María Ángeles Fernández-Izquierdo
María Jesús Muñoz-Torres
Raúl León (Eds.)

LNBIP 145

Modeling and Simulation in Engineering, Economics, and Management

International Conference, MS 2013
Castellón de la Plana, Spain, June 2013
Proceedings

Springer

Lecture Notes
in Business Information Processing 145

María Ángeles Fernández-Izquierdo
María Jesús Muñoz-Torres
Raúl León (Eds.)

Modeling and Simulation in Engineering, Economics, and Management

International Conference, MS 2013
Castellón de la Plana, Spain, June 6-7, 2013
Proceedings

Springer

Volume Editors

María Ángeles Fernández-Izquierdo
University Jaume I
Department of Finance and Accounting
Castellón de la Plana, Spain
E-mail: afernand@uji.es

María Jesús Muñoz-Torres
University Jaume I
Department of Finance and Accounting
Castellón de la Plana, Spain
E-mail: munoz@uji.es

Raúl León
University Jaume I
Department of Business Administration and Marketing
Castellón de la Plana, Spain
E-mail: rleon@uji.es

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-38278-9 e-ISBN 978-3-642-38279-6
DOI 10.1007/978-3-642-38279-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013938608

ACM Computing Classification (1998): I.6, J.1, G.1, I.2

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Association for the Advancement of Modeling and Simulation Techniques in Enterprises (AMSE) and the Jaume I University are pleased to present the main results of the International Conference on Modeling and Simulation in Engineering, Economics, and Management, held in Castellón de la Plana, Spain, June 6–7, 2013, through this book of proceedings published by Springer in the series *Lecture Notes in Business Information Processing.*

MS 2013 Castellón was co-organized by the AMSE Association and the Jaime I University through the SoGReS Research Group, Castellón de la Plana, Spain. It offered a unique opportunity for students, researches and practitioners to present and exchange ideas concerning modeling and simulation and related topics, and see how they can be implemented in the real world.

In this edition of the MS international Conference, we gave special attention to the area of modeling and simulation in corporate social responsibility and sustainable development, an emerging and interdisciplinary field in business management. The title of the book, *Modeling and Simulation in Engineering, Economics and Management,* refers to a broad research area and we have tried to systematize the resulting collection into a reasonable number of cohesive themes. In this sense the book has been organized in six tracks that comprise the most demanded topics of the congress. The importance of modeling and simulation techniques in business and economic research is reflected through a great selection of papers related to the use of modeling and simulation in diverse fields of business management, economics, and engineering This congress has as main incentive the interaction among different disciplines, and this multidisciplinary cooperation contributes to improve the availability and quality of models and techniques.

The MS 2013 Castellón Proceedings include 28 papers from 15 countries, selected from 65 submissions. The book is arranged according to six general topics of the conference: Modeling and Simulation in CSR and Sustainable Development, Modelling and Simulation in Finance and Accountability, Modeling and Simulation in Management and Marketing, Modeling and Simulation in Economics and Politics, Knowledge-Based Expert and Decision Support Systems and Modeling and Simulation in Engineering.

We would like to thank all the contributors, reviewers, and the scientific and honorary committees for their kind cooperation with MS 2013 Castellón, the whole of the Organizing Committee including Elena Escrig-Olmedo, Idoya Ferrero-Ferrero, José M. Merigó Lindahl, Juana María Rivera-Lirio, and Emilio Vizuete, and the rest of the Jaume I team for their support regarding the logistics of the conference. We would also like to thank Ralf Gestner, Christine Reis, and Viktoria Meyer (Springer) for their kind advice and help in publishing this volume.

Finally, we would like to express our gratitude to Springer for their support in the preparation of this book and in particular to the editors of the book series *Lecture Notes in Business Information Processing,* Wil van der Aalst, John Mylopoulos, Michael Rosemann, Michael J. Shaw, and Clemesns Szyperski.

June 2013

María Ángeles Fernández-Izquierdo
María Jesús Muñoz-Torres
Raúl León Soriano

Organization

Honorary Committee

Jaime Gil-Aluja — President of *AMSE* and President of the *Spanish Royal Academy of Financial and Economic Sciences*

Scientific Committee

Bernard De Baets, Belgium
Maria Glòria Barberà Mariné, Spain
Luciano Barcellos de Paula, Brazil
Christian Berger-Vachon, France
Luis Callarisa Fiol, Spain
Christer Carlsson, Finland
Patricia Crifo, France
Dong-Ling Xu, UK
Nashaat El-Khameesy, Egypt
Kurt Engemann, USA
Elena Escrig-Olmedo, Spain
María Ángeles Fernández Izquierdo, Spain
Joan Carles Ferrer Comalat, Spain
Idoya Ferrero-Ferrero, Spain
Anna M. Gil-Lafuente, Spain
Jaime Gil-Lafuente, Spain
Federico González Santoyo, Mexico
Madam Gupta, Canada
Nart Hameed, Jordan
Rainer Hampel, Germany
Korkmaz Imanov, Azerbaijan
Jian-Bo Yang, UK
Janusz Kacprzyk, Poland
Yuriy Kondratenko, Ukraine
Ravidoje Konstantinovic, Serbia
Viktor Krasnoproshin, Belarus
Rajko Kuzmanovic, Republic of Srpska
Leonid A. Kuznetsov, Russia
José María Labeaga, Spain
Reza Langari, USA
Rafael Lapiedra Alcamí, Spain
Raúl León Soriano, Spain
Vicente Liern Carrión, Spain
José M. Merigó Lindahl, Spain
Radko Mesiar, Slovakia
Peter Meža, Slovenia
Dragoljub Mirjanic, Republic of Srpska
José Mariano Moneva Abadía, Spain
Francesco C. Morabito, Italy
María Jesús Muñoz Torres, Spain
Vladimir S. Neronov, Kazakhstan
Eduardo Ortas Fredes, Spain
Witold Pedrycz, Canada
Emmanuel Perrin, France
Gary Pletcher, USA
Juana María Rivera-Lirio, Spain
Jose Miguel Rodríguez Fernandez, Spain
Petr Stemberk, Czech Republic
Antonio Terceño Gómez, Spain
Emili Vizuete Luciano, Spain
Ronald R. Yager, USA
Egon Zizmon, Slovenia
Constantin Zopounidis, Greece

Organizing Committee

Organizing Committee Chairs

Anna M. Gil-Lafuente, Spain
María Jesús Muñoz-Torres, Spain
María Ángeles Fernandez-Izquierdo, Spain

Organizing Committee Co-chair

Raúl Leon Soriano, Spain

Organizing Committee

Elena Escrig-Olmedo, Spain
Idoya Ferrero-Ferrero, Spain
José M. Merigó Lindahl, Spain
Juana María Rivera-Lirio, Spain
Emilio Vizuete, Spain

External Reviewers

Artis Aizstrauts
Ramón Alemany
Francisco-Javier Arroyo
Mercedes Ayuso
Gloria Barberá
Abdelhakim Begar
Jaume Belles-Sampera
Moussa Benyoucef
Catalina Bolancé
Herminia I. Calvete
Cristina Campos
Arnis Cirulis
Alexandra Maria Constantin
Djordje N. Dihovicni
Abdelhakim Elouadih
Elena Escrig-Olmedo
María Ángeles Fernandez-Izquierdo
Aurelio Fernández
Idoya Ferrero-Ferrero
Carmen Galé
Isabel Gallego-Álvarez
Reyes Grangel
Montserrat Guillén
Angel A. Juan
Raúl León
Vicente Liern
Salvador Linares-Mustarós
José M. Merigó
José Mariano Moneva
María Jesús Muñoz-Torres
Eduardo Ortas
Esteban Otto
Ahmed Oulad-Said
Juana María Rivera-Lirio
Seweryn Spalek
Noureddine Slimane
Stelian Stancu
Antonio Terceño
Wilfred I. Ukpere
Jian-Bo Yang
Egon Zizmond

Sponsors

A.M.S.E.
Association for the advancement of Modelling & Simulation techniques in Entreprises

UNIVERSITAT JAUME•I

GENERALITAT VALENCIANA
CONSELLERIA D'EDUCACIÓ, CULTURA I ESPORT

Facultat de Ciències Jurídiques i Econòmiques
UNIVERSITAT JAUME•I

UNIVERSITAT JAUME I
Departamento de Finanzas y Contabilidad

MÁSTER UNIVERSITARIO
Sostenibilidad y Responsabilidad Social Corporativa

bp

Table of Contents

Can Board Diversity Enhance the Integration of ESG Aspects in Management System?

Idoya Ferrero-Ferrero,
María Ángeles Fernández-Izquierdo,
and María Jesús Muñoz-Torres

Finance and Accounting Department, Universitat Jaume I, Av. de Vicent Sos Baynat s/n,
12071 Castelló de la Plana, Spain
{ferrero,afernand,munoz}@uji.es

Abstract. Board of directors is one of the most significant governance issues under review by the corporate governance initiatives. A recent recommendation to improve board effectiveness is concerning diversity of the boards. However, further research to assess the impact of board diversity in the business world and to better understand how board diversity interacts with sustainability is clearly needed. Building on this research gap, this study aims to explore empirically how board diversity affects the adoption of ESG management practices. The results reveal generational diversity enhances the integration of ESG aspects in the management process. This study has important implications for theory, business practice and public policy on sustainability and corporate governance fields.

Keywords: Board diversity, ESG (Environmental, social and governance) aspects, management system, generations, sustainability.

1 Introduction

New initiatives of corporate governance are necessary in order to promote sustainable companies. We understand a sustainable company as a company that integrate stakeholder's interest in the management process, taking into account economic, social, and environmental concerns in the strategies, practices and activities of the firm. For that end, board of directors and managers have to integrate the concerns and interests of stakeholder into decision-making processes in a balanced way. In this point, the role of the board of directors, as the pinnacle of performance management systems of the organization, is extremely important to integrate the sustainability into corporate agendas.

Nowadays, the board composition and structure is one of the governance issues under review. A common recommendation to improve board effectiveness is concerning diversity of the boards. For instance, European Commission in the communication entitled "Action Plan: European company law and corporate governance – a modern legal framework for more engaged shareholders and sustainable companies" has recently announced new developments on board diversity

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 1–9, 2013.

Knowledge-Based Expert and Decison Support Systems

Modeling and Simulation in Engineering

that require future action [1]. The recent corporate governance initiatives argue that a greater diversity enhances information resources and broadens the cognitive and behavioural range of the board. In this context, boards examine different policies, means of implementation and control mechanisms with the aim of making effective decisions. Therefore, board diversity may be an important element that positively affects corporate performance.

Consistent with the abovementioned reasoning, a stream of board literature [2] states that the characteristics of the board of directors allow a stronger explanation of strategic decisions and organizational performance. However, the empirical studies that explored the effect of board composition on financial and environmental, social and governance (ESG) outcomes found mixed results. A better understanding of this relationship represents one important question which requires further research in order to assess the impact of board diversity recommendation on the business world. Building on this research gap, this study aims to examine whether generational board diversity positively affects sustainable performance. This paper is divided into five sections. After this introduction, a review of the theoretical framework is provided. The third section includes information on the sample, variables, and the methodology used in estimating the model. The fourth section presents the findings and empirical analysis. The final section discusses the main results and concludes the study.

