

WWW.JAQM.RO

**JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS**

**Evaluation of Academic
Activities in Universities**

**Vol. 3
No. 2
Summer
2008**

ISSN 1842-4562

JAQM Editorial Board

Editors

Ion Ivan, University of Economics, Romania

Claudiu Herteliu, University of Economics, Romania

Gheorghe Nosca, Association for Development through Science and Education, Romania

Editorial Team

Cristian Amancei, University of Economics, Romania

Sara Bocaneanu, University of Economics, Romania

Catalin Boja, University of Economics, Romania

Eugen Dumitrascu, Craiova University, Romania

Nicu Enescu, Craiova University, Romania

Giani Ionel Gradinaru, University of Economics, Romania

Irina Maria Isaic-Maniu, University of Economics, Romania

Miruna Mazurencu Marinescu, University of Economics, Romania

Daniel Traian Pele, University of Economics, Romania

Marius Popa, University of Economics, Romania

Mihai Sacala, University of Economics, Romania

Cristian Toma, University of Economics, Romania

Erika Tusa, University of Economics, Romania

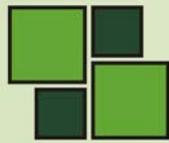
Adrian Visoiu, University of Economics, Romania

Manuscript Editor

Lucian Naie, IBM Romania

JAQM Advisory Board

Ioan Andone, Al. Ioan Cuza University, Romania
Tudorel Andrei, University of Economics, Romania
Ovidiu Artopolescu, Microsoft Romania
Gabriel Badescu, Babes-Bolyai University, Romania
Catalin Balescu, National University of Arts, Romania
Constanta Bodea, University of Economics, Romania
Ion Bolun, Academy of Economic Studies of Moldova
Recep Boztemur, Middle East Technical University Ankara, Turkey
Constantin Bratianu, University of Economics, Romania
Irinel Burloiu, Intel Romania
Ilie Costas, Academy of Economic Studies of Moldova
Valentin Cristea, Politehnica University of Bucharest, Romania
Marian-Pompiliu Cristescu, Lucian Blaga University, Romania
Victor Croitoru, University Politehnica of Bucharest, Romania
Cristian Pop Eleches, Columbia University, USA
Bogdan Ghilic Micu, University of Economics, Romania
Anatol Godonoaga, Academy of Economic Studies of Moldova
Alexandru Isaic-Maniu, University of Economics, Romania
Ion Ivan, University of Economics, Romania
Radu Macovei, University of Medicine Carol Davila, Romania
Dumitru Marin, University of Economics, Romania
Dumitru Matis, Babes-Bolyai University, Romania
Adrian Mihalache, University Politehnica of Bucharest, Romania
Constantin Mitrut, University of Economics, Romania
Mihaela Muntean, Western University Timisoara, Romania
Ioan Neacsu, University of Bucharest, Romania
Peter Nijkamp, Free University De Boelelaan, The Netherlands
Stefan Nitchi, Babes-Bolyai University, Romania
Gheorghe Nosca, Association for Development through Science and Education, Romania
Dumitru Oprea, Al. Ioan Cuza University, Romania
Adriean Parlog, National Defense University, Bucharest, Romania
Victor Valeriu Patriciu, Military Technical Academy, Romania
Perran Penrose, Independent, Connected with Harvard University, USA and London University, UK
Dan Petrovici, Kent University, UK
Victor Ploae, Ovidius University, Romania
Gabriel Popescu, University of Economics, Romania
Mihai Roman, University of Economics, Romania
Ion Gh. Rosca, University of Economics, Romania
Gheorghe Sabau, University of Economics, Romania
Radu Serban, University of Economics, Romania
Satish Chand Sharma, Janta Vedic College, Baraut, India
Ion Smeureanu, University of Economics, Romania
Ilie Tamas, University of Economics, Romania
Nicolae Tapus, University Politehnica of Bucharest, Romania
Daniel Teodorescu, Emory University, USA
Dumitru Todoroi, Academy of Economic Studies of Moldova
Nicolae Tomai, Babes-Bolyai University, Romania
Viorel Gh. Voda, Mathematics Institute of Romanian Academy, Romania
Victor Voicu, University of Medicine Carol Davila, Romania
Vergil Voineagu, University of Economics, Romania



Evaluation of Academic Activities in Universities

**Margaret MACDOUGALL, Simon C. RILEY,
Helen S. CAMERON, Brian MCKINSTRY**
Halos and Horns in the Assessment of Undergraduate Medical Students: A
Consistency-Based Approach 116

Mihaela Cornelia PREJMEREAN, Simona VASILACHE
A Three Way Analysis of the Academic Capital of a Romanian University 129

A. INDIRA
Evaluation of Training Programs for Rural Development 139

Gholamreza JANDAGHI, Fatemeh SHATERIAN
Validity, Reliability and Difficulty Indices for Instructor-Built Exam Questions 151

Quantitative Methods Inquires

Ion DOBRE, Adriana AnaMaria ALEXANDRU
Modelling Unemployment Rate Using Box-Jenkins Procedure 156

Ercan TIRTIROGLU, Matt ELBECK
Qualifying Purchase Intentions Using Queueing Theory 167

Bogdan OANCEA
Improving the Performance of Sparse LU Matrix Factorization Using a Supernodal
Algorithm 179

Mariana KAZNOVSKY
Money Demand in Romanian Economy, Using Multiple Regression Method and
Unrestricted VAR Model 187

Book Review

Bogdan Vasile ILEANU
Book Review on Econometric Theory and Practice ("Teorie si Practica
Econometrica") by Vergil VOINEAGU, Emilia TITAN, Radu SERBAN,
Simona GHITA, Daniela TODOSE, Cristina BOBOC, Daniel PELE
Editura Meteor Press, Bucharest, 2007 196

HALOS AND HORNS IN THE ASSESSMENT OF UNDERGRADUATE MEDICAL STUDENTS: A CONSISTENCY-BASED APPROACH^a

Margaret MacDOUGALL^b

PhD, Medical Statistician
Community Health Sciences, Public Health Sciences Section,
College of Medicine and Veterinary Medicine, University of Edinburgh
Teviot Place, Edinburgh EH8 9AG, Scotland, UK

E-mail: Margaret.MacDougall@ed.ac.uk



Simon C. RILEY^c

PhD, Senior Lecturer in Obstetrics and Gynaecology (Non-Clinical)
Centre for Reproductive Biology, Queen's Medical Research Institute,
University of Edinburgh
47 Little France Crescent, Edinburgh EH16 4TJ, Scotland, UK

E-mail: Simon.C.Riley@ed.ac.uk



Helen S. CAMERON^d

BSc, MBChB, Senior Lecturer and Archie Duncan Fellow in Medical Education
Director Medical Teaching Organisation, College of Medicine and Veterinary Medicine,
University of Edinburgh
Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB, Scotland, UK

E-mail: Helen.Cameron@ed.ac.uk



Brian MCKINSTRY^e

PhD, Senior Research Fellow
Community Health Sciences, General Practice Section, University of Edinburgh
20 West Richmond Street, Edinburgh EH8 9DX, Scotland, UK

E-mail: Brian.Mckinstry@ed.ac.uk



Abstract: *The authors introduce a consistency-based approach to detecting examiner bias. On comparing intra-class correlation coefficients on transformed data for supervisor continuous performance and report marks (ICC1*) with those for supervisor continuous performance and second marker report marks (ICC2*), a highly significant difference was obtained for both the entire cohort (ICC1* = .72, ICC2* = .30, F = 2.47, p < .0005 (N = 1085)) and the subgroup with high supervisor ratings for continuous performance (ICC1* = .62, ICC2* = .24, F = 1.97, p < .0005 (n = 952)). A strong halo effect was detected and preliminary evidence was obtained for the presence of a strong horn effect for students with lower scores, thus providing a basis for future research.*

Key words: *halo effect; horn effect; intra-class correlation coefficient; second marker; supervisor bias; undergraduate assessment; Zegers-ten Berge general association coefficient*

Introduction

The tendency for good or bad performers over one dimension to deliver consistently good or bad performances overall is already recognized (Dennis 2007, Fiscaro & Lance 1990, Pike 1999, Pulakos et al. 1986). Thus, in an ideal assessment setting where ratings are untainted by examiner bias, one would expect there to be a detectable level of consistency in individual student performance across various assessment dimensions. It is this particular type of consistency, representative of true consistency and hence, illusory bias, which we choose to refer to henceforth in this study as natural consistency.

The need to detect and eliminate examiner bias is clearly a critical one if marks allocated to students are to be representative of performance, particularly in contexts where students are ranked against one another for future selection purposes. Moreover, assessment procedures must be rigorously monitored if the reputational quality of academic programmes is to be maintained and justified. Our specific aim here, therefore, is to introduce new methodology for testing examiner bias where examiners have prior exposure to student performance in one dimension and are required to objectively mark students in a separate but related dimension. Through use of a case study involving undergraduate medical students, this methodology will test for supervisor bias in report marking where supervisors have prior exposure to student continuous performance. The procedure adopted will also explicitly correct for natural consistency as defined above by identifying supervisor bias as that specific contribution to consistency in supervisor ratings across continuous performance and written report performance which is explicitly over and above that of natural consistency. Where this type of bias is found to coincide with the attribution of high or low marks to student assignments, we shall refer to it as a *halo* or *horn effect*, respectively.

Two similar tendencies are apparent in the literature wherever the term 'halo effect' is adopted. The first of these tendencies is a non-prescriptive use of language (as in Wakeford et al. 1995) which suggests that the halo effect is merely the existence of evidence for the rating of one attribute influencing the rating of another. The second, and more common, tendency is to use the term 'halo effect' to refer to a phenomenon akin to *any one* of the two forms of bias considered in this study whilst, with some exceptions (for example, Brown 1965, Pulakos et al. 1986, Fiscaro & Lance 1990), leaving the problem of natural consistency unchallenged.

The latter tendency originates with the inception of the term 'halo effect' to discuss phenomena in measurement data under the auspices of Thorndike (1920); thus those who choose to assume this interpretation (see, for example, Bowden 1933, Anastasi 1988, Fairweather 1988 and Streiner & Norman 2003) may be referred to as his followers. Nevertheless, it makes a great deal of sense to keep the original everyday use of this notion, with its positive connotation, in mind when passing from the material world to the world of measurement theory (Dudycha 1942), we suggest not least because of the greater opportunity this affords to differentiate between different kinds of examiner bias.

The above two generalizing tendencies have the effect that the terms 'horn effect' and 'stigma effect' occur much more rarely in the literature than that of 'halo effect' as their interpretation is already subsumed within the intended notion of halo effect. Nevertheless, confusion can arise in this area too. For example, Marshall (2003) appears to use the terms 'stigma effect' and 'negative stigma' interchangeably to refer to negative bias in examiners

where pupils are known to be repeating a grade. Moreover, he omits to provide a definition for either of these terms at the outset and the reader is left to interpret their meaning either implicitly or based on the hidden assumption that their meaning is in some sense obvious. Further, Evans (2002) appears to make a distinction by referring to 'The "halo" effect and the opposing "horns" effect,' but in the absence of any supporting definitions for either of these effects. By contrast, Rubin (1982) uses the term 'horn effect' to refer simply to the tendency to limit the overall assessment of an individual to a single negative attribute.

It is interesting to note, however, that within the context of employee appraisal, Arnold and Pulich (2003) make the interesting distinction between the 'horn' and 'halo' effects, whereby, for example, the horn effect is specifically that 'which occurs when a manager perceives one negative aspect about an employee or his or her performance and generalizes it into an overall poor appraisal rating.'

In seeking to make a similar distinction, the notions of halo and horn effect which we define in this paper (both intuitively and mathematically) are contrary to the two tendencies outlined above. Moreover, these notions make a substantial contribution to addressing Pike's 'critical [problem] for assessment research' (Pike 1999) of differentiating between supervisor bias and true 'regularities' in performance across different dimensions. Our study also benefits from there being a meaningful standard *against which* to measure examiner bias. Precisely, we utilize second marker ratings with second markers having been blinded to the student's identity (and hence their participation in the project) and to the continuous performance rating allocated by their supervisor. As such, our study avoids the potential for uncertainty in other studies (Pulakos et al. 1986, Fisticaro & Lance 1990) wherein correlations across ratings for multiple attributes assigned by expert or trained markers are assumed as surrogates for measures of natural associations (or, associations based on student abilities which are uncontaminated by examiner bias). Moreover, due to constraints on staff time, inclusion of at most a second marker (that is, 'double-marking') is by far a more common choice of assessment regime across different disciplines and places of learning than those involving further markers. Thus, we consider our approach to detecting bias a pragmatic one in so far as, realistically speaking, it may be replicated to test for bias in a wide variety of real-life assessment scenarios.

Method

Background to Participants

Within the 4th year of the undergraduate medical curriculum at the University of Edinburgh, all students are required to identify a supervisor and field of interest to enable them to participate in a 14-week research project known as the 4th year Student Selected Component (SSC4).

During the SSC4 period, the students must prepare a project report, usually in the form of a medical or scientific article of up to 3000 words, which reports on their research findings. The project supervisor allocates a total of two percentage marks to each of their students. The two marks constitute a continuous performance rating measuring overall performance throughout the duration of the project and a report rating measuring the quality of the final written report. The quality of the written report is also allocated a percentage mark by a second examiner with concurrent experience of supervising and marking SSC4 projects within the same student cohort. In their capacity as a second marker,

this rater is, however, also blinded to the continuous performance rating allocated to the student concerned and to the identity of that student.

All supervisors are advised to use the same detailed list of performance indicators to assist them in allocating continuous performance ratings to their students. In the allocation of ratings for written reports, all supervisors and second markers are recommended to use a separate comprehensive but shorter list of marking criteria, this list being identical for all markers.

Each of the above three percentage marks is then converted to a grade (A – F), with grades A, B, C, D, E and F corresponding to marks 90 - 100, 80 - 89, 70 - 79, 60 - 69, 50 - 59 (marginal fail) and 0 - 49 (fail), respectively. In the majority of cases, there is no need to call in a third marker to correct for mismatch between supervisor and second marker ratings and the final grade assigned to the student is that obtained from combining the supervisor continuous performance, supervisor report and second marker report ratings.

Whilst continuous performance and report writing are intended to constitute two separate dimensions of SSC4 student performance, that is not to say that student abilities across these two dimensions should differ markedly. Thus, we assumed that there was natural consistency best assessed by the correlation between the supervisor continuous performance mark and the second marker report mark and that supervisor bias could be evaluated by looking for additional consistency between the supervisor continuous performance and report marks. We therefore used intra-class correlation coefficients (ICCs) to assess the evidence that consistency between supervisor continuous performance and written report ratings was significantly greater than that between supervisor continuous performance ratings and the corresponding second marker report ratings.

Data Preparation

All SSC4 continuous performance and report performance data corresponding to the period July 2001 to June 2006 ($N = 1096$) were extracted in an anonymized format from internal undergraduate medical student examination records at the University of Edinburgh and stored in an MS Excel database. Ethical approval to use these data for the current study was formally granted by the University of Edinburgh College of Medicine and Veterinary Medicine Committee on the Use of Student Volunteers.

Statistical Analyses and Underlying Theory

Calculations and data analyses were performed using MS Excel 2003 and the statistical packages Minitab (Version 14.12) and SPSS (Version 14.0).

The model we assumed for this study was a two-way mixed effects model (McGraw & Wong 1996) in which examiners were recognized as fixed effects and students as random effects. In calculating ICCs for consistency rather than absolute agreement, we chose to measure the extent to which corresponding sets of marks agreed according to an additive transformation rather than in absolute terms. Thus, in the notation of Fagot (1993), we used the consistency-based intra-class correlation coefficient $ICC(3, 1)$ for a two-way mixed model in which raters are fixed and subjects are random.¹

In testing for a halo effect, two ICCs were calculated over the period 2001 – 2006. The first of these, $ICC1$, measured consistency between supervisor continuous performance and report marks and the second, $ICC2$, measured consistency between supervisor continuous performance and second marker report marks. In our study, these ICCs represent

the proportion of the total variance in marks (inclusive of error variance) which can be explained purely in terms of variation between the students in the study. As is well known, ICCs range from -1 to 1. However, within the current context, they are understood to converge towards 1 as the association between the two corresponding sets of marks increases, with negative ICCs indicating the extreme case where on examination of ratings, error variance is greater than that across individual students.

Using the above terminology, in testing for a halo effect, our preliminary null hypothesis was as follows:

$$ICC1 = ICC2. \quad (1)$$

The hypothesis test which we used was based on the method of Alsawalmeh and Feldt (1994). Alsawalmeh and Feldt already allow for the comparison of two ICCs based on the same sample, although in the absence of any application to educational data or any allowance for the possibility that ratings for different ICCs might violate the assumption of rater independence. Our sample size for subjects was much greater than that assumed by Alsawalmeh and Feldt. We were therefore able to apply the asymptotic properties of the mean square terms to simplify the algebra used in the calculation of the degrees of freedom whilst allowing for the non-independence of raters across $ICC1$ and $ICC2$.

Nevertheless, the original requirement of Normality for the Alsawalmeh-Feldt test still required to be met. Thus, we sought an optimal transformation for ensuring that the data for each of supervisor continuous performance mark, supervisor report mark and second examiner report mark approximated to Normality. With the aid of the Box-Cox transformation procedure (Box & Cox 1964), we therefore assumed the polynomial transformation

$$transformed\ mark = (original\ mark)^5 \quad (2)$$

as the single choice of transformation to be applied in each case. Consequently, in practice, it was necessary for us to apply our hypothesis test to refute the null hypothesis,

$$ICC1^* = ICC2^*, \quad (3)$$

with $ICC1^* = ICC1^5$ and $ICC2^* = ICC2^5$.

In testing the null hypothesis for the transformed data, we used the property (Alsawalmeh & Feldt 1994) that the test statistic $F = \frac{1 - ICC2^*}{1 - ICC1^*}$ approximates to a central F -distribution with degrees of freedom d_1 and d_2 defined as strictly positive integers in accordance with the method of Satterwaite (1941). One notable impact of our use of the asymptotic properties of the mean square in our adaptation of the hypothesis test for larger samples was that of decreasing the degrees of freedom d_1 and d_2 , above for the sample sizes we assumed. This made our test more conservative (with the effect that the probability of a Type I error was reduced).

In order to differentiate between halo and horn effects, we divided the data into two cohorts according to the grades corresponding to the percentage marks for continuous performance assigned by supervisors. Thus, the high grade cohort referred to those

percentage marks corresponding to grades A and B, whilst the lower grade cohort referred to those percentage marks corresponding to grades C – F.

Using the raw percentage data, we determined the ICCs and corresponding confidence intervals for both grade cohorts. On the basis of the Box-Cox transformation procedure, we found that the transformation defined under (2) was also the optimal one for Normalization of data for the high grade cohort. On application of this transformation, we tested hypothesis (3) as previously. For the lower grade cohort, on the other hand, it was not possible to find a Normalizing transformation for the data. Thus, in adherence to the assumptions of our hypothesis test, we did not test hypothesis (3) for these data.

For each application of our hypothesis test, we assumed a significance level of .05.

In interpreting our choice of ICC as a measure of examiner consistency, it is useful to consider Zegers and ten Berge's notion of a general association coefficient (Zegers & ten Berge 1985). The latter coefficient was designed to measure the level of absolute agreement between two variables in terms of the mean squared distance once each of these two variables has undergone a specific admissible transformation (ibid.) in accordance with the type of data under consideration. Later, Stine (1989) coined the useful term 'relational agreement' rather than 'association' to refer to the type of measurement represented by Zegers and ten Berge's coefficient. In adopting this term, Stine recognized absolute agreement under the *identity* transformation to be the strictest of a *family* of possible types of agreement which are meaningful in a measurement theoretic sense, the appropriate transformation being dependent on the particular measurement scale represented by the data.

Fagot (1993) has already established a useful identity between a particular case of the Zegers-ten Berge general association coefficient and $ICC(3,1)$ for continuous ratings when they are understood to be representative of Normally distributed data on an additive scale. In particular, for a study involving k examiners and N subjects, let X_i denote the variable ranging over all N ratings for examiner i ($i = 1, 2, \dots, k$), \bar{X}_i denote the arithmetic mean of all ratings for examiner i and let V_i be defined according to the admissible transformation $V_i = X_i - \bar{X}_i$ ($i = 1, 2, \dots, k$). Then the general association coefficient for the transformed variables is precisely equal to $ICC(3,1)$ for the corresponding untransformed variables.

This result is particularly useful because it informs us that, within the context of our study in which two sets of ratings are being compared at any one time, $ICC(3,1)$ is a measure of the extent to which the distribution of the marks about the mean for one set of data is the same as that for the other. For the case in which two sets of marks are being compared at any one time, this interpretation of relational agreement can be understood graphically in terms of the degree of scatter of the data points (V_1, V_2) about the line $V_2 = V_1$. Moreover, as any one of our $ICC1^*$ and $ICC2^*$ approaches 1, the two corresponding sets of marks should tend towards perfect agreement in the above sense.

On applying the above admissible transformation to supervisor and second marker Normalized ratings, we therefore used scatter plots to address the challenge of providing a visual representation of the contrasting relationships between supervisor continuous performance and report marks and supervisor continuous performance and second marker report marks which had previously come to light by means of the ICCs. We carried out this procedure separately for the data in its entirety and for the high grade cohort but not for the

lower grade cohort, on account of the absence of a suitable Normalizing transformation for the corresponding data.

Results

The ICCs used to assess examiner bias together with their corresponding 95% CIs are provided in Table 1 both for the raw data and for the transformed data, where appropriate.

Table 1. ICC-Based Consistency Between a) Supervisor Continuous Performance Mark and Supervisor Report Mark (*ICC1* and *ICC1**) and b) Supervisor Continuous Performance Mark and Second Marker Report Mark (*ICC2* and *ICC2**)

Grade cohort	<i>ICC1</i> (95% CI)	<i>ICC1*</i> (95% CI)	Grade cohort	<i>ICC2</i> (95% CI)	<i>ICC2*</i> (95% CI)
All grades (<i>N</i> = 1085) ^a	.76 (.74, .79)	.72 (.69, .75)	All grades (<i>N</i> = 1085) ^a	.33 (.28, .38)	.30 (.25, .36)
High grades: A - B (<i>n</i> = 952)	.59 (.55, .63)	.62 (.58, .65)	High grades: A - B (<i>n</i> = 952)	.22 (.16, .28)	.24 (.18, .30)
Lower grades: C - F (<i>n</i> = 133)	.72 (.63, .80)		Lower grades: C - F (<i>n</i> = 133)	.42 (.27, .55)	

Note. ICC = intra-class correlation coefficient

^a All ICCs were calculated only for those students for whom all three percentage marks, corresponding to supervisor continuous performance and supervisor and second marker report ratings, were available. *ICC1** and *ICC2** denote the consistency measures for the data further to the transformation defined under (2), above. Marks were incomplete for 11 out of 1096 (1.0%) of the students within the 2001 - 2006 dataset.

On testing hypothesis (3) for the data in their entirety and in particular, for the high grade cohort, a highly significant difference was found between *ICC1** and *ICC2** in each case ($F = 2.47, p < .0005 (N = 1085)$, and $F = 1.97, p < .0005 (n = 952)$, respectively).

The relationships between supervisor continuous performance and report marks and supervisor continuous performance and second marker report marks are represented in Figure 1 for all of the data and separately for those data corresponding only to students who received high grades for continuous performance.

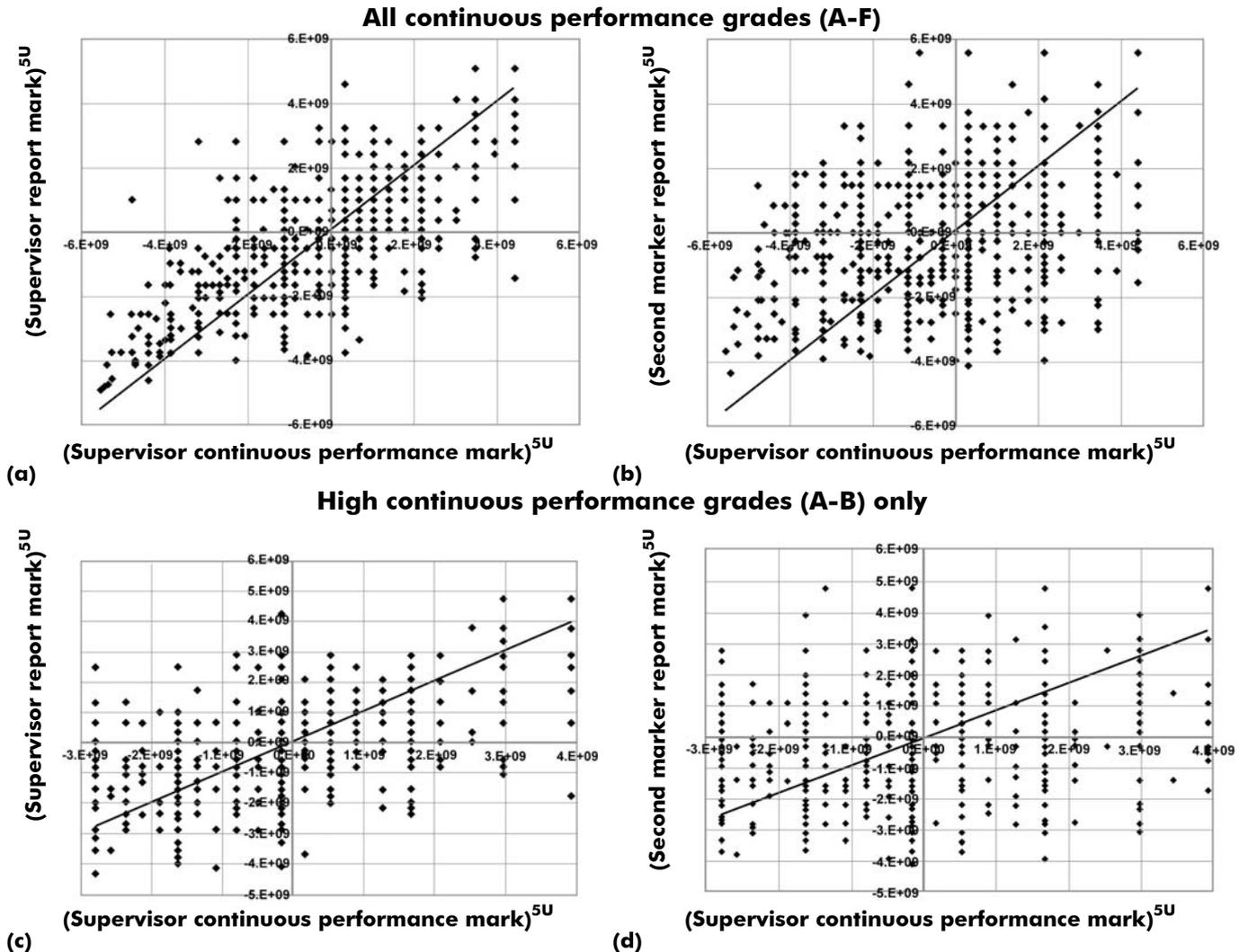


Figure 1. Relationship between continuous performance marks and report marks relative to the 45° line of perfect agreement through the origin following Normalization and subsequent application of the Zegers-ten Berge uniforming transformation $V_i = X_i - \bar{X}_i$, where X_i ranges over all ratings for a given type of measurement i and \bar{X}_i denotes the arithmetic mean of all ratings for measurement type i ($i = 1, 2$).

Note. The notation '(Supervisor continuous performance mark)^{5U}', '(Supervisor report mark)^{5U}' and '(Second marker report mark)^{5U}' is used here to denote that the expression in brackets has been transformed, first through Normalization by exponentiation to the power 5 and subsequently through application of the above Zegers-ten Berge uniforming transformation.

Discussion

Having tested hypothesis (3) for our data, according to our definitions, there is extremely compelling evidence for the existence of supervisor bias and more specifically, for the existence of a strong halo effect in supervisor assessment of SSC4 reports.

