualitative public services.
3. In order to enable the achievement of satisfactory levels of human development the quality of life should be increased. It is for this reason that an increased rate of awareness is considered to be of paramount importance. The community should put

pressure on the decision making authorities so that they improve their policies aiming at the betterment of all the afore-mentioned indicators.

References

1. Agresti, A. **An introduction to Categorical Data Analysis,** New York, Wiley, 2002
2. Argyle, M. **The Social Psychology of Leisure,** Penguin, London, 1996
3. Fallowfield, L. **The Quality of Life: The Missing Measurement in Health Care,** Souvenir Press (E & A), London, 1990
4. Fox, J. **Applied Regression Analysis, Linear Models and Related Methods,**. Sage publication, Thousand Oaks, CA,1997
5. Fox, J. **Logistic Regression, Maximum-Likelihood Estimated and Generalized Models,** York Summer Programme in Data Analysis, May 2005
6. Lamau, M.L. **The Idea of Quality of Life in the Health Field,** The Quality of Life in the Mediterranean Countries, First Mediterranean Meeting on Bioethics, Instituto Siciliano di Bioetica, 1992, pp. 47-68
7. Mano, R. **Regresi Logjistik dhe një zbatim në microfinance,** Mikrotezë Master, Departamenti I Matematikës, FSHN, UT, 2005
8. Moon, B. E. **The Political Economy of Basic Human Needs,** Cornell University Press, London, 1991
9. Nagamine, H. **Human Needs and Regional Development,** Nagoya, Japan, 1981
10. Noll, H.-H. **Societal Indicators and Social Reporting: The international Experience,** The Quality-of-Life Research Center, Copenhagen, 1998
11. Sen, A. **Capability and Well-being in the Quality of Life,** Nussbaum, M.C. and Sen, A. (eds.). Oxford University Press, 1993
12. * * * **Easy-to-use statistical software,** MedCalc, 2002
13. * * * **Quality of Life – A Concept Paper. Defining, Measuring and Reporting Quality of Life for Canadians,** Treasury Boeard of Canada Secretariat, 2000

STATISTICAL INDICATORS FOR RELIGIOUS STUDIES: INDICATORS OF LEVEL AND STRUCTURE[1]

Claudiu HERTELIU[2]

PhD, Assistant Professor, Department of Statistics and Econometrics,
University of Economics, Bucharest, Romania
(Co)Author of the books: Metode cantitative in studiul fenomenului religios (2009), Strategia universității: Metodologie şi studii de caz (2007)

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

Alexandru ISAIC-MANIU

PhD, University Professor, Department of Statistics Econometrics
University of Economics, Bucharest, Romania
(Co)Author of the books: Abordarea Sase Sigma. Interpretari, controverse, proceduri (2008), Proiectarea statistica a experimentelor (2006), Enciclopedia calitatii (2005), Dictionar de statistica generala (2003)

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

Abstract: *Using statistic indicators as vectors of information relative to the operational status of a phenomenon, including a religious one, is unanimously accepted. By introducing a system of statistic indicators we can also analyze the interfacing areas of a phenomenon. In this context, we have elaborated a system of statistic indicators specific to the religious field, which highlights the manner of positioning, evolving, interfering and manifesting of this field in society.*

Key words: *statistical indicators; religious studies; indicators of level; indicators of structure; indicators' descriptors; indicator card*

Context

Statistic indicators systems have been build to be applied in various areas: education, environment, etc. Depending on the destination of these systems, their complexity and dimension can be smaller or larger. In the followings, we present a few descriptive elements for a component of the system of statistic indicators proposed for the analysis of the religious phenomenon. The diagram below presents the general context and structure of the system of indicators for the analysis of the religious phenomenon. The indicators system is structured in six sub-systems (some of them divided in other categories), and has multiple approaches of the religious phenomenon: i) the use of different reporting coordinates, ii) the analysis based on various aggregation/depth levels, iii) the use of various measurement scales, iv) the definition of both primary (absolute as well as relative) and derived indicators, etc.