2 Theoretical Framework

The Upper Echelon Theory [2] argues that complex decisions, such as strategic choices, are largely the outcome of behavioural factors rather than a mechanical quest for economic optimization. Given the great difficulty to obtain conventional psychometric data on top executives, Hambrick and Mason [2] suggested to use observable –demographic- managerial characteristics as valid proxies of executives' cognitive frames. Therefore, the characteristics of the top management team may be an important factor of strategic choices.

A board characteristic often investigated in stakeholder literature is board diversity. Although the meaning of board diversity is still unclear, Harrison and Klein [3] defined three types of diversity -diversity as separation, variety and disparity- and examined their implications on corporate performance. Diversity as separation refers to differences in position or opinion among unit members and it is assumed that greater separation presumably lead to disagreements and conflicts. Diversity as variety indicates differences in king, category, knowledge or experience among unit members and it can translate greater information richness in the team and, therefore, they can make more effective decisions. Diversity as disparity represents differences in concentration of power, prestige and status among unit members. The basic idea is that in teams where few members have a marked influence over the group decision, they control the flow of information, impose their views and limit a democratic participation in the team and, consequently, the decisions are made in worse conditions.

The majority of studies focused on board diversity and corporate performance have adopted a definition of diversity close to diversity as variety. In fact, these previous studies expect that different knowledge domains, perspectives, values, and ideas are considered in the decision-making process [4] and consequently board diversity, by means of the management quality, enhances corporate performance. Moreover, Hafsi and Turgut [5] argued that diversity is desired by customers and other stakeholders for whom it is a demonstration of management sensitivity to stakeholder's preferences, aspirations, and concerns. Likewise, Fernández Sánchez et al. [6] suggested that board diversity is more sensitive to some specific social practices and therefore it improves social responsibility taking into account stakeholders' interests in the corporation.

The empirical studies, [4], [5], [7], and [8], have used a broad variety of director attributes, such as gender, age, or nationality, to measure board diversity. However, the research on age diversity of the board of directors is much less developed and needs new paradigms and new approaches to explain how it may affect corporate performance. Prior findings of limited empirical studies, [4] and [5], on the effect of age diversity on CSR performance did not support theoretical assumptions on the relationship between the composition of boards and sustianability. Consistent with Harrison and Klein's [3] diversity typology, a possible explanation of the previous confusing findings could be due to researchers specified the diversity as variety, which is related to greater information richness and knowledge, however, they operationalised the diversity as separation, which implies disagreements and conflicts. Ferrero-Ferrero et al. [9] explored how the different types of age diversity affect economic performance and, following Harrison and Klein [3] for research design, found that age diversity as variety positively impacts on corporate performance. In this sense, diversity as variety has to be a categorical attribute that represents differences in personality, traits, skills, attitudes, mental health, work values and behaviours. Focusing on age, these differences may be categorized according to the generations, since the social and historical experiences and circumstances from a respective generation have influenced the individuals' behaviour [9]. In line with this approach, this study uses generational diversity as a consistent construct of age diversity as variety.

Today's boards of directors consist of members from three generations: Veterans (1922-1942), Baby Boomers (1943-1960), and Generation X (1961-1981). The Veterans has exhibited a high concern for security and a desire to avoid the risks and disasters witnesses during their early years. Veterans are viewed as hard-working, dependable, and supportive of conservative values that emphasize the importance of loyalty, duty, conformity, and security [10]. Likewise, Post et al. [4] and Hafsi and Turgut [5] and argue that older individuals exhibit higher moral reasoning and may be more sensitive to society and more willing to contribute to its welfare. The Baby Boomers was active in radical social changes, including the emergence of the civil rights changes, protests against the Vietnam War, and the women's movement. These changes coincided with greatly liberalized attitudes toward gender roles. Generation X was influenced by the financial, family and societal insecurities that dominated their childhoods. They lack solid traditions but are highly mobile and are accustomed to rapid change. This generation place less importance on job security and status, but more on personal freedom and challenging work, which allows for a balance work-personal life style. They are supportive of social liberalism and environmentalism, and fight against climate change because they perceive a higher vulnerability to its consequences [11].

Therefore, a greater generational diversity enhances information resources and broadens the cognitive and behavioural range of the board. In this context, boards are more sensitive to stakeholder's preferences, aspirations, and concerns which integrate in the corporate policies, means of implementation and control mechanisms in order to make effective decisions and achieve better corporate performance. Following this arguments, the following hypothesis is introduced:

Hypothesis: Generational diversity positively impacts on the adoption of ESG management practices

3 Data and Methodology

In this section it is presented the sample, the variables used and the model estimated to test the hypothesis developed.

3.1 Sample

The sample consists of companies listed in FTSE 100, DAX 30, and CAC 40 for year 2009. Observations were deleted if information was missing. The final sample contained 146 companies: 80 were in FTSE 100 (54.80%), 30 were in DAX 30 (20.55%), and 36 were in CAC 40 (24.66%). The information was obtained from Asset4 database.

3.2 Variables

In this section, the different variables used in the estimation are presented.

Dependent Variable. This study presents three different proxies to measure the adoption of ESG management practices, which has been constructed using Asset4 dataset. Asset4 contains over 250 indicators on financial, environmental, social and corporate governance management. This study has selected drivers indicators, which provide information on management quality, by looking at the different policies, means of implementation and control mechanisms a corporation has in place. The first measure, "Vision and Strategy" (V&S – ESG management practices), is an indicator provided by Asset 4 that assesses a company's management commitment and effectiveness towards the creation of an overarching vision and strategy integrating financial and extra-financial aspects into its day-to-day decision-making processes. The second measure, "Integrated rating" is an equal-weighted rating created by Asset4 based on the information in economic, environmental, social and corporate governance pillars. It reflects a balanced view of a company's performance in these four areas. "IR-ESG management practices" proxy is calculated from 72 drivers indicators. The third measure is a new index, which has been developed following the approach of principal component analysis. Similarly to Goss and Gordon [12], this study extracts the first PC of the 18 categories of Asset4 database to calculate the index "PC -ESG management practices", excluding the category referred to board structure.

Independent Variables. According to Ferrero Ferrero et al. [9], this study uses Blau's Index [13] to measure "generational diversity" by means of three categories based on generations - the Veterans (67 - 87), Boomers (49 - 66), and Xers (28 - 48). This index has been divided by its theoretical maximum with the aim of standardising the results and making the interpretation of the index easier.

Control Variables. Harrison and Klein [3] argued that diversity as separation and diversity as disparity may affect the decision-making process of a team. Therefore, in order to examine wheter generational diversity, i.e. diversity as variety, encourage the adoption of ESG aspects in the management system, it is necessary to control by the other types of diversities (diversity as separation and diversity as disparity). Following Ferrero Ferrero et al. [3], this study measures diversity as separation ("Board separation") by standard deviation of the age of directors that are members of the board and diversity as disparity ("Board disparity") by the coefficient of variation of the age of the board members. The presence of a dual Chairman and Chief Executive Officer (CEO) could impact the effectiveness of board monitoring [7]. In this vein, Quigley and Hambrick [14] argued that the effectiveness of the board also may be affected when the chair is the former CEO, since he restricts the new CEO's behaviour, limiting strategic changes. Consistent with this argument, this study includes as a control variable the variable "CEO duality". This variable is equal to 1 when the chairman of the board is CEO or was the former CEO. Other variable that could affect the corporate performance and has been included as a control variable is the size of the firm (SIZE) measured as the natural log of total assets. Additionally, dummy variables are considered to reflect differences between countries (COUNTRY), and industries (INDUSTRY) using one-digit SIC.

3.3 Model

This study tests the impact of generational diversity on ESG management practices. For that end, this study estimates the linear regression model presented in equation (1) and (2). The issue of collinearity is explored by means of the Variance Inflation Factors (VIFs) for the independent variables. The result of this analysis show that "Board Separation" and "Board Disparity" present multi-collienarity problems. For this reason, both variables are not jointly regressed.

$$
\begin{aligned}
ESG\_Management\_\Pr actices_i = {} & \alpha_0 + \alpha_1 \cdot GenerationalDiversity_i + \alpha_2 \cdot BoardSeparation_i \\
& + \alpha_4 \cdot CEODualiry_i + \alpha_5 \cdot SIZE_i + \sum_{J=0}^{7} \varphi_J \cdot Industry_i \\
& + \sum_{J=1}^{2} \varphi_J \cdot Country_i \cdot + \mu_i
\end{aligned}
\tag{E.1}
$$

$$ESG\_Management\_Practices_i = \alpha_0 + \alpha_1 \cdot GenerationalDiversity_i + \alpha_3 \cdot BoardDisparity_i + \alpha_4 \cdot CEODualiry_i + \alpha_5 \cdot SIZE_i + \sum_{J=0}^{7} \varphi_J \cdot Industry_i + \sum_{J=1}^{2} \varphi_J \cdot Country_i \cdot + \mu_i \quad \text{(E.2)}$$

Given the feature of the sample, the equations are regressed by means of OLS. The estimator process uses a robust variance matrix, in particular, White-corrected standard errors in presence of heteroskedasticity.

4 Empirical Results

Table 1 reports the results of the regression analysis using the three proxies for ESG Management Practices.

Table 1. Results of regression analysis

Dependent Variables	V&S-ESG Management Practices		IR -ESG Management Practices		PC-ESG Management Practices	
Independent Variables	(E.1)	(E.2)	(E.1)	(E.2)	(E.1)	(E.2)
Generational diversity	0.17**	0.17**	0.09*	0.09*	1.09*	1.04*
	(0.06)	(0.06)	(0.04)	(0.04)	(0.50)	(0.48)
Board Separation	-0.01		-0.01		-0.07	
	(0.01)		(0.01)		(0.06)	
Board Disparity		-0.34		-0.33		-3.34
		(0.47)		(0.29)		(3.55)
CEO Duality	-0.03	-0.03	-0.01	-0.01	-0.17	-0.18
	(0.02)	(0.03)	(0.02)	(0.02)	(0.20)	(0.20)
Firm size	0.03***	0.03***	0.03***	0.03***	0.36***	0.36***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.05)	(0.05)
CONSTANT	0.44***	0.45***	0.40***	0.40***	3.36***	3.39***
	(0.09)	(0.09)	(0.06)	(0.06)	(0.66)	(0.67)
Country and Industry Dummies Included						
R^2	0.28	0.28	0.30	0.30	0.33	0.33
F- test	4.80***	4.85***	6.31***	6.12***	6.47***	6.33***
N. obs.	146	146	146	146	146	146

Standard errors are in brackets. † p<0.10; *p<0.05; **p<0.01; ***p<0.001.

The hypothesis of this study predicts that generational diversity positively affects ESG management practices. Table 1 shows that the coefficient for generational diversity is positive and statistically significant in all proxies used for ESG Management Practices. Accordingly, the results support the hypothesis of the study. This study finds evidence that generational board diversity positively affects the creation of an overarching vision and strategy integrating financial and extra-financial aspects into the day-to-day decision-making processes and it leads to integrates financial and extra financial factors in the management discussion and to adopt ESG practices in the management system.