The result (Table 1) that none of the ICCs for the grade subgroups attain or exceed the corresponding values for the entire cohort is an inevitable consequence of the increase in the ratio of error variance to true variance across students, which occurs when sample size is reduced. Nevertheless, the ICCs for the subgroups can be considered in their own right, together with their corresponding confidence intervals. The failure to Normalize the data for the lower grade cohort was undoubtedly influenced by the relatively small cohort size ($n = 133$). Even in the presence of a suitable transformation, it is doubtful that n would have been sufficiently large here to satisfy the underlying asymptotic assumptions of our hypothesis test.

On comparing $ICC1$ and $ICC2$ with $ICC1^*$ and $ICC2^*$, respectively in Table 1, it is clear that the Normalizing transformation has had very little impact on the level of consistency as represented by these indices. The transformation is also appreciably conservative of the original confidence intervals. These observations support the testing of hypothesis (3) as a surrogate for (1) in satisfying the requirements of our F -test. Further, they are supportive of the meaningfulness of the idea of comparing the untransformed values of the ICCs for the lower grade cohort with a view to finding preliminary evidence for the existence of a horn effect. Notice in particular that for this cohort the ICC for consistency between first examiner continuous performance and report marks (.72) is a great deal higher than that for consistency between first examiner continuous performance and second examiner report marks (.42). This discrepancy in ICCs is of the same order of magnitude as that for the corresponding ICCs for each of the complete cohort and high grade cohort prior to and subsequent to transformation, suggesting the possibility of a strong horn effect.

On moving from part a) to b) and from part c) to d) of Figure 1, an increase in the visual spread of the data about the 45° line is recognizable in each case, indicating a tendency for greater agreement in dispersion from the mean when comparing supervisor continuous performance and report marks than when comparing supervisor continuous performance and second marker report marks. These findings are consistent with those which would be expected on examination of the corresponding raw and transformed ICCs in Table 1. However, the more rigorous analysis afforded by hypothesis testing serves to provide a more precise indication of the level of supervisor bias suggested by these discrepancies. For example, given the large sample sizes considered in each case, differences in spread between corresponding figures are masked through overlapping of multiple points and it is difficult to overcome this effect, even through jittering.

Our use of ICCs and the corresponding graphical representation of varying levels of relative agreement illustrated in Figure 1 ought to be distinguished from efforts based on the Pearson Correlation Coefficient to establish the level of conformity of data to merely any straight line. It is already recognized (Streiner & Norman 1985) that the latter coefficient has a tendency to inflate true agreement levels.

Our findings suggest that the report mark allocated by SSC4 supervisors is not purely based on written performance but that prior knowledge of continuous performance has a highly significant role to play. They are also supportive of the more general view that the provision of detailed descriptors does not suffice to remove examiner bias. As has been observed elsewhere (Eric et al. 1998), in improving the reliability of assessment procedures, there is the additional challenge of the successful training of examiners in the use of these descriptors.

Thus, whilst it has been traditionally assumed (Evans 2002) that the provision of detailed objective descriptors counters examiner bias, much more may need to be done to make this type of intervention sufficiently effective.

We acknowledge that in being blinded to student continuous performance, second examiners may be restricted in terms of their knowledge of the subject matter of the student projects they are marking. This could have led to an attenuation of the values of the transformed and untransformed ICCs which we used to represent natural consistency in this study and an inflation of the corresponding measures of halo and horn effects. Nevertheless, given that second markers are usually selected for their expertise in the field of study covered by the project reports which they mark, we assume here that the above confounding effect on level of supervisor bias is minimal.

Future research

It is very clear from this study and from our ongoing work with assessment data that there is a tendency for continuous performance marks for SSC4 students to be heavily skewed towards those representative of high grades. Such behaviour in assessment data is not unique to SSC4 data (see, for example, Dennis 2007 and Phelps et al. 1986). Additionally, whilst we have benefited from the availability of assessment data on a continuous scale and a successful Normalizing transformation as a means of ensuring that the assumptions of our hypothesis test have been satisfied, these conditions are not guaranteed within the context of the analysis of assessment data in general. In particular, successful Normalizing transformations may not be forthcoming or studies may be limited to the consideration of Likert scale data.

In future work, therefore, we anticipate using bootstrap sampling on existing educational data to assess the robustness of our hypothesis test to Type I errors following departures from Normality, in a manner akin to Hsu and Feldt (1969). This work would prove particularly valuable where the intention is to find a reliable procedure for testing for a horn effect with smaller sample sizes. Furthermore, such testing should be extended to the consideration of Likert scale data. Investigations of these types would have applications not only within the context of the current study but wherever it is of interest to quantify agreement for non-parametric assessment data.

Bibliography

1. Alsawalmeh, Y.M. and Feldt, L.S. **Testing the Equality of Two Related Intraclass Reliability Coefficients**, *Applied Psychological Measurement* 18, no. 2, 1994, pp. 183-190
2. Anastasi, A. **Psychological Testing**, 6th edn, New York, MacMillan Publishing Company, 1998
3. Arnold, E. and Pulich, M. **Personality Conflicts and Objectivity in Appraising Performance**, *Health Care Manager* 22, no. 3, 2003, pp. 227-232
4. Bowden, A.O., Caldwell, F.F. and West, G.A. **Halo Prestige**, *Journal of Abnormal and Social Psychology* 28, no. 4, 1934, pp. 400-406
5. Brown, E.M. **Influence of Training, Method and Relationship on the Halo Effect**, *Journal of Applied Psychology* 52, no. 3, 1965, pp. 195-199
6. Dennis, I. **Halo Effects in Grading Student Projects**, *Journal of Applied Psychology* 92, no. 4, 2007, pp. 1169-1176
7. Dennis, I., Newstead, S.E. and Wright, D.E. **A New Approach to Exploring Biases in Educational Assessment**, *British Journal of Psychology* 87, 1996, pp. 515-534

8. Dudycha, G.J. **A Note on the "Halo Effect" in Ratings**, *Journal of Social Psychology; Political, Racial and Differential Psychology* 15, Short Articles and Notes, 1942, pp. 331-333
9. Evans, A.W. **Facing the Challenges of Competency-Based Assessment of Postgraduate Dental Training**, *Medical Education* 36, Letters to the Editor, 2002, p. 586
10. Fagot, R.F. **A Generalized Family of Coefficients of Relational Agreement for Numerical Scales**, *Psychometrika* 58, no. 2, 1993, pp. 357-370
11. Fairweather, J.S. **Reputational Quality of Academic Programs: The Institutional Halo**, *Research in Higher Education* 28, no. 4, 1988, pp. 345-355
12. Fisicaro, S.A. and Lance, C.E. **Implications of Three Causal Models for the Measurement of Halo Error**, *Applied Psychological Measurement* 14, 1990, pp. 419-429
13. Holomboc, E.S. and Hawkins, R.E. **Methods for Evaluating the Clinical Competence of Residents in Internal Medicine: A Review**, *Annals of Internal Medicine* 129, no. 1, 1998, pp. 42-48
14. Hsu, T.-C. and Feldt, L.S. **The Effect of Limitations on the Number of Criterion Score Values on the Significance Level of the F-test**, *American Educational Research Journal* 6, no. 4, 1969, pp. 515-527
15. Marshall, J.H. **Grade Repetition in Honduran Primary Schools**, *International Journal of Educational Development* 23, 2003, pp. 591-605
16. McGraw, K.O. and Wong, S.P. **Forming Inferences about Some Intraclass Correlation Coefficients**, *Psychological Methods* 1, no. 1, 1996, pp. 30-46
17. McKinstry, B.H., Cameron, H.S., Elton, R.A. and Riley, S.C. **Leniency and Halo Effects in Marking Undergraduate Short Research Projects**, *BMC Medical Education* 4, no. 28, 2004, pp. 1-5
18. Phelps, L.A., Schmitz, C.D. and Boatright, B. **The Effects of Halo and Leniency on Cooperating Teacher Reports Using Likert-Type Rating Scales**, *Journal of Educational Research* 79, no. 3, 1986, pp. 151-154
19. Pike, G.R. **The Relationship between Perceived Learning and Satisfaction with College: An Alternative View**, *Research in Higher Education* 34, no. 1, 1993, pp. 23-40
20. Pike, G.R. **The Constant Error of the Halo in Educational Outcomes Research**, *Research in Higher Education* 40, no. 1, 1999, pp. 61-86
21. Pulakos, E.D., Schmidt, N. and Ostroff, C. **A Warning about the Use of a Standard Deviation Across Dimensions Within Rates to Measure Halo**, *Journal of Applied Psychology* 71, no. 1, 1986, pp. 29-32
22. Ram, P., Grol, R., Rethans, J.J., Schouten, B., van der Vleuten, G. and Kester, A. **Assessment of General Practitioners by Video Observation of Communicative and Medical Performance in Daily Practice: Issues of Validity, Reliability and Feasibility**, *Medical Education* 33, 1993, pp. 447-454
23. Rubin, S. **Performance Appraisal: a Guide to Better Supervisor Evaluation Processes**, Panel Resource Paper No. 7, in "National Society for Internships and Experiential Education" Washington, DC US, 1982
24. Satterwaite, F.E. **Synthesis of Variance**, *Psychometrika*, 6, 1941, pp. 309-316
25. Spike, N., Alexander, H., Elliot, S., Hazlett, C., Kilminster, S., Prideaux, D. and Roberts, T. **In Training Assessment - Its Potential in Enhancing Clinical Teaching**, *Medical Education* 34, 2000, pp. 858-861
26. Stine, W.W. **Influence of Training, Method and Relationship on the Halo Effect**, *Journal of Applied Psychology* 52, no. 3, 1965, pp. 195-199
27. Streiner, D.L. and Norman, G.R. **Health Measurement Scales: A Practical Guide to Their Development and Use**, 3rd edn., Oxford University Press, 2003.
28. Thorndike, E.L. **A Constant Error in Psychological Ratings**, *Journal of Applied Psychology* 4, 1920, pp. 25-29

29. Wakeford, R., Southgate, L. and Wass, V. **Improving Oral Examinations: Selecting, Training, and Monitoring Examiners for the MRCGP**, *BMJ*, 311, Education and Debate, 1995, pp. 931-935
30. Zegers, F.E. and ten Berge, J.M.F. **A Family of Association Coefficients for Metric Scales**, *Pyschometrika* 50, no. 1, 1985, pp. 17-24

^a **Acknowledgements**

We express our gratitude to Ronnie Grant, Senior Graphic and Web Designer, University of Edinburgh for his dedication, expertise and patience in assisting with the preparation of the figures for this paper.

^b **Dr Margaret MacDougall, BSc (Hons), PGCE, PhD, FRSS University of Edinburgh**

Personal web page: <http://www.chs.med.ed.ac.uk/people/staffProfile.php?profile=mmacdoug>

Margaret MacDougall is the statistician at the University of Edinburgh responsible for providing advice on statistical design and analysis to undergraduate medical students involved in 4th year Student Selected Component (SSC4) projects. In 2004, she helped design the 3rd year EBM station for Edinburgh's first OSCA (Online System for Clinical Assessment). She is a tutor on the international Msc/Dip/Cert Course in Pain Management, which is run by the University of Edinburgh in partnership with the University of Sydney, Australia.

At an external level, she served for 2 consecutive years (2005 – 2006) as Module Organizer and Lecturer for the Statistics and Clinical Epidemiology module of the Royal College of Physicians of Edinburgh Scottish Oncology Course for SpRs. Her principal area of research is Medical Education. She has been awarded three consecutive one-year grants from the University of Edinburgh Principal's e-learning Fund.

Recently, she was awarded funding by the Maths, Stats & OR Network to lead a project entitled 'Statistics in medicine: a risky business?' and funding by the MEDEV subject centre of the Higher Education Academy to lead a project entitled 'Research-based SSCs: a pragmatic approach to advancing the research-teaching nexus' (http://www.medev.ac.uk/resources/funded_projects/show_mini_project?reference_number=549).

^c **Dr Simon C Riley, BSc (Hons), PhD, FHEA, MRCOG, University of Edinburgh**

Personal web page:

<http://www.rds.mvm.ed.ac.uk/RDS%20Team%20Pages/Simon%20Riley%20Main%20and%20Linked%20Page/Simon%20Riley%20Linked%20Page.html#current>

Simon Riley is a Senior Lecturer in Obstetrics and Gynaecology. He is also Programme Director for the undergraduate self-selected student projects, known as Student Selected Components (SSCs), within the undergraduate medical curriculum at the University of Edinburgh.

He has been involved in the development of SSCs since their inception by the General Medical Council in 1993, when they were previously referred to as Special Study Modules (SSMs). He is Convenor for the SSC Directors Liaison Sub-Group of the Scottish Teaching Deans' Scottish Doctor Medical Curriculum Group (<http://www.scottishdoctor.org>) and External Examiner for Student SSCs for the Undergraduate programme in Medicine at Queen's University, Belfast. He is also a Fellow of the Higher Education Academy.

His research interests in Medical Education include the effective assessment of professional and generic skills and attitudes of students performing self-selected research projects across a wide range of specialties within medicine. He is also interested in how well medical students use the optional components of their curriculum to optimize their own personal professional development in terms of research and professional skills and as a means of gaining a better insight into future career options. His research activities also involve collaboration as a Project Partner in the Higher Education Academy funded project 'Research-based SSCs: a pragmatic approach to advancing the research-teaching nexus' (http://www.medev.ac.uk/resources/funded_projects/show_mini_project?reference_number=549).

^d **Dr Helen S Cameron BSc (Med Sci) MBChB MRCP FHEA**

Helen Cameron has a clinical background but since 1994, the focus of her work has been in medical education. Since 2003, she has been Director of the Medical Teaching Organisation (MTO), the centre for educational innovation, development and research in the Edinburgh Medical School.

Her principal academic interest is the investigation of assessments to inform the development of the assessment strategy for the undergraduate medical curriculum. Key emphases include maximising the validity and reliability of assessment and developing students' learning through good feedback and self-appraisal. Her other interests within medical education include developing a portfolio approach to learning and assessment within the undergraduate curriculum, and exploring the use of technology to develop and adapt teaching and assessment to the imperatives of today's healthcare. The latter has involved major contributions to online teaching materials and an online, media-rich system for clinical assessment.

She has published and presented papers across a range of topics in medical education, including *supervisor leniency in assessment, comparison of methods for standard-setting, development and assessment of clinical skills, portfolio learning and assessment, use of patient recordings and images in medical education, peer-assisted learning, and learning outcomes in medical education*.

She is a member of the Scottish Deans Medical Education Group, which collaborates on projects to develop medical education across Scotland, and is a Foundation member of the Academy of Medical Educators. She is also a Fellow of the Higher Education Academy. As a member of the Steering Group of the Tuning Workforce for the EU funded MEDINE Network (Medical Education in Europe) she helped develop the project which produced a consensus on the

learning outcomes for a medical graduate in Europe. She is also reviewer for the academic journals *Medical Teacher* and *Medical Education*.

° **Dr Brian McKinstry, MBChB, MRCPUK, MRCP (Dist), MD, FRCGP, FRCPE, University of Edinburgh**

Personal web page: <http://www.chs.med.ed.ac.uk/people/staffProfile.php?profile=bmckinst>

Brian McKinstry is Reader in General Practice at the University of Edinburgh a General Practitioner and holds a Fellowship from the Chief Scientist Office of the Scottish government. His main interests are in tele-health and medical education.

¹ This notation is derived from Table 4 of McGraw and Wong 1996, where a two-way mixed effects model is specified as Case 3 and the additional '1' was introduced in order to distinguish from the case where agreement is being assessed for ratings taken as averages over c independent ratings ($c > 1$). The formula for $ICC(3,1)$ is also available from the last row of this table as a special case of formulae, $ICC(C,1)$, for ICCs based on consistency between measurements.

A THREE WAY ANALYSIS OF THE ACADEMIC CAPITAL OF A ROMANIAN UNIVERSITY¹

Mihaela Cornelia PREJMEREAN²

PhD, Assistant Professor
UNESCO Department for Business Administration,
University of Economics, Bucharest, Romania

E-mail: mihaela.prejmerean@fabiz.ase.ro



Simona VASILACHE³

PhD Candidate, University Assistant
UNESCO Department for Business Administration,
University of Economics, Bucharest, Romania

E-mail: simona.vasilache@fabiz.ase.ro



Abstract: *The paper applies three-way analysis (Kroonenberg, 1982, 2008) to the components of academic capital of a Romanian university, over a five-year period, showing the biases and the relations between the various components. The influences from inside the academia are being discussed, together with analyzing their positive or not so positive effect on the variables related to academic capital. The model of intellectual capital, thoroughly discussed in literature, is being adapted to this particular situation, of a university in the situation of "leasing" its academic capital for the sake of academic capitalism. The tensions between the two concepts are underlined, in a framework which, being applicable to every university, relies on data collected from a Romanian economic university.*

Key words: *academic capital; three-way analysis; university intellectual capital*

1. The concept of academic capital

If we speak about academic capitalism (Deem, 2001; Slaughter and Leslie, 1997), then we need to speak also about academic capital. At first sight, the two concepts are complementary, if not opposed. The academic capital includes academic freedom and collegiality (Kinman and Jones, 2004), defined as consensual decision-making, by means of cooperation and value-sharing. The *Lehrfreiheit* and the *Lernfreiheit* of the Humboldtian, collegial model, which guaranteed the freedom of teaching and the freedom of studying are, according to Cobban (1975), the persistent hallmark of the European idea of university. Paradoxically, as the university suits the free-market paradigm, its freedom is progressively lost (Solly, 1996), because it has to become accountable to its stakeholders (Tapper and Salter, 1995). The strength gained by non-academic criteria which prevail in university decision-making leads to managerialism (Harvey, 1995), the power of the experts, which may turn the university into a more dynamic, ready to react structure, but may also carry the

risk of not applying properly business principles to non-business entities (Michael, Sower and Motwaki, 1997). If we add to this the critical nature of the academics (Davies, Douglas and Douglas, 2007), which prevents them for readily accepting exterior standards, that violate the collegial model, we have the picture of the conflict between accountability and autonomy (Bridgman, 2007).

Still, according to Hanna (1989), universities have to maintain both an internal and an external image. Caught between the internal need to cut down costs and the external need to build a strong reputation, universities experience multiple pressures (Dickenson, 2003). The academic capital they transfer to their students becomes, then, an umbrella concept, including the organizational culture, particular to universities, and the indicators which make the university accountable in the eyes of its stakeholders.

We can, then, define academic capital as a transformed instance of intellectual capital, taking into account the specificities of the academic climate. Following Leal (1991), the identification and management of the intellectual capital leads to a sense of cohesion in the organizational culture, thus responding to the collegiality paradigm. On the other hand, the benchmarking and scorecarding opportunities that the intellectual capital measurement offers (Martins Rodriquez and Viedma Martí, 2006; Kaplan and Norton, 1996, 2001) contribute to rendering the university more accountable.

Back in 1895, List speaks of mental capital, which forms a "hidden" part of the economy. In 1969, Galbraith (in Bontis, 1998), defines intellectual capital as intellectual action, pointing at its dynamic nature. In the forty years which followed, researchers of the field provided around fifty definitions and ways of systematizing this concept whose final purpose is to prove meaningful and useful for enhancing organizational performance. One of the classics of the domain, Steward (1998) equals intellectual capital with intellectual material, consisting of knowledge, information and intellectual property, which, altogether, create organizational wealth. In the same period, Edvinsson and Malone (1997) see intellectual capital as "the possession of knowledge, applied experience, organizational technology, customer relationships and professional skills that provide the firm with a competitive edge in the market" (p.44). These definitions give a systematic view on intellectual capital, as a dynamically structured macro-asset of the organization, which has an internal, an external and an interface component.

In the works of Saint-Onge (1996), Roos and Roos (1997), Sveiby (1997), Smith and Parr (2000), Sullivan (2000) these components are, with little variation from researcher to researcher, human capital, structural (organizational) capital, and customer (relational) capital. In fact, intellectual capital is constructed by integrating the flows of knowledge circulating between these three compartments. Mouritsen and Larsen (2005) regard intellectual capital management as a second wave of knowledge management, after the first one, which supposed that knowledge is embodied in individuals, while Viedma Martí (2001) postulates that intellectual capital management implies a strategic and global perspective, while knowledge management takes a tactical or operational perspective of the same transfer processes.

In our model of academic capital, adapted from the proposed models of intellectual capital, we define three compartments: human capital, relational capital, and process capital. The human capital is considered in a dynamic perspective, taking into account both the existing human capital and the inflows and outflows of human capital to and from the academic organization. The structural capital is replaced by process capital, because

universities are loosely coupled systems, which are at the same time “open and closed, indeterminate and rational, spontaneous and deliberate” (Orton and Weick, 1990, p.4; Fusarelli, 2002), having fuzzy structures which depend more on the relational web entering the knowledge transmission process, than on the organizational routines and procedures (Elkin, Farnsworth and Templer, 2008).

By considering the modifications occurring over time in the human capital of the university – practically, although tenure was long regarded as the extreme form of career stability, new career literature tends to undermine this myth (Wicks, 2004), and the university human capital is continuously moving –, the process instead of the structure and the complex web of relationships surrounding the university and proliferating, also, inside, due to the numerous and diverse university stakeholders (Neave, 2002), we place academic capital in a dynamic perspective and we are able to identify not merely the indicators related to this capital, but the trends which are significant for the university management, in the sense of being able to forecast its future performance. The link between intellectual capital and performance being already proved (Bontis and Fitz-enz, 2002; Choo and Bontis, 2002; Sveiby, 1997; Kaplan and Norton, 1996, 2001), we take the indicators we propose for academic capital as representative for evaluating the overall efficiency of the academic activities.

2. Methodology

We studied the intellectual capital indicators in the University of Economics, Bucharest, on a period including the academic years 2003/2004, 2004/2005, 2005/2006, 2006/2007 and 2007/2008. The sources of data were the 2004-2008 Report of the Senate of the Academy, secondary data from the Economic Research Department, and survey results synthesized in the intermediary report, on 2007, of the CEREX research project (authors: Al. Isaic-Maniu, C. Bratianu, C. Herteliu, A. Dima, S. Vasilache, I. Jianu). The system of indicators used is presented in Table 1 below:

Table 1. Academic capital indicators

ACADEMIC CAPITAL COMPONENTS		
HUMAN CAPITAL	RELATIONAL CAPITAL	PROCESS CAPITAL
H ₁ : staff to student ratio	R ₁ : number of academic exchanges (Fulbright, AUF, etc.)	P ₁ : number of taught disciplines
H ₂ : drop-out rate in PhD programmes	R ₂ : number of incoming research visits	P ₂ : number of specialties
H ₃ : Rookie ratio (staff with less than two years in the organization over total staff)	R ₃ : number of outgoing research visits	P ₃ : number of research centers
H ₄ : retention rate (former students choosing an academic career)	R ₄ : co-tutorship PhDs	P ₄ : number of doctoral domains
H ₅ : staff turnover (staff leaves over staff recruitments)	R ₅ : external partnerships (other universities and business environment)	P ₅ : number of research projects.

We recorded, in a SPSS database, the yearly variation of these indicators, on a 1 to 5 Likert scale, where:
 1 = decrease; 2 = slight decrease (less than 20%); 3 = unchanged; 4 = slight increase (more than 20%); 5 = increase.

Then, we imported the Fixed ASCII database in the 3WayPack programme, and we performed a three way analysis in the following system of coordinates:

- Mode 1: scales (5)
- Mode 2: years (5)
- Mode 3: components (3).

The results were presented per slices of cases, according to the distribution of components per scales and per monitored years.

3. Results

The goodness of fit for the analysis was of .76. For the human capital indicators the Cronbach alpha value was of .56, for the relational capital indicators of .71, and for the process capital indicators of .66. Between H_2 , drop-out rate in PhD programs, and H_3 , Rookie ratio, there is a correlation of -.48, which means that, at least partly, the students graduating from a PhD program become assistant professors, and a high drop-out rate reduces the proportion of recently recruited staff. Between H_3 , Rookie ratio, and H_5 , staff turnover, there is a correlation of -.25, which proves that not only recently hired and presumably young staff leaves the system. Between R_3 , outgoing research visits, and R_5 , external partnerships, there is a .219 correlation, signifying that, in part, the contacts with foreign universities result in research projects with those universities. Between P_4 , number of doctoral domains, and P_5 , number of research grants, there is a .495 correlation, which shows a significant correspondence between doctoral specializations and research interests.

The distribution of human capital as compared to relational capital, across scales and years, is presented in Figure 1:

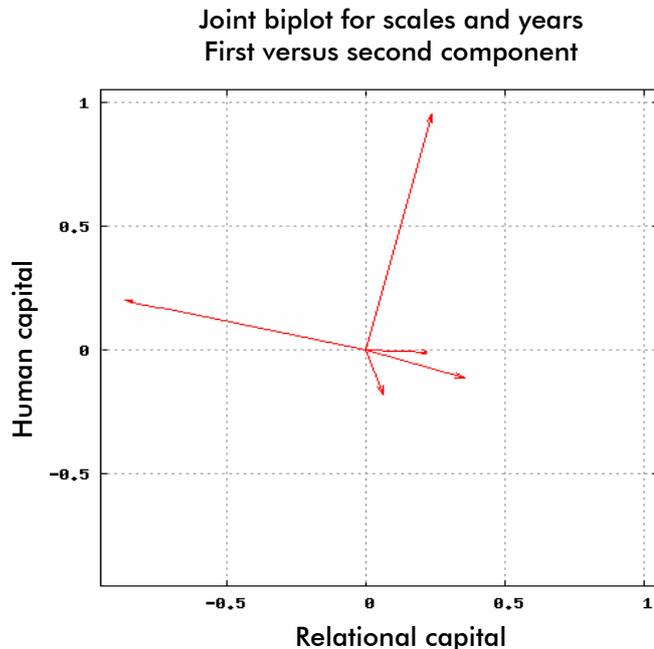


Figure 1. The distribution of human and relational capital across years and scales

The first year we monitored exhibits more the influence of the human capital indicators, while the second year shows a sudden growth in relational indicators (due to the significant increase in research visits of the staff). Afterwards, the next years are still predominantly relational, but with visibly lower scores. This decrease in relational capital, after an initial boom, may be explained by the appearance of selection mechanisms, of filters which limit research visits, academic exchanges, external partnerships, etc. Some of these filters are intrinsic, and take into account the quality of the relationships and the outcomes they might bring, some others are extrinsic, institutional, referring to the number of approved research visits per academic year, to the value of the external research contracts, etc. While the first filters are beneficial, the second category artificially decreases the relational capital, which has indirect effects on the variation of the human capital indicators.

Figure 2 shows the distribution of human and process capital:

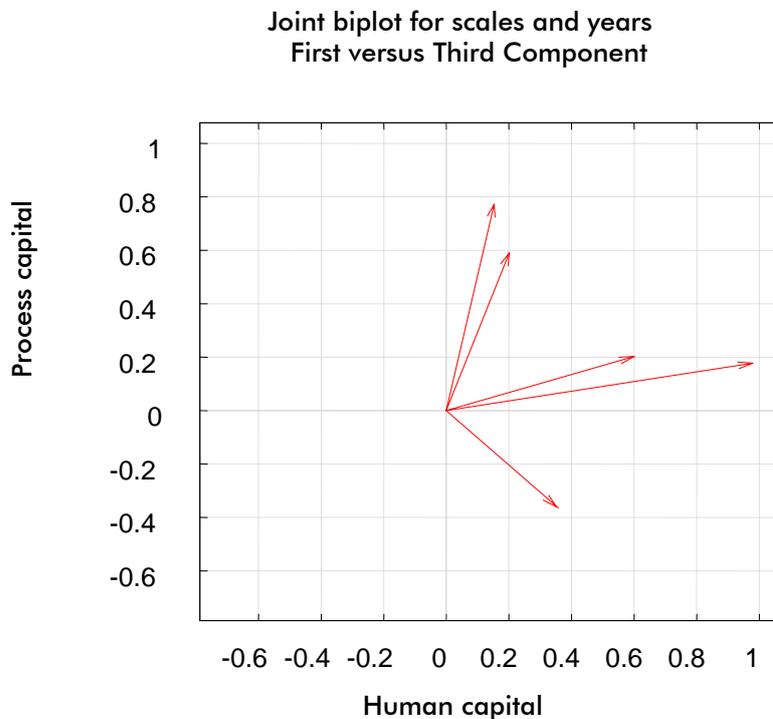


Figure 2. The distribution of human and process capital across years and scales

The competition between human and process capital is neatly favourable to the former, in all investigated years, especially in the last three. Whereas human capital was thought to influence structural capital (Bontis, 1998; Sveiby, 1997), it can be seen that in the university considered, the equilibrium is pulled to the human capital, which is the main asset behind research centers, doctoral programs, scholarly disciplines, etc, disputed by both the research and teaching processes. The development of the human capital with little or no focus on structural capital bears a risk for the university, the risk that the human capital is not particularly trained and co-interested in specific programmes, which have the capability

to absorb and to structure what the human resources can offer and, thus, promote a sense of institutional stability, no matter how dynamic and fluid the human capital is.