Not all presented indicators are innovative. Some of them (e.g. the structure of the population by declared religious affiliation) are generally used by most of the researchers, and contribute to obtaining the general image of the pursued phenomenon.

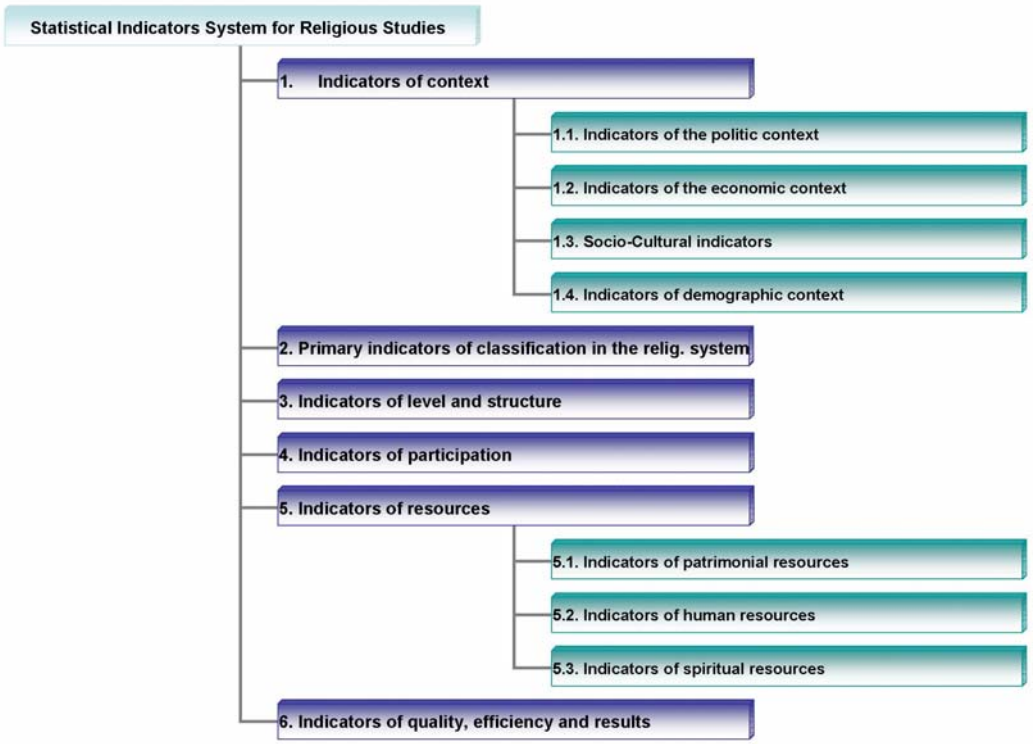


Figure 1. The structure of indicators' system for religious studies

This article is intended to detail the third sub-system of indicators: Indicator of level and structure, the indicators of this sub-system are the following: **Indicators of level an structure.** The indicators of this subsystem are the following:

1. Number of existing religious groups (NRG)
2. Structure of the population from viewpoint of the religious affiliation (RLG)
3. Minimum number of religious groups for the establishment of religious oligopoly (OLIG)
4. The structure of the population from viewpoint of the active religious affiliation (ACTRLG)
5. The balance of legally constituted families (LEGFAM)
6. The balance of multi-confessional families (MCONFAM)

Indicators' descriptors

Indicator / Characteristics	1. Number of existing religious groups
Definition	Represents the total number of religious groups legally constituted and recognized "as is" in a distinct geographical area (usually a country).
Scope	The indicator measures the level of flexibility and easiness in obtaining a "license" for a certain communion. It reflects the degree of religious freedom, compatibility and interest on which raises a certain communion within a country.
Symbol	**NRG**
Calculus method	It counts the number of religious groups recognized by an authority.