Regarding control variables, the coefficients present the expected signs, although, in the case of "board separation", "board disparity", and "CEO Duality" variables, the coefficients are not significant. This result could be explained because these proxies are inaccurate to fit the distribution of the differences of opinions and powers of the directors in the board. Therefore, future studies can develop better proxies to measure the behavior of diversity as separation and diversity as disparity and explore their influence on the management system.

5 Conclusion and Discussion

This study aims to examine whether board diversity affects the adoption of ESG management practices. This research question is tested using 146 companies listed in FTSE 100, DAX 30 and CAC 40 for year 2009. This study uses three different measures of ESG management practices and generational diversity as a consistent construct of age diversity as variety.

The main results reveal that generational diversity positively affects the creation of an overarching vision and strategy integrating financial and extra-financial aspects into the day-to-day decision-making processes and it leads to integrate financial and extra financial factors in the management discussion and to adopt ESG aspects in the management process.

This study has important implications for theory, business practice and public policy. First, this study presents the generational diversity as an interesting area of literature that has been largely unexplored in the board of directors context. Second, this study empirically assesses the recommendation of corporate governance that encourage companies to enhance board diversity in order to lead to successful governance of the company. Third, the findings suggest that generational diversity enables more effective design of vision and strategies to address financial and extra-financial aspects and, consequently, it encourages companies to adopt a sustainable approach to their business.

As with all empirical research, the findings presented should be viewed in light of some potential limitations that open new areas for future research. A limitation of this study is that empirical findings are conditioned by sample and availability of information. Larger samples are clearly needed to test the robustness of the results. The results of this study may also be limited by some variables used. The integrated rating and the indexes developed following the Principal Component Analysis may present offsetting effect among different CSR dimensions [15]. Moreover, the proxies used to measure separation and disparity in the boards could be inaccurate to fit the distribution of the differences of opinions and powers of the directors in the board. Future studies should use more fine-grained measurements of integral ESG management practices, as well as, the separation and disparity dimensions of the boards.

To sum up, this study offers new and interesting insights on the question whether generational board diversity enhances the adoption of ESG practices in the management system and encourages to future research to explore generational diversity as a factor of sustainable development of the organizations.

References

1. COM: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions. Action Plan: European company law and corporate governance – a modern legal framework for more engaged shareholders and sustainable companies. COM (2012); 740/2. European Commission (2012)
2. Hambrick, D.C., Mason, P.A.: Upper echelons: the organization as a reflection of its top managers. Academy of Management Review 9, 193–206 (1984)
3. Harrison, D., Klein, K.: What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Academy of Management Review 32, 1199–1228 (2007)
4. Post, C., Rahman, N., Rubow, E.: Green governance: Boards of directors' composition and environmental corporate social responsibility. Business & Society 50(1), 189–223 (2011)
5. Hafsi, T., Turgut, G.: Boardroom Diversity and its Effect on Social Performance: Conceptualization and Empirical Evidence. Journal of Business Ethics (2012), doi:10.1007/s10551-012-1272-z (forthcoming)
6. Fernández Sánchez, J.L., Ladislao Sotorrío, L., Baraibar Díez, E.: The relationship between corporate governance and corporate social behavior: a structural equation model analysis. Corporate Soccial Responsibity and Environmental Management 18, 91–101 (2011)
7. Bear, S., Rahman, N., Post, C.: The impact of board diversity and gender composition on corporate social responsibility and firm reputation. Journal of Business Ethics 97(2), 207–221 (2010)
8. Walls, J., Berrone, P., Phan, P.H.: Corporate Governance and Environmental Performance: Is there really a link? Strategic Management Journal 33, 885–913 (2012)
9. Ferrero-Ferrero, I., Fernández-Izquierdo, M.Á., Muñoz-Torres, M.J.: Age Diversity in the Boardroom: Measures and Implications. In: Engemann, K.J., Gil-Lafuente, A.M., Merigó, J.M. (eds.) MS 2012. LNBIP, vol. 115, pp. 134–143. Springer, Heidelberg (2012)
10. Egri, C.P., Ralston, D.A.: Generation Cohorts and Personal Values: A Comparison of China and the United States. Organization Science 15(2), 210–220 (2004)
11. Ciocirlan, C., Pettersson, C.: Does Workforce Diversity Matter in the Fight against Climate Change? An Analysis of Fortune 500 Companies. Corporate Soccial Responsibity and Environmental Management 19, 47–62 (2012)
12. Goss, A., Gordon, S.R.: The impact of corporate social responsibility on the cost of bank loans. Journal of Banking & Finance 35(7), 1794–1810 (2011)
13. Blau, P.M.: Inequality and heterogeneity. Free Press, New York (1977)

14. Quigley, T.J., Hambrick, D.C.: When the Former CEO Stays on as Board Chair: Effects on Successor Discretion, Strategic Change, and Performance. Strategic Management Journal 33, 834–859 (2012), doi:10.1002/smj.1945
15. Escrig, E., Rivera, J.M., Muñoz, M.J., Fernández, M.A.: Measuring Sustainability Performance through Fuzzy Logic. World Scientific Proceedings Series in Computer Engineering and Information Science 3: Computational Intelligence in Business and Economics. Word Scientific Publishing Co. Pte. Lte. (2010)

Application of the HJ Biplot Methodology to Variation Greenhouse Gas Emissions in International Companies*

Jennifer Martínez-Ferrero and Isabel Gallego-Álvarez

Department of Business Administration
Faculty of Economics and Business Administration
Campus Miguel de Unamuno. Edificio FES
University of Salamanca
37007-Salamanca- Spain
{jenny_marfe,igallego}@usal.es

Abstract. The quantifying and reporting of greenhouse gas emissions is one of the most important tools for monitoring and auditing proposed to mitigate climate change, and it also directly affects business. It is thus vital at this time that we learn in detail whether firms actually report on greenhouse gas emissions and make the account entries that must be included within it. This research has a twofold objective: first to analyse the report on greenhouse gas emissions of international firms in the 2007, 2008 and 2009 period and to see what kind of variation occurs in CO_2 emissions between 2006-2007, 2007-2008 and 2008-2009. Secondly we shall use the biplot methodology to represent emissions variations in firms grouped into geographic areas. To do so we group only the 89 firms in our study into the geographical areas of the Europe (EU), North America (NA), Asia (AS), and South America (SA). As regards the variation in CO_2 emissions, it is noteworthy that variation in CO_2 emissions for the 2006/2007 period are located closer to companies in North America, variation in CO_2 emissions for 2007/2008 are located closer to European Union companies and very strikingly, the variation in CO_2 emissions for the 2008/2009 period are located closer to companies in Asia and South America. This leads us to conclude that companies located in developing countries are nowadays the most aware of climate change and the need to reduce emissions.

Keywords: HJ-Biplot methodology, international companies, variation greenhouse gas emission.

1 Introduction

In 1987, the World Commission on Environment of the United Nations Organization defined sustainable development as development that could satisfy current needs without endangering the capability of future generations to attend to their own needs. This definition comes under the polemical and important issue that for many years has

* International Conference on Modeling and Simulation in Engineering, Economics and Management. Modeling and Simulation in CSR and Sustainable Development.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 10–22, 2013.

been of interest to the UN and which is an integral part of the economic and social development of all of humanity: the natural environment.

Among the problems associated with the environment we can highlight those related to forests, protection of the ozone layer, climate change, water, energy and natural resources, biodiversity, etc. Although all of these topics affect the environment and are of maximum priority, here we would like to concentrate on the climate change that in recent years has become a major concern for humanity.

Companies, as part of society, are now faced with not only the challenge of how to reduce emissions to mitigate climate change [1] [2] but also how climate change will impact their operations, since the increase in atmospheric temperature has given rise to an accumulation of greenhouse gas emissions, especially of carbon dioxide (CO_2) [3].

To deal with this problem, the Kyoto Protocol to the United Nations Framework Convention on Climate Change was signed in 1997. The Kyoto Protocol is aimed at lowering greenhouse gas emissions and favours a redistribution of the costs associated with climate change, by moving them from citizens of poor countries that are not producers of greenhouse gas emissions to the firms that are really responsible for the emissions and who profit from them. The Kyoto Protocol has been ratified by different countries and companies will have to adapt to the new regulations contained in the Protocol, as well as those developed in the European Union and other countries.

Kyoto Protocol emissions trading forms an umbrella under which national and regional trading schemes operate [4]. Entity-level trading uses Kyoto units and needs to be reflected in the accounting. The European Union emissions trading scheme (EU ETS) launched in 2005 is one example of a regional trading system operating under the Kyoto Protocol umbrella. EU emissions trading scheme has opened up a public debate that affects both government and business, and although some years have now passed since the first stages of the debate, the question of the impact of emissions trading on the competitiveness of firms has not diminished in importance, thus showing the need to continue analysing the situation of different companies [5].

Considering the above, some authors [6] propose a series of measures for the mitigation of climate change that affect businesses and that focus on the following: direct quantification of greenhouse gas emissions from operations, reporting greenhouse gas emissions and comparing them across the sector, assessing greenhouse gas emissions from the value chain (including suppliers and use of products), locating the position of the organization within the system of production and consumption and evaluating the effect of the organization on other systems.

Some of these measures are taken into account in our research: quantification of greenhouse gas emissions from operations and reporting greenhouse gas emissions and comparing them across the sector. As regards how these measures can affect companies, some authors [7] indicated that in the annual report of WMC Ltd., one of Australia's largest firms, the directors stated that: "good environmental management reduces costs, meets the expectations of the communities in which the company operates, and enables the company to be more profitable". Moreover, "greenhouse gas emission may be an opportunity to enhance a corporation's reputation and can have important benefits, i.e. investors who consider environmental strategies in making investments"[8].

This research has a twofold objective: first to analyse the report on greenhouse gas emissions of international firms in the 2007, 2008 and 2009 period and to see what kind of variation occurs in CO_2 emissions between 2006-2007, 2007-2008 and 2008-2009. Secondly we shall use the biplot methodology to represent emissions variations in firms grouped into geographic areas. To do so we group only the 89 firms in our study into the geographical areas of the Europe (EU), North America (NA), Asia (AS), and South America (SA).

In order to carry out this study, we selected companies from different countries worldwide. Of the 162 firms selected originally according to their activity sector 138 firms present CO_2 emissions reports for 2007 but only 89 present data for 2006 and therefore the variation in CO_2 emissions could only be determined with the latter number of firms (89).

The database used was that of the Fortune 500, as it compiles the largest international companies worldwide, classified by activity sectors. The activity sectors selected to undertake this research are consistent with those established in the *Green Paper on Greenhouse Gas Emissions Trading* within the European Union (EU, 2000) and in the *Kyoto Protocol*. With respect to the activity sectors to be included in an emissions trading system, the Green Paper mentions certain sectors that are large point sources of carbon dioxide emissions. such as electricity and heat production, iron and steel, refining, chemicals, glass, pottery and building material (including cement), paper and printing (including paper pulping). For our study, we chose the following from among all the activity sectors included in the Fortune 500: aerospace and defence; airlines; chemicals; energy; forest and paper products; industrial and farm equipment; metals; mining, crude-oil production; motor vehicles and parts; petroleum refining and utilities.