Figure 3 relates relational and process capital:

Joint biplot for scales and years
Second versus Third Component

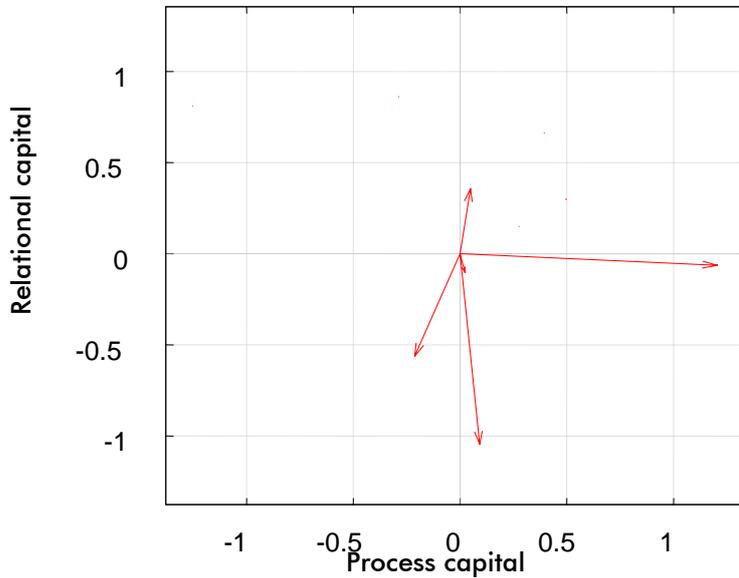


Figure 3. Relational and process capital distribution across scales and years

The relation between process and relational capital is more equilibrated than in the case of the previous categories, showing that the grow at approximately the same pace. The number of specialties, of taught disciplines, and of doctoral programs depends, ultimately, on the degree of openness of the university, on its relationships with external partners, either universities or business entities. External partnerships provide opportunities for doctoral research to gain depth, while research visits and academic exchanges create the normal tendency to equalize curricula and, thus, to modify the number of disciplines, specialties, etc.

Finally, in the three-dimensional plot presented in Figure 4, we brought together the three considered components, taking into account multi-annual averages:

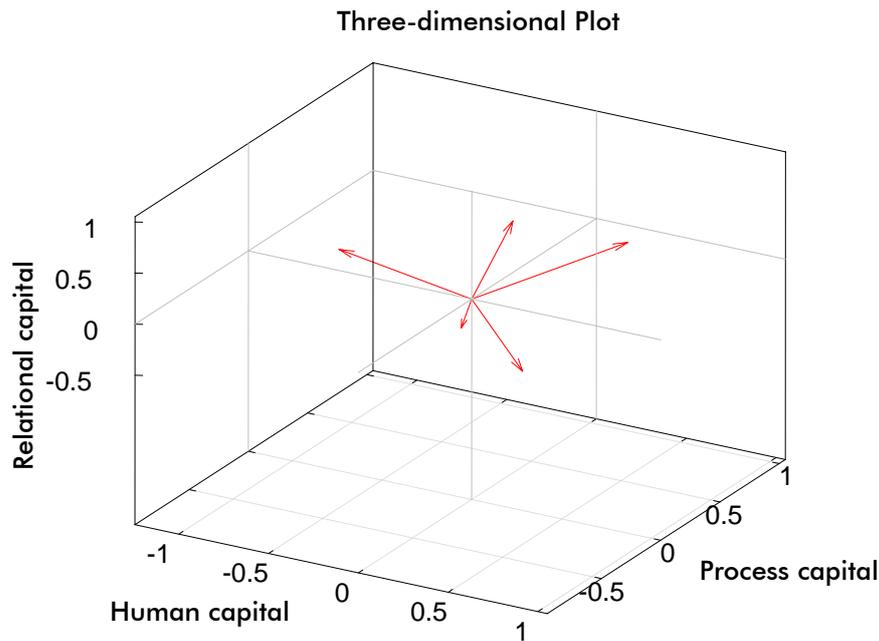


Figure 4. The distribution of human, relational and process capital

As it can be seen, considering the five scales, the human capital is underrepresented in the considered university, while process capital seems to be in progress, as the number of taught disciplines increases and so does the number of research grants and of research centers. Still, if this increase in process capital is not paralleled by a corresponding increase in human capital, which means that the same human capital will have to maintain more processes, this will result in organizational wear and tear and dramatic decrease of the quality of the human capital.

4. Conclusions

The study reveals that the three components of academic capital, which we adapted after the classical constituents of intellectual capital, referred to in literature, are asymmetrically distributed. While studies have been made on the influence of human capital on structural and relational capital, our plots show that there are biases towards one or other of the components, some of them explainable by means of internal regulations or changes in the way each of the components is regarded. Some components are underrepresented, and the apparent compensation in other components is illusory, since each of the three components needs, for its proper functioning, the collaboration with the other two. For sure, the short period, of only five academic years, on which these indicators were monitored, may introduce significant seasonal variations – if we take into account the revisions of the sets of criteria for staff recruitment, staff promotion, research activities recognition, etc. occurring during this period, we may have the environmental explanations for most of the disequilibria. Still, these biases have to be followed over larger intervals and ways to eliminate them and to correlate the evolution paces of the three interrelated components are to be sought for. If this doesn't happen, the danger of converting academic capital into academic capitalism, interested only in indicators, without taking a close look to the processes underlining them, is obvious.

References

1. Bontis, N. **Intellectual capital: an exploratory study that develops measures and models**, *Management Decision*, 2, 1998, pp. 63-76
2. Bontis, N. and Fitz-enz, J. **Intellectual capital ROI: a causal map of human capital antecedents and consequences**, *Journal of Intellectual Capital*, 3, 2002, pp. 223-247
3. Bridgman, T. **Freedom and autonomy in the university enterprise**, *Journal of Organizational Change Management*, 4, 2007, pp. 478-490
4. Choo, C.W. and Bontis, N. **The Strategic Management of Intellectual Capital and Organizational Knowledge**, Oxford University Press, New York, 2002
5. Cobban, A.B. **The Medieval Universities: Their Development & Organization**, Methuen Co, Ltd, London, 1975.
6. Davies, J., Douglas, A. and Douglas, J. **The effect of academic culture on the implementation of the EFQM Excellence Model in the UK Universities**, *Quality Assurance in Education*, 4, 2007, pp. 382-401
7. Deem, R. **Globalization, new managerialism, academic capitalism and entrepreneurship in universities: is the local dimension still important?**, *Comparative Education*, 1, 2001, pp. 7-20
8. Dickenson, R.C. **Foreword to unintended consequences of tuition discounting**, LUMINA, Foundation for Educators, New Agenda Series, 1, 2003
9. Elkin, G., Farnsworth, J. and Templer, A. **Strategy and internationalization of universities**, *Journal of Educational Management*, 3, 2008, pp. 239-250
10. Edvinsson, L. and Malone, S.M. **Intellectual Capital: Realizing Your Firm's True Value by Finding its Hidden Brainpower**, Harper Collins Publishers, New York, 1997
11. Fusarelli, L.D. **Tightly coupled policy in loosely coupled systems: institutional capacity and organizational change**, *Journal of Educational Administration*, 6, 2002, pp. 561-575
12. Hanna, D.E. **Planning programs to enhance institutional image**, in Simerly, R.E. (ed.) *Handbook of Marketing Continuing Education*, pp. 89-102, Hossey Bass, San Francisco, 1989
13. Harvey, L. **Quality Assurance Systems, TQM and the New Collegialism**, QHE, Birmingham, 1995
14. Kaplan, R. and Norton, D. **Using the balanced scorecard as a strategic management system**, *Harvard Business Review on Measuring Corporate Performance*, pp. 183-211, HBS Press, Boston, 1996
15. Kaplan, R. and Norton, D. **The Strategy Focused Organization**, HBS Press, Boston, 2001
16. Kinman, G. and Jones, F. **Working to the limit: stress and work-life balance in academic and academic-related employees in the UK**, Association of University Teachers, London, 2004
17. Kroonenberg, P. M. **Annotated bibliography of three-mode factor analysis**, *British Journal of Mathematical and Statistical Psychology*, 36, 1983, pp. 81-113
18. Kroonenberg, P.M. **Applied three-way data analysis**, Willey, New York, 2008
19. Leal, A. **Cultura organizativa y orientación al Mercado: un análisis multisectorial en pymes**, paper presented at XIV Congreso Acede, Murcia, 1991
20. List, F. **The National System of Political Economy**, 1895 [translated in English by S.S. Lloyd, Longmans, 1904]
21. Martins Rodriguez, B. and Viedma Martí, J.M. **The region's intellectual capital benchmarking system: enabling economic growth through evaluation**, *Journal of Knowledge Management*, 5, 2006, pp. 41-54

22. Michael, R.K., Sower, V.E. and Motwaki, J. **A comprehensive model for implementing total quality management in higher education**, *Benchmarking for Quality Management & Technology*, 2, 1997, pp.104-120
23. Mouritsen, J. and Larsen, T.H. **The 2nd wave of knowledge management: the management control of knowledge resources through intellectual capital information**, *Management Accounting Research*, 3, 2005, pp. 371-394
24. Neave, G. **Anything Goes: Or: How the Accommodation of Europe's Universities to European Integration Integrates an Inspiring Number of Contradictions**, *Tertiary Education and Management*, 3, 2002, pp. 178-191
25. Orton, J.D. and Weick, K.E. **Loosely coupled systems: a reconceptualization**, *Academy of Management Review*, 2, 1990, pp. 203-223
26. Roos, G. and Roos, J. **Measuring your company's intellectual performance**, *Long Range Planning*, 3, 1997, pp. 413-426
27. Saint-Onge, H. **Tacit knowledge: the key to the strategic alignment of intellectual capital**, *Strategy & Leadership*, 2, 1996, pp. 10-14
28. Slaughter, S. and Leslie, L.L. **Academic Capitalism: Politics, Policies and the Entrepreneurial University**, The Johns Hopkins University Press, Baltimore, 1997
29. Smith, G.V. and Parr, R.L. **Valuation of intellectual property and intangible assets**, 3rd edition, John Wiley & Sons, New York, 2000
30. Soley, L.C. **Leasing the Ivory Tower: The Corporate Take-Over of Academia**, South End Press, Boston, 1996
31. Steward, T.A. **The new wealth of organizations**, Currency, New York, 1998
32. Sullivan, P.H. **Value-driven intellectual capital: how to convert intangible corporate assets into market value**, Wiley, New York, 2000
33. Sveiby, K.E. **The new organizational wealth. Managing and measuring knowledge-based assets**, Berrett Koehler Publishers, San Francisco, 1997
34. Tapper, E. and Salter, B. **The changing idea of university autonomy**, *Studies in Higher Education*, 1, 1995, pp. 59-72
35. Viedma Martí, J.M. **ICBS – Intellectual capital benchmarking system**, *Journal of Intellectual Capital*, 2, 2001, pp.148-165
36. Wicks, D. **The institution of tenure: freedom or discipline?**, *Management Decision*, 5, 2004, pp. 619-627

¹ The authors would like to thank Professor Pieter M. Kroonenberg, Leiden University, The Netherlands, and the organizers of the Spring Seminar (March 2008, GESIS - ZA, Cologne), for the training in using three mode data analysis.

² **Co-author/ editor of the books:**

Prejmerean, M., Vasilache, S. *What's a university worth. Changes in the lifestyle and status of post-2000 European graduates*. Grin Verlag, Munich, Germany, 2008

Olaru, M., Prejmerean, M., Nita, I., Vasilache, S. (eds.) *Calitate-management-integrare europeana*, Editura ASE, Bucuresti, 2007

Author of the articles and chapters in books:

Vasilache, S., Prejmerean, M. Estimation of Romanian Hospitals Efficiency in Relation to Hospital Market Competition, *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 42, No.3-4, 2008

Prejmerean, M., Vasilache, S. Strategies for University Originated Organizational Intelligence, *University of Munich Paper Series*, 2007

Bogdan A., Prejmerean M.C., Vasilache S., Research Activity as an Indicator of Market Competitiveness. The Case of Post-adhesion Romanian Business Education, in Pieter Bielik (ed.), *The Path of Internationalisation and Integration in the Europe of Regions*, Slovak University of Agriculture Press, 2007, pp. 60-67

Bratianu, C., Prejmerean, M., Jianu, I., Vasilache, S. Competitiveness through University Intellectual Capital on European Education Market, in Pieter Bielik (ed.), *The Path of Internationalisation and Integration in the Europe of Regions*, Slovak University of Agriculture Press, 2007, pp. 60-67

³ **Co-author of the books:**

Bratianu, C., Curaj, A., Vasilache, S., Baltei, C.C. *Managementul cercetarii stiintifice universitare*, Editura Economica, Bucuresti, 2007

Bratianu, C., Jianu, I., Vasilache, S. *Business Management*, Editura ASE, Bucuresti, 2006

Author of the articles and chapters in books:

Bratianu, C., Jianu, I., Vasilache, S. Integrators for Organizational Intellectual Capital, *International Journal of Learning and Intellectual Capital*, 2008, Inderscience Publishers

Brătianu, C., Vasilache, S. Procese fundamentale în dezvoltarea inteligenței organizaționale, in *Management & Marketing*, 4, 2007.

Bratianu, C., Tantau, A., Vasilache, S., Jianu, I., Bogdan, A. Analiza capitalului intelectual de cercetare al ASE Bucuresti comparativ cu alte universitati, *Calitatea. Acces la succes*, Numar special, septembrie 2007, pp.384-392.

Bratianu, C., Vasilache, S., Prejmorean, M. Universitatea – organizatie care invata, *Calitatea. Acces la succes*, Numar special, septembrie 2007, pp.402-410

EVALUATION OF TRAINING PROGRAMS FOR RURAL DEVELOPMENT¹

A. INDIRA²

PhD, Associate Professor, Xavier Institute of Management, Bhubaneswar, India



E-mail: aindira1@rediffmail.com; indira@ximb.ac.in

Abstract: An Evaluation of the “Impact Assessment of the Training Programs” of a National Level Training Institution in India was conducted using the Kirkpatrick Method (KP Method). The studied Institution takes up research, provides training, offers consultancy and initiates action in the rural sector of India. The evaluation study used a detailed questionnaire for conducting a survey on the entire population of participants who attended the training programs in the selected study period. Personal interviews and workshops were also conducted with respondents to understand the behavioural changes and results seen in the work environment. The study brought out the need to understand the training evaluation as a continuous process, requiring periodic review and analysis of the needs of the various sectors of rural development. It reiterated the need to develop a systematic evaluation process within the institution. It also showed that the rural development professionals undergoing training themselves were keen to participate in the evaluation process so as to help in the process of self-learning and bring about sustainable changes.

Key words: Kirkpatrick Model; Survey; Rural Development; Evaluation; Training

1. Introduction

There is much ado about training as an important component of the overall capacity building of any employee in an organization. Training is a planned process to acquire knowledge, skill, to modify attitude and/or behaviour. Training is done using different kinds of techniques, which include Lecture-cum-Discussion, Case Discussions, Group Discussions, and Exercises/Hands on Sessions, and Field visits. It is seen that sharing of learning experiences during training helps achieve a more effective performance in an activity or range of activities among the trainees.

Efforts are made to evaluate training programs by both the training institution and the participating organizations, so that there can be an effective learning process. Where training is in-house the three most common reasons for evaluating the training program is: (1) to see how future programs can be improved; (2) to determine whether training should be continued and (3) to justify the existence of the training department itself. However most organizations continue to perceive training as a ‘philanthropic’ activity whose results is intangible in nature, and hence do not consider evaluation as a necessary activity. But over

the years, evaluation has become an important component as training requires enormous amount of investment in terms of human, financial and other resources.

Donald Kirkpatrick in 1959 formulated the four Levels of Evaluation. The four levels represent a sequence of steps to evaluate training programs. Each level is important and has an impact on the next level. As one moves from one level to the next, the process becomes more difficult and time consuming, but it also provides more valuable information.

Level 1: Evaluation at this level measures how those who participate in the program react to it and is taken immediately after the completion of the training program. It can also be called a '*measure of customer satisfaction*'. It is observed that a positive and favourable reaction from a few key persons in the group influences the future of a program. Infact a less than favourable reaction affects the motivation to learn among the participants. A negative reaction could greatly reduce chances of continuity for further programs.

Level 2: A simple standardized paper and pencil test is administered (same test) before and after the programs as part of the evaluation process. This helps in understanding the extent to which participants change attitudes, improve knowledge and /or increase skill as a result of attending the program.

Level 3: At this stage, it would be possible to assess the extent to which the behaviour is changed because of training and ideally assessed between six to nine months after the training is completed.

Level 4: At this level the final impact results are taken into consideration. The final results can be in the form of increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, decreased costs, increased sales, reduced turnover and higher profits. It would help if the final objectives of the training program can be stated in these terms, for seeing improvement in the long-term. This assessment can be made between one to three years after completion of training, because otherwise there is a danger of 'lack of recall'. It is also that changes cannot always be singularly identified with the training received.

The Kirkpatrick (KP) model has been used in different situations but predominantly in industrial settings, because, the final results are more quantifiable in nature both for the trainers and the trainees. This helps the authorities to take decisions about the continuation of training in a very proactive manner. Some of the companies which have used this evaluation method very rigorously include Motorola Corporation, University of Wisconsin, USA which evaluated the training for developing supervisory skills in the staff, and Intel Corporation which used the method to evaluate a corporation wide performance improvement system (Kirkpatrick 1998).

Increasingly, the KP model is being used by training organizations to understand the impact of the training programs even where the results are not very tangible in nature (Marcotte, Bakker-Dhaliwal and Bell, 2002).

The present paper is based on an evaluation study conducted for a training institution in India. It brings forth the method in which the training programs were evaluated following the four levels of the KP model. The studied institution over the years has emerged as a premier training and research institute in the field of rural development in India. Training rural development professionals is complex as it encompasses many areas of knowledge and myriad skills which are sought to be transferred through various methods. The composition of the participants is heterogeneous in nature. The studied Institution imparts training mainly to lower and middle level government officials working in the

Departments relating to rural development and related sectors, which include Water Supply and Sanitation, Health, Local Self Governments, Public Works Departments, Roads and Buildings, among others.. The personnel are nominated by the respective departments for training every year based on the proposed internal requirements.

The training institution went in for the evaluation study to assess the impact of the training programs and get a macro picture across all the training programs and all participants across the country. It covered a three year period for the survey. The main objectives of the study were:

- (i) to understand the impact of training on the participant's knowledge, skills and attitudes;
- (ii) to understand whether the training programs had affected participant's performance in the work environment back home and (iii) to understand the efficacy of institutionalizing an evaluation process using the KP model.

The paper is divided into three sections. The second section gives the methodology of the study based on the KP model, the third section discusses the survey results with respect to the four levels of the KP model and the last section gives the conclusions drawn from the study.

2. Methodology

A primary survey was done using a detailed questionnaire as a tool. The survey helped in establishing an understanding of all the four levels of evaluation – reaction, learning, changes and results. The survey used the entire population of participants who attended the training programs of the Institution over the selected three years. The institution on an average trained 3000 participants every year from across the country in its 100 training programs per year.

The questionnaire had three main parts –

- I. Personal details - to build the profile of the participants;
- II. 'Effectiveness of Program' was studied with key questions on whether the objectives of rural development were met within the program. The participants were asked to rate the program content and design on the basic inputs of knowledge, skills and attitudes.
- III. 'Professional relevance of training' was evaluated with key questions asking how relevant the program content was for meeting the local needs and whether there was enough practical application which could be used for working or transferring the knowledge to functionaries further down the line. It also probed whether the learning could be shared with other colleagues in the organization and lastly whether the course had helped in the organizational performance.

Database:

In the first instance, the database of 9000 participants was cleaned for missing names and incomplete addresses. The questionnaire was then posted to all the participants together with a stamped self-addressed envelope. Three reminders were also posted over a period of three months to the trainees who had not replied. Questionnaires were also posted to e-mail ids wherever available. The replies received were tabulated in the SPSS format and analysed.

The exercise with the database highlighted the weakness in the system where the names of all the participants of the programs, names of the resource faculty and feedback received were not maintained properly. This seemed to have happened because of a lack of a central system to maintain records of the various programs. The institution, recognizing this lacuna, had recently developed a new digital format to capture the information correctly. With this, a good electronic database would be in place which while giving a clear picture of all training programs, could be used as a ready input for any future evaluation.

The survey received a 16% response rate from a total population of 9000 participants. In the second phase of the study, personal interviews were also conducted to understand the perceptible change in knowledge, skills and attitudes after the training, for Levels 3 and 4.

Further the survey findings were shared in four regional workshops held in the north, east, west and southern parts of India respectively. A background paper was prepared with a regional focus, and circulated among all participants of the workshop. The workshops helped the study team to understand the impact of training programs and the future changes and interventions that could potentially improve the reach and relevance of training programs for rural development. The workshops also helped validate the findings of the survey and obtain suggestions for improving the quality of the training programs. Suggestions were also received about new ways for evaluating training programs in future.

3. Survey Results

Profile of the participants:

The survey had a response of 87% from men when compared to 13% from women. This result was clearly in conformity to the pattern of the database. Fewer women were trained as compared to men. The respondents were mostly in the age group of 41-50 years. The educational profile however showed a very interesting insight across the country, that the majority of the trainees were post-graduates.

Majority of the responses came from those participants who were trained in the most recent year of the study period. This confirmed the Kirkpatrick model, which states that the evaluations conducted within a shorter span of time after the actual training has been given, is more relevant. The rate of recall reduced as time passed by.

The survey showed that the majority of the participants trained were those having experience for about 10 years or more in the rural sector. Another very interesting fact emerged from the study that nearly 70% of those trained remained in the same position after receiving the training.

Level 1

In the studied institution, feedback was taken at the end of every program on important heads such as realization of program objectives, program content, and training methods on a 5-point scale. This is the Level 1 process of the KP model. The institution over the years had continually received a rating of over 84% at this level, which reflects the quality of performance of training activities carried out by the Institute.

The survey also sought to understand the rating of the programs. The participants were asked to rate the program content and design, teaching methodology and style, material used and distributed on a 3-point scale - (a) innovative and effective, (b) repetitive but effective, and (c) ineffective.

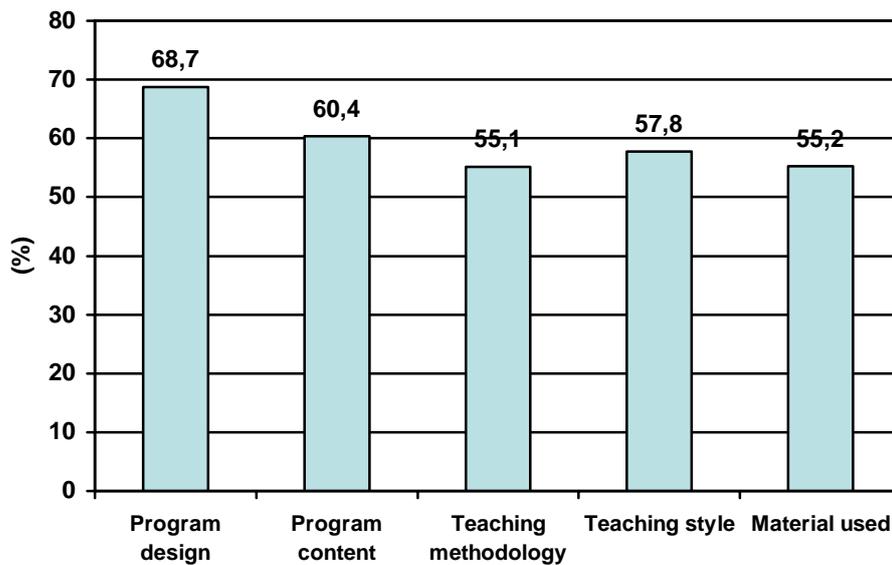


Figure 1. Rating of Programs – Innovative and Effective

The survey showed that nearly 55% of the respondents observed that overall the training components were innovative and effective. But when seen closely, program design stood first followed by program content, as shown in Figure 1. All the other three attributes, namely, teaching style and delivery, teaching methodology and the material used and distributed were ranked lower showing that there was much scope for improvement in these areas. 37% of the total respondents said that the programs were repetitive but effective and the rest said that they were ineffective.

The survey reaffirmed the initial reaction of the participants even after a time gap of three years. However the respondents mentioned that more could be learnt through interactions with the trainers. Clearly there was a need for “hands-on” knowledge of resolving issues on ground through local case studies. It was also important for the respondents that the lead coordinators or trainers were available throughout the period of the program.

Another important question under the “effectiveness of the program”, was the “extent to which objectives of rural development” were achieved in the various training programs. The respondents on a scale of 1 to 5 measured the objectives given below, with 5 standing for ‘completely’ to 1 standing for ‘minimally’.

- Understanding of concepts, definition, planning, and implementation of rural development programs;
- Awareness of personal and institutional role in Rural Development;
- Bringing about attitudinal change towards Rural Development;
- Develop new skills for a sustainable approach to Rural Development;
- Appraisal of rural development projects;
- Understanding institutional mechanisms;
- Analysis of strengths/ weaknesses of various anti poverty programmes;
- Role of Information Technology in Rural Development.

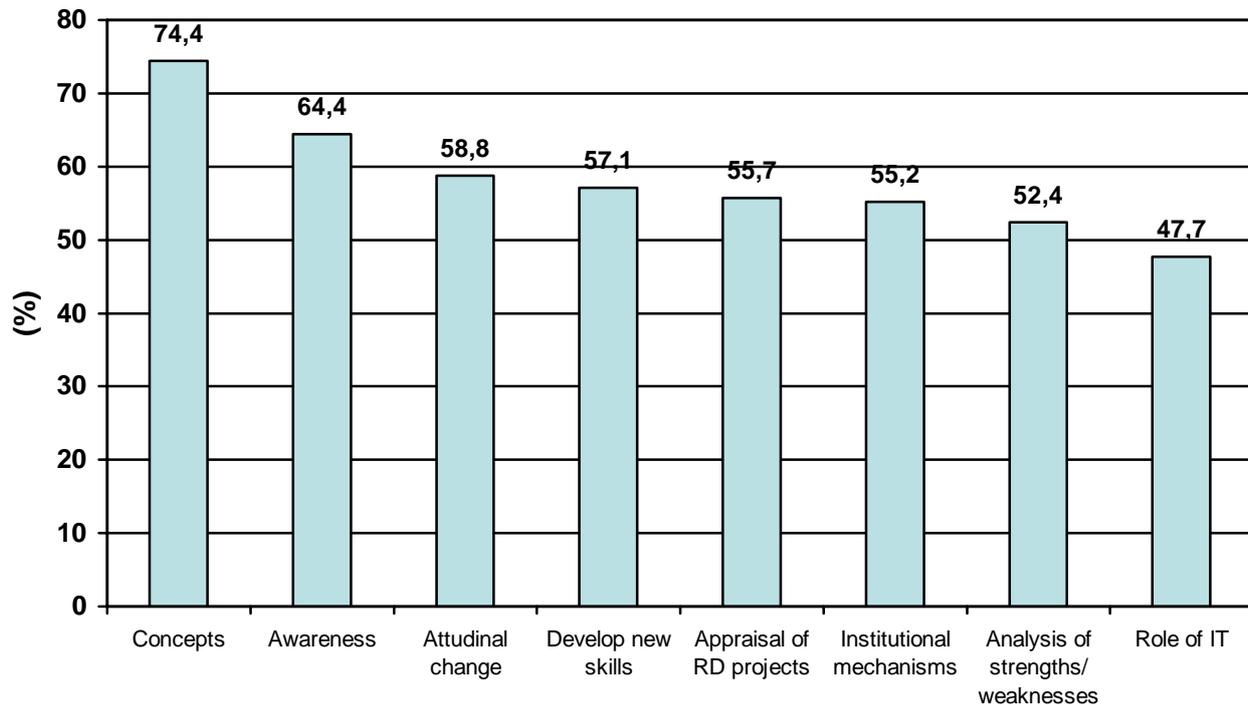


Figure 2. Objectives of Rural Development Achieved

For simpler analysis, the scales of 4 and 5 marked by the respondents against each of the objectives were taken together, and similarly 1 and 2 were clubbed to understand minimal achievement. The scores in Figure 2 shows the percentage of respondents who felt the respective objective was achieved on the five point scale.