Formula	NRG=max(i) (1.) where: i – the numbers of order in the list of confessions legally recognized by the authorities
Necessary data	The list of legal confessions as recognized by the authorities.
Data source	The Ministry of Culture and Cults
Type	From the data type point of view : absolute From the measurement scale point of view: interval From the calculus point of view: primary From the time evolution point of view: moment
Class/ category	Indicators of level and structure
Aggregation level	National, regional, continental, worldwide, etc
Interpretation	It characterizes the degree of attractiveness and freedom within a country or region from the religious activities point of view.
Quality standards	One must consider the differentiation between the various phases a cult/church/movement must cover until the complete recognition or until the stage of legally accepted or tolerated group.

Indicator / Characteristics	2. Structure of the population from viewpoint of the religious affiliation
Definition	The weights of religious institutions from the perspective of the number of self-declared believers, in a certain geographical region.
Scope	The indicator is a classical one and maybe the most commonly used indicator for the study of religious phenomena. Based on its results, international comparisons can be made or it can characterize the existent type of "religious market".
Symbol	**RLG**
Calculus method	It is calculated by relating the total number of persons affiliated to a religious denomination to the total population investigated. Facultative, the result can be expressed as a percentage by multiplication with 100.
Formula	$RLG_i = \frac{B_i}{\sum B_i} \cdot [100]$ (2.) Where: B_i – the number of believers affiliated to the o confession $\sum B_i$ - the volume of the investigated population
Necessary data	The number of believers for each existing confession in the studies population (the C indicator within the indicators characterizing human resources).
Data source	NIS[3], Religious Institutions, Censuses[4]
Type	From the data type point of view : relative From the measurement scale point of view: continuous From the calculus point of view: derived From the time evolution point of view: moment
Class/ category	Indicators of level and structure
Aggregation level	Local, by county, regional, national, continental, worldwide, by gender, by residence environments, etc.
Interpretation	We can also measure and make comparisons between the existing situation regarding religious affiliation in various geographical areas or in population structures by various aggregation criteria (gender, education, nationality, etc).
Quality standards	In order to give a good interpretation of the results, one must also consider the fact that at a certain moment in time, a person can belong to a single church. Also, the denominator and the numerator of the fraction must be specific to a population synchronized spatially and timely. Another issue is given by the fact that this indicator is almost always based on the free statement of the individuals.

Indicator / Characteristics	3. Minimum number of religious groups for the establishment of religious oligopoly
Definition	The number of confessions, calculated on a descending sorted scale depending on the ratios obtained, assuring a limit percentage (ε) of the total number of believers. The limit percentage can be set, for example, to a level of 80%.
Scope	The indicator measures the degree of saturation/ diversification existent on the "religious market" in a certain location.
Symbol	**OLIG**
Calculus method	The weights obtained for each confession are determined and are cumulated increasingly. OLIG is the rank of the confession for which the value of ε is outran for the first time.
Formula	$OLIG = \min(i)\left\|F_i \in B; B = \{F_i \geq \varepsilon\}\right.$ (3.) Where: F_i – are the cumulative frequencies sorted decreasingly ε - is the minimum level set for constituting oligopoly.
Necessary data	The number of believers for each confession in the studied population.
Data source	NIS, Religious Institutions
Type	From the data type point of view : absolute From the measurement scale point of view: interval From the calculus point of view: primary From the time evolution point of view: moment
Class/ category	Indicators of level and structure
Aggregation level	Local, by county, regional, national, continental, worldwide, by gender, by residence environments, by ethnicity, etc.
Interpretation	If the minimum limit (ε) can be assured by 2-3 confessions a church – cartel can become the representatives in the relations between State and Church. If OLIG=1 than we are in the situation of monopoly.
Quality standards	The indicator characterizes the religious market in a certain location. Even if at a national or regional level the situation is in a certain way (one confession dominates the market) there can be other locations where another confession dominates the religious market, locally.