The results obtained show that there was a reduction in CO_2 emissions in the 2007-2008 period and also in the 2008-2009 period in percentage points (-3.048, -2.226, respectively).

As regards the variation in CO_2 emissions, it is noteworthy that variation in CO_2 emissions for the 2006/2007 period are located closer to companies in North America, variation in CO_2 emissions for 2007/2008 are located closer to European companies and very strikingly, the variation in CO_2 emissions for the 2008/2009 period are located closer to companies in Asia and South America. This leads us to conclude that companies located in developing countries are nowadays the most aware of climate change and the need to reduce emissions.

2 Research Method

2.1 Sample Description

In order to carry out this study, we selected companies from different countries worldwide. Of the 162 firms selected originally according to their activity sector, 138 firms present CO_2 emissions reports for 2007, but only 89 present data for 2006, and therefore the variation (increase or decrease) in CO_2 emissions could only be determined with the latter number of firms (89). These firms comprise the final sample, as they present CO_2 emissions for the years 2006 and 2007. Moreover, 2007 was the most recent year for which complete emissions data were available (the sustainability report or corporate social responsibility report of each firm).

The database used was that of the Fortune 500, as it compiles the largest international companies worldwide, classified by activity sectors. The activity sectors selected to undertake this research are consistent with those established in the *Green Paper on Greenhouse Gas Emissions Trading* within the European Union [9] and in the *Kyoto Protocol*. With respect to the activity sectors to be included in an emissions trading system, the Green Paper mentions some sectors that are large point sources of carbon dioxide emissions, such as electricity and heat production, iron and steel, refining, chemicals, glass, pottery and building material (including cement), paper and printing (including paper pulping). For our study, we chose the following from among all the activity sectors included in the Fortune 500: aerospace and defence; airlines; chemicals; energy; forest and paper products; industrial and farm equipment; metals; mining, crude-oil production; motor vehicles and parts; petroleum refining and utilities.

2.2 Analysis Technique

The analysis of several economics problems require the storage of large volumes of data. In order to exploit the data to get a better understanding of the behaviour of several processes, it is important to identify the salient features underlying them. The reduction in the dimensionality of the problem enables to summarize the information captured in a large number of variables by a smaller number of variables.

The technique chosen in this research is Biplot, which has been used in other environmental studies [10, 11] but it has not been applied to emission variation of greenhouse gas. Then, it will allow us to check if the variation of greenhouse gas emission of companies are similar in different countries, in other words, if the environmental concerns are similar in different geographic zones.

This technique consists of depicting graphically a data matrix X (nxp) derived from analysing n individuals according to p numerical characteristics. In this study, the n individuals are the 89 companies worldwide presented in the Appendix 1, grouped in 4 geographic areas; the p numerical characteristics are the variation of greenhouse gas emission that present companies in different years from 2006 to 2009 [12].

The Biplot offers a visual representation (usually in two or three dimensions), based on two types of vectors derived from two types of information: individuals (in rows) and variables (in columns). Hence, the vectors represent graphically individuals or rows and variables or columns. [13] the method of obtaining the vectors is not specified, thereby using the method of least squares and the decomposition in vectors and singular values of X. However, it is argued that although it reflects the statistical and geometric properties of the variables adequately, the individuals are not appropriately represented.

Some authors like [14] generalises the concept of simultaneous representation, by creating a new type of Biplot, HJ Biplot, applied to all the data set, which allows to represent individuals and variables with the same quality of representation. This type improves other approaches [12, 13].

The HJ Biplot as a multivariate graphical representation of the matrix X (nxp) through vectors for their rows and for columns, so that both vectors can be depicted in the same reference systems with the highest quality of representation [14, 15]. HJ-Biplot as a technique derived from principal components analysis with an important objective: to reduce the volume of data in order to obtain information. To achieve this aim, it is necessary to analyze the initial points cloud in the hyperspace by a simplified configuration in a less dimension space [16].

As for the interpretation of Biplots, points represent individuals (in our study, the companies grouped by geographic areas) and axes to reflect variables (in our study, variation greenhouse gas emission). The interpretation is based on the angles among the vectors: variables with vectors displaying a small angle show similar behaviours, points of close individuals correspond to similar individuals and points of remote individuals have to do with non similar individuals.

Also, if there is a small angle between an individual and a variable, it means that the individual is significant in order to explain the variable and that the variable has a high value for the individual. The distance among the points reflects the variability of those points in the study. By analysing the length of the variables, we obtain the variability of the variables, providing the researcher with an idea of the dispersion in the graphic. When the variables are close, it is said that they are highly correlated, with similar behaviour; when they take different directions, they are highly correlated in an inverse sense; if they are perpendicular, they are independents [17].

Regarding the angles, the lower the angle between two vectors that join the centre of gravity with the points that representing the variables, the more concentrated the characters are; finally the covariance of variables is obtained observing the angle.

The software used to implement the HJ Biplot has been developed by [18], which contains classical biplot, HJ biplots and simple correspondence analysis of a contingency table.

3 Results of Empirical Analysis

The results are shown in different tables and figures obtained from the software multbiplot. The eigenvalues and variance explained are exhibited in Table 1.

Table 1. Eigenvalues and variance explained

Axis	Eigenvalue	Expl. Var	Cummulative
Axis 1	11.627	51.205	51.205
Axis 2	10.034	38.134	89.339
Axis 3	5.305	10.661	100

From this table, it can be deduced that there is a dominant axis that takes the 51.205 per cent of the total inertia of the system. The tendency in the eigenvalues is truncated in the second axis, achieving an accumulative inertia of 89.339. The remaining factors provide a lower load of information; therefore, we opt for retaining the two first factorial axes for the classification.

Table 2 contains the contribution of each factor to the element.

Table 2. Relative contribution of the factor to the element

Variables	Axis 1	Axis 2	Axis 3
VGAS 2006/2007	412	491	97
VGAS 2007/2008	266	651	82
VGAS 2008/2009	858	1	141

Hence, the variables VGAS 2008/2009 has a high contribution to the Axis 1 and a low contribution to the remaining axes. On the contrary, VGAS 2006/2007 and VGAS 2007/2008 have a higher contribution to Axis 2. When analysing the contributions to the different axes, the first axis is explained by variation in greenhouse gas emissions 2008/2009 (858). The second factorial axis is determined by variation in greenhouse gas emissions 2006/2007 (491) and by variation in greenhouse gas emissions 2007/2008 (651).

Regarding the graphic representation, in Figure 1, the five geographic areas which include the countries analysed (see Appendix 1; in our Biplot, individuals) are located.

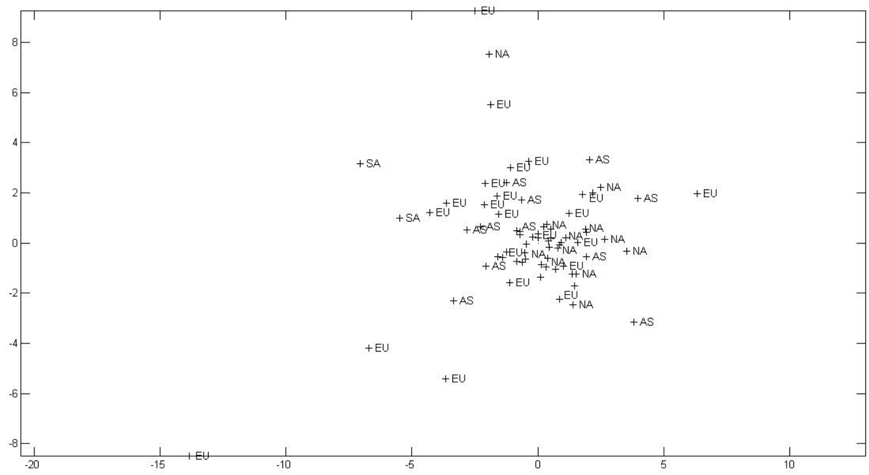


Fig. 1. Geographical areas that include the 89 companies

All the companies grouped in four geographic areas are represented by points (in this Figure, +) in four quadrants. In the quadrant 1 the companies located in North America (NA) are mainly represented, whereas the quadrants 2 the companies located in Europe (EU) are mainly represented.

In Figure 2, the following variables are displayed: VGAS 2006/2007, VGAS 2007/2008 and VGAS 2008/2009.

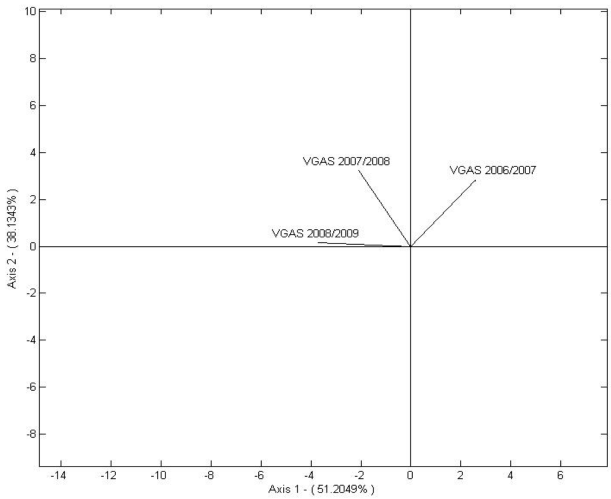


Fig. 2. Representation of variation of greenhouse gas emission of companies

As can be observed from the Figure 2, the variable VGAS 2006/2007 is located in the quadrant 1 and next to the second factorial axes, VGAS 2007/2008 is situated in the quadrant 2 and next to the second factorial axis and VGAS 2008/2009 is located in the quadrant 2 and next to the first factorial axis. This situation can also be deduced when analysing the contributions to the different axes, the first axis is explained by variation in greenhouse gas emissions 2008/2009 (858). The second factorial axis is determined by variation in greenhouse gas emissions 2006/2007 (491) and by variation in greenhouse gas emissions 2007/2008 (651).

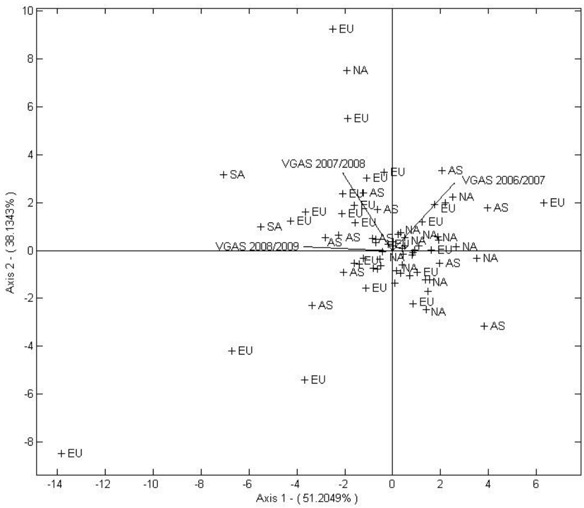


Fig. 3. Representation of geographical areas and variation in greenhouse gas emissions

In Figure 3, the geographic areas and the variables representing variation in greenhouse gas emissions are displayed jointly.

According to the interpretation, if there is a small angle between an individual and a variable, it means that the individual is significant to explain the variable and that the variable is of great value for the individual.