The survey showed that the training programs were able to achieve a near sense of completeness of objectives (Figure 2) in terms of 'understanding of concepts'- as shared by 74% of the respondents across the various categories of trainees, followed by 'awareness of role in Rural Development' and 'attitudinal change towards Rural Development', clearly shown by the percentage scores. 'Development of new skills' clearly received a lower ranking.

Level 2:

Measurement of 'learning' at Level 2 requires administration of tests at the beginning and the end of training programs. This was not taken up in the study because: (1) the institution had not conducted such tests for the training programs being evaluated and (2) the tests could not be conducted as part of the survey as the training programs being evaluated were spread over a period of three years in the past. Level 2 measurement is useful where the training programs predominantly target conceptual understanding or transfer of skills. As one of the objectives of the study was to look at the need for a regular evaluative process, the need for a Level 2 assessment was discussed with the participants in the workshops.

It came out clearly through the interviews that the respondents were open to having a test on concepts both before and after the training program. The respondents even suggested that the participant's performance should be assessed during the course of the program for both knowledge and skill inputs using a simple questionnaire. It was strongly recommended that the results of the test should be communicated directly to the head of the department where the participant worked. In cases where the results could not be shared on a one to one basis with the participants, it was suggested that the institution could think of having a collective group grading system both pre and post training.

Level 3

The questionnaire had a separate part on "Professional relevance of training" which tried to understand the Level 3 and Level 4 attributes of change and results of training. The components included the relevance of Program content, extent of practical application of the program, benefits of training, ability to share the information received and change in organisational performance.

The survey results were rechecked through personal interviews conducted with 16% of the total respondents spread across the four regions of the country. The focused interviews helped in understanding whether the participant observed any perceptible change in his/her knowledge, skill, and attitudes after the training program. The trainee was also asked to identify how the learning could be utilized in the work situation.

The survey clearly showed that the program content had high relevance. But in comparison to the extent of practical application of the program, content showed medium relevance. As the participants were heterogenous in nature, hailing from different parts of the country and sometimes working in different departments, they could not implement what they had been trained in. The most common reasons for this shortcoming was that: (i) the training was not always relevant to the specific areas of work, (ii) tools and techniques (information technology and GIS) shown during the training did not exist in the work environment, and /or (iii) the training program did not cover issues of 'how to resolve the problem' which arose in the field and (iv) after training, the trainee was shifted to another area of work, even if working in the same department.

With regards to the benefit of training, it is interesting to note that the respondents had an increased awareness through the program. The respondents were asked to rank the benefits of training, on a scale of 1 to 5, with 5 standing for 'extremely useful' to 1 for 'not at all useful'. The benefits being:

1. Developing concern for rural development
2. Improvement in inter personal relations
3. New concepts / ideas introduced
4. Networking
5. Increased efficiency in delivery of service
6. Improved service levels and
7. Efficient utilization of financial resources

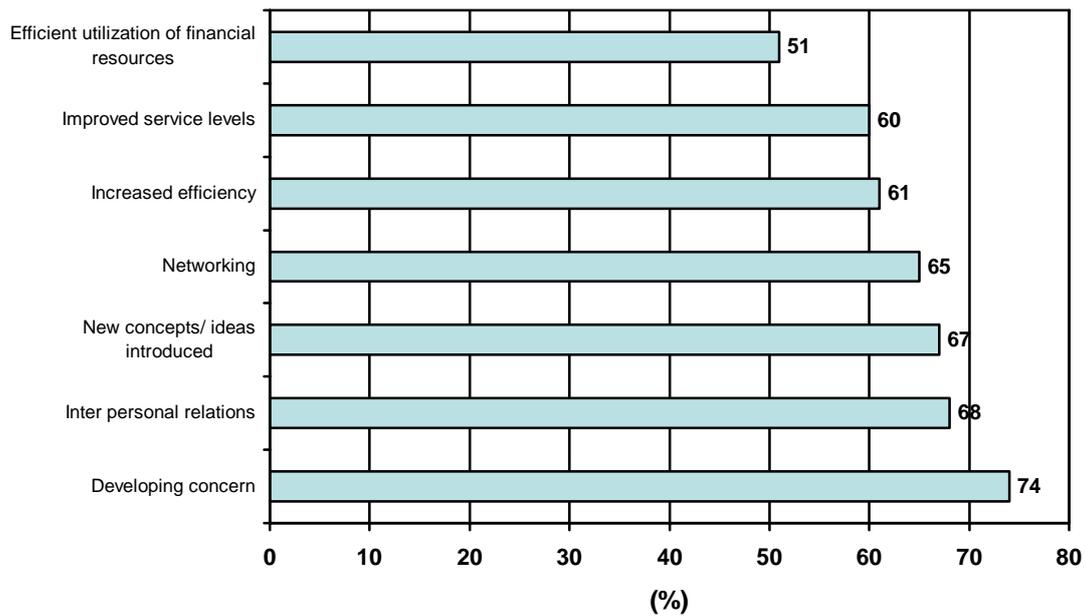


Figure 3. Benefits of Training

Here again the scales of 4 and 5 and scales of 1 and 2 were put together for simpler understanding. The Figure 3 shows that 74% of the respondents said that the greatest benefit that they had derived from the training program was in 'developing concern' for rural development. The other two main benefits were seen in improvement of inter-personal relations and learning of new concepts/ ideas introduced in the field of rural development.

It was observed that the changes and the ensuing results in the organizational context were the more difficult steps of evaluation. No final results could be expected unless a positive change in behaviour occurred. It was also seen that at times, the change was not fully under the control of the participant. Participants could not change their behaviour unless they had a chance to do so. Also, it was impossible to predict when change occurred, and sometimes, the participant may have decided against the change after contemplation because (s)he felt the earlier style was better or that the environment was not conducive for change. Infact it was seen that in many cases, changes could not be singly attributed to the training received.

At such times it became imperative to see how the participants came to be nominated for the training program. From the survey it came out that the most common reasons were: (i) the participants considered themselves to be the most 'Suitable person' for training. But most could not elaborate further on this. Under the circumstances, it could not be inferred further whether any specific mode of selection was identified or was in place within the departments when identifying the suitable person; (ii) Nominated – This in most cases meant that the participant had not really thought about his/her nomination. The trainees had attended because they had been sent for training. From the trainer's point of view, it seemed that many times the most 'dispensable' person was sent for training. This was true in cases where the nominations were received in the last minute; (iii) As per procedure – this was also a reason similar to that of the suitable person, where the trainees could not clearly define what the procedure was for being selected to the training program;

and (iv) Don't know – the trainees did not simply know why they had been sent for training, even after a lapse of nearly 1 to 3 years.

This seemed to give a clear picture of how training was not considered serious business by the sponsoring authorities. The point was reiterated during interviews with the Heads of the Departments. The Departmental heads many times were trying to use the budget allotted under capacity building. No clear guideline had been followed before the nominations were made for the training program.

It was seen that the impact of the training was best seen where the needs of the training was clear right from the beginning. But under the circumstances, one could venture to say that the impact of the training was limited by the participant who was nominated. Because of this the studied institution was compelled to design a more common, base level program catering to a heterogenous group rather than a specific issue based training program.

All the same, once selected for training, the respondents (77%) came with some expectations before the training, though they were not able to clearly express what their expectations had been. The respondents (75%) also went on to say that their expectations had been met. 23% of the respondents however said that the question of expectations being met was not applicable to them, as they had not really understood what they wanted initially from the program nor realized the objectives of the program even by the end of the training.

59% of the participants had become aware of new needs and the personal handicaps in terms of skills required for meeting the needs after attending the training. Such participants wanted to attend repeat programs at an advanced level. In some cases they even wanted to attend a few basic courses for better understanding of the concepts. Research Methodology and Communication Skills were two such popular programs.

When asked about the 'perceptible changes' in knowledge, skills and attitudes after the training, 99.5% said they had observed that there was some learning, though unable to pinpointedly mark it in their work.

Overall it came out that the training programs had created awareness, and helped them in 'being more concerned' about rural development. It was found that the middle level functionaries working in the field, who were working very grudgingly earlier, realized that they could contribute positively in the work environment. Training had helped them become aware of the problems and needs of rural development. And the already 'aware' participants became more articulate in their expression. The training helped them in thinking in a more sequential manner to help find solutions locally.

These findings directly related to the training programs which had good program content, and design, but was not sufficiently performance oriented in terms of practical application. This perhaps could be one of the main reasons for the participants inability in realizing a perceptible change, or in they having a continued niggling feeling that something was missing, which restrained them from completing the 'last mile' successfully.

This only confirmed that in the training institution, the trainer would have to take note of the work situation that the participant returned to while designing the course. And for a more effective result, an enabling atmosphere would need to be created in the Departments where change was expected.

Level 4

For Level 4, the study tried to get the respondent to quantify through figures, the sharing of learning in terms of meetings held, number of beneficiaries influenced and such others. The Heads of Departments were contacted to get their views on what was needed for rural development.

When asked about how the knowledge had benefited the organization, respondents found it difficult to quantify the benefits derived. 96% of the respondents were emphatic that they had utilized the learning from the training in their work situation. But a large number, were unable to explain how it had been translated into work. It was seen that only 68% said that they had actually attempted knowledge sharing while the rest had made no attempts in this direction.

The common ways in which the participants utilized the learning were (i) to implement day-to-day work and prepare reports, planning and evaluation, (ii) design and implement new schemes effectively and (iii) Learned skills to train the field staff.

One of the main reasons why no quantification could be made, because there were no attempts prior to the training or later to document the work done to help fill in the gaps where needed. Hence no comparative analysis could be done.

However the survey results under the section 'Professional relevance' of the program helped gain insight into this aspect. The question was administered on a 5-point likert scale. The results showed a definite change among the trainees in (a) new concepts / ideas were introduced, (b) increased productivity, (c) networking with different people, (d) increased efficiency in delivery of services, (e) improvement in interpersonal relations, (f) improved service levels and (g) career advancement. However reduction of execution time of projects and improvement in revenue/project viability got relatively low scores, meaning perceptible changes could not be seen in these two parameters. It can be seen that these heads are such which need a systemic organizational change with more awareness among all concerned. Infact it was interesting to note that many respondents shared the view that their 'bosses' and Heads of Departments needed to be trained for a more systemic change and overall better performance in the organization.

In the current scenario, the departments in Government had no regular process whereby the learning could be transferred to the work environment. Most times the back home work environment remained neutral with no expectations from the Departments.

The respondents also suggested ways to improve the training programs so that more changes could be seen at Level 3 and Level 4 in the long run, which were as follows -

- 1) Most participants were unaware of the gamut of training programs offered by the Institution. It was suggested that if the calendar of the training programs were communicated to the concerned Department a year in advance then the participants could choose those training programs which would be most relevant to them.
- 2) Selection of participants had to be done in an objective manner by the sponsoring department, which was sending the participants; based on the relevance of the training program. Younger personnel with more years in service would have to be sent for training.
- 3) The training Institution would have to scrutinize the nominations sent, through its own rigorous protocol for selection of the participants. This way the right participant who was well experienced in a relevant field would get enough inputs for betterment.

- 4) Every group of participants needed a program guide from the parent Department to help them through the training program.
- 5) Impact of training programs to be evaluated within 3-6 months after training program to help reiterate the learning.

4. Conclusions

Overall, the studied institution had delivered through its numerous training programs. The participants were happy about the training they had received, and looked forward to learning more from the Institute. There were however shortcomings in trainings in terms of the nominations of the participants, and the problems faced by the participants in utilizing the learning in their home environment.

It was interesting to note that through the study, the respondents came up with the same kind of suggestions for evaluation as suggested under the KP model, even when the study team had never mentioned that the study was using such a model for analysis.

From the study it came out quite clearly that:

- a) Level 1 is the best evaluation method of the four levels with respect to cost and efficiency. Level 1 evaluation was being done at the institution and needed to be continued.
- b) Level 2 was presently not done at the institution. The various respondents were keen on being evaluated during training as also at the start and the end.
- c) The Level 3 evaluation clearly brought out that there were perceptible behavioural changes in the respondents after training, though not measurable in all cases. The study clearly showed that nearly 70% of those trained remained in the same position after training, which offered great hope for active policy interventions and changes that could be brought through the training programs.
- d) For Level 4, sponsoring departments had to clearly put down the expected results. The study showed that there was need for cooperation between sponsoring authorities and the trainers to design programs which could become more functional and useful for changes to happen at the ground level. The respondents were clearly keen to cooperate and take the efforts forward.

Bibliography

1. Kirkpatrick, L. D. and Kirkpatrick, D. J. **Evaluating Training Programs – The Four Levels**, Benett Koehler Publishers Inc, December 2005
2. Marcotte, P. L., Bakker-Dhaliwal, R. and Bell, M. **Assessing Training Impact : IIRI's New Evaluation Approach**, Paper No.10, Occasional Papers: Issues in Training, Training Center, IIRI, International Rice Research Institute, January 2002
3. Saxena, R.N. **Impact Evaluation of Personal Growth Training: A Quasi-experimental Design**, ASCI Journal of Management, Vol. 35(1&2), March 2006, pp. 28-38
4. Reay, D.G. **Evaluating Training**, Kogan Page, London, 1994
5. Phillips, J. and Stone, R. D. **How to measure training results: A practical guide to tracking the six key indicators**, McGraw-Hill, New York, 2002
6. Spitzer, D. R. **Embracing Evaluation Training**, Training, Vol.36(6), 1999, pp. 42-47.

¹ The paper is based on the study for the training institution. I wish to acknowledge the help of my colleagues in the study - Saxena, Hemnath, Sahoo and Amarnath for their inputs and the participants of the survey during the consultation and also the support from the training institution in completing the study on time.

² A. INDIRA is Coordinator-Training: Management Development Programs (Development). She is also involved in teaching Operations Research, Production and Operations Management, Macroeconomic Analysis for Rural Society. She holds a PhD (1995) in Economics granted by Institute for Social and Economic Change, Bangalore. She received also a MA in Economics from University of Hyderabad (1988), a BSc in Economics, Mathematics and Statistics (1986) and a Bachelor in Law (2000) from Bangalore University.

Selective list of publications

1. **Democracy and Decentralisation : Zilla, Taluk and Grama Panchayats**, A.Indira, et.al., March 2000, -- a CBPS monograph
2. **Budget Analysis: for whose sake?**, a mimeograph presented in an International Conference on Budget Analysis, Mumbai, November 6-9, 2000, with Dr.L.C.Jain.
3. **Rural Industries in India: Lessons from a Survey**, a background paper brought out by South Asia Multidisciplinary Team (SAAT), International Labour Organisation (ILO), New Delhi, August 2000, with Vinod Vyasulu and B.P.Vani
4. **A Health Budget in Karnataka : a preliminary study**, A.Indira and Vinod Vyasulu,
5. **The Budget for Education – A study at the District Level in Karnataka**, A.Indira and Vinod Vyasulu, November 2001, a CBPS monograph
6. **Data systems at Local Levels- experience from an "IT" state**, A.Indira, commentary in Economic and Political Weekly, March 30 2002.
7. **Karnataka: Secondary Education and The New Agenda for Economic Growth**, Human Development Sector Unit, South Asia Region, Report No.24208-IN, Document of the World Bank with Dr.Vinod Vyasulu.

VALIDITY, RELIABILITY AND DIFFICULTY INDICES FOR INSTRUCTOR-BUILT EXAM QUESTIONS¹

Gholamreza JANDAGHI²

PhD, Associate Professor Faculty of Management, University of Tehran, Qom Campus, Iran

E-mail: jandaghi@ut.ac.ir



Fatemeh SHATERIAN³

MSc, Academic member of Islamic Azad University, Saveh Branch, Iran

E-mail: shaterian@yahoo.com



Abstract: *The purpose of the research is to determine college Instructor's skill rate in designing exam questions in chemistry subject. The statistical population was all of chemistry exam scripts for two semesters in one academic year from which a sample of 364 exam scripts was drawn using multistage cluster sampling. Two experts assessed the scripts and by using appropriate indices and z-test and chi-squared test the analysis of the data was done. We found that the designed exams have suitable coefficients of validity and reliability. The level of difficulty of exams was high. No significant relationship was found between male and female instructors in terms of the coefficient of validity and reliability but a significant difference between the difficulty level in male and female instructors was found ($P < .001$). It means that female instructors had designed more difficult questions. We did not find any significant relationship between the instructors' gender and the coefficient of discrimination of the exams.*

Key words: *instructor-built exam; content validity; face validity; reliability; coefficient of discrimination; coefficient of difficulty*

1. Introduction

Examination and testing is an important part of a teaching-learning process which allows instructors to evaluate their students during and at the end of an educational course. Many instructors dislike preparing and grading exams, and most students dread taking them. Yet tests are powerful educational tools that serve at least four functions. First, tests help you

evaluate students and assess whether they are learning what you are expecting them to learn. Second, well-designed tests serve to motivate and help students structure their academic efforts. Crooks (1988), McKeachie (1986), and Wergin (1988) report that students study in ways that reflect how they think they will be tested. In last 40 years the most exams used to evaluate the students have been designed by instructors. Some may have used tests which have been designed by outsider exam designers. These tests have not had enough efficiency (Seif 2004). Regarding the importance of instructor-designed test in evaluation process of the students, many researches have been done in this area (Lottfabadi 1997). In theory, the best test for a subject is a test that includes all educational objectives of the course. But if the test is too long, its preparation is impractical. Therefore, instead of including all content and objectives, one may choose some questions which are representative of the whole subject to achieve all objectives. Such a test is said to have content validity (Seif 2004).

Content validity of a instructor-designed test can be assessed by a sample of the test questions. When a test does not have content validity two possible outcomes may occur. First, the students can not present the skills that are not included in the test when they need. Second, instead some unrelated question may be included in the test that are answered wrongly. The important point here is that we should not mistake the face validity with content validity. Basically the face validity is a measure that determines whether a test is measuring logically and whether students think the test questions are appropriate (Lottfabadi 1997).

Based on what is said, an ideal test in addition to measuring what is supposed to measure, must be consistently constant in different times. This characteristic is called reliability. Other measures of an ideal test are difficulty level and discriminant index. The total percent of the individuals who answer the question correctly is known as difficulty coefficient denoted by P (Seif 2004). The discriminant index is a measure of discrimination between strong and weak groups. In this study, we intend to evaluate the extent of ideal quality measures (validity, reliability,...) in instructor-designed test for first year college.

Materials and methods

The statistical population in this study consisted of all chemistry exam papers for final chemistry exams in first and second semester for first year of college in Qom province of Iran of which a sample of 364 was taken. A twostage cluster sampling was used to draw samples. In first stage three colleges was randomly selected. In second stage a number of exam papers from each college was selected according to the number of students in each college.

In this study the content validity of the exam questions was assessed in two ways. In the first method we used a two dimensional table. One dimension was educational goals and the second dimension was the content of the course materials(Seif 2004). The second method applied for assessing content validity was a questionnaire with Likert scale in which two chemistry education expert evaluated the extent of compatibility of exam questions with course contents. For assessment of face validity of instructor-built exams we used a 12-item questionnaire answered by two chemistry experts.

Reliability

To assess the reliability of the tests, we needed to use a number of experts to mark the exam papers in order that the marking does not affect the marker's opinion(seif 2004). In this study, we asked two instructors to mark the exam papers separately and used Kendal agreement coefficient to check the agreement of the two markings.

Difficulty Coefficient and Discriminant Coefficient

Because all of chemistry exam questions were open questions, we used the following formula for calculating the difficulty coefficient(DifCo).

$$DifCoef_{question(i)} = \frac{M_{S(i)} + M_{W(i)}}{N_B * m_i}$$

Where

$M_{S(i)}$ = sum of marks for Strong group in question i

$M_{W(i)}$ = sum of marks for Weak group in question i

N_B = number of students in both groups

m_i = total mark of question i

And the Discriminant Coefficient(DisCo) was calculated based on the following formula(Kiamanesh 2002).

$$DisCoef_{question(i)} = \frac{M_{S(i)} - M_{W(i)}}{n_g * m_i}$$

Where

$M_{S(i)}$ = sum of marks for Strong group in question i

$M_{W(i)}$ = sum of marks for Weak group in question i

n_g = number of students in one group

m_i = total mark of question i

Results

The percentages of papers were almost equal in terms of students' sex(49% males and 51% females). The characteristics of the exam questions is summarized in Table 1.

Table 1. Exam characteristics by book chapters

Characteristic chapter	knowledge		concept		application		total
	mark	percent	mark	percent	mark	percent	
1	41.5	11.5	78	21.7	8.25	2.3	127.75
2	36	10	85.5	23.7	2	0.6	123.5
3	30.75	8.5	20.5	5.7	2.25	0.6	53.5
4	32.5	9.1	22.75	6.3	0	0	55.25
total	140.75	39.1	206.75	57.4	12.5	3.5	360

Table 1 shows that most chemistry questions were on concept (57.4%) and percentages on knowledge (39.1%) and small percentage on application (3.5%). There were no questions on analysis, combination and evaluation in the exams.

As stated before, the agreement of instructors evaluations was calculated using Kendal's agreement coefficient. The value of the coefficient was 0.49 which was significant at p-value of 0.05. The Kendal's agreement coefficient for face validity of the questions based on the evaluation of expert instructors was 0.42 and significant at p-value < 0.05). The reliability coefficient based on markers' evaluations was 0.971 and significant (p < 0.0001). The minimum and maximum difficulty coefficients estimated were DifCoef(min) = 0.14 and DifCoef(max) = 1 with standard error of 0.16 which indicates that the questions have moderate difficulty level. The minimum and maximum discriminant coefficients were DisCoef(min) = 0.07 and DisCoef(max) = 0.98 with standard error of 0.20 indicating that the questions have good discriminant coefficient.

We also found no significant difference for content validity and reliability between female and male instructors. Then we compared the Difficulty coefficient and discriminant coefficient between two sexes of instructors. The test results are shown in Tables 2 and 3.

Table 2. Chi-square test for comparison of difficulty coefficients between female and male instructors

Difficulty level	# of questions from female instructors	# of questions from male instructors	Chi-squared value	Degrees of freedom	p-value
0-0.2	4	7	28.230	4	0.000
0.21-0.4	19	22			
0.41-0.6	21	41			
0.61-0.8	45	28			
0.81-1	60	20			

Table 2 shows that there is a significant relationship between difficulty level of the questions and the sex of instructors. Female instructors tend to design more difficult chemistry questions than males.

Table 3. Chi-square test for comparison of discriminant coefficients between female and male instructors

discriminant level	# of questions from female instructors	# of questions from male instructors	Chi-squared value	Degrees of freedom	p-value
0-0.2	23	26	5.212	4	0.266
0.21-0.4	54	37			
0.41-0.6	36	22			
0.61-0.8	17	21			
0.81-1	19	12			

Table 3 shows no relationship between the instructor's sex and the discriminant level of the questions.

Discussion and Conclusion

One of the important issues in any teaching and learning system is the quality of the students. There should be some standards for exam questions so that we have the same and high level of quality among all educational organizations' output. Although the achievement of students in their course of study is important, the performance of instructors is also of great importance. One of the factors in the performance of instructors is good examination and good marking. Exam questions plays a vital role in students' achievement. The level of difficulty, discrimination, validity and reliability of exam questions must be ensured in order to have good outputs. In this study, we concluded that some of these factors can differ among different instructors in terms of instructor's sex. Female instructors tend to design more difficult questions than males. This may be because of the performance of the female students (Jandaghi 2008). We also found that a high percentage of exam questions concentrate on concept(57.4%) and knowledge(39.1%) whereas the small percentages on other characteristics such as applications. This may be because of the nature of chemistry. These percentages may of course change when the topic of the course changes. In summary, instructors need to be assessed and evaluated during their teaching process to ensure the quality of their performance.

References

1. Crooks, T. J. **The Impact of Classroom Evaluation Practices on Students**, Review of Educational Research, 58 (4), 1988, pp. 438-481
2. Jandaghi, Gh. **The Relationship between Undergraduate Education System and Postgraduate Achievement in Statistics**, International Journal of Human Sciences, 5 (1), 2008
3. Kiamanesh, A. R. **Assessment and Evaluation in Physics**, Ministry of Education Publications, Tehran, Iran, 2008
4. Lotfabadi, H. **Assessment and Evaluation in Psychological Sciences**, Samt Publication, Tehran, Iran, 1997
5. McKeachie, W. J. **Teaching Tips** (8th ed.), Lexington, Mass., Heath, 1986
6. Seif, A. A. **Assessment and Evaluation in Education**, Doran Publication, Tehran, Iran, 2004
7. Wergin, J. F. **Basic Issues and Principles in Classroom Assessment**, in McMillan, J. H. (ed.), „Assessing Students' Learning New Directions for Teaching and Learning“, no. 34, San Francisco, Jossey-Bass, 1988

¹ Acknowledgement:

This research was funded by Education and Teaching Organization of Qom Province of Iran.

² Gholamreza Jandaghi Graduated from University of Toronto in Canada in 1994.

He has published as an author and co-author over 50 articles in journals and more than 30 articles on national and international conferences. He is also author of 3 books(Descriptive Statistics for Human Sciences, Probability and Statistics for Engineers and Applied Probability and Statistics for Management and Economics).

He teaches Mathematics and Statistics courses and supervises master and Ph.D. students at Qom Campus of University of Tehran.

His current research areas are EFQM , Statistics Education and Simulations.

³ Fatemeh Shaterian was born in 1978 in Qom Province of Iran. She is holding B.Sc. in Clinical Psychology, M.Sc. in Assessment and Measurement and is doing her Ph.D. in Behavioral Psychology at Islamic Azad University of Iran. Shaterian is now an academic member at Saveh branch of Islamic Azad University. Her research interests include "job satisfaction", "meta analysis" and "validity and reliability".

MODELLING UNEMPLOYMENT RATE USING BOX-JENKINS PROCEDURE

Ion DOBRE

PhD, University Professor, Department of Economic Cybernetics
University of Economics, Bucharest, Romania



E-mail: dobrerio@ase.ro

Adriana AnaMaria ALEXANDRU

PhD Candidate, University Assistant, Department of Statistics and Econometrics
University of Economics, Bucharest, Romania



E-mail: adrianaalexandru@yahoo.com

Abstract: *This paper aims to modelling the evolution of unemployment rate using the Box-Jenkins methodology during the period 1998-2007 monthly data. The empirical study relieves that the most adequate model for the unemployment rate is ARIMA (2,1,2). Using the model, we forecasts the values of unemployment rate for January and February 2008. Therefore, the unemployment rate for January 2008 is 4.06%.*

Key words: *Unemployment rate; Box-Jenkins methodology; ARIMA models; Romania*

1. Theoretical Background

The pioneers in this area was Box and Jenkins who popularized an approach that combines the moving average and the autoregressive models in the book¹. Although both autoregressive and moving average approaches were already known (and were originally investigated by Yule), the contribution of Box and Jenkins was in developing a systematic methodology for identifying and estimating models that could incorporate both approaches. This makes Box-Jenkins models a powerful class of models.