Indicator / Characteristics	4. The structure of the population from the viewpoint of active affiliation
Definition	The weights that religious institutions have within a population active from a religious point of view.
Scope	It is calculated from the balance of religious institutions and the active religious population, taking into consideration the IRI[5].
Symbol	**ACTRLG**
Calculus method	It is calculated by dividing the total number of persons affiliated to a religious confession to the total investigated population. Facultative, the result can be ecpressed as percentage, by multiplication with 100. The declaration of a person as belonging to a category of active believers s accomplished by using the AB[6] indicator (presented in the human resources indicators section).
Formula	$ACTRLG_i = \frac{AB_i}{\sum AB_i} \cdot [100]$ (4.) Where: AB_i – the number of active believers belonging to confession ranking i. $\sum AB_i$ - The volume of the population of active believers.
Necessary data	AB indicator– active believers.
Data source	Statistical investigations, Religious institutions, NIS.
Type	From the data type point of view : relative From the measurement scale point of view: continuous From the calculus point of view: derived

	From the time evolution point of view: moment
Class/ category	Indicators of level and structure
Aggregation level	Local, by county, regional, national, continental, worldwide, by gender, by residence environments, etc.
Interpretation	Measurements and comparisons can be made between the existing situation regarding active religious affiliation in various geographical locations, or in population structured according to several aggregation criteria (gender, education, nationality, etc).
Quality standards	The illustration precision of the existing situation for a certain location increases if this indicator is utilized. Taking into consideration the secularization manifested in religion, sometimes there can be situations when a religious minority (according to RLG) becomes the leader of a certain religious market (using ACTRLG as study method).

Indicator / Characteristics	**5. The balance of legally constituted families**
Definition	The specific weight of families legally registered as against to the total number of families.
Scope	The indicator shows, from a quantitative perspective, how the concept of "family –basic cell of society" is perceived in the daily practice.
Symbol	**LEGFAM**
Calculus method	The total number of legally constituted families is divided to the total number of families. The result can also be expressed as percentage, by multiplication with 100.
Formula	$LEGFAM = \frac{LF}{TNF} \cdot [100]$ (5.) Where: LF – the total number of legally constituted families. TNF – the total number of families.
Necessary data	The number of legally constituted families, as well as the total number of families.
Data source	Statistical investigation, NIS, censuses.
Type	From the data type point of view : relative From the measurement scale point of view: continuous From the calculus point of view: derived From the time evolution point of view: moment
Class/ category	Indicators of level and structure
Aggregation level	Local, by county, regional, national, continental, worldwide, by confessions, by residence environments, etc.
Interpretation	The more the indicator is closer to 1 (or 100%) the more the studied population has the tendency to register legally constituted families.
Quality standards	Various interpretations can be given to the way of registration. Most of the times, religious institutions don't recognize a family unless it has also been registered by a religious institution. There are also situations in which legal registration is sufficient for religious registration[7].

Indicator / Characteristics	**6. The balance of multi-confessional families**
Definition	The number of families that consist of at least two believers of another religious institution compared to the total number of families.
Scope	It intends to illustrate the degree of opening, freedom, and overcome of confessional barriers within a family.
Symbol	**MCONFAM**
Calculus method	The number of multi-confessional families is divided to the total number of families. The result can also be expressed as percentage by multiplication with 100.

Formula	$MCONFAM = \frac{MCF}{TNF} \cdot [1000]$ (6.) Where: MCF – the number of multi-confessional families TNF – the total number of families
Necessary data	The number of multi-confessional families, as well as the total number of families.
Data source	Statistical investigations, NIS
Type	From the data type point of view : relative From the measurement scale point of view: continuous From the calculus point of view: derived From the time evolution point of view: moment
Class/ category	Indicators of level and structure
Aggregation level	Local, by county, regional, national, continental, worldwide, by residence environments, by confessions, by ethnic groups, etc.
Interpretation	The interpretation can be performed by comparison to other countries or by study of confessions where the multi-confessional families are more frequent.
Quality standards	Religious institutions usually tolerate but don't encourage multi-confessional families. This is the reason why the indicator will not register a high level. The analysis can be refined by using the information regarding the religion chosen by the children of the multi-confessional families.