In the Figure 3, it can be observed that the variable related to variation in greenhouse gas emissions 2006/2007 are closer to the companies of countries located in the geographic area of North America. Variation in greenhouse gas emissions 2007/2008 are located closer to European companies and very strikingly, the variation in greenhouse gas emissions for the 2008/2009 period are located closer to companies in Asia and South America. This leads us to conclude that companies located in developing countries are nowadays the most aware of climate change and the need to reduce emissions.

4 Conclusions

One of the first conclusions that we obtain from our research is the great importance in recent years that it has had the reporting of greenhouse gas emission by companies at the international level. This shows that good environmental management reduces costs, meets the expectations of the communities in which the company operates and enables the company to be more profitable [7]. At the same time, it reaffirms the above of greenhouse gas emission may be an opportunity to enhance a corporation's reputation and can have important benefits, i.e. investors who consider environmental strategies in making investments[8].

In the same line, other authors, suggest that a firm may be green and competitive [19], and that strategic position may jointly cause both lower pollution levels and better financial performance [20], which may explain the great relevance both for companies as for transmitters and addressees of reporting about greenhouse gas emission.

Regarding the emission variation, again, it is evidenced the previous above of it has occurred a reduction in in CO_2 emissions in the 2007-2008 period and also in the 2008-2009 period in percentage points (-3.048, -2.226, respectively), which reaffirms that companies, as part of society, are now faced with not only the challenge of how to reduce emissions to mitigate climate change [1] [2] but also how climate change will impact their operations, since the increase in atmospheric temperature has given rise to an accumulation of greenhouse gas emissions, especially of carbon dioxide (CO_2) [3].

Furthermore, as a result of the analysis, it can be deduced that there exist two well-defined areas, one compiling more developed countries and another area with less developed countries: Europe (EU) and North America (NA), on the one hand; and Asia (AS) and South America (SA), on the other.

We have used the HJ-Biplot, a technique derived from principal components analysis with an important objective: to reduce the volume of data in order to obtain information. To achieve this aim, we analyzed the initial points cloud in the hyperspace using a simplified configuration in a smaller space.

As regards the variation in CO_2 emissions, it is noteworthy that variation in CO_2 emissions for the 2006/2007 period are located closer to companies in North America, variation in CO_2 emissions for 2007/2008 are located closer to European companies and very strikingly, the variation in CO_2 emissions for the 2008/2009 period are located closer to companies in Asia and South America.

This leads us to conclude that companies located in developing countries are nowadays the most aware of climate change and the need to reduce emissions. Specifically, companies of these countries are the companies that in the last reporting period are conducting environmental-friendly best practices. This is the case of Indian companies, which in the last year have increased their greenhouse gas emission responses and climate change by 7%.

As to the limitations of the present study, it can be said that further study should be made using data on greenhouse gas emissions from later years in order to verify the variations in emissions from companies in different countries with a view to seeing whether the trend found here remains over time.

Future lines of research could increase the number of years studied and the involvement of key stakeholders in the process, as well as the establishment of clear goals and targets for improvement, since the no-for-profit organizations forecasts for the long term that all issues relating to climate change (emissions variations being just one of them) will have major financial implications, although these will vary according to economic sector and geographical area.

The importance of environmental-friendly practices is also considered by the [21], when it establishes among its strategies for overcoming the crisis a low growth in carbon emissions, that policies should foster adaptation to climate change in an efficient way and that investment in environmental-friendly new technologies can be an important new source of growth and opportunity without a downside for either the environment or the economy.

We must also analyze both internal and external factors of firms that may influence CO_2 emissions variations as well as the stakeholder reaction to emissions variation, considering that stakeholders are interested in knowing whether a firm is acting in an environmentally friendly way.

To end, given the importance of the topic at the present time and the effort being made by companies to reduce their CO_2 emissions, as seen in our study, it is to be hoped that their effort will be compensated in the near future once the crisis period is over.

References

1. Okereke, C.: An exploration of motivations, drivers and barriers to carbon management: the UK FTSE 100. European Management Journal 25(6), 475–486 (2007)
2. Weinhofer, G., Hoffmann, V.: Mitigating climate change – How do corporate strategies differ? Business Strategy and the Environment 19, 77–89 (2010)

3. Botzen, J.W., Gowdy, J.M., Van de Bergh, J.M.: Cumulative CO_2 emissions: shifting international responsibilities for climate debt. Climate Policy 8(6), 569–576 (2008)
4. United Nations Framework Convention on Climate Change, Kyoto protocol reference manual on accounting of emissions and assigned amount, UNFCC, Germany (2008)
5. Smale, R., Hartley, M., Hepburn, C., Ward, J., Grubb, M.: The impact of CO_2 emissions trading on firm profits and market prices. Climate Policy 6, 31–48 (2006)
6. Wittneben, B., Kiyar, D.: Climate change basics for managers. Managament Decision 47(7), 1122–1132 (p. 1124) (2009)
7. Brown, N., Deegan, C.: The public disclosure of environmental performance information-a dual test of media agenda setting theory and legitimacy theory. Accounting and Business Research 29(1), 21–41 (1998)
8. Hoffman, A.J.: Climate change strategy: the business logic behind voluntary greenhouse gas reductions. California Management Review 47(3), 21–46 (2005)
9. European Commission, 2000. Green Paper on Greenhouse Gas Emissions Trading within the European Union, Brussels (March 8, 2000)
10. González-Cabrera, J., Hidalgo-Martínez, M., Martín-Mateos, E., Vicente-Tavera, S.: Study of the evolution of air pollution in Salamanca (Spain) along a five-year period (1994-1998) using HJ-Biplot simultaneous representation analysis. Environmental Modelling & Software 21, 61–68 (2006)
11. Aerni, P.: What is sustainable agriculture? Empirical evidence of diverging views in Switzerland and New Zealand. Ecological Economics 68, 1872–1882 (2009)
12. Gabriel, K.R.: The Biplot graphic display of matrices with applications to principal components analysis. Biometrika 58, 453–467 (1971)
13. Gabriel, K.R., Odoroff, C.L.: Biplot in biomedical research. Statistics in Medicine 9, 469–485 (1990)
14. Galindo, M.P.: Contribución a la representación simultanea de datos multidimensionales. Doctoral Thesis, University of Salamanca (1985)
15. Galindo, M.P.: Una alternativa de representación simultánea: HJ Biplot. Questíio 10(1), 13–23 (1986)
16. Bachero, J., Esteban, J., Ivars, A.: Aplicación del Biplot al análisis comparativo de las economías de las comunidades autónomas: el caso de Asturias. In: Anales de Economía Aplicada, XIV Reunión Asepelt, Spain, June 22-23 (2000)
17. Blasius, J., Eilers, P., Gower, J.: Better biplots. Computational Statistics and Data Analysis 5, 3145–3158 (2009)
18. Vicente-Villardón, J.L.: Multbiplot. A package for Multivariante Analysis using Biplots. Departamento de Estadistica, University of Salamanca, Salamanca, Spain (2010), `http://biplot.usal.es/ClassicalBiplot/index.html`
19. Porter, M., Van der Linde, C.: Green and competitive. Harvard Business Review, pp. 121–134 (September-October 1995)
20. King, A., Lenox, M.: Exploring the locus of profitable pollution reduction. Mangement Science 48(2), 289–299 (2002)
21. OECD, OECD strategic response to the financial and economic crisis (2011), `http://www.oecd.org/crisisresponse` (accessed June 2012)

Appendix 1. Variation of CO_2 emission

Geographical zones	Country	VGAS 2006/2007	VGAS 2007/2008	VGAS 2008/2009
EU	UK	0.035	0.200	0.158
NA	USA	-0.040	-0.046	-0.008
NA	CANADA	-0.029	-0.100	-0.073
EU	NETHERLANDS	0.140	0.006	-0.118
EU	ITALY	0.160	0.396	0.169
NA	USA	0.009	-0.058	-0.066
NA	USA	0.100	-0.200	-0.287
NA	USA	0.093	-0.135	-0.216
EU	FRANCE	0.025	-0.070	-0.094
NA	USA	0.045	-0.065	-0.107
EU	FRANCE	-0.038	-0.065	0.055
NA	USA	0.016	-0.110	0.000
EU	UK	-0.066	-0.006	0.060
AS	JAPAN	0.060	0.171	0.090
EU	GERMANY	0.003	0.102	0.104
EU	NETHERLANDS	0.038	-0.080	-0.912
EU	GERMANY	-0.050	-0.247	-0.898
EU	GERMANY	0.010	-0.013	-0.003
NA	USA	-0.005	-0.216	-0.043
NA	USA	0.020	-0.651	-0.169
EU	GERMANY	-0.016	-0.038	-0.023
EU	GERMANY	-0.014	-0.014	0.000
AS	JAPAN	0.032	-0.067	-0.098
NA	USA	-0.004	-0.164	-0.167
EU	GERMANY	-0.076	0.216	0.309
EU	RUSIA	-0.010	-0.345	0.075
EU	GERMANY	0.150	0.000	-0.160
EU	SWITZERLAND	0.090	-0.011	-0.101
NA	USA	0.036	0.000	-0.037
AS	SOUTH KOREA	-0.060	0.113	0.189
NA	USA	0.080	-0.086	-0.161
AS	JAPAN	0.052	0.100	0.042
AS	JAPAN	-0.030	-0.148	-0.011
NA	USA	0.018	-0.008	-0.027
AS	CHINA	-0.100	0.000	0.104
AS	JAPAN	-0.060	-0.410	0.006
AS	JAPAN	0.040	-0.145	-0.179

Appendix 1. (*continued*)

AS	JAPAN	-0.030	-0.086	-0.058
EU	NORWAY	-0.026	-0.258	-0.041
EU	UK	0.290	-0.193	-0.431
AS	AUSTRALIA	-0.003	0.035	0.038
SA	BRAZIL	-0.078	0.440	0.553
NA	CANADA	0.230	0.535	0.179
NA	USA	-0.006	0.019	0.025
SA	MEXICO	-0.120	0.248	0.401
EU	UK	0.000	0.148	0.148
EU	SWITZERLAND	-0.380	0.000	0.385
AS	JAPAN	-0.040	0.099	0.150
EU	GERMANY	0.018	-0.076	-0.093
EU	GERMANY	-0.050	-0.049	0.000
EU	GERMANY	0.026	-0.039	-0.064
EU	ITALY	-0.350	-0.177	0.112
NA	USA	0.085	-0.077	-0.156
NA	USA	0.170	0.000	-0.179
AS	SOUTH KOREA	0.200	0.087	-0.132
AS	JAPAN	-0.014	-0.374	-0.366
NA	USA	0.020	-0.023	-0.013
AS	JAPAN	0.270	0.682	0.217
EU	FRANCE	0.018	-0.042	-0.060
EU	GERMANY	-0.070	0.004	0.076
AS	JAPAN	0.200	-0.094	-0.284
AS	JAPAN	-0.012	-0.152	-0.163
AS	JAPAN	-0.020	-0.197	-0.157
EU	SWEDEN	0.090	0.200	0.091
EU	UK	0.002	-0.102	-0.039
EU	SPAIN	0.004	0.021	0.037
NA	USA	0.010	-0.018	-0.031
AS	CHINA	0.043	0.000	-0.044
NA	USA	-0.001	0.027	0.029
EU	ITALY	-0.100	-0.075	0.019
NA	USA	-0.030	-0.071	-0.910
EU	PORTUGAL	0.026	-0.050	-0.023
AS	JAPAN	0.043	-0.015	0.028
EU	AUSTRIA	0.120	-0.008	-0.032
NA	CANADA	0.035	-0.999	0.035

Appendix 1. (*continued*)

SA	BRAZIL	0.016	-0.009	0.073
EU	POLAND	0.040	-0.018	-0.060
EU	SPAIN	-0.030	-0.031	0.002
EU	NETHERLANDS	0.060	-0.109	-0.128
AS	SOUTH KOREA	-0.050	0.000	0.055
NA	USA	-0.019	-0.295	-0.083
EU	UK	-0.740	-0.016	0.918
EU	FRANCE	0.017	0.170	0.087
EU	FRANCE	-0.005	-0.126	-0.121
AS	SOUTH KOREA	-0.200	-0.025	0.170
EU	UK	0.120	0.187	0.039
EU	UK	-0.020	-0.122	-0.099
EU	SWEDEN	0.003	-0.024	-0.027
EU	FRANCE	-0.030	0.195	0.293

Source: The authors.