The Box-Jenkins ARMA model is a combination of the AR and MA models as follows:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} - b_1 u_{t-1} - b_2 u_{t-2} - \dots - b_q u_{t-q} + u_t$$

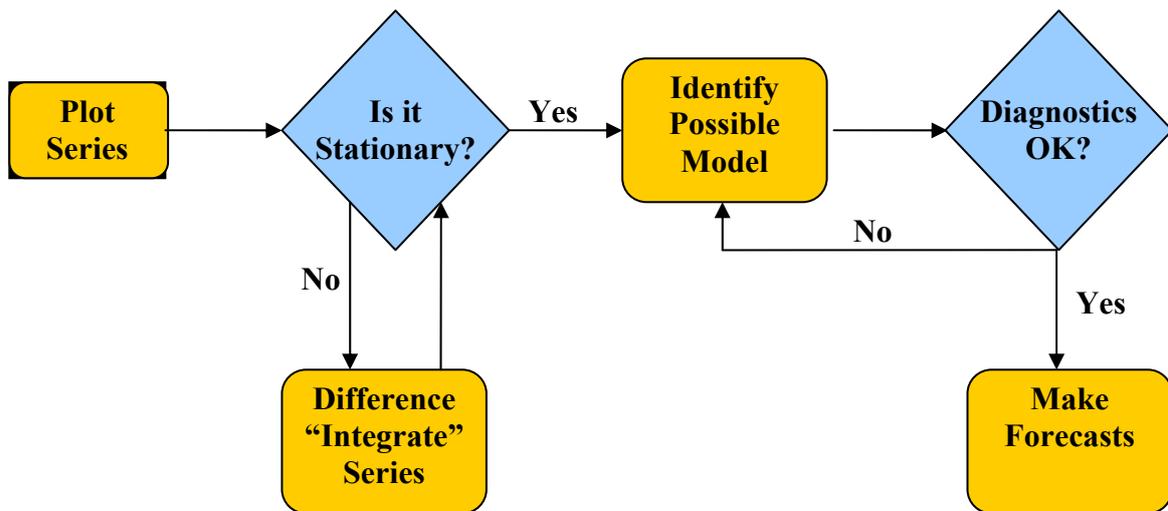


Figure 1. Box-Jenkins procedure

There are three primary stages in building a Box-Jenkins time series model:

1. Model Identification
2. Model Estimation
3. Model Validation

1.1. Box-Jenkins Model Identification

The identification stage is the most important and also the most difficult: it consists to determine the adequate model from ARIMA family models. The most general Box-Jenkins model includes difference operators, autoregressive terms, moving average terms, seasonal difference operators, seasonal autoregressive terms, and seasonal moving average terms². This phase is founded on the study of autocorrelation and partial autocorrelation.

The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modelled.

Stationarity in Box-Jenkins Models

The Box-Jenkins model assumes that the time series is *stationary*. A stationary series has:

1. Constant mean
2. Constant variance
3. Constant autocorrelation structure

Regression with nonstationary variables is a spurious correlation. The random walk $y_t = y_{t-1} + u_t$ $u_t \sim N(0, \sigma^2)$ is not stationary, since its variance increases linearly with time t . Stationarity can be assessed from a run sequence plot. The run sequence plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay.

Box and Jenkins recommend differencing non-stationary series one or more times to achieve stationarity. Doing so produces an ARIMA model, with the "I" standing for

"Integrated". But its first difference $\Delta y_t = y_t - y_{t-1} = u_t$ is stationary, so y is „integrated of order 1“, or $y \sim I(1)$.

Testing for non-stationarity

1. Autocorrelation function (Box-Jenkins approach)-if autocorrelations start high and decline slowly, then series is nonstationary, and should be differenced.

2. Dickey-Fuller test

$y_t = a + by_{t-1} + u_t$ would be a nonstationary random walk if $b = 1$. So to find out

if y has a "unit root" we regress: $\Delta y_t = a + cy_{t-1} + u_t$ where $c = b-1$ and test hypothesis that $c = 0$ against $c < 0$ (like a "t-test").

Seasonality in Box-Jenkins Models

Box-Jenkins models can be extended to include *seasonal* autoregressive and seasonal moving average terms.

Model identification: seasonality of order s is revealed by "spikes" at $s, 2s, 3s$, lags of the autocorrelation function.

Model estimation: to make series stationary, may need to take s -th differences of the raw data before estimation. These seasonal effects may themselves follow AR and MA processes.

At the model identification stage, our goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For Box-Jenkins models, it isn't necessary remove seasonality before fitting the model. Instead, it can include the order of the seasonal terms in the model specification to the ARIMA estimation software.

Once stationarity and seasonality have been addressed, the next step is to identify the order (the p and q) of the autoregressive and moving average terms. The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behaviour of these plots when the order is known.

Order of Autoregressive Process (p)

Specifically, for an AR (1) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order AR processes are often a mixture of exponentially decreasing and damped sinusoidal components. For higher-order autoregressive processes, the sample autocorrelation needs to be supplemented with a partial autocorrelation plot. The partial autocorrelation of an AR (p) process becomes zero at lag $p+1$ and greater, so we examine the sample partial autocorrelation function to see if there is evidence of a departure from zero. This is usually determined by placing a 95% confidence interval on the sample partial autocorrelation plot (most software programs that generate sample autocorrelation plots will also plot this confidence interval). If the software program does not generate the confidence band, it is approximately $\pm 2/\sqrt{N}$, with N denoting the sample size.

The data is AR (p) if: ACF will decline steadily, or follow a damped cycle and PACF will cut off suddenly after p lags.

Order of Moving Average Process (q)

The autocorrelation function of a MA (q) process becomes zero at lag q+1 and greater, so we examine the sample autocorrelation function to see where it essentially becomes zero.

The following table summarizes how we use the sample autocorrelation function for model identification.

Table 1. The type of the model

<i>Shape</i>	<i>Indicated Model</i>
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

The data is MA (q) if: ACF will cut off suddenly after q lags and PACF will decline steadily, or follow a damped cycle.

It's not indicated to build models with:

- Large numbers of MA terms
- Large numbers of AR and MA terms together

You may well see very (suspiciously) high t-statistics. This happens because of high correlation ("colinearity") among regressors, *not* because the model is good.

1.2. Box-Jenkins Model Estimation

The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation. Maximum likelihood estimation is generally the preferred technique³.

1.3. Box-Jenkins Model Diagnostics

Model diagnostics for Box-Jenkins models is similar to model validation for non-linear least squares fitting. Model diagnostics for Box-Jenkins models is similar to model validation for non-linear least squares fitting.

That is, the error term u_t is assumed to follow the assumptions for a stationary unvaried process. The residuals should be white noise (or independent when their distributions are normal) drawings from a fixed distribution with a constant mean and variance.

If the Box-Jenkins model is a good model for the data, the residuals should satisfy these assumptions. If these assumptions are not satisfied, we need to fit a more appropriate model. That is, we go back to the model identification step and try to develop a better

model. Hopefully the analysis of the residuals can provide some clues as to a more appropriate model. The residual analysis is based on:

1. Random residuals: the Box-Pierce Q-statistic: $Q(s) = n \sum r(k)^2 \approx \chi^2(s)$ where $r(k)$ is the k -th residual autocorrelation and summation is over first s autocorrelations.
2. Fit versus parsimony: the Schwartz Bayesian Criterion (SBC):
 $SBC = \ln \{RSS/n\} + (p+d+q) \ln (n)/n$, where RSS = residual sum of squares, n is sample size, and $(p+d+q)$ the number of parameters.

2. The data

The variable used in the analysis is the unemployment rate that ran from 1998 to the end of 2007 and its available monthly. The source of data is the Monthly Bulletins of National Bank of Romania.

Stage 1: The time series analysis

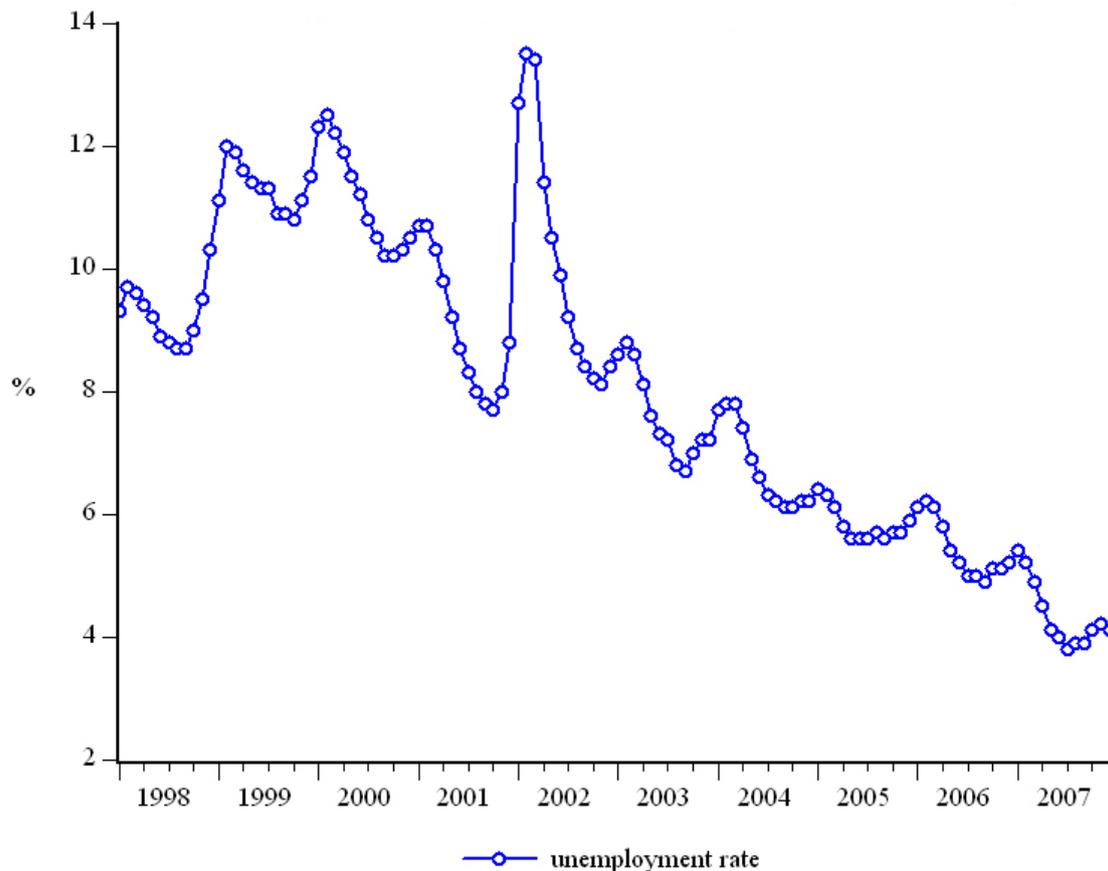


Figure 2. The unemployment rparte evolution during the period 1998-2007

Source: Montly Bulletins of National Bank of Romania

The data presents some seasonal fluctuations and that is the reason for with data has been seasonally adjusted, using the moving average method implemented in Eviews program.

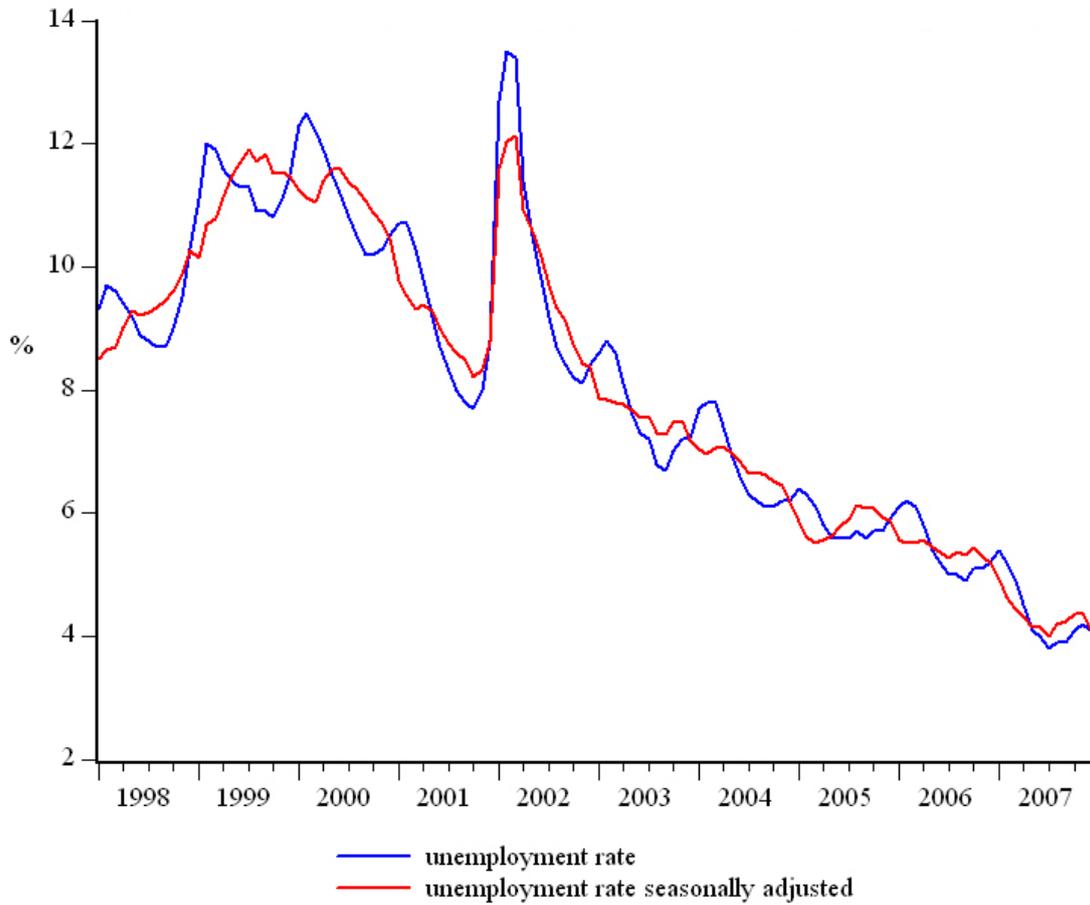


Figure 3. The unemployment rate and the unemployment rate seasonally adjusted

The first step in developing a Box-Jenkins model is to determine if the series is stationary. For this, we use the autocorrelation function (ACF) and Augmented Dickey-Fuller test (ADF).

Because the autocorrelation (ACF) start high and decline slowly, then series is nonstationary, and should be differenced. We have analyzed the data series stationarity by using the Augmented Dickey-Fuller (ADF) test, who reveals the fact that the zero hypotheses is accepted, the series has a root unit and it is non stationary. It becomes stationary by first order differences.

Date: 03/18/08 Time: 00:29
 Sample: 1998M01 2007M12
 Included observations: 120

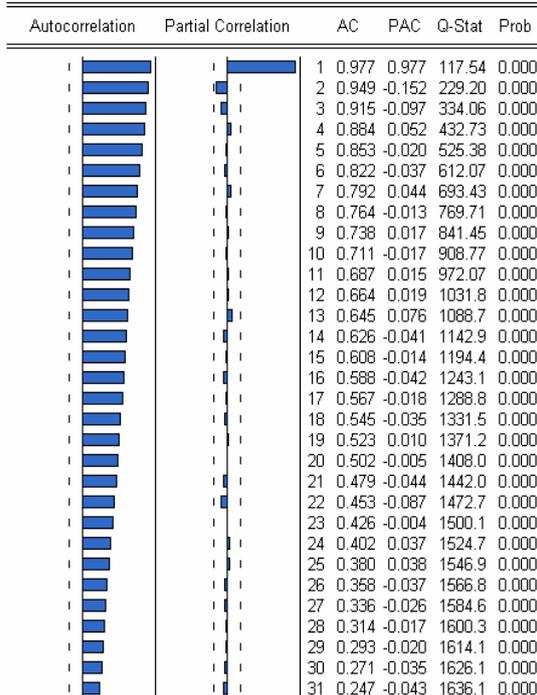


Figure 4. The correlogram of unemployment rate seasonally adjusted

Null Hypothesis: RSSA98 has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 0 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.037670	0.1266
Test critical values:		
1% level	-4.036983	
5% level	-3.448021	
10% level	-3.149135	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(RSSA98)
 Method: Least Squares
 Date: 03/26/08 Time: 16:07
 Sample (adjusted): 1998M02 2007M12
 Included observations: 119 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RSSA98(-1)	-0.085244	0.028062	-3.037670	0.0029
C	1.056913	0.336113	3.144520	0.0021
@TREND(1998M01)	-0.006726	0.001941	-3.464857	0.0007
R-squared	0.093806	Mean dependent var		-0.037015
Adjusted R-squared	0.078182	S.D. dependent var		0.354721
S.E. of regression	0.340572	Akaike info criterion		0.708507
Sum squared resid	13.45477	Schwarz criterion		0.778569
Log likelihood	-39.15619	F-statistic		6.003934
Durbin-Watson stat	1.312423	Prob(F-statistic)		0.003302

Figure 5. The Augmented Dickey-Fuller test results

Stage 2: The identification - the autocorrelation is computed on the first differences series

Sample: 1998M01 2007M12
 Included observations: 119

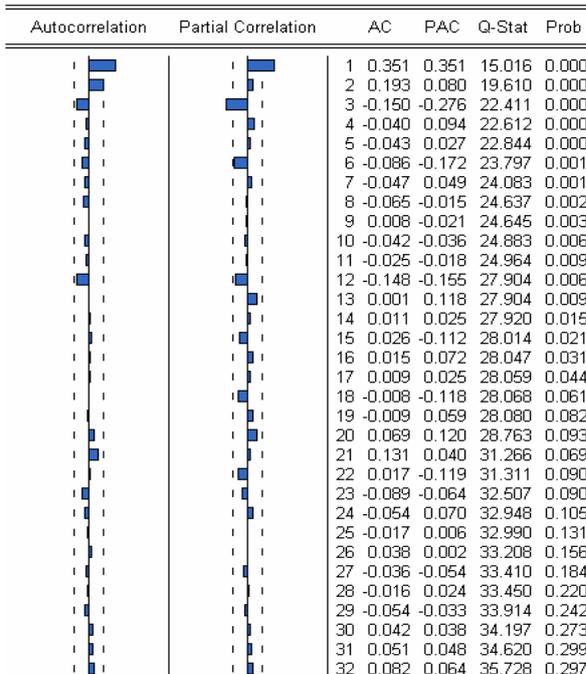


Figure 6. The Correlogram of first differences of unemployment rate

By applying the ADF test for the series of the first order differences one can observe that the series becomes stationary, so the initial series of the monthly unemployment rate is integrated by first order.

As a result, we have applied the Box- Jenkins procedure on the stationary data series and we want to identify the corresponding ARIMA (p, q) process. The series corelogram has allowed us to choose appropriate p and q for the data series. We have estimated more models in order to determine the right specification, by choosing from both the different models estimated on the informational criteria Akaike and by generating predictions on the basis of estimated models. The series corelogram suggests the necessity of introduction in the process estimation of both the analyzed variable lags and the lags of the error. We have started with an AR (1) process and further analyzed the residual corelogram in order to catch the correlations and autocorrelations from lags bigger that 1. From Akaike criteria's point of view, the proper model to best adjust the data is ARIMA (2, 1,2).

Stage 3: The Estimation

Dependent Variable: D(RSSA98)
 Method: Least Squares
 Date: 03/27/08 Time: 15:16
 Sample (adjusted): 1998M04 2007M12
 Included observations: 117 after adjustments
 Convergence achieved after 21 iterations
 Backcast: 1998M02 1998M03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.443316	0.077808	-5.697588	0.0000
AR(2)	-0.515496	0.072797	-7.081316	0.0000
MA(1)	0.813233	0.009985	81.44546	0.0000
MA(2)	0.977082	0.013197	74.03841	0.0000
R-squared	0.281409	Mean dependent var	-0.039311	
Adjusted R-squared	0.262331	S.D. dependent var	0.357265	
S.E. of regression	0.306846	Akaike info criterion	0.508651	
Sum squared resid	10.63948	Schwarz criterion	0.603084	
Log likelihood	-25.75609	Durbin-Watson stat	1.848075	
Inverted AR Roots	-.22+.68i	-.22-.68i		
Inverted MA Roots	-.41+.90i	-.41-.90i		

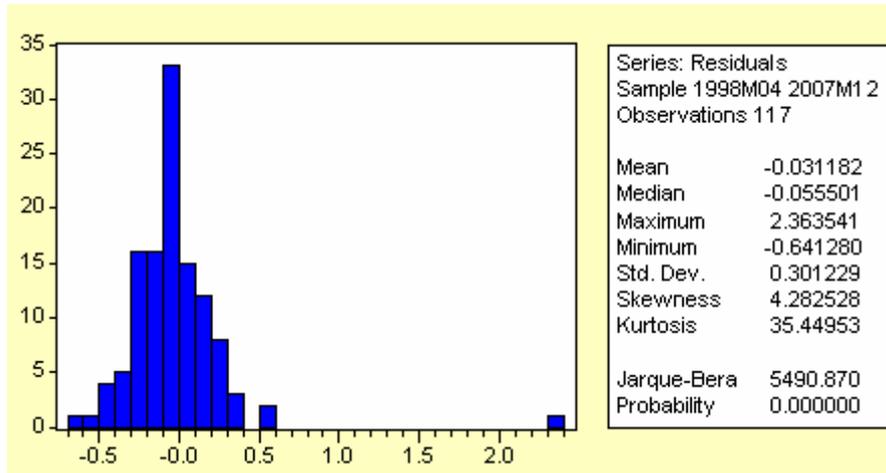
Figure 7. The ARIMA model estimation

Stage 4: The Model's Adaptation

The coefficients of the model are significantly different of 0 (the t-test). The others statistics (DW, F-stat) let portend a good fitting. The determination coefficient R-squared is 28.14%.

The residual analysis is based on two criterions:

- The normality test point out that the average of residuals is approximately 0.



- The residual is a white noise, analysing the autocorrelation. Any term isn't exterior to the confidence intervals and the Q-statistic has a critical probability near to 1. The residue it may be assimilate to a white noise process.

Date: 03/27/08 Time: 15:52
 Sample: 1998M04 2007M12
 Included observations: 117
 Q-statistic probabilities adjusted for 4 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.072	0.072	0.6292	
		2	-0.002	-0.007	0.6297	
		3	0.026	0.026	0.7097	
		4	0.010	0.006	0.7214	
		5	0.017	0.016	0.7555	0.385
		6	-0.002	-0.005	0.7558	0.685
		7	-0.005	-0.005	0.7593	0.859
		8	0.002	0.002	0.7599	0.944
		9	-0.017	-0.018	0.7974	0.977
		10	-0.018	-0.015	0.8384	0.991
		11	-0.014	-0.012	0.8655	0.997
		12	0.083	0.087	1.7869	0.987
		13	0.001	-0.011	1.7870	0.994
		14	-0.024	-0.021	1.8626	0.997
		15	-0.014	-0.015	1.8896	0.999
		16	-0.013	-0.012	1.9128	1.000
		17	-0.023	-0.024	1.9875	1.000
		18	-0.021	-0.017	2.0483	1.000
		19	-0.023	-0.018	2.1214	1.000
		20	-0.022	-0.019	2.1935	1.000
		21	-0.011	-0.004	2.2113	1.000
		22	-0.027	-0.022	2.3163	1.000
		23	-0.020	-0.013	2.3749	1.000
		24	-0.015	-0.020	2.4072	1.000

Figure 8. The Correlogram of Residuals Squared

Therefore, the estimation of ARIMA (2,1,2) model is validated, the time series can be described by an ARIMA(2,1,2) process. The unemployment rate seasonally adjusted times series and in first differences (DRSSA) is described by the process:

$$DRSSA = -0.4433 \cdot RSSA_{t-1} - 0.5154 \cdot RSSA_{t-2} + 0.8132 \cdot u_{t-1} + 0.9777 \cdot u_{t-2}$$

Stage 5: The forecasting

The forecasting is computed by reaggregation of different components. The residual values for the months of December and November are: $u_{2007:12} = -0.21465$,

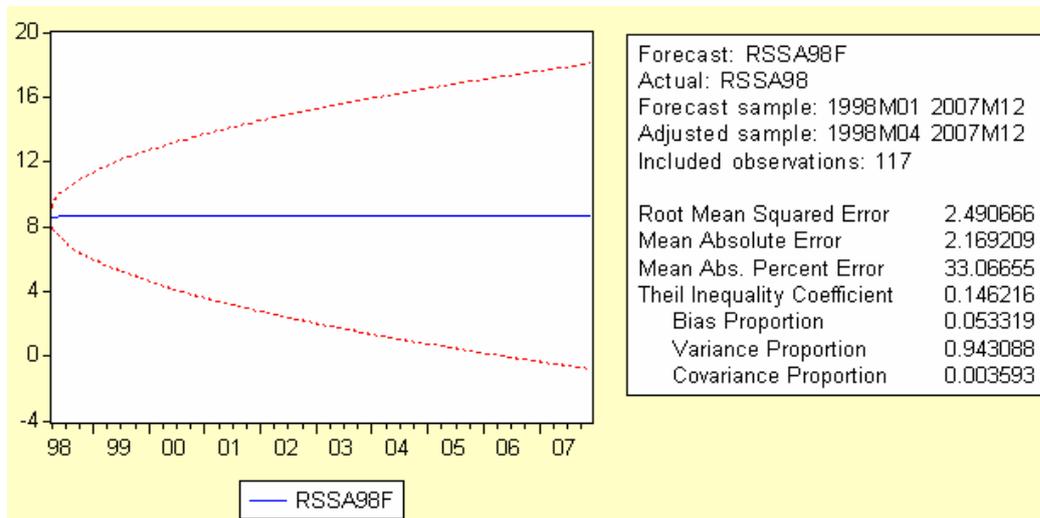
$$u_{2007:11} = -0.15538.$$

The fitting values of ARIMA model for unemployment rate are:

$$RSSA_{2007:12} = -0.06526, \quad RSSA_{2007:11} = 0.15115.$$

Table 2. The unemployment rate forecasts

	u_t	DRSSA	RSSA	Seasonal Coefficients	Unemployment rate(%)
November 2007	-0,21465				
December 2007	-0,15538		4,08929		
January 2008		-0,37544	3,71384	1.0948	4,06
February 2008		-0,0098	4,15817	1.1226	4,15
March 2008		0,197846	4,10594	1.1048	4,35



Using an ARIMA (2,1,2) model of monthly values series of unemployment rate we can predict the value of unemployment rate for January and February 2008. In January 2008 the unemployment rate forecasted by the model was 4, 06% and for February 4,15%. The result troves sustainability into the monthly bulletin of National Institute of Statistics. According to this publication the unemployment rate is 4.3% for January 2008.