Conclusions and future work

Without a doubt, the description within a standardized system and the creation of a unitary and comprehensive conceptual context is merely the methodological foundation necessary for the following phases: i) implementing improvements (after receiving feed-back from the interested academic community), ii) collecting statistic information for the alimentation and operationalization of this system, iii) analyzing, processing and disseminating the results towards the scientific community and interested public.

We hope that this methodological process will be followed by the implementation of the other phases necessary to finalize the scientific process initiated by this material.

References

1. Frunza, S. and Frunza, M **Etica, superstitie si laicizarea spatiului public,** Journal for the Study of Religions and Ideologies, nr. 23, Spring 2009, pp. 13-35
2. Herteliu, C **Metode cantitative in studiul fenomenului religios,** Ed. Napoca Star, Cluj Napoca, 2009
3. Herteliu, C. **Statistical Indicators System regarding Religious Phenomena,** Journal for the Study of Religions and Ideologies, no. 16, Spring 2007, pp. 111-127
4. Iannaccone, L., **Introduction to the Economics of Religion**, Journal of Economic Literature, Vol. 36, No. 3, September 1998, pp. 1465-1495
5. Idel, M. **Abordari metodologice in studiile religioase,** Journal for the Study of Religions and Ideologies, volume 6 (16), 2007, pp. 5-20
6. Isaic-Maniu, Al. (coord.), Pecican, E., Stefănescu, D., Voda, V., Wagner, P. **Dictionar de Statistica Generala**, Ed. Economica, Bucharest, 2003
7. Isaic-Maniu, Al. and Herteliu, C. **Ethnic and Religious Groups in Romania – Educational (Co)Incidences,** Journal for the Study of Religions and Ideologies, No. 12, Winter 2005, pp. 68-75

8. Jonhnstone, J.N. **Indicators, Educational**. In: T.Husén, T.N. Postlethwaite (editors) **The International Encyclopedia of Education**, Pergamon Press, Oxford, vol. V, 1985, p. 2432-2438
9. Smidt, C.E., Kellstedt, L.A. and Guth, J.L. (eds.) **The Oxford Handbook of Religion and American Politics,** Oxford University Press, New Yourk, 2009, pp. 3-42

---

[1] **Acknowledgements**
This paper was designed with support from the following research project: „The Influence of Religion on the Social Processes from Romania– A Quantitative New-Weberian Approach Using Statistical Samples Techniques", code 1809, Project Manager Claudiu Herteliu. Grant won in 2008 "Ideas" Competition from PNCD II and financial supported by UEFISCSU.

[2] **Claudiu Herteliu** currently holds the position of Assistant Professor within the Chair of Statistics and Econometrics, at the Bucharest University of Economics. In 2007, he earned his PhD degree at the same university. His main areas of academic interest are: theoretical and macroeconomic statistics, religion statistics, education statistics, econometrics. He attended many scientific conferences both in Romania and abroad (USA, Germany, Holland, France, Italy, Hungary etc.) He published (as co-author and single author) over 20 articles in scientific journals, and contributed to a number of four books published in Romania.

[3] National Institute of Statistics

[4] In Romania, in the communist period, the censuses in 1956, 1966 and 1977 have no record of religious affiliations, which makes time comparisons difficult.

[5] It is about the Intensity of Religious Implication indicator.

[6] It is about Active Believer indicator.

[7] The case of the USA procedures, where in many states, the marriage license is given after the religious ceremony.