Introducing Integrated Acceptance and Sustainability Assessment of Technologies: A Model Based on System Dynamics Simulation

Dace Aizstrauta and Egils Ginters

Sociotechnical Systems Engineering Institute
Vidzeme University of Applied Sciences
Valmiera, Latvia
{dace.aizstrauta,egils.ginters}@va.lv

Abstract. This article introduces the cornerstones of Integrated Acceptance and Sustainability Assessment Model used for new and existing technologies evaluation. It combines the user acceptance metrics used by UTAUT model with additional socio-technical factors influencing development and running of technology. The model is developed using system dynamics approach and consists of four main flows – management, quality of technology, technology acceptance and domain development.

Keywords: Acceptance and sustainability, Integrated Acceptance and Sustainability Assessment Model (IASAM), System dynamics simulation, STELLA, CHOReOS.

1 Introduction

Recent studies focus on behavioral aspects of technology acceptance or adoption. There have been plenty of researches on different factors that influence information technology acceptance – individual, organizational aspects, cultural, gender and professional differences. The most prominent model to be mentioned is Technology Acceptance Model (TAM)[1]. It has been criticized for focusing on initial adoption and not on continuous use [2]. There are also other approaches, for example Expectation-Confirmation Theory (ECT) [3] that initially originated in marketing sphere and Unified Theory of Acceptance and Use of Technology (UTAUT) that tries to consolidate eight approaches into [4].

The above mentioned adoption/acceptance theories focus mainly on exploitation stage and deal with prediction and modeling of the behavior of users that make the decision to adopt the technology or reject it. But to invest for elaboration of new technologies, one has to be sure that the possibility of failures has been diminished also in the development stage or during testing and maintenance, as different socio-technical factors influencing these stages might also lead to failure of the whole project. Therefore this article introduces the Integrated Acceptance and Sustainability Assessment Model (IASAM) and addresses the question, how to evaluate technology

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 23–30, 2013.

acceptance and sustainability at any chosen point of time of the technology life cycle and forecast the chances of technology to attract users and achieve the aims of its developers? What are the main elements and factors that influence the acceptance and sustainability of technology? IASAM suggests integrating the UTAUT approach for acceptance evaluation with other socio-technical factors thus framing united multi-level framework for technology assessment (see Fig. 1).

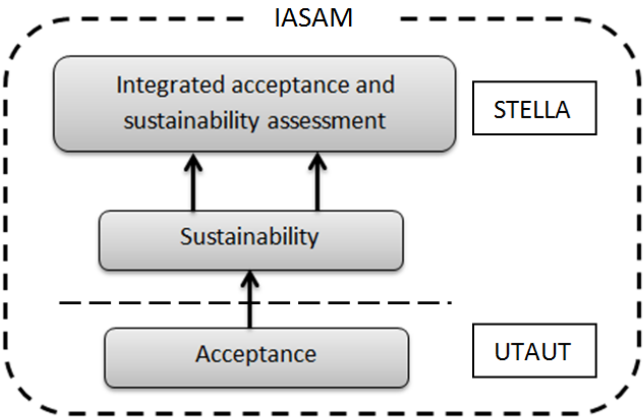


Fig. 1. Technology acceptance and sustainability assessment framework

By introducing IASAM, the authors also propose the concept of technology sustainability for evaluation of the set of socio-technical factors that let the technology to be developed, implemented, maintained properly (according to the needs of all stakeholders) and attract long-term users and create positive output and/or outcome according to the purpose of the technology and initial intentions of its developers (financial, social, etc).

At this phase of the research the authors concentrate on Information, Communication Technologies and Electronics (ICTE) systems, but it is possible to broaden this issue by including different kinds of technologies and other products.

2 Previous Studies

There are several theories that partly reflect the issues of this research, but none of them gives full understanding about the factors influencing acceptance and sustainability combined. Moreover, only a few theories analyze system sustainability, although this parameter is critically important for decisions about investments in technology development and exploitation.

This article already mentioned several theories that question the factors behind the intentions and behaviors of users from psychological perspective. Different variations of TAM, UTAUT model, ECT are just some to mention in the discussion of technology acceptance and adoption research.

Technology life-cycle approach concentrates on defining universal stages that can be applied to technology and innovation research. In comparison with acceptance research, this approach focuses rather on market forces, management decisions. In the literature, it is common to see the terms industry life cycle, product life cycle and technology life cycle used interchangeably, ambiguously and often inappropriately. Moreover the discourse is dominated by the product life cycle while the technology life cycle has largely been neglected [5]. The unit analysis for technology life cycle is broader that a specific product or a process innovation, which applies to products sold in different markets [6]. Taylor&Taylor [5] point out that this is only the tip of the iceberg since there are also disconnects and inconsistencies pertaining to the various perspectives on the technology life cycle.

This approach does not answer the questions mentioned above as it concentrates rather on commercial/managerial problems and views technology as separate item and does not analyze the differences of the technologies themselves.

3 Technology Sustainability Explained

Generally speaking, sustainability is the capacity to endure. But it differs from viability, as it includes additional meaning. The terms “sustainable” and “sustainability” have recently gained wide popularity in different domains of life. Being one of the most discussed topics these terms have also gained many meanings and conceptual interpretations. Different definitions range over such concepts as resource use, long-term existence, responsible management, in different application situations they have environmental, economic and social dimensions.

The sustainability of technology is mostly analyzed within the context of environmental issues that the technology itself creates or problems it helps to tackle [7, 8]. Speaking about sustainable technology, one refers to environmentally, socially and economically responsible technology that eases promotes or creates some kind of benefits. Only a few authors broaden the understanding of sustainability of technology and especially ICTE (for example, Gubrod &Wiele write about sustainable software [9]).

In the domain of information system research, sustainability has been addressed in the research of digital sustainability [10], sustainable competitive advantage from IT usage [11] and software sustainability.

The authors define a framework for technology acceptance and sustainability – four basic flows of the model are described in more detail including the factors that impact acceptance and sustainability.

4 Technology Acceptance and Sustainability

Within this approach, there are four main flows that shape ICTE sustainability.

$$S = < M, Q, A, D > . \quad (1)$$

S – Sustainability;
M – Management;
Q – Quality of technology;
A – Acceptance;
D – Domain development.

Two internal flows are – Management of ICTE development and exploitation and Quality of technology. And two external flows are – Technology acceptance and Domain development. Each of them has several socio-technical factors that all together constitute the Integrated Acceptance and Sustainability Assessment Model. See Fig. 2 for interoperability of four flows and factors.

Continuing the previous studies on IASAM [12, 13] also here the authors use system dynamics approach for process specification and interactive simulation, which is becoming more and more popular in studying different sociotechnical systems, where engaging in a real system is not possible. In short, system dynamics modeling approach provides the opportunity to simulate a time-varying system with multiple feedback links and analyze quantitative and qualitative factors [12].

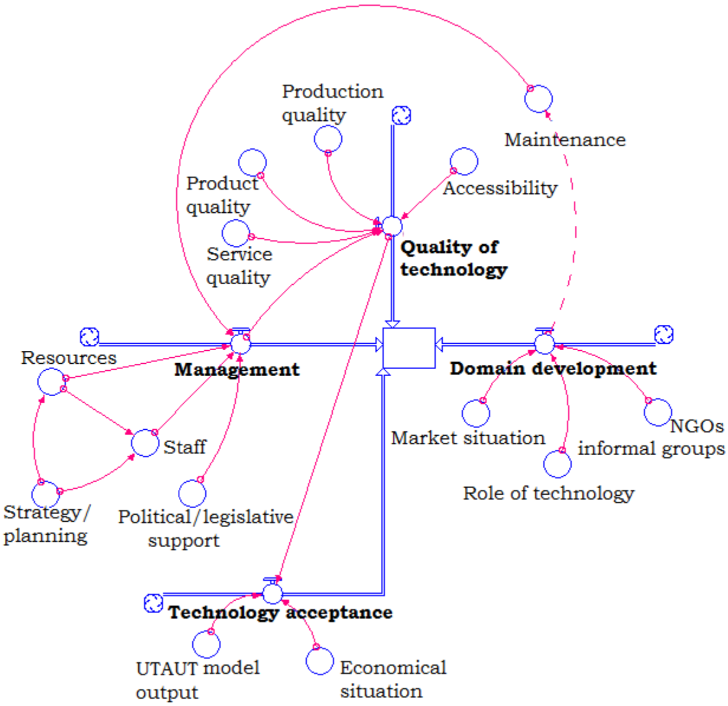


Fig. 2. IASAM model in STELLA notation

This approach allows describing technology development as a set of parallel processes. This set is characterized by:

- Socio-technical features of the system (dual nature: technical plus social and/or environmental factors;
- Development in a specific period of time;
- Involvement of multiple decision making entities, such as companies, institutions and individual consumers;
- Set of relevant internal and external socio-technical factors that impact the trends of individual change processes;
- Possibility to append or replace parameters [12].

Integrated technology sustainability and acceptance assessment model is created using the system dynamics simulation environment Stella [14].

The first of the four flows is the Management. This includes the management of technology development, as well as management of resources.

According to Gubrod&Wiele sustainability covers all aspects that potentially impact the use of any limited resource [19]. Therefore it is important whether and how the development of new technology is organized and managed. The main factors influencing the technology sustainability are as follows.

$$M = < R, S, Ps, L, Mt > . \quad (2)$$

M – Management;
R – Available resources, including financial and technical resources, etc.
S – Staff;
Ps – Strategic and managerial principles and approaches used to manage the resources ;
L – political/legislative support;
Mt – maintenance.

Technology acceptance and sustainability is also affected by its quality. Quality has multiple dimensions – technical quality, content and output quality, service quality. The approach in defining quality is similar to ICT success theory developed by DeLone and McLean in 1992 that was revised in 2003 [15]. They use a threefold understanding of quality – information quality, systems quality and service quality. As their model is aimed at information systems, the IASAM broadens the constructs.

$$Q = < M, P, Pr, Sv, As > . \quad (3)$$

Q – Quality of technology;
M – Management (flow described in section IV.A);
O – Product (output) quality;
Pr – Production quality;
S – Service (support) quality;
As – Accessibility.