Bibliography

1. Bourbonnais, R. **Econométrie**, 6nd. ed., Dunod, Paris, 2005
2. Box, G. E. P. and Jenkins, G. M. **Time Series Analysis, Forecasting and Control**, Holden Day, San Francisco, 1970
3. Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. **Time Series Analysis, Forecasting and Control**, 3rd ed., Prentice Hall, Englewood Clifs, 1994
4. Brockwell, P.J. and Davis, R. A. **Introduction to Time Series and Forecasting**, 2nd. ed., Springer Verlag, 2002
5. Chatfield, C. **The Analysis of Time Series**, 5th ed., Chapman & Hall, New York, NY, 1996
6. Cuthbertson, K. and Hall, S. **Applied Econometrics Techniques**, Whreatoms Ltd., 1995
7. Dobre, I. and Alexandru, A. **Scenarios regarding the unemployment rate evolution evolution in Romania during the period 2006-2009**, Economic Computation and Economic Cybernetics Studies and Research, vol. 41, No. 2, 2007, pp. 39-52
8. Enders, W. **Applied Econometric Time Series**, Wiley, New York, 2004
9. Gujarati, D. **Basic Econometrics**, McGraw-Hill Inc., N. Y., 1995
10. Hamilton, J. **Time Series Analysis**, Princeton University Press, Princeton, 1994
11. Johnston, J. **Econometric Methods**, McGraw-Hill, N. Y., 1991
12. Maddala, G. S. **Introduction to Econometrics**, McMillan Public., 1988
13. Pecican, E., S. **Econometrie**, Ed. All, Bucharest, 1994
14. Pecican, E. S. and Tanasoiu, O. **Modele Econometrice**, Ed. ASE, Bucharest, 2001
15. Pecican, E. S. **Econometrie**, Ed. C.H. Beck, Bucharest, 2006
16. Pecican, E. S. **Econometria pentru...economisti: Econometrie-teorie si aplicatii**, Ed. Economica, Bucharest, 2005
17. www.bnr.ro, National Bank of Romania, **Monthly Bulletins**

¹ Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. **Time Series Analysis, Forecasting and Control**, 3rd ed., Prentice Hall, Englewood Cliffs, 1994

² Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. **Time Series Analysis, Forecasting and Control**, 3rd ed. Prentice Hall, Englewood Cliffs, 1994
 Chatfield, C. **The Analysis of Time Series**, 5th ed., Chapman & Hall, New York, NY, 1996
 Brockwell, P. J. and Davis, R. A. **Introduction to Time Series and Forecasting**, 2nd. ed., Springer-Verlag, 2002

³ Bourbonnais, R. **Econométrie**, 6e éd, Dunod, Paris, 2005

QUALIFYING PURCHASE INTENTIONS USING QUEUEING THEORY

Ercan TIRTIROGLU

PhD, Professor of Decision Sciences and Operations Management, Charlton College of Business
University of Massachusetts Dartmouth, USA
285 Old Westport Road, North Dartmouth, MA 02747, USA

E-mail: etirtiroglu@umassd.edu

Matt ELBECK

PhD, Professor of Marketing, College of Business Administration, Troy University
500 University Drive, Dothan AL 36303, USA

E-mail: melbeck@troy.edu

Abstract: *We propose that a consumer's purchase behaviour emulates a cognitive framework such that an appropriately selected queueing model can be used to assess the credibility of a consumer's expression of purchase intention to proceed into purchase action. The modelling idea is to envision that the customer has a mental queue of needs, and the surfacing of a need represents a new arrival to this queue while a purchase satisfying a need represents a departure from the queue. Using queueing formulae (and/or readily available tables on queueing), several statistics known as "operating characteristics" can be computed to capture and describe the individual's current need status. Hence, a description of the customer's state of readiness or prospect of actually executing the expressed purchase intention for a product (good or service) whose future sales are of interest in a market survey may be obtained. In this paper, such a description is termed Individual Buyer Profile (IBP), and IBP is offered as a purchase intention qualifier.*

Key words: IBS; queueing; purchase behaviour; purchase intention qualifier

Marketers of consumer goods and services have an understandably strong interest to predict the purchase behaviour of customers. In turn, these predictions contribute to market forecasts and related generalizations (Bird and Ehrenberg, 1966, Lipman, 1988,) for both existing (Morrison, 1979) and new products (Urban and Hauser, 1993).

In its simplest form, predicting purchases rests on the stage preceding actual purchase, and is referred to as "intention to purchase" (Howard and Sheth, 1967). According to various theories of buyer behavior, purchase intention helps predict subsequent purchase (Bagozzi, 1983; Engel, Blackwell and Kollat, 1978; Fishbein and Ajzen, 1975; Howard and Sheth, 1969; Warshaw, 1980). Information about purchase intention is typically drawn

from a purchase intent scale (McDaniel and Gates, 1991) or an 11-point purchase probability scale (Juster, 1966; Wright et al, 2002) which are designed to elicit a response to the question how likely an item will be purchased within a specific time period. Both purchase intent and probability scales are reported as empirically unbiased, with the latter offering greater precision (Wright and MacRae, 2007).

The intention-purchase relationship has attracted a number of empirical studies highlighting significant inconsistencies between purchase intention and purchase behaviour (Ferber and Priskie 1965; Juster, 1966; Kalwani and Silk, 1982; Mullett and Karson, 1985; Namias, 1959; Pickering and Isherwood, 1974; Warshaw, 1980). For example, Tull and Hawkins (1987) note that results for private surveys have shown moderate discrepancy between intent and purchase for industrial goods. The results for consumer goods are more disappointing (Kalwani and Silk, 1982). In public surveys, the results have not been any more encouraging. The United States Bureau of the Census Consumer Buying Expectations Survey has been discontinued (McNeil, 1974). The Canadian Buying Intentions Survey was also questioned as noted by Murray (1969) who states "...buying intentions, when used alone, have limited predictive ability..."

Nonetheless, most studies do report a significant and positive relationship between intention and behaviour, albeit with varying strengths of association (Bennaor, 1995; Clawson, 1971; Granbois and Summers, 1975; Morwitz and Schmittlein, 1992; Morowitz et al, 1996; Newberry, Kleinz and Boshoff, 2003; Sheppard, Hartwick and Warshaw, 1988; Taylor, Houlahan and Gabriel, 1975).

Purchase Intenders and Non-Intenders

As with any other survey, the predictive ability of a purchase intention survey lies heavily with the truthfulness, or fidelity, of each survey participant's response. There are two situations when an intention will fail to be truthful: 1) if a positive response --i.e., intention to purchase-- is not followed by a purchase action, or 2) if a negative response --i.e., intention not to purchase-- is followed by a purchase action (each, during the time period as specified in the survey). The former situation corresponds to an "**intender-nonbuyer**," denoted herein by $(B^- | I^+)$, which represents (i.e., given that) the intention by the respondent to purchase the product (I^+). However, what is observed within the study period (B^-) is that the respondent did not follow through with his or her intention and actually buys the product (for whatever reasons, and whether the respondent consciously or unconsciously introduces this discrepancy to be observed eventually). On the other hand, the latter situation corresponds to a "**nonintender-buyer**," $(B^+ | I^-)$, which, though representing a negative intention by the respondent, (I^-), nevertheless, the respondent buys it, (B^+) during the observation period of the survey.

It should be immediately evident that these two cases have their respective complements; viz., $(B^- | I^-)$, and $(B^+ | I^+)$, corresponding to "truthful" respondents; i.e., "**nonintender-nonbuyers**" and "**intender-buyers**," respectively. In other words, these complementary cases represent "fidelity" to the intentions, where the expressed purchase or no-purchase intentions are true/truthful reflections of the purchase or no purchase actions that will prevail during the observation period of the survey. (Here, the term "truthful" is not necessarily being used strictly in the moral sense.)

The intent-purchase discrepancy varies in magnitude as reported by a number of studies. For example, Belk (1985) notes that an average of about 74 percent of nonintenders act in accord with their intentions, while only 43 percent of intenders are consistent with their intentions to purchase. Hence, as probabilistic estimates, we can write $Pr(B^- | I^-) = 0.74$ (i.e., given a respondent who expresses a negative purchase intention, the probability that the respondent is a nonbuyer-to-be is 0.74), and $Pr(B^+ | I^+) = 0.43$ (i.e., given a respondent who expresses a positive purchase intention, the probability that the respondent is a buyer-to-be is 0.43). Accordingly, $Pr(B^+ | I^-) = 0.26$, and $Pr(B^- | I^+) = 0.57$; i.e., there is a 26% chance that the respondent will buy, given that s/he may have said otherwise, and there is a 57% chance that a respondent will not buy, although s/he may have responded positively, expressing intention to buy. Thus, for a sample of respondents surveyed, the intention and purchase relationship may be represented as shown in Figure 1:

As further examples, Urban, Hauser and Dholakia (1987) estimate that responses to "definitely will buy" will be realized 70% of the time, while Infosino's (1986) work erodes this to 48%, which compares with Belk's $Pr(B^+ | I^+) = 0.43$. In an earlier study, Longman (1968) reported that, regarding automobiles, 60% of intenders and 17% nonintenders eventually purchased. These are considerably large departures! And it is unfortunate that actual purchase predictions, market forecasts, and related generalizations often rely heavily on these misleading and/or untrue intentions to buy. Hence, it is important to reduce the magnitude of the intent-purchase discrepancy facing the decision maker.

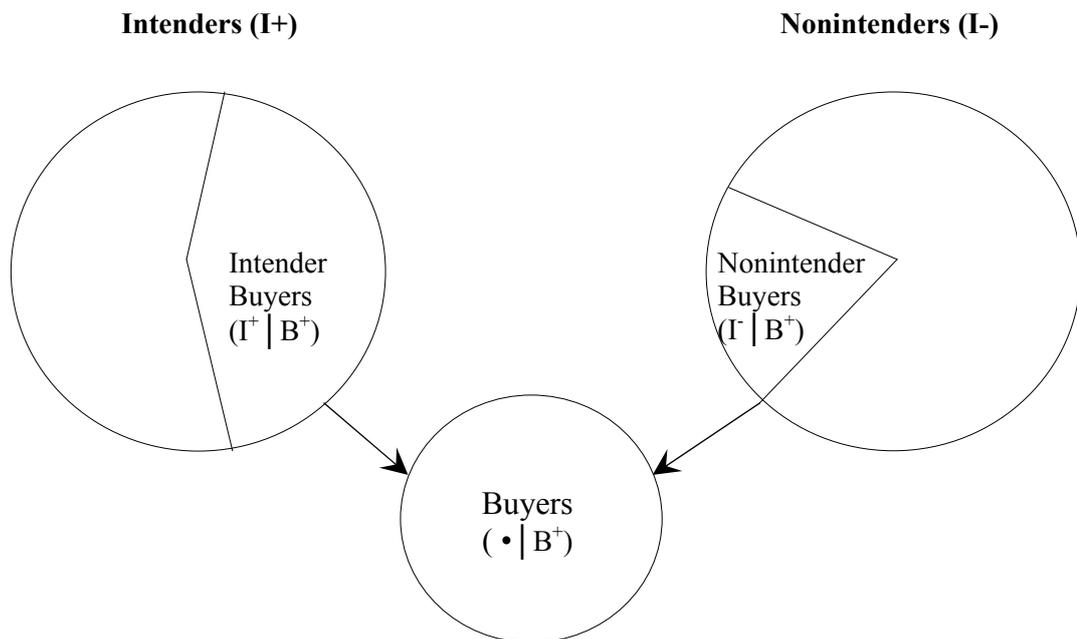


Figure 1. A Representation of Intention-Purchase Relationship

Possible causes for the gap between purchase intention and actual purchase include the survey itself (Chardon et al, 2005), failure to predict future preferences (Loewenstein and Adler, 1995), effect of information acquisition (Manski, 1990), bias in measuring and reporting intentions (Fisher, 1993; Mittal and Kamakura, 2001, Fitzsimons and Morowitz, 1996), perceived quality of the product (Boulding et al, 1993; Parasuraman et al, 1996), and product satisfaction (Cronin and Taylor, 1992; Sweeny et al, 1999).

Purpose of the Study

To date, studies examining the gap between stated and actual purchase have resulted in remarkable depth, sophistication and complexity, in lieu of parsimony much desired by the marketing practitioner. The practitioner is faced with quite a dilemma when conducting an intentions study. One approach might be to develop a mathematically complex model able to incorporate every researched influence and so offer a result in terms of a range of purchase probability. Alternatively, to encourage researchers to uncover an integrated and comprehensive model able to reduce the stated to actual purchase gap. It is the latter which forms the basis for this paper. That is, the purpose of this paper is to describe how queueing theory might be used to help improve the credibility, and hence, the predictive quality of consumer intention to purchase.

We envision that every consumer has a cognitive queue of needs. The arrivals to this queue are the needs that surface or emerge (i.e., the needs the individual becomes aware of), and the departures are the needs that are satisfied by appropriate purchase action.¹ Hence, a properly chosen queueing model can capture and reflect the individual's current need status (e.g., whether the individual already has a lengthy queue of unsatisfied needs; what the average waiting time for a need is, etc.). Based on the values of such statistics, we can come up with a description of the individual's readiness for --or prospect of-- actually completing the intended purchase within the survey period, and so, assess the fidelity of the response s/he gives. In this paper, we call such a description **Individual Buyer Profile (IBP)**, and offer **IBP** as a purchase intention qualifier. This qualification or assessment of response fidelity (or truthfulness) across all respondents would facilitate and lead to an improvement in the predictive quality of the survey.

While queueing theory is studied in the stochastic processes literature (e.g., see Cooper, 1981; Heyman and Sobel, 1982), its use in marketing as suggested in this paper for qualifying purchase intentions is, to the best of our knowledge, a new attempt. However, various existing stochastic and other buyer behaviour models and related topics can be found in Howard and Sheth (1969), and Lilien and Kotler (1983).

Individual Intentions and Forecasts

Morrison (1979) emphasizes the gathering of actual purchase data as a follow-up to stated intentions to facilitate an evaluation of stated intentions versus actual purchases. This would be useful for estimating future purchase rates, as well. Clearly, there is an understandable anticipation of a difference between the intention to purchase and the actual behaviour. There might be a variety of reasons for this difference or discrepancy — factors such as intervening time, unforeseen event(s), influence of others, new information, etc., may well suppress or enhance the intention resolution (Peter and Olson, 1996, p.183). The question is whether one can qualify or evaluate the individual-level purchase intentions in light of additional information about, and mostly from, the individual at the time of initial data collection, allowing for a-priori identification of the inherent potential for the outcome of no purchase.

By identifying those purchase intentions unlikely to be realized, and so eliminating them from further consideration in the estimation of actual purchases, one can reduce the intention-purchase discrepancy due to the intender-nonbuyer group; the main group

responsible for the discrepancy. Forecasts may be improved in the conservative sense; that is, the lower bound of the forecast representing the purchases expected from intenders can be sharpened. As the actual sum of purchases is likely to include a (small) portion from the nonintender-buyer group, a forecast adjusted/corrected for the intention-purchase discrepancy from the intender-nonbuyer group should offer a realistic lower bound on sales.

Queueing Mechanism in a Cognitive Context

Suppose that in an intention study, respondent **R** honestly indicates a positive response to buy the product (good or service) in question; call this **product P**. However, the investigator would not really know whether the response is honest and realistic. After allowing a reasonable period of time² for the purchases to occur, let us assume the investigator establishes that **R** did not make the purchase. From the investigator's point of view, **R's** intention was, in one sense or another, not truthful. Perhaps, respondent **R's** intention to purchase, although genuine or honest, could not be realized. How likely or ready was **R** to buy product **P** at the time the intention to purchase was expressed? It is possible that respondent **R** already had too many incumbent needs waiting for their turn to be satisfied (i.e., waiting to be serviced) via the purchase of relevant product(s) for the satisfaction of each need to occur.

In other words, **R's** need-satisfaction mechanism or capacity might have been in a state of backlog or congestion (i.e., in a state with too many needs to satisfy within a time period) when indicating an honest positive response to buy **P**. That is, respondent **R** may not have been fully cognizant of the underlying self-need-status at the time the intention was solicited. And later, a variety of these inadequately-perceived needs versus perhaps overestimated means --such as financial expectations-- may have prevented **R** the opportunity (i.e., the "room") to service the need for product **P** by a purchase action. Such a situation is analogous to the case of a fully occupied (or busy) queueing system that is unable to accommodate any more customers. Hence, it is possible to visualize the presence of a queueing mechanism couched within a cognitive framework. Needs are recognized by the consumer as they emerge or surface, constituting the arrivals to this cognitive queue, while the consumer's purchases which successfully satisfy these needs represent service completions and departures from this queueing system. Upon satisfaction, each need disappears, opening up room or the opportunity for another (the next) need to be satisfied.

In the likely event a survey respondent is not in full appreciation of the juxtaposition of incumbent personal needs and means, such an individual may honestly, but perhaps unrealistically, indicate a positive response to buy **P** within the period the intention study covers. It is certainly desirable to control or account for this intention-purchase discrepancy at the outset.

Queueing and Individual Buyer Profiles

With regard to the preceding discussion, assume that products belonging to **category C** are purchased by a customer at an average rate of μ (μ). Furthermore, suppose that **P** is a product classifiable into this category. Let us also assume there exists an average need emergence of λ (λ) for the customer concerning category **C** products. With regard to the queueing construct outlined earlier, an appropriate model can be selected, and its standard formulae and/or tabulated results can be employed to obtain certain statistics that will capture the customer's current outlook of needs, and reveal his/her

prospect for actually executing the purchase of **P** which was responded to with a positive intention.

Obviously, characterizing the product concept into its groups (convenience, shopping, specialty, unsought) may require separate treatment --such as using a different queueing model and/or λ and μ parameter values-- for each product based on the classification provided by the respondent.

Model Implementation

In numerous purchase behaviour models, the Poisson process appears as a standard assumption (Chatfield and Goodhart, 1973; Dalal, et al., 1984; Gupta and Morrison, 1991; Herniter, 1971; Schmittlein and Morrison, 1985; Wagner and Taudes, 1986). In keeping with this mainstream assumption and following Kendall's (1953) standard queueing notation for model designation, consider the M/M/1 queueing model. It is very likely that, due to its "completely" random nature, the M/M/1 model would be appropriate for products that are likely to invoke impulse purchases (for staples, the D/M/1 model is suggested).

Based on some statistics (known as "operating characteristics") that describe a queueing model's expected (i.e., average) behaviour, the following **IBP** description can be obtained for the M/M/1 model. First consider:

$$\text{Pr}(0) = 1 - (\lambda/\mu) ,$$

Where $\text{Pr}(0)$ is the probability that the individual's cognitive queueing system is empty (i.e., idle). An individual with a high $\text{Pr}(0)$ value who also states an intention to buy **P** is more likely to make the purchase than if the respondent's $\text{Pr}(0)$ value were low. This is because when $\text{Pr}(0)$ is high, the respondent's set of needs is near empty (or idle), and consequently, a respondent with a positive purchase intention and an idle or near empty set of needs is apt to act and purchase product **P** without having to worry about satisfying many other incumbent needs first, which would be the case with a respondent whose $\text{Pr}(0)$ value is low. Now consider:

$$L_q = \lambda^2 / \mu (\mu - \lambda) ,$$

Where L_q is the average number of unsatisfied incumbent needs that are already waiting in the individual's cognitive queueing mechanism (i.e., needs in line). A large L_q value would qualify a consumer who indicates an intention to buy product **P** a less likely buyer-to-be (with high potential to join the intender-nonbuyer group) than if the L_q value were small. This is simply because when L_q is large, the currently unsatisfied and queued up needs for other products would take precedence, thereby delaying, precluding and/or averting the purchase of product **P** (in a timely way). Next, consider:

$$W_q = \lambda / \mu (\mu - \lambda) ,$$

Where W_q is the average time a need waits before it is served. For a purchase intention to be more credible, it would be desirable to have a low W_q value. The reason is that a large W_q value implies a lengthy waiting time, on average, until a need gets its turn to

be serviced by the purchase of a relevant product. There are two immediate problems associated with a large W_q regarding the credibility of a respondent's indication of an intention to buy **P** (i.e., positive response). One of these problems has to do with the purchase time, in that the customer's purchase of **P** may occur far too late to be of any use to the survey. The other problem is that the purchase may not occur at all due to a long waiting time, and this may cause forgetfulness and/or create various other obstacles to purchase action such as loss of interest.

Finally), we could also consider Δ (ρ), the so-called "traffic intensity factor" in queueing terminology, for further characterization of IBP:

$$\Delta = (\lambda/\mu)$$

This will indicate if the respondent's queueing mechanism to serve his/her needs is at a high or low state of being busy. If the Δ value is high, this should be taken as yet another indicator that the respondent's ability to satisfy and dispense with his/her product needs is slow in coming. Hence, such a respondent's positive intention to purchase the product in question in the survey should not be viewed as being quite realistic and/or credible.

A Numerical Example and Comparison:

To give a more concrete picture of the foregoing discussion, consider the scenario presented in Table 1 for purposes of **IBP** description and comparison. It should be noted that a data set such as the one in Table 1 has to be product category specific. Also note it is assumed that both respondents in Table 1 have expressed a positive intention to purchase product **P** in a purchase-intention survey.

It is also important to keep in mind that λ and μ are average rates (average input parameter values). As such, they are measured in "number of needs" — that is, "average number of needs surfacing" and "average number of needs that can be serviced," respectively, per unit time (per week, month, etc.). Consequently, L_q is measured in number of needs and W_q is measured, for each need, in the time unit used for λ and μ (e.g., see McMillan and Gonzalez, 1973, chapter 9 for further reference on units of measure). On the other hand, $Pr(0)$ and Δ are unitless quantities.

Table 1. Hypothetical Data for Numerical Comparison

	Respondent 1	Respondent 2
Δ	9.000	5.000
μ	10.000	30.000
$Pr(0)$	0.100	0.830
N_q	8.100	0.030
W_q	0.900	0.007
Δ	0.90	0.17

Based on the foregoing discussion, we can develop the **IBP** descriptions of these two respondents, and qualify the credibility of each respondent's intention to buy **P**. Accordingly, we see that Respondent 1 with $Pr(0)=0.10$ (or $\Delta=0.90$) already has a "very

busy" cognitive queue compared to Respondent 2 whose mental queue of needs has a high $Pr(0)=0.83$ value for being idle (i.e., for having no incumbent need), or alternatively, a low Δ value of 0.17 for being busy. Indeed, Respondent 1 already has 8.1 unsatisfied needs waiting in the queue for their turn for service to come, while Respondent 2, with a low L_q of 0.03, has practically no need waiting (as supported by the high probability, $Pr(0)=0.83$, of zero incumbent need for this respondent). As a result, Respondent 2 is much more likely to make the purchase of product **P** (on average, within $W_q=0.007$ time units), whereas Respondent 1, with $W_q=0.9$, has to keep the servicing of each need waiting almost a complete time unit. The obvious conclusion resulting from this comparative **IBP** descriptions is that, Respondent 1's purchase intention is not very credible; and certainly far less credible than Respondent 2's.

Caveats and Discussion

The appropriateness of standard queueing formulae and/or tabulated values is subject to a general stationarity and steadiness of purchase behaviour in a given marketing environment. In most cases, this is unlikely to be compromised since customers included in an intention study have probably purchased a variety of products over a lengthy period of time. However, for an innovative product which is totally new in the experience of customers, it could be difficult to contend with the appropriateness of this assumption.

It is also worthwhile to note that the arrivals (or calling units) in this cognitive queueing framework are the needs which are inanimate/abstract beings. That is, they are one's (product) needs that the individual is aware of (at some level of awareness). As such, the needs are "patient customers" in the terminology of queueing. However, renegeing on intentions might still occur as a result of such factors as the consumer's forgetfulness and change in need(s) over time. Indeed, this point relates to the "reasonable" time requirement mentioned in footnote 2 (also see Schmittlein and Morrison 1985).

The estimation of rates λ and μ would require empirical investigation that is beyond the scope and purpose of this conceptual paper. We acknowledge that, of these two parameters, it seems λ is the more challenging to estimate being the more abstract one in nature. However, in today's world of extensive consumer databases, scanner data, and data mining techniques, together with an increasingly viable microsegmentation, the estimation of these parameters should be possible with a reliable degree of accuracy. Writing on a related topic, Schmittlein and Morrison, (1985) indicate that in most real settings, it is impractical to estimate parameter μ separately for each individual. Instead, Schmittlein and Morrison (1985) suggest an estimation procedure based on a sample of customers. The context of the Schmittlein and Morrison (1985) article is on actual purchases rather than intentions, and is not couched in a cognitive framework. However, the approach they suggest would be helpful. Note that rate μ obtained from a sample can be adjusted up or down, capable of being individualized based on a consumer's demographic and income data.

An argument on the dependence and/or independence of λ and μ might be raised. We tend to think that need recognition occurs independently of purchase action. On the other hand, the requirement of arrival rate being smaller than the service rate, ($\lambda < \mu$), needs careful verification for the $M/M/1$ model to be used. This might be a vulnerable constraint in a number of cases, especially if the independence of λ and μ assumption is to

be accepted. But such a violation presents no problem for us, because if λ is found to be greater than μ for a respondent, then this finding could be viewed as a definite signal to discredit such an individual's positive purchase intention. This is because, if $\lambda > \mu$, then the respondent would already be experiencing serious difficulty in servicing his/her ever-increasing backlog of incumbent needs, let alone address the new needs that surface (except for emergency needs). Therefore, from the marketing point of view, the **IBP** interpretations based on this queueing approach preserve integrity by being robust to the violation of the M/M/1 requirement of ($\lambda < \mu$).

The approach suggested in this paper is not limited to the Poisson process case, although the Poisson category of queueing models --such as the M/M/1 model-- would be versatile, convenient and simple in addressing nonlinearities that exist in buyer behaviour (Laroche and Howard, 1980). In fact, as we pointed out earlier, for different product categories, different queueing models could be adopted with careful selection. Furthermore, given the well-grounded presumption that consumers classify products/services (convenience, shopping, etc.) into processing categories (routinized response, high and low involvement), we can view --at any given moment in time-- an individual's overall purchase disposition for several different products as a complex collection --or perhaps, a network-- of queues, where each queue is considered appropriate for a particular category of product and/or need. To capture the different levels of urgency of needs, this collection may also include "priority queue" models. That is, the queueing approach, in general, offers richness. Nonetheless, we advocate that single server queueing model should be appropriate in most cases, since a respondent would ordinarily service his/her own needs by his/her own means (i.e., the occasional receipt of gifts, contributions, etc. should not compromise the use of single server models).

Finally, the approach can readily lend itself to an evaluation of managerial variables and marketing decisions, as they can have a direct effect on μ . For example, price reductions and/or promotional activities might lead to changes in the **IBPs**. This is considered by Zufryden (1980), and Wagner and Taudes (1986), when examining the saliency of managerial variables and the effect of changes in them. However, as Katona (1951) points out, willingness to spend is the ultimate determining factor for a purchase to occur.

Concluding Remarks

Although the purchase intention discrepancy is well documented, there does not appear a method addressing its a-priori control at the time of data collection. Available works advocate follow up (ex-post) studies to enhance the experiential value of a survey that was completed, rather than add a discrepancy-reducing practicality to a survey that is currently under way.

The contribution of queueing theory as outlined in this paper also helps to overcome a problem that is part and parcel of existing purchase-intention assessment. The determination and availability of **IBP** avoids the highly subjective practice of adjusting downward purchase-intention survey results (Jamieson and Bass, 1989) by some rule-of-thumb (e.g., 50% of "will definitely purchase" will purchase, etc.). That is, **IBP** provides a quantified basis in describing/assessing a respondent's prospect of being truthful to his/her

response (particularly when the response is positive which is the main source of purchase-intention discrepancy).

The remarkable contribution of consumer databases together with an increasingly viable microsegmentation strategy (particularly for electronic commerce over the Internet, and use of scanner data) makes the step from intention to purchase ever more important. That is, marketers have access to individual consumers that significantly helps to facilitate the estimation of λ and μ , and enhance the use of the queueing approach we offer in this paper.

In summary, queueing and **IBP** descriptions may be a valuable tool to assess a credible versus an unrealizable purchase intention. Such assessments may lead to the elimination of unlikely purchase intentions in advance (i.e., reduce intention gap), and so, improve market forecasts drawn from intention studies, thereby resulting in cost savings in market research and promotion. Queueing models seem to offer potential for this purpose. Furthermore, the versatility of this approach in view of the rich collection of queueing models and the availability of numerous results (i.e., formulae and/or tabulated values) suggests a new and useful application of queueing theory. Therefore, it is hoped that this work will stimulate further research in an area where need for improvement is unquestionably evident. Purchase-intention discrepancy is of serious dimensions, and requires some (a-priori) control and reduction.