JOURNAL OF APPLIED QUANTITATIVE METHODS

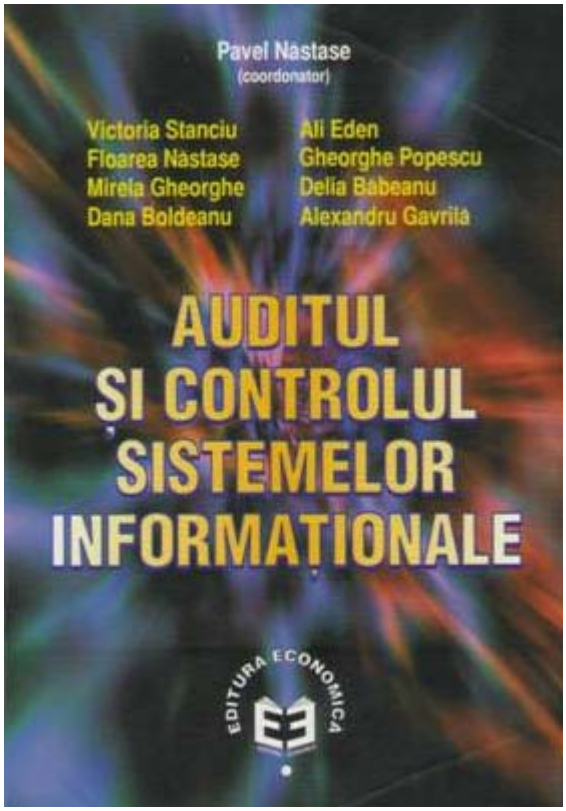
Valerica MARES
Ph.D, Lecturer, Faculty of Accounting and Management Information Systems,
University of Economics, Bucharest, Romania

E-mail: maresvalerica@yahoo.com **Web page:** www.cig.ase.ro

Key words: *information systems audit concepts; corporative governance; systems development life cycle; information system infrastructure and security; encrypt systems; business continuity and disaster recovery*

Book Review on
AUDIT AND INFORMATION SYSTEMS CONTROL
("AUDITUL SI CONTROLUL SISTEMELOR INFORMATIONALE"),

by Pavel NASTASE (coord.), Eden ALI, Floarea NASTASE, Victoria STANCIU, Gheorghe POPESCU, Mirela GHEORGHE, Delia BABEANU, Dana BOLDEANU, Alexandru GAVRILA, Ed. Economica, Bucharest, 2007



The necessity to adapt the audit to the new technological reality, having in the first plan the advantages generate by these technologies and the risks of the environments, has double impact: the information technology is subject and object of the audit.

The authors propose to identify the vulnerabilities regarding the information systems security and to elaborate different forms to eliminate the threats.

The *first chapter* presents standards and guides for information systems audit. It is also present the planning, management and deployment of the information systems audit processes.

Chapter three is dedicated to the information system life cycle management and application controls. There are define the life cycle, projects management, development methods, including systems

development life cycle, risks associated, Business Process Reengineering (BPR).

Technical infrastructure of the information systems, *chapter four*, contain: hardware and software architecture, operating in information systems, data management and network infrastructure.

Chapter five has information systems security basic elements as security management, logical access control, Internet and Intranet security, intruders' detection systems, security systems and electronic equipments protection.

Encrypt systems are the most important in data protection. In this *chapter* there are presented encryption with secret and public key, digital signature, infrastructure of the public keys and how we use the encryption in OSI protocols.

Business Continuity and Disaster Recovery are also important. In *chapter seven*, there are presented the BCP components, information systems continuity and planning for disaster recovery.

All seven chapters are finalized by questionnaires to determine the knowledge gathered regarding the problems presented.

The paper work is necessary to the auditors, specialists in information systems security, managers, but not the last to those who want to prepare CISA exam.

The preface of this book is presented by Alan T Lord, Professor of Accounting, Bowling Green State University, Ohio USA, and he recommends:

"As the Chair of the global task force from ISACA, I think the authors have done a commendable job with this text. The organizational structure and content of this text allow it to serve as a preference primer for studying the information systems audit and control discipline."