The third IASAM flow turns to potential Technology acceptance by its users. Sustainability of technology cannot be explained without acceptance.

This flow is measured using basic constructs of UTAUT model. UTAUT is a definitive model that synthesizes what is known and provides a foundation to guide future research in this area. By encompassing the combined explanatory power of the individual models and key moderating influences, UTAUT advances cumulative theory while retaining a parsimonious structure [4].

Recent study on articles citing UTAUT revealed that current research on UTAUT constructs are impacted upon by many external variables across different studies [16]. That corresponds also to the strategy used in this article.

$$A = < U, E, Q > . \tag{3}$$

A – Acceptance;
U – UTAUT output;
E – Economical situation;
Q – Quality of technology.

The model also includes Domain development impacts. Despite the positive impact of technology development on the society overall, looking from the technology creators perspective at the same time, every innovation endangers its current position within the technology market. The main factors influencing this flow are as follows.

$$D = < Tm, G, R > . \tag{4}$$

D – Domain development;
Tm – Technology among other competitors in the market;
G – Nongovernmental activists/informal groups;
R – The role of technology.

5 Assessing the Technology Sustainability

The assessment according to the model is carried out using a set of criteria that were described in previous section. Each criterion is evaluated within a scale. This section reflects on the evaluation process and the scale.

The choice for a particular rating scale format can be broken down into two major components: the number of response categories to be offered, including the choice for an odd or even number of categories, and the labeling of response categories [17].

According to the developed IASAM model, each factor is measured with certain set of criteria. After examination of different types of scales and their characteristics, it has been chosen to use a 7 point Likert scale. Dawes concludes that either 5, 7 or 10-point scales are all comparable for analytical tools. Empirical studies have generally concurred that reliability and validity are improved by using five to seven-point scales rather than coarser ones (those with fewer scale points). But more finely graded scales do not improve reliability and validity further [18].

The result gained from evaluation of all criteria indicates the forecast of integrated technology sustainability and acceptance, measured as the % of expected maximum for the number of criteria that have been evaluated. The bigger the result, the more the technology satisfies the criteria of IASAM.

The methodology should be usable at any point of technology life cycle, so if there are questions that cannot be answered at the time of evaluation, then it should be marked with NA. The total number of questions marked with NA and expressed as percentage of total number of questions indicates the inner credibility of IASAM forecast. Accordingly, the smaller the percentage of inner credibility, the less credible is the forecast of technology sustainability and acceptance based on IASAM. Thus the model gives two numbers and they both characterize the evaluation of technology acceptance and sustainability.

6 Conclusion

Previous research focus separately on psychological or socio-economical aspects of technology acceptance, specifically on success of ICTE projects, on management systems, or diffusion on innovations on the whole. This paper presents a new approach for evaluation of technologies that combines socio-economical aspects and socio-technical characteristics of technology development and exploitation.

Technologies are changing rapidly and software is becoming larger and more complex. In addition, large-scale, distributed development poses new challenges [19]. The same is happening also with hardware and in the fields of Future Internet development and Cloud computing.

IASAM consists of four groups of factors that have an impact on integrated technology acceptance and sustainability – Management, Quality of technology, Acceptance and Domain development. Acceptance is measured using UTAUT methodology and other factors are evaluated using a set of pre-defined criteria. The model serves as a framework for successful technology development and assessment. By using system dynamics the model allows its users to monitor the variation of the IASAM index over time.

Approach practically has no limitations of application and it is intended to develop it to apply in other fields of economy. The validation of approach will be realised under the framework of other FP7 projects.

Acknowledgments. The IASAM model described above is under development and will be tested under the framework of FP7-ICT-2009-5 CHOReOS project No. 257178 (2010-2013) *“Large Scale Choreographies for the Future Internet (IP)”*to assess CHOReOS project pilots and whole project.

References

1. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35(8), 982–1003 (1989)
2. Premkumar, G., Bhattacherjee, A.: Explaining Information Technology Usage: A Test of Competing Models. Omega 36, 64–75 (2008)
3. Bhattacherjee, A.: Understanding Information Systems Continuance: and Expectation-Confirmation Model. MIS Quarterly 25(3), 351–370 (2001)

4. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Tward a Unified View. MIS Quarterly 27(3), 425–478 (2003)
5. Taylor, M., Taylor, A.: The Technology Life Cycle: Conceptualization and Managerial Implications. International Journal of Production Economics 140(1), 541–553 (2012)
6. Cohen, S.K.: Innovation-Driven Industry Life Cycles. In: Narayanan, V.K., O'Connor, G.C. (eds.) Encyclopedia of Technology and Innovation Management, pp. 69–76. John Wiley & Sons, Wiltshire (2010)
7. Weaver, P., Jansen, L., Van Grootveld, G., Van Spiegel, E., Vergrat, P.: Sustainable technology development. Greenleaf Publishing Ltd. (2000)
8. Hilty, L.M.: Information technology and sustainability, p. 180. Books on Demand (2008)
9. Gudbrod, R., Wiele, C.: The Software Dilemma: Balancing Creativity and Control on the Path to Sustainable Software, p. 281. Springer, Berlin (2012)
10. Bradley, K.: Defining Digital Sustainability. Library Trends 56(1), 148–163 (2007)
11. Daneshvar, P., Ramesh, H.N.: Review of Information Technology Effect on Competitive Advantage – Strategic Perspective. International Journal of Engineering Science and Technology 2(11), 6248–6256 (2010)
12. Barkane, Z., Ginters, E.: Introduction to Technology Acceptance and Sustainability Modelling. In: Annual Proceedings of Vidzeme University of Applied Sciences. ICTE in Regional Development, Valmiera, pp. 62–90 (2009/2010)
13. Ginters, E., Barkane, Z., Vincent, H.: System Dynamics Use for Technology Assessment. In: Proc. of the 22th European Modeling & Simulation Symposium (EMSS) (2010)
14. ISEE Systems, `http://www.iseesystems.com/softwares/Education/StellaSoftware.aspx`
15. Petter, S., DeLone, W., McLean, E.: Measuring Information Systems Success: Models, Dimensions, Measures, and Interrelationships. European Journal of Information Systems 17, 236–263 (2008)
16. Williams, M., Rana, N., Dwivedi, Y., Lal, B.: Is UTAUT Really Used or Just Cited for the Sake of It? A Systematic Review of Citations of UTAUT's Originating Article. In: ECIS 2011 Proceedings, Paper 231 (2011)
17. Weijters, B., Caboote, E., Schillerwaert, N.: The Effect of Rating Scale Format on Response Styles: the Number of Response Categories and Response Category Labels. Universiteit Gent, Working Paper, D/2010/7012/07 (2010), `http://feb1.ugent.be/nl/Ondz/wp/Papers/wp_10_636.pdf`
18. Dawes, J.: Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5 Point, 7 Point and 10 Point Scales. International Journal of Market Research 51(1), 61–77 (2008)
19. Magdaleno, A.M., Werner, C.M.L., Mendes de Araujo, R.: Reconciling Software Development Models: A Quasi-Systematic Review. The Journal of Systems and Software 85, 351–369 (2012)

Incorporating Stakeholders' Preferences into CSR Ratings: Fuzzy TOPSIS Approach to Evaluate Agri-Food Companies

Elena Escrig-Olmedo*, María Jesús Muñoz-Torres,
María Ángeles Fernández-Izquierdo, and Juana María Rivera-Lirio

Department of Finance and Accounting, University Jaume I, Campus de Riu Sec – Avda. Vicent Sos Baynat s/n, 12071 Castellón, Spain
{eescrig,munoz,afernand,jrivera}@uji.es

Abstract. Sustainability ratings agencies evaluate companies in economic, social, environmental and corporate governance terms. However, the scoring modes employed by the rating agencies to evaluate the social and environmental performance are associated with some problems such as the lack of adaptation of rating criteria and rating evaluation methodologies to the preferences of investors and companies. To overcome this difficulty a multi-criteria decision-making method is proposed: Fuzzy TOPSIS method. The subjectiveness and imprecision of the evaluation process are modeled as fuzzy numbers by means of linguistic terms. The proposed method is applied to measure the environmental performance in a sample of European, North American, and Australasian agri-food companies.

Keywords: agri-food sector, Fuzzy TOPSIS, Multiple criteria evaluation, sustainability.

1 Introduction

In the Brundtland Report [1] the concept of sustainable development is built on three pillars: economic prosperity, social equity and environmental protection. Currently, civil society and media request companies to consider the social and environmental impacts of their activities and to provide more information with respect to their actions [2].

In that sense, sustainability or corporate social responsibility (CSR) rating agencies study businesses and make evaluations in social, environmental and corporate governance (ESG) terms – using their own research methodologies. However, the scoring modes employed by CSR rating agencies to evaluate the social and environmental performance are associated with some problems such as the lack of adaptation of rating criteria and rating evaluation methodologies to the needs and preferences of investors and companies.

* Corresponding author.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 31–41, 2013.

It is desirable to address a large number of stakeholders with CSR ratings evaluations in order to enhance their acceptance and to promote sustainable development [3]. In this sense, Graafland et al. [4] remark the importance of disclosing methods and assumptions of benchmarks to stakeholders. However, currently, CSR ratings do not incorporate the investors and companies' knowledge to carry out analysis, comparisons and rankings of the alternatives. This incorporation ensures that CSR ratings are more balanced and accepted and also increases transparency.

For this reason, the first objective of this paper is to evaluate the performance in organizations by means of Fuzzy TOPSIS methodology to generate aggregate scores selection of best alternative, according to the preferences of investors and companies.

The proposed method is tested to measure the environmental performance in European, North American and Australasian agri-food companies.

In the last years, quality and safety of raw and processed food production is becoming increasingly important in terms of human health and competition [5], and the European institutions have adopted specific policies to ensure control over food production and to promote food quality [6]. Concerns about the environmental impact and social consequences of food production and consumption have grown as well, making people more interested in the way in which their food is produced [7]. Moreover, reduction in environmental emissions, disposition of harmful wastes and adoption of green technology are among the objectives of several organizations [8].

In this context, the second objective of this paper is to identify good environmental practices in the agri-food sector and to rank agri-food companies, according to their environmental performance, using a multi-criteria decision-making (MCDM) method.

This paper is structured as follows: the introduction is followed by a brief analysis of the theoretical background of this research. Subsequently, we analyse the main characteristics of the methodology used and explain the design of the study. After the presentation of the results obtained, the main conclusions are provided.

2 Theoretical Background

2.1 Agri-Food Industry and Environment

Following Global Industry Classification Standard (GICS) - that is an industry taxonomy developed by MSCI and Standard & Poor's- the food processing industry covers different subsectors quite different among then, such as food and staples retailing or food, beverage and tobacco.

Agri-food sector is highly dependent on natural resources and produces considerable and different impacts on the environment [9, 10]. Bearing in mind this, it is necessary a sustainable agri-food sector. In this sense, SustainAbility Organization defines a sustainable food system as "*one that is reliable, resilient and transparent, which produces food within ecological limits, empowers food producers, and ensures accessible, nutritious food for all*".