Bibliography

1. Bagozzi, R.P. **A Holistic Methodology for Modelling Consumer Response to Innovation**, *Operations Research*, 31, 1983, pp. 128-176
2. Belk, R.W. **Issues in the Intention-Behavior Discrepancy**, in Sheth, J.N. (Ed.), "Research in Consumer Behavior", 1, Greenwich, CT, JAI Press, 1985, pp.1-34
3. Bemmaor, A.C. **Predicting Behaviour from Intention-to-Buy Measures: The Parametric case**, *Journal of Marketing Research*, 32, 1995, pp. 176-191
4. Bird, M., and Ehrenberg, A.S.C. **Intentions-to-Buy and Claimed Brand Usage**, *Operations Research Quarterly*, 17, 1996, pp. 27-46
5. Boulding, W., Karla, A., Staelin, R. and Zeithaml, V.A. **A Dynamic Process Model of Service Quality: From Expectations to Behavioural Intentions**, *Journal of Marketing Research*, 36, 1993, pp. 171-186
6. Chardon, P., Morowitz, V.G. and Reinartz, W.J. **Do Intentions Really Predict Behaviour? Self-Generated Validity Effects in Survey Research**, *Journal of Marketing*, 69, 2005, pp. 1-14
7. Chatfield, C. and Goodhart, G.J. **A Consumer Purchasing Model with Erlang Inter-Purchase Times**, *Journal of the American Statistical Association*, 68, 1973, pp. 828-835
8. Clawson, C.J. **How Useful are 90 Day Purchase Probabilities?**, *Journal of Marketing*, 35, 1971, pp. 43-47
9. Cooper, R.B. **Introduction to Queueing Theory**, New York, North-Holland, 1981
10. Cronin, J. and Taylor, S. **Measuring Service Quality: A Re-examination and Extension**, *Journal of Marketing*, 56, 1992, pp. 55-58
11. Dalal, S.R., Lee, J.C. and Sabavala, D.J. **Prediction of Individual Buying Behavior: A Poisson-Bernoulli with Arbitrary Heterogeneity**, *Marketing Science*, 3, 1984, pp. 352-367
12. Engel, J., Blackwell, R. and Kollat, D. **Consumer Behavior**, 3rd Edition, Hinsdale, Illinois, Dryden Press, 1978
13. Ferber, R. and Priskie, R. **Subjective Probabilities and Buying Intentions**, *Review of Economics and Statistics*, 47, 1965, pp. 322-325
14. Fishbein, M. and Ajzen, I. **Belief, Attitude, Intention, and Behaviour**, Reading, MA, Addison-Wesley, 1975

15. Fisher, R.J. **Social Desirability Bias and Validity of Indirect Questioning**, *Journal of Consumer Research*, 20, 1993, pp. 303-326
16. Fitzsimons, G.J. and Morowitz, V.G. **The Effect of Measuring Intent on Brand-Level Purchase Behaviour**, *Journal of Consumer Research*, 23, 1996, pp. 1-11
17. Granbois, D., and Summers, J.O. **Primary and Secondary Validity of Consumer Purchase Probabilities**, *Journal of Consumer Research*, 1, 1975, pp. 31-38
18. Gupta, S. and Morrison, D.G. **Estimating Heterogeneity in Consumers' Purchase Rates**, *Marketing Science*, 10, 1991, pp. 264-269
19. Herniter, J. **A Probabilistic Market Model of Purchase Timing and Brand Selection**, *Management Science*, 18, 1971, pp. 102-113
20. Heyman, D.P. and Sobel, M.J. **Stochastic Models in Operations Research: Stochastic Processes and Operating Characteristics**, New York, McGraw-Hill, 1982
21. Howard, J.A. and Sheth, J.N. **A Theory of Buyer Behavior**, in Moyer, R. (ed.) "Changing Marketing System", *Proceedings of the 1967 Winter Conference of the American Marketing Association AMA*, 1967, pp.253-262
22. Howard, J.A. and Sheth, J.N. **The Theory of Buyer Behavior**, New York, Wiley, 1969
23. Infosino, W.J. **Forecasting New Product Sales from Likelihood of Purchase Ratings**, *Marketing Science*, 5, 1986, p. 375
24. Jamieson, L.F. and Bass, F.M. **Adjusting Stated Intention Measures to Predict Trial Purchases of New Product: A Comparison of Models and Methods**, *Journal of Marketing Research*, 26, August 1989, pp. 336-345
25. Juster, F.T. **Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design**, *Journal of the American Statistical Association*, 61, 1966, pp. 658-696
26. Kalwani, M.U. and Silk, A.J. **On the Reliability and Predictive Validity of Purchase Intention Measures**, *Marketing Science*, 1, 1982, pp. 243-287
27. Katona, G. *Psychological Analysis of Economic Behavior*, New York, McGraw-Hill, 1951
28. Kendall, D.G. **Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain**, *Annals of Mathematical Statistics*, 24, 1953, pp. 338-354
29. Laroche, M. and Howard, J.A. **Nonlinear Relations in a Complex Model of Buyer Behavior**, *Journal of Consumer Research*, March, 1980, pp. 377-388
30. Longman, K.A. **Promises, Promises**, in Adler, L. and Crespi, L. (eds.) "Attitude research on the rocks", *American Marketing Association*, 1968, pp. 28-37
31. Lilien, G.L. and Kotler, P. **Marketing Decision Making: A Model Building Approach**, New York, Harper and Row, 1983
32. Lipman, J. **Single-Source Ad Research Heralds Detailed Look of Households Habits**, *The Wall Street Journal*, Tuesday, February 10, 1988, p. 39
33. Loewenstein, G., and Adler, D. **A Bias in Predictions in Tastes**, *The Economic Journal*, 105, 1995, pp. 929-937
34. Manski, C.F. **The Use of Intentions Data to Predict Behaviour: A Best-Case Analysis**, *Journal of the American Statistical Association*, 85, 1990, pp. 934-940
35. Mittal, V., and Kamakura, W.A. **Satisfaction, Repurchase Intent, and Repurchase Behaviour: Investigating the Moderating Effect of Customer Characteristics**, *Journal of Marketing Research*, 38, 2001, pp. 131-142
36. McDaniel, C. and Gates, R. **Contemporary Marketing Research**, St. Paul, MN, West Publishing Company, 1991
37. McNeil, J. **Federal Programs to Measure Consumer Purchase Expectations, 1946-73: A Post-Mortem**, *Journal of Consumer Research*, December 1974, pp. 1-15
38. McMillan, C. and Gonzalez, R.F. **Systems Analysis: A Computer Approach to Decision Models**, Homewood, IL, Irwin, 1973
39. Morowitz, V.G. and Schmittlein, D. **Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which Intenders Actually Buy?**, *Journal of Marketing Research*, 29, 1992, pp. 391-405
40. Morowitz, V.G., Steckel, J. and Gupta, A. **When do Purchase Intentions Predict Sales?**, (Working Paper, Stern School of Business, New York University, New York, 1996)

41. Morrison, D.G. **Purchase Intentions and Purchase Behaviour**, *Journal of Marketing*, 43, 1979, pp. 65-74
42. Mullett, G.M. and Karson, M.J. **Analysis of Purchase Intent Scales Weighted by Probability of Actual Purchase** *Journal of Marketing Research*, 22, February 1985, pp. 93-96
43. Murray, J. **Canadian Consumer Expectations Data: An Evaluation**, *Journal of Marketing Research*, 6, February 1969, p. 60
44. Namias, J. **Intentions to Purchase Compared with Actual Purchases of Household Durables**, *Journal of Marketing*, 24, 1959, pp. 26-30
45. Newberry, C.R., Kleinz, B.R. and Boshoff, C. **Managerial Implications of Predicting Purchase Behaviour from Purchase Intentions: A Retail Patronage Case Study**, *Journal of Services Marketing*, 17, 2003, pp. 609-618
46. Parasuraman, A., Zeithaml, V. and Berry, L. **The Behavioural Consequences of Service Quality**, *Journal of Marketing*, 60, 1996, pp. 31-46
47. Peter, J.P. and Olson, J.C. **Consumer Behavior and Marketing Strategy**, Homewood, IL, Irwin, 1996
48. Pickering, J.F. and Isherwood, B.C. **Purchase Probabilities and Consumer Durable Buying Behaviour**, *Journal of the Market Research Society*, 16, 1974, pp. 203-226
49. Schmittlein, D.C. and Morrison, D.G. **Is the Customer Still Active?**, *American Statistician*, 39, 1985, pp. 291-295
50. Sheppard, B.H., Hatwick, J. and Warshaw, P.R. **The Theory of Reasoned Action: A Meta Analysis of Past Research with Recommendations for Modifications and Future Research**, *Journal of Consumer Research*, 15, 1988, pp. 325-343
51. Sweeny, J.C., Soutar, G.N. and Johnson, L.W. **The Role of Perceived Risk in the Quality-Value Relationship: A Study in a Retail Environment**, *Journal of Retailing*, 75, 1999, pp. 77-105
52. Taylor, J.W., Houlahan, J.J. and Gabriel, A.C. **The Purchase Intention Question in New Product Development**, *Journal of Marketing*, 39, 1975, pp. 90-92
53. Tull, D.S. and Hawkins, D.I. **Marketing Research, Measurement and Method**, New York, MacMillan Publishing Company, 1987
54. Urban, G.L., and Hauser, J.R. **Design and Marketing of New Products**, 2nd Edition, Englewood Cliffs, NJ, Prentice-Hall, 1993
55. Urban, G.L., Hauser, J. and Dholakia, N. **Essentials of New Product Management**, Englewood Cliffs, N.J., Prentice-Hall, 1987
56. Wagner, U. and Taudes, A. **A Multivariate Polynomial Model of Brand Choice and Purchase Incidence**, *Marketing Science*, 5, 1986, pp. 219-244
57. Wright, M. and MacRae, M. **Bias and Variability in Purchase Intention Scales**, *Journal of the Academy of Marketing Science*, 35, 2007, pp. 617-624
58. Wright, M., Sharp, A. and Sharp, B. **Market Statistics for the Dirichlet Model: Using the Juster Scale to Replace Panel Data**, *International Journal of Research in Marketing*, 19(1), 2002, pp. 81-90
59. Warshaw, P.R. **Predicting Purchase and Other Behaviours from General and Contextually Specific Intentions**, *Journal of Marketing Research*, 17, 1980, pp. 26-33
60. Zufryden, F.S. **A Multivariate Stochastic Model of Brand Choice and Market Behaviour**, in Sheth, J.N. (ed.) "Research in Consumer Marketing", 3, Greenwich, CT, JAI Press, 1980, pp. 273-303

¹ No attempt has been made to discuss wants along with needs in this paper as it would unduly complicate the presentation due to both the differences and similarities.

² The word "reasonable" is used with caution. Undoubtedly, what is a reasonable period of time for a purchase to occur is a topic of discussion unto itself.

IMPROVING THE PERFORMANCE OF SPARSE LU MATRIX FACTORIZATION USING A SUPERNODAL ALGORITHM

Bogdan OANCEA

PhD, Associate Professor, Artifex University, Bucharest, Romania

E-mail: oanceab@ie.ase.ro

Abstract: *In this paper we investigate a method to improve the performance of sparse LU matrix factorization used to solve unsymmetric linear systems, which appear in many mathematical models. We introduced and used the concept of the supernode for unsymmetric matrices in order to use dense matrix operations to perform the LU factorization for sparse matrices. We describe an algorithm that uses supernodes for unsymmetric matrices and we indicate methods to locate these supernodes. Using these ideas we developed a code for sparse LU matrix factorisation. We conducted experiments to evaluate the performance of this algorithm using several sparse matrices. We also made comparisons with other available software packages for sparse LU factorisation.*

Key words: sparse matrices; linear algebra; LU factorisation

1. Introduction

The numerical solution for large sparse linear systems lies at the heart of many engineering and scientific applications like macroeconomic models, linear programming, semiconductor device simulations, computational fluid dynamics.

The problem of solving sparse symmetric positive definite systems of linear equations on sequential processors is fairly well understood. Normally, the solution process is performed in two phases:

- First, the matrix A is factorized, $A = LU$ where L is a lower triangular matrix with 1s on the main diagonal and U is an upper triangular matrix; in the case of symmetric positive definite matrices, we have $A = LL^t$.
- Second, we have to solve two linear systems with triangular matrices: $Ly = b$ and $Ux = y$.

While the problem of solving sparse symmetric positive definite systems of linear equations is well understood, for unsymmetric systems it is difficult to design high performance algorithms because the pivoting needed in LU factorization generates dynamic and unpredictable amount of work and intermediate results.

For positive definite systems, the solution is computed in three phases:

- Symbolic factorization to determine the nonzero structure of the Cholesky factor;
- Numeric factorization;

- Solution of two triangular systems;

Elimination trees (Liu, 1990) are the standard way to reduce the time and space for symbolic factorization. For numeric factorization there are two high performance solutions: the supernodal method and the multifrontal method (Duff, 1983). Supernodal and multifrontal methods allow the use of dense vector operations for nearly all of the floating-point computation, thus reducing the symbolic overhead in numeric factorization. Overall, the Megaflop rates of modern sparse Cholesky codes are nearly comparable to those of dense solvers.

2. The generalization of supernodes for unsymmetric LU factorization

For unsymmetric systems, where pivoting is required to maintain numerical stability, the performances of the software packages are below the ones for symmetric systems.

The research has concentrated on two basic approaches for unsymmetric systems: submatrix-based methods and column-based methods.

Submatrix methods typically use some form of Markowitz ordering with threshold pivoting, in which each stage's pivot element is chosen from the uneliminated submatrix by criteria that attempt to balance numerical quality and preservation of sparsity.

Column methods typically use ordinary partial pivoting. The pivot is chosen from the current column according to numerical considerations alone; the columns may be reordered before factorization to preserve sparsity. In column methods, the reordering for sparsity is completely separate from the factorization, just as in the symmetric case. This is an advantage when several matrices with the same nonzero structure but different numerical values must be factored. However, symbolic factorization cannot be separated from numeric factorization, because the nonzero structures of the factors depend on the numerical pivoting choices. Thus column codes must do some symbolic factorization at each stage.

In this paper we describe a solution of sparse LU factorization using a left looking column method with partial pivoting.

In our method, reordering for preserving sparsity is completely separate from the factorization process, but symbolic factorization cannot be separated from numeric factorization because the structure of factors changes dynamically due to the pivoting. In order to speedup the numerical factorization we use a generalized version of supernodes for unsymmetric matrices.

The idea of a supernode is to group together columns with the same nonzero structure, so they can be treated as a dense matrix for storage and computation. Supernodes were originally used for symmetric sparse Cholesky factorization. In the factorization $A = LL^T$ (or $A = LDL^T$), a supernode is a range $(r:s)$ of columns of L with the same nonzero structure below the diagonal; that is, $L(r:s; r:s)$ is full lower triangular and every row of $L(s:n; r:s)$ is either full or zero.

Using the supernodes improves the LU factorisation because of the following (we considered the influence of supernodes on the left-looking LU factorization) (Gilbert, 1993):

1. The inner loop has no indirect addressing and thus sparse Level 1 BLAS is replaced by dense Level 1 BLAS;
2. The outer loop can be unrolled to save memory references and Level 1 BLAS can be replaced by level 2 BLAS operations;

3. Elements of the source supernode can be reused in multiple columns of the destination supernode to reduce cache misses - level 2 BLAS is replaced by Level 3 BLAS.

For a symmetric positive definite matrix, a formal definition of a supernode can be given in terms of elimination tree (Liu, 1990). A supernode is a maximal set of contiguous nodes $\{j, j+1, \dots, j+w\}$ such that

$$Adj_G(T[j]) = \{j+1, j+2, \dots, j+w\} \cup Adj_G(T[j+w])$$

In matrix terms, $Adj_G(T[j])$ indicates the nonzero elements in column j of the factor L . In other words, a supernode is a maximal block of contiguous columns in the Cholesky factor L where these columns have identical nonzero structure below the diagonal and the corresponding block diagonal is full triangular.

It is possible to use a generalization of the supernode concept for the unsymmetric matrices. In this case, a supernode is a range $(k:t)$ of columns of L with the triangular diagonal block full and the same structure below the diagonal block. In figure 1 we have represented a sparse matrix A and its factors $F = L + U$. Using the above definition of supernodes, matrix A has 7 supernodes : $\{1, 2, 3\}$, $\{4\}$, $\{5\}$, $\{6\}$, $\{7, 8\}$, $\{9\}$, $10, 11, 12\}$.

The effect of supernodes in Cholesky factorization is that all updates from columns belonging to a supernode are aggregated into a dense vector before the sparse update of the current column. This process of columns update can be implemented using Level 2 BLAS matrix-vector multiplications. It is even possible to use supernode-supernode update which can be implemented with Level 3 BLAS operations.

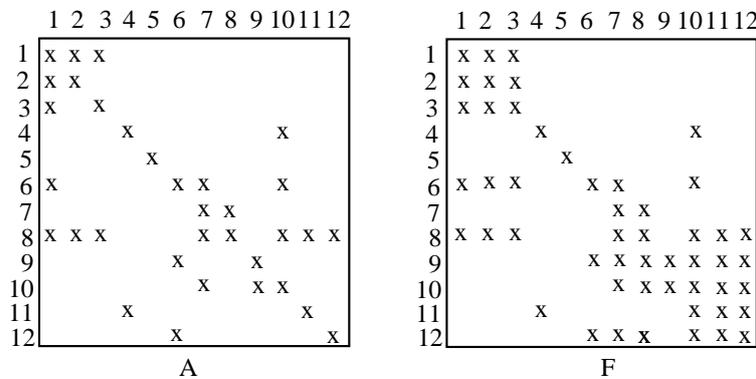


Figure 1. An example of a sparse matrix and its factors L and U

Supernodes are only useful if they actually occur in practice. The occurrence of symmetric supernodes is related to the clique structure of the chordal graph of the Cholesky factor, which arises because of fill during the factorization. Unsymmetric supernodes are harder to characterize, but they also are related to dense submatrices arising from fill.

In sparse Cholesky factorization, supernodes can be determined during the initial symbolic factorization which is not the case of unsymmetric LU factorization where the nonzero structure cannot be determined before numeric factorization. In order to obtain larger supernodes we must permute the columns of L to bring together columns with the same nonzero structure.

In Cholesky factorization this reordering can be obtained by a postorder traversal of the elimination tree. The postorder etree is used to locate the supernodes before numerical factorization. The analog for etree in the case of an unsymmetric matrix is the *column elimination tree* (Gilbert, 1993). The column etree of A is the symmetric elimination tree of the column intersection graph of A , or equivalently the elimination tree of $A^T A$ provided there is no cancellation in computing $A^T A$ (Gilbert, 1993). The column etree can be computed from A in time almost linear in the number of nonzeros of A by a variation of an algorithm of Liu (Liu, 1990).

It can be shown that column elimination tree represents the dependence between columns of U and L , and for strong Hall matrices there is no other information obtainable from the structure of the original matrix A (Gilbert, 1993). Before we factor the matrix A we determine its column elimination tree and permute its columns according to a postorder on the tree.

In figure 2 we present a description of the algorithm used for unsymmetric LU factorization. The advantage of the supernode-column updates consists in using efficient Level 2 BLAS operations, while in the simple column algorithm there are only Level 1 BLAS operations. The presence of the pivoting process precludes the separation of symbolic from the numeric factorization. Thus, the symbolic factorization for each column is computed just before the numeric factorization of that column. The structure of the column j can be determined by a traversal of the directed graph G associated with matrix $L(:, 1:j-1)^T$ (Gilbert, 1993). The depth-first traversal of G determine the structure of $U(:, j)$ and the columns of L that updates the column j . At the end of the symbolic factorization of column j we can determine very easy if column j is part of the same supernode as column $j-1$.

```

for each column j = 1 to n do
  c = A(:,j); c is the current column
  Determine the set S of the supernodes from L which
  updates column j;
  Verify if column j is part of the same supernode as
  column j-1;
  for each supernode in S do
    c(k:t) = L(k:t, k:t)-1c(k:t);
    c(k+1:n)=c(k+1,n)- L(k+1:n, k:t) c(k:t);
  endfor
  find the index i of the pivot, c(i) = maxi=j:nc(i)
  swap c(i) and c(j);
  U(1:j,j)= c(1:j), L(j:n) = c(j:n)/c(j);
endfor

```

Figure 2. The supernodal LU factorization algorithm

3. Experimental results

We have implemented the supernodal LU factorization using the C programming language and we have conducted a series of experiments on several sparse matrices. We have used a system with an INTEL CORE 2 processor at 1.6 GHz with 1 GB of main memory. The experimental matrices are from the Harwell Boeing collection. These matrices are presented in table 1 and they are factored using the natural order and using an approximate minimum degree ordering algorithm (Davis, 2002).

Table 1. Test matrices: **n** is the number of lines/columns and **nnz** is the number of nonzero elements

Matrix	n	nnz(A)	nnz(A)/n
SHERMAN3	5005	20033	4
SHERMAN5	3312	20793	6.27
FS7601	760	5976	7.86
FS7602	760	5976	7.86
FS7603	760	5976	7.86
ORSIRR2	886	5970	6.73

In our implementation, we used the ATLAS (Automatically Tuned Linear Algebra Software) library as a high performance implementation of the BLAS.

Table 2 presents the performance of the supernodal LU factorization with natural ordering of the matrix A and table 3 presents the case with matrix A reordered with the approximate minimum degree algorithm. This results show that the approximate minimum degree ordering improves the sparsity of factors as we expected. The *mflops* rate of sparse supernodal factorization is about 70% of the dense LU factorization. The dense factorization performance was measured with the standard ATLAS - BLAS implementation and LAPACK package.

Table 2. Performance of supernodal LU factorization using natural ordering

Matrix	MFLOPS	nnz(L)	nnz(U)
SHERMAN3	212.98	548571	548571
SHERMAN5	280.06	409095	998665
FS7601	291.77	201627	207257
FS7602	291.78	207378	208907
FS7603	296.08	207378	208024
ORSIRR2	200.01	66027	79763

Table 3. Performance of supernodal LU factorization using the approximate minimum degree ordering

Matrix	MFLOPS	nnz(L)	nnz(U)
SHERMAN3	279.57	184086	184086
SHERMAN5	251.80	93357	118436
FS7601	161.9	16535	18206
FS7602	161.08	17094	19479
FS7603	157.80	17532	52003
ORSIRR2	200.01	66027	79763

We compared the performance of the supernodal LU factorization implementation with a multifrontal factorization package – UMFPACK 5.2. UMFPACK uses a multifrontal algorithm. Where the outer loop of a left-looking algorithm like supernodal LU is over columns of the factors being computed, the outer loop of a multifrontal algorithm is over pivots being eliminated. All the updates created when a block is eliminated are computed at once and stored as a dense update matrix. Before a block of pivots is eliminated, all the update matrices contributing to that block are summed into a frontal matrix. The elimination step can use Level 2 or Level 3 BLAS because the arithmetic is carried out on the dense frontal matrix. Some extra intermediate storage is needed to record update matrices that have not yet been assembled into frontal matrices, and some extra data movement is needed for the assembly. UMFPACK does not use a column reordering; rather, it chooses

row and column pivots to balance considerations of stability and sparsity by using approximate Markowitz counts with a pivot threshold. In principle, the pivot threshold can lead to a less accurate solution than strict partial pivoting; in practice, the lost accuracy can usually be retrieved by iterative refinement of the solution. In principle, the freedom to choose both row and column pivots dynamically could lead to sparser factors than strict partial pivoting; in practice, some matrices have sparser factors by one method and some by the other.

UMFPACK does not include an initial column ordering step. For the initial column ordering in supernodal LU factorization, we ran the approximate minimum degree algorithm (Davis, 1997) on the structure of $A^T A$. We reported times for ordering and factorization separately. In applications where many matrices with the same nonzero structure but different values are factored, the cost of column ordering can be amortized over all the factorizations; in applications where only a single matrix is to be factored, preordering is part of the solution cost.

Table 4 and 5 shows the time needed for LU factorisation using our supernodal algorithm and the UMFPACK. Table 6 shows the memory requirements for the two approaches.

Table 4. Performance of supernodal LU factorization and UMFPACK using natural ordering

Matrix	Supernodal Factorization (msec)	UMFPACK Factorization (msec)
SHERMAN3	560	710
SHERMAN5	670	780
FS7601	48	57
FS7602	51	55
FS7603	45	72
ORSIRR2	54	53

Table 5. Performance of supernodal LU factorization and UMFPACK using the approximate minimum degree ordering

Matrix	Supernodal Factorization (msec)	UMFPACK Factorization (msec)
SHERMAN3	832	1002
SHERMAN5	932	1210
FS7601	75	89
FS7602	78	92
FS7603	70	81
ORSIRR2	80	93

Table 6. Memory requirements for LU factorization using the supernodal approach and UMFPACK.

Matrix	Supernodal Factorization (MB)	UMFPACK Factorization (MB)
SHERMAN3	3.67	5.87
SHERMAN5	3.83	6.01
FS7601	1.22	2.23
FS7602	1.25	2.32
FS7603	1.22	2.24
ORSIRR2	1.44	2.98

It can be observed from this figures that the supernodal approach performs slightly better for five from the six matrices used for tests. Although there is not a clear difference between two approaches, the supernodal algorithm can be used with succes for large sparse systems.

The supenodal LU factorization algorithm can be improved further on machines with a memory hierarchy by changing the data access pattern. The data we are accessing in the inner loop include the destination column j and all the updating supernodes to the left of column j . Column j is accessed many times, while each supernode is used only once. In practice, the number of nonzero elements in column j is much less than that in the updating supernodes. Therefore, the access pattern given by this loop provides little opportunity to reuse cached data. In particular, the same supernode may be needed to update both columns j and $j+1$. But when we factor the $(j+1)$ st column, we will have to fetch the same supernode again from memory, instead of from cache (unless the supernodes are small compared to the cache). To exploit memory locality, we factor several columns (say s of them) at a time in the outer loop, so that one updating supernode can be used to update as many of the s columns as possible.

4. Conclusions

In this paper we presented an algorithm for sparse LU factorization using a generalization of the supernode concept. The supernode for the unsymmetric matrices is a range $(k:t)$ of columns of L with the triangular diagonal block full and the same structure below the diagonal block. We developed an algorithm for matrix factorization using supernodes based on the clasical left-looking LU factorization.

We implemented this algorithm in a software package using the C programming language and ATLAS library as a high performance implementation of BLAS. We tested our implementation using some sparse matrices from the Harwell-Boeing collection. The results of the tests shows that the sparse LU factorization using supernodes can achieve about 70% of the performance for the dense LU factorization.

We also compared the sparse LU factorization using supernodes with a multifrontal factorization package –UMFPACK. The results shows that the supernodal approach performs slightly better than the multiforntal approach.

References

1. Davis, T.A., Gilbert, J.R., Larimore, S.I. and Ng, E. **Approximate minimum degree ordering for unsymmetric matrices**, SIAM Symposium on Applied Linear Algebra, October, 1997
2. Davis, T.A. **A column pre-ordering strategy for unsymmetric-pattern multifrontal method**, Technical Report TR-02-001, University of Florida, January, 2002
3. Demmel, W.D., Eisenstat, S.C., Gilbert, J.R., Li, X.S. and Liu, J.W.H. **A supernodal approach to sparse partial pivoting**, SIAM J. Matrix, Anal. Appl. Vol 20, no. 3, 1999
4. Duff, I., and Reid, J. **The multifrontal solution of indefinite sparse symmetric linear equations**, ACM Trans. Mathematical Software, 9, 1983
5. Eisenstat, S.C., Gilbert, J.R. and Liu, W. **A supernodal approach to sparse partial pivoting code**, Householder Symposium XII, 1993

6. George, J.A. and Liu, J. W. H. **Computer Solution of Large Sparse Positive Definite Systems**, Prentice-Hall, Englewood Cliffs, 1981
7. Golub, G. H. and Van Loan, Ch., **Matrix Computations**, Third Edition, The Johns Hopkins University Press, 1996
8. Liu, J. W. H. **The role of elimination trees in sparse factorization**, SIAM J. Matrix Analysis and applications, 11,1990, pp. 134-172
9. Gilbert, J. R. and Ng, E. **Predicting structure in nonsymmetric sparse matrix factorizations**, in George, A., Gilbert, J.R. and Liu, J. W. H. (editors) "Graph Theory and Sparse Matrix Computation", Springer-Verlag, 1993
10. Gilbert, J. R., and Liu, J. W. H. **Elimination structures for unsymmetric sparse LU factors**, SIAM J. Matrix Anal. Appl., 14, 1993
11. Grigori, L. and Li, X.S. **Performance Analysis of Parallel Right-Looking Sparse LU Factorization on Two Dimensional Grids of Processors**, Para'04 workshop on state-of-the-art in scientific computing, 2004
12. ***, <http://math-atlas.sourceforge.net/>

MONEY DEMAND IN ROMANIAN ECONOMY, USING MULTIPLE REGRESSION METHOD AND UNRESTRICTED VAR MODEL

Mariana KAZNOVSKY

PhD Candidate, University of Economics, Bucharest, Romania
Monetary and Financial Statistics Division, National Bank of Romania, Bucharest, Romania

E-mail: Mariana.Kaznovsky@bnro.ro

Abstract: The paper describes the money demand in Romanian economy using two econometrics models. The first model consist in a multiple regression between demand money, monthly inflation rate, Industrial production Index and the foreign exchange rate RON/Euro. The second model (Unrestricted Vector AutoRegressive model) is applied for the same variables used in the first model. Identifying a statistically strong model, capable of stable estimations for the money demand function in Romania's economy constitutes a prerequisite to the application of an efficient monetary policy.

Key words: money demand; unrestricted VAR model; Romania

Multiple regression estimation of Romanian money demand function

The theory underlying money demand function is based on the classical macroeconomic model of Hicks & Hansen IS-LM, specifically LM curve. The theoretical hypothesis (assumptions) of the dual equilibrium model for the money market in an open economy are: the perfect mobility of capital, uncovered interest rate parity principle, monetary policy conducted by the central bank are using the short term interest rate variable as the operational one without affecting the stability of the exchange rate of the national currency.

The LM curve is defined by the all possible combinations of interest rate and income levels for which the demand of money is equal with money supply (Figure 1.).

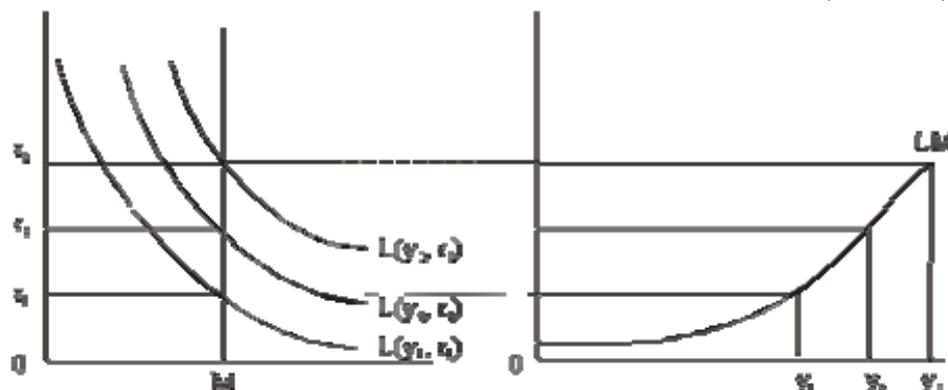


Figure 1. LM curve

The money demand function is a synthetic way of measuring the dependence between, on the one side, the monetary aggregates - as the money issued by the monetary financial institutions: credit institutions and money market funds, and used as financial resource by the non-banking entities: non-issuing money institutions, and, on the other side, the money consumers in the economy.

The classic model [3;4]¹ estimates this correlation by the degree of explanation of the endogenous variable "monetary aggregate" by the following exogenous variables: monthly price growth rate, value of the economic output (GDP, industrial production value), average passive interest rate practiced by the credit institutions as an expression of the "price of money" and other variables expressing the cost of opportunity for possessing the currency - like exchange rate, the dynamics of the domestic capital index or a foreign capital index related to the analyzed economy. Taking into account changes in the international oil markets as an indicator of foreign restrictions could be useful in explaining the money demand pattern.

The specific choice of variables used to estimate the demand of money depends on the working hypotheses, on the availability of data with adequate frequency, as well as on conclusions of previous studies and research works regarding the significance of correlations that point to one indicator being more reliable than others in approximating the variable.

In order to express the monetary aggregate in the Romanian economy, the choice has been made for the broad money indicator M2 (known as broad money up to 2007, after which M3 was introduced, M2 becoming the intermediary monetary aggregate). The explanation of the use of M2 resides in the higher degree of coverness of the financial instruments by this indicator. Narrow money M1 is almost designed to be a proxy measure of the exchange transaction incentives of money only, while broad money M2 is designed to quantify also the accumulation of value purpose of holding money.

Although, the exogenous variables have to be restricted to the most significant ones, thus avoiding multicollinearity. Out of purely practical reasons, the industrial production index has been selected to measure the economic output, whereas for the cost of money we considered significant the average interest rate for one month as a liability of monetary financial institutions. For medium and long term maturity we used the interest rate of the one year government bonds. Longer maturities have been left out because of the discontinuity in issuance, in relation to the investor lack of preference for medium and long term maturities.

As an indicator of price increases, we used the monthly Consumer Price growth rate, as the GDP deflator is available, at best, quarterly, starting with 1998.

Our study has been compensating for the inflationary component by studying the dependence between the deflated monetary aggregate and the real money demand factors.

For the following regressors the 't' statistic significance of the coefficients of the money demand function has been confirmed: industrial production index, real money balances as the log level recorded three months ago, monthly inflation rate and the foreign exchange rate (leu/Euro). The money demand elasticity in respect of interest rate (as the average cost of monetary financial institutions for the borrowed resources and, implicitly, as the rate of return of deposits made by non-banking entities with the banks) was not being confirmed at the 10% level of significance. Thus, the conclusions of some previous work papers that the interest rate channel is not efficiently working in the Romanian economy are confirmed by the statistical data [1]. The weak sensitivity of the real variables block could be

explained by the rigidity of the economy to the monetary impulses due to the specific structural changes in our emerging economy. On the other side, the National Bank of Romania's monetary policy was focused on the monetary aggregates (base money) as the operational target, the exogeneity of interest rate being a practical issue in the nominal variables block.

Thus, we have estimated the equation of the money demand for the Romanian economy using the following specific version:

$$m_t = a_0 + a_1 m_{t-1} + a_2 \text{prodind}_{t-3} - a_3 \text{rinfl}_t + a_4 \text{curs_eur}_t + \varepsilon_t$$

where:

m is the real monetary aggregate, deflated using the CPI (real M2) and seasonally adjusted, decimal logarithm values being considered; seasonally adjustment has been performed using the Census X11 method, considering the multiplicative method. The need to isolate and detach the seasonal component from the series of the monetary aggregate was imposed by the known peak effect during summer and holidays. Introducing unadjusted series would have led to rejection of the coefficients, the statistical significance being infirmed with a probability of 90%.

rinfl is the monthly CPI rate, logarithm values;

prodind is the Industrial Production Index, logarithm vales

curs_eur is the foreign exchange rate RON/Euro, logarithm values.

Table 1 presents the estimators of the regression coefficients, obtained using Eviews 4 software, for the time horizon for January 1992-December 2005, statistical series having monthly frequency.

Table 1. Estimators of the regression coefficients

Dependent Variable: LOG(M2_SA/p)

Method: Least Squares

Date: 02/27/08 Time: 15:22

Sample(adjusted): 1992:04 2005:12

Included observations: 165 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.192377	0.031722	6.064535	0.0000
LOG(m2_SA(-1)/p(-1))	0.951809	0.008594	110.7478	0.0000
LOG(rinfl)	-0.012963	0.002819	-4.599288	0.0000
LOG(PROD_IND(-3))	0.054574	0.027316	1.997866	0.0474
LOG(CURS_EUR)	0.048924	0.011085	4.413621	0.0000
R-squared	0.999853	Mean dependent var		4.075405
Adjusted R-squared	0.999850	S.D. dependent var		1.911885
S.E. of regression	0.023439	Akaike info criterion		-4.639032
Sum squared resid	0.087899	Schwarz criterion		-4.544912
Log likelihood	387.7201	F-statistic		272760.0
Durbin-Watson stat	1.789483	Prob(F-statistic)		0.000000

A powerful influence of the autoregressive component upon the deflated broad money has been detected (the estimated coefficient being 0.95); the value of the industrial production as an approximate value of the real aggregate supply, is positively correlated with monetary aggregate, but transmission of this influence is produced with a time lag of 3 months. Thus, the changes in real variables is reflected in values of the nominal variables

after 3 months, but the influence is not strong (estimated elasticity is 0.05: at a change in the industrial production index with 1%, the reaction of the broad money over 3 months is of size 0.05%). Estimation of simultaneous correlation of this link has been infirmed by the "t" test, at a probability level of 90%.

The opportunity cost of holding the money has been approximated by means of introducing the leu/Euro exchange rate: the equation confirms the positive correlation between the exchange rate and the real broad money. The national currency depreciation influences the growth of money demand in real terms, as a consequence of the considerable weight of the foreign currency denominated part of the monetary aggregate. The shifting from local currencies to USD/Euro, as a process of substitution of the national currency, is characteristic for emerging markets, marked by significant changes in economic structure, and for which the tax of holding the money (inflation rate) and inflationary expectations are high.

The inverse correlation between the inflation rate and the real broad money is statistically confirmed by the "t" test. Interesting to observe is the small influence of prices upon the real broad money.

Stationarity of the data has been verified with the ADF (Augumented Dicky-Fuller) test, for the case of a liniar trend, a constant and eight lags, corresponding to the timespan of january 1992-december 2005 (results are presented in table 2).

The degree of explanation brought by the exogenous variables in their entirety, contributes in proportion of 99% (adjusted R² coefficient) to the obtaining of values for the adjusted series of money demand, as is visible from figure 2.

The errors terms resulted from the regression, represented as a blue line has been tested for autocorrelation: the Durbin-Watson statistic confirms the rejection of the autocorrelation in the residual series; as a consequence, the regression parameters are relevant and statistically significant.

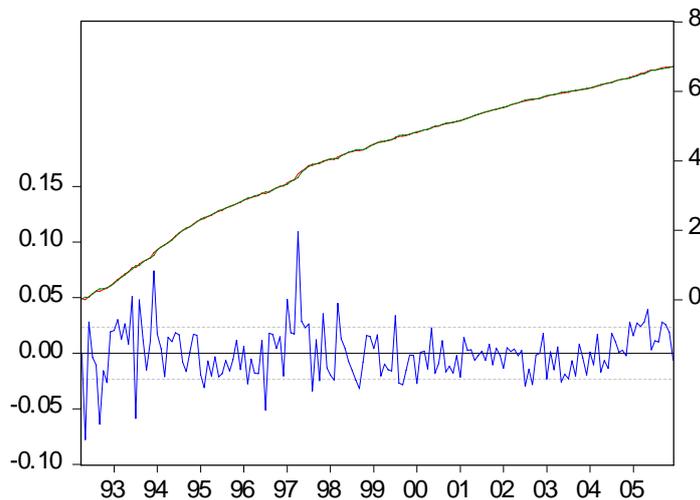


Figure 2. Adjusted versus real money demand

Tabel 2. Augmented Dicky-Fuller

Augmented Dickey-Fuller Test for real broad money M2/p				
Sample(adjusted): 1992:06 2005:12				
ADF Test Statistic	7.864602	1%	Critical Value*	-3.4715
		5%	Critical Value	-2.8792
		10%	Critical Value	-2.5761
*MacKinnon critical values for rejection of hypothesis of a unit root.				
Augmented Dickey-Fuller Test for industrial production index prod_ind				
Sample(adjusted): 1992:06 2005:12				
ADF Test Statistic	-7.769099	1%	Critical Value*	-3.4715
		5%	Critical Value	-2.8792
		10%	Critical Value	-2.5761
*MacKinnon critical values for rejection of hypothesis of a unit root.				
Augmented Dickey-Fuller Test for exchange rate leu/euro				
Sample(adjusted): 1992:06 2005:12				
ADF Test Statistic	-3.205847	1%	Critical Value*	-3.4715
		5%	Critical Value	-2.8792
		10%	Critical Value	-2.5761
*MacKinnon critical values for rejection of hypothesis of a unit root.				
Augmented Dickey-Fuller Test for inflation rate				
Sample(adjusted): 1992:06 2005:12				
ADF Test Statistic	-3.205847	1%	Critical Value*	-3.4715
		5%	Critical Value	-2.8792
		10%	Critical Value	-2.5761

* MacKinnon critical values for rejection of hypothesis of a unit root.

Estimating the reaction of the broad money to shocks in real economy variables (VAR model for the money demand in the economy)

An estimation of the correlation between the real exogenous variables and money aggregates based on the UVAR (Unrestricted Vector AutoRegressive) with three lags has been applied for the same variable used in the multiple regression. The system of simultaneous equations comprises thus, the following variables: real broad money M2 (seasonally adjusted levels), IPI (Industrial production index), inflation rate, leu/Euro foreign exchange rate. Series are monthly and covers the years 1992-2005; ADF stationarity tests have shown the stationarity of the series of broad money, foreign exchange rate and inflation rate with a probability of 95%.

The model specification is as follows:

$$\begin{aligned} \log(M_2_SA(t)/p(t)) &= a_0 + \sum_{i=1}^3 a_{1i} \log(M_2_SA(t-i)/p(t-i)) + \\ &+ \sum_{i=1}^3 a_{2i} \log(\text{prod_ind}(t-i)) + \sum_{i=1}^3 a_{3i} \log(\text{curs}(t-i)) + \sum_{i=1}^3 a_{4i} \log(\text{rinf l}(t-i)) + u_1(t) \\ \log(\text{prod_ind}(t)) &= b_0 + \sum_{i=1}^3 b_{1i} \log(M_2_SA(t-i)/p(t-i)) + \sum_{i=1}^3 b_{2i} \log(\text{prod_ind}(t-i)) + \\ &+ \sum_{i=1}^3 b_{3i} \log(\text{curs}(t-i)) + \sum_{i=1}^3 b_{4i} \log(\text{rinf l}(t-i)) + u_2(t) \\ \log(\text{curs}(t)) &= c_0 + \sum_{i=1}^3 c_{1i} \log(M_2_SA(t-i)/p(t-i)) + \sum_{i=1}^3 c_{2i} \log(\text{prod_ind}(t-i)) + \\ &+ \sum_{i=1}^3 c_{3i} \log(\text{curs}(t-i)) + \sum_{i=1}^3 c_{4i} \log(\text{rinf l}(t-i)) + u_3(t) \\ \log(\text{rinf l}(t)) &= d_0 + \sum_{i=1}^3 d_{1i} \log(M_2_SA(t-i)/p(t-i)) + \sum_{i=1}^3 d_{2i} \log(\text{prod_ind}(t-i)) + \\ &+ \sum_{i=1}^3 d_{3i} \log(\text{curs}(t-i)) + \sum_{i=1}^3 d_{4i} \log(\text{rinf l}(t-i)) + u_4(t) \end{aligned}$$

where $i=1,3$.

u_i ($i=1,4$) are the regression residuals called innovations or shocks. The corresponding innovation is, thus, that part of the evolution of the variable that neither be explained by its past values (own history), nor by other variables of the model.

The VAR method concentrates mainly on studying the impact of every shock upon every variable of the system of equations; this analysis is being performed by impulse response functions, by factorial decomposition of variance.

Table 3. comprises estimated coefficient of the VAR model, obtained with Eviews 4 software.

The impulsion response functions graphically represent the evolution of these shocks in time across 10 months, identifying the maximum impact upon variables taken into account by the model; the sizing of these dependencies between innovations and model variables is expressed in relative terms, that is, standard deviations of the shocks.

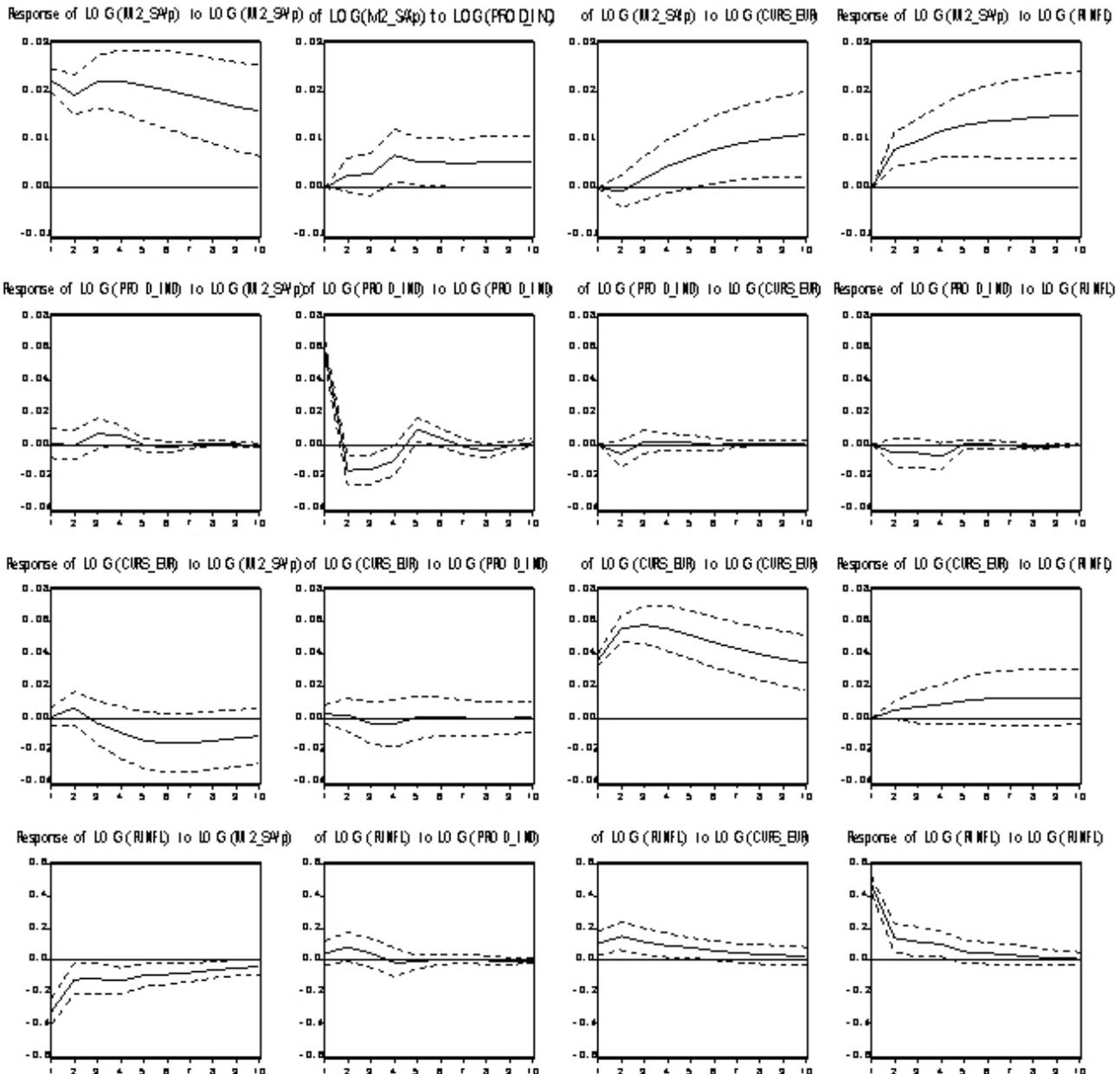
Table 3. The estimated coefficients of VAR model

Sample(adjusted): 1992:04 2005:12
Included observations: 165 after adjusting endpoints
Standard errors and t-statistic in brackets

	LOG(M2_SA/p)	LOG(PROD_IND)	LOG(CURS_EUR)	LOG(RINFL)
LOG(M2_SA(-1)/p(-1))	1.107870 (0.09407) (11.7771)	-0.151211 (0.24728) (-0.61150)	0.376171 (0.15506) (2.42604)	-1.267625 (2.45033) (-0.51733)
LOG(M2_SA(-2)/p(-2))	0.107563 (0.13958) (0.77064)	0.322834 (0.36690) (0.87989)	-0.960039 (0.23006) (-4.17291)	-1.195572 (3.63570) (-0.32884)
LOG(M2_SA(-3)/p(-3))	-0.215949 (0.09335) (-2.31338)	-0.246611 (0.24538) (-1.00501)	0.616877 (0.15387) (4.00918)	2.005835 (2.43154) (0.82492)
LOG(PROD_IND(-1))	0.034269 (0.02889) (1.18632)	-0.261478 (0.07593) (-3.44352)	-0.040848 (0.04761) (-0.85791)	1.036583 (0.75244) (1.37764)
LOG(PROD_IND(-2))	-0.017588 (0.02858)	-0.315671 (0.07512)	-0.105047 (0.04710)	0.614014 (0.74434)

	(-0.61547)	(-4.20240)	(-2.23021)	(0.82491)
LOG(PROD_IND(-3))	0.047984 (0.02894) (1.65781)	-0.313506 (0.07608) (-4.12050)	0.018482 (0.04771) (0.38740)	0.014463 (0.75393) (0.01918)
LOG(CURS_EUR(-1))	-0.073281 (0.04646) (-1.57730)	-0.128218 (0.12213) (-1.04987)	1.496446 (0.07658) (19.5411)	3.231345 (1.21018) (2.67013)
LOG(CURS_EUR(-2))	0.133757 (0.07548) (1.77219)	0.263857 (0.19840) (1.32991)	-0.706407 (0.12441) (-5.67820)	-3.294675 (1.96600) (-1.67583)
LOG(CURS_EUR(-3))	-0.061791 (0.04544) (-1.35983)	-0.050815 (0.11945) (-0.42542)	0.157700 (0.07490) (2.10550)	0.493682 (1.18363) (0.41709)
LOG(RINFL(-1))	0.016839 (0.00376) (4.47250)	-0.010827 (0.00990) (-1.09394)	0.011044 (0.00621) (1.77960)	0.295220 (0.09807) (3.01028)
LOG(RINFL(-2))	-0.001959 (0.00452) (-0.43299)	-0.006298 (0.01189) (-0.52963)	-0.011551 (0.00746) (-1.54912)	0.148286 (0.11784) (1.25840)
LOG(RINFL(-3))	-0.003161 (0.00389) (-0.81272)	-0.020569 (0.01023) (-2.01157)	0.011172 (0.00641) (1.74250)	0.153754 (0.10132) (1.51746)
C	0.073499 (0.03838) (1.91519)	0.173660 (0.10088) (1.72144)	-0.071412 (0.06326) (-1.12892)	0.423674 (0.99964) (0.42382)
R-squared	0.999864	0.237022	0.999321	0.668904
Adj. R-squared	0.999854	0.176787	0.999268	0.642765
Sum sq. resids	0.081358	0.562183	0.221043	55.20174
S.E. equation	0.023136	0.060816	0.038134	0.602635
Log likelihood	394.0993	234.6297	311.6410	-143.7914
Akaike AIC	394.2569	234.7872	311.7986	-143.6338
Schwarz SC	394.5016	235.0320	312.0433	-143.3891
Mean dependent	4.075405	-0.000154	-0.123119	-3.762357
S.D. dependent	1.911885	0.067029	1.409360	1.008271
Determinant Residual Covariance		4.92E-10		
Log Likelihood		831.6832		
Akaike Information Criteria		832.3135		
Schwarz Criteria		833.2923		

The responses of the variables studied to a standard deviation of innovations (variation interval ± 2 standard deviations) are graphically represented, for a timespan of 10 monhs.



Conclusions

Identifying a statistically strong model, capable of stable estimations for the money demand function in Romania's economy constitutes a prerequisite to the application of an efficient monetary policy.

Obtaining by econometric means, the series of adjusted money demand, for which the statistical stability tests are confirmed, allows for the formalization of the link between the real-sector and monetary block, as well as the impact assessment of the levels of monetary variables upon the economy.

Bibliography

1. Antohi, D., Udrea, I. and Braun, H. **Mecanismul de transmisie a politicii monetare in Romania**, Caiete studii BNR nr.13/2003
2. Boughton, J. M. **Long run money demand in large industrial countries**, International Monetary Fund, 1991
3. Friedman, M. **The optimum quantity of money and other essays**, Aldine, Chicago, 1969
4. Sriram, S. S. **A survey of recent empirical money demand studies**, International Monetary Fund, 2001
5. Treichel, V. **Broad money demand and monetary policy in Tunisia**, International Monetary Fund, 1997

¹ Codification of references:

[1]	Antohi, D., Udrea, I. and Braun, H. Mecanismul de transmisie a politicii monetare in Romania , Caiete studii BNR nr.13/2003
[2]	Boughton, J. M. Long run money demand in large industrial countries , International Monetary Fund, 1991
[3]	Friedman, M. The optimum quantity of money and other essays , Aldine, Chicago, 1969
[4]	Sriram, S. S. A survey of recent empirical money demand studies , International Monetary Fund, 2001
[5]	Treichel, V. Broad money demand and monetary policy in Tunisia , International Monetary Fund, 1997

Bogdan Vasile ILEANU

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

E-mail: bogdan.ileanu@csie.ase.ro; ileanub@yahoo.com

Web page: http://www.bodo.ase.ro

Key words: econometrics; applied econometrics; Vergil Voineagu; Emilia Titan; Radu Serban; Simona Ghita; Daniela Todose; Cristina Boboc; Daniel Pele

**Book Review on
ECONOMETRIC THEORY AND PRACTICE
("TEORIE SI PRACTICA ECONOMETRICA")
by Vergil VOINEAGU, Emilia TITAN, Radu SERBAN,
Simona GHITA, Daniela TODOSE, Cristina BOBOC, Daniel PELE
Editura Meteor Press, Bucharest, 2007**

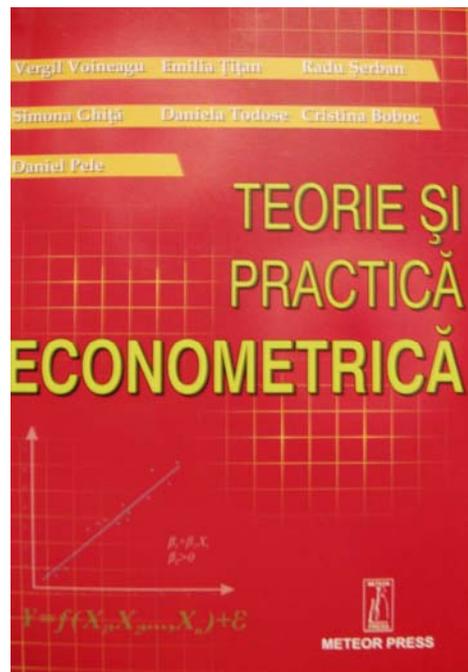
The high level of economical actual development, the speed and amplitude of changes and the future evolutions implies the storage, the processing and analyses of huge amount of data. The multitude and variation of information from present, completed by the evolution of storage and processing technologies represent an important factor for the approval or rejection of economical theories and in the same time could be a start for the elaboration of new theories.

Practice econometrics could be a connection between data and facts registered by statistics, theory of probabilities founded by mathematics and the final goal of economic theories.

Ragnar Frisch said that

*„Experience has shown that each of these three view-points, that of statistics, economic theory and mathematics is a necessary, but not by itself a sufficient condition for a real understanding of the quantitative relations in the modern economic life. It is the unification of all three that is powerful. And it is this unification that constitutes econometrics“.*¹

Starting from this point of view I may consider that the book elaborated by the seven authors represent an unification of mathematics and statistics knowledge applied on



the economic theory having the main goal: complete and easy understanding of mechanism of the actual economy

The information value found in this book is given in the same measure by the experience accumulated by the authors in the university medium and by their interest from many domains concluded by international stages of research and also by their activities in the research institutes from Romania.

The book is a very good mixture of theory and practice. Good examples, solved problems, remarks and the multitude of methods and tests used are making a complete approach and contributes in the same time to understand econometric theory in a simpler but complete way.

A good idea is represented by the usage of different software to solve applications. Using Statistica, SPSS and EXCEL the book introduce the reader in the middle of modern approach of the econometric problems and gives him the possibility to compare and finally to achieve the most efficient methods to practice econometrics.

In the end I may consider that the presented book could represent the best start for the beginners and an useful material for the intermediate researchers.

¹ Econometrica, Vol.1, No.1 (January 1933) pp. 1-4