2.2 Socially Responsible Investing vs. Corporate Social Responsibility

Socially responsible investing (SRI) is the financial investment process that takes into account social, environmental and corporate governance impacts and/or investment in the community – as well as shareholder activism [11] and, it is the best practice for investors to incorporate CSR in their investment decision [12].

The integration of the sustainability concept in the companies' initiatives is accomplished via CSR. It can be defined as "*an aggregate of the voluntary policies and subsequent management processes arising from corporate decisions aimed at achieving actions in the field of economics, environment, social development, and corporate governance*" [13]. Consequently, CSR may be considered as an instrument for implementing the concept of business sustainability [14].

Companies and investors perceive the value of CSR differently. Companies want to obtain a competitive advantage and long-term value by working strategically with CSR [15], whereas investors do not have properly methods to include ESG factors in an evaluation of the company. Successful active investor should have policies and processes that shape a desired impact, which usually results in a combination of improved financial and ESG performances [16].We consider that it is necessary generate an integrated decision model. CSR should be linked to business strategy, core business and day-to-day processes in all organizational units [17].

2.3 Corporate Social Responsibility Ratings and Rankings

Specialised rating agencies have appeared developing new rating typologies in response to changing attitudes towards responsible behaviour. These ratings study companies according to economic and ESG indicators. CSR Ratings can push companies to a more sustainable behavior, foster the institutionalization of information management, and stimulate competition between companies [18, 4]. However some issues such as the lack of standardization and transparency, bias, weighting tradeoffs, credibility of information and independence should be improve [19]. In general terms, the final overall rating is the result of adding the scores in each of the key domains or indicators but this overall rating has several limitations: (i) higher scores for one domain may hide very low scores in another domain [4, 19, 20]; (ii) the balance is not reflected between the various aspects of management necessary for a firm to be listed as socially responsible; and (iii) the varying assessments that different stakeholders may give to each criterion are not included.

2.4 Fuzzy TOPSIS

TOPSIS has been used to rank the preference order of alternatives and determine the optimal choice. It was first developed by Hwang and Yoon [21]. It helps decision maker(s) (DMs) organize the problems to be solved and carry out analysis, comparisons and rankings of the alternatives. The essence of these problems is the processing of information to consider different attributes with different importance degrees in the process for proposing a solution to the problem analysed [22].

The basic concept of TOPSIS is that the chosen alternative should have the shortest distance from the positive-ideal solution (PIS) i.e., the solution that maximizes the benefit criteria and minimizes the cost criteria; and the farthest distance from the negative-ideal solution (NIS) i.e., the solution that maximizes the cost criteria and minimizes the benefit criteria [23]. A more realistic approach may be to use linguistic assessments instead of numerical values [24, 25, 26]. In this sense, Fuzzy Set Theory presents a framework for modeling the investor selection problem in an uncertainty environment; and it can be combined with other techniques to improve the quality of the final tools like TOPSIS.

Fuzzy TOPSIS can be expressed in a series of steps:

Step 1: *Determine the weighting of evaluation criteria*. The importance weights of evaluation criteria and the ratings of alternatives are considered as linguistic terms.

Step 2: *Establish of decision matrix* $\tilde{D}$, where the number of criteria is *n* and the count of alternatives is *m*. Fuzzy decision matrix will be obtained with *m* rows and *n* columns.

Step 3: *Calculate the normalized decision matrix* in order to make each criterion value is limited between 0 and 1, so that criterion is comparable.

Step 4: *Calculate the weighted normalized fuzzy decision matrix*. The weighted normalized value $\tilde{v}_{ij}$ is calculated by multiplying the weights $\tilde{w}_j$ of criteria with the normalized fuzzy decision matrix $\tilde{r}_{ij}$. The weighted normalized decision matrix $\tilde{R}$ for each criterion is calculated through the following relations:

$$\tilde{R} = \left[\tilde{v}_{ij}\right]_{m\times n} \quad i = 1,...,m,\ j = 1,...,n$$
$$Where \quad \tilde{v}_{ij} = \tilde{r}_{ij}(\cdot)\tilde{w}_i \tag{1}$$

Step 5: *Determine the fuzzy positive-ideal solution (FPIS) and fuzzy negative-ideal solution (FNIS).*

$$A^+ = \left(\tilde{v}_1^+, \tilde{v}_2^+, ..., \tilde{v}_n^+\right) = \{\left(\max \tilde{v}_{ij} \,|\, i = 1,...,m\right), j = 1,...,n\}$$
$$A^- = \left(\tilde{v}_1^-, \tilde{v}_2^-, ..., \tilde{v}_n^-\right) = \{\left(\min \tilde{v}_{ij} \,|\, i = 1,...,m\right), j = 1,...,n\} \tag{2}$$

Here $\tilde{v}_j^+$ (1,1,1) and $\tilde{v}_j^-$ (0,0,0)

Step 6: *Calculate distance from the FPIS and FNIS for each alternative*. According to Bojadziev and Bojadziev [27], the distance between two triangular fuzzy numbers (Ã) and ($\tilde{B}$) is calculated as:

$$d(\tilde{A},\tilde{B}) = \sqrt{\frac{1}{3}\left[(a_1-b_1)^2 + (a_2-b_2)^2 + (a_3-b_3)^2\right]} \tag{3}$$

The separation of each alternative from the ideal solution is given as

$$d_i^+ = \sum_{j=1}^{k} d\left(\tilde{v}_{ij}, \tilde{v}_j^+,\right), i = 1,...,m \tag{4}$$

Similarly, the separation from the negative ideal solution is given as

$$d_i^- = \sum_{j=1}^{k} d\left(\tilde{v}_{ij}, \tilde{v}_j^-,\right), i = 1,...,m \tag{5}$$

Step 7: *Calculate the relative closeness coefficient to the ideal solution*. The alternative with the highest closeness coefficient (CC) value will be the best choice.

$$CC \quad or \quad \overline{R}_i = \frac{\overline{d}_i^-}{\overline{d}_i^+ + \overline{d}_i^-}, i = 1,...,m \tag{6}$$

Step 8: *Rank preference order*. The ranking of the alternatives can be determined according to the CC in descending order.

This method present some advantages [28]: it provides the ranks of available alternatives according to the best and worst combinations of characteristics, and the weights capture the tacit knowledge of experts involved.

3 Methodology

3.1 Sample and Sustainability Key Indicators Selection

Data have been collected from ASSET4 Database to the year 2010. The final sample is composed of 54 European companies, 57 North American companies and 55 Australasian companies belonging to the agri-food industry (Table 1).

Table 1. Classification of the sample according to GICS

Consumer Staples *(166 companies)*	**Food & Staples Retailing** *(41 Companies)*	**Food & Staples Retailing** *(41 companies)*	**Drug Retail** *(6 companies)*
			Food Distributors *(6 companies)*
			Food Retail *(27companies)*
			Hypermarkets & Super Centers *(2 comp)*
	Food, Beverage & Tobacco *(125 companies)*	**Beverages** *(35 companies)*	**Brewers** *(12 companies)*
			Distillers & Vintners *(8 companies)*
			Soft Drinks *(15 companies)*
		Food Products *(80 companies)*	**Agricultural Products** *(13 companies)*
			Packaged Foods & Meats *(67 companies)*
		Tobacco *(10 companies)*	**Tobacco** *(10 companies)*

The indicators were selected according to the information that ASSET4 declares to use in assessing environmental performance. In order to select the most relevant environmental indicators applicable to agri-food industry, Global Reporting Initiative Guidelines -GRI- 2006 [29] as well as specific tools and protocols of the agri-food sector, such as Sustainability Reporting Guidelines & Food Processing Sector Supplement 2011 [30], were matched with ASSET4 indicators (Table 2). The final number of indicators was reduced to 62.

Table 2. Example Environmental Indicators by agri-food industry

Code	Indicator name	GRI Guidelines	Accountability Rating
En_En_ER_D01_1	Emission Reduction/ Policy_1	GRI-EN18	SI
En_En_ER_O11	Emission Reduction/ Waste Recycling Ratio	GRI-EN24 GRI-EN22	OP
En_En_ER_O16	Emission Reduction/ Environmental Partnerships	Disclosure on Management approach	E
En_En_RR_O11_1	Resource Reduction/ Environmental Supply Chain Management	Monitoring and Follow-up	GM

After that, the environmental indicators were classified in four domains: strategic intent (SI), governance and management (GM), engagement (E), and operational performance (OP), according to the methodology of Accountability Rating 2008. Of the total number of final indicators 13 were classified into SI domain, 11 into GM domain, 26 into E domain and 12 into OP domain.

3.2 Fuzzy TOPSIS

We want to establish the best agri-food company in terms of environmental performance taking into account the criteria strategic intent (SI = C_1), governance and management (GM= C_2), engagement (E= C_3), and operational performance (OP= C_4). All the variables (C_1, C_2, C_3 and C_4) are called positive, ie, must be maximized.

To work with the data in the Fuzzy TOPSIS, we must normalize them. An example of the normalized decision matrix for this problem is presented in Table 3.

Table 3. Normalized decision matrix

Companies	SI	GM	E	OP
A_1	0.692	0.545	0.423	0.501
A_2	0.692	0.727	0.500	0.500
A_3	0.615	0.455	0.346	0.500
A_4	0.769	0.545	0.462	0.500
…	…	…	…	…
A_{165}	0.692	0.455	0.346	0.500
A_{166}	0.000	0.000	0.000	0.500

The importance weights of various criteria and the ratings of qualitative criteria are considered as linguistic variables. These linguistic variables can be expressed in positive triangular fuzzy numbers as Tables 4 and 5. We have used triangular fuzzy numbers because it is the most common in the literature [31].

Table 4. Linguistic variables for the relative importance weights of four criteria

Linguistic Variables	Membership Functions
Very Low (VL)	(0, 0, 0.25)
Low (L)	(0, 0.25, 0.5)
Medium (M)	(0.25, 0.5, 0.75)
High (H)	(0.5, 0.75, 1)
Very High (VH)	(0.75, 1, 1)

Table 5. Linguistic variables for the ratings

Linguistic Variables	Membership Functions
Poor Sustainability (PS)	(0, 0, 0.25)
Low Sustainability (LS)	(0, 0.25, 0.5)
Medium Sustainability (MS)	(0.25, 0.5, 0.75)
High Sustainability (HS)	(0.5, 0.75, 1)
Top Sustainability (TS)	(0.75, 1, 1)

Criteria importance is a reflection of the decision maker's subjective preferences (expert knowledge) as well as the objective characteristics of the criteria themselves [32]. In our study decision makers have been represented by one socially responsible investor (DM_1) and one company (DM_2), and their preferences are according to the literature on this field and the results of a survey among Spanish investors [33].

The decision-making process of a socially responsible consumer is more complex than the process for consumers of traditional financial products. This is because social and environmental criteria are sometimes more important than financial criteria [34]. According to Escrig et al. [33], engagement and operational performance are the most important criteria for investors today. Meanwhile, 79% of investors consider natural resources and corporate governance feature as the most relevant types of extra-financial information in company analysis according to the report "The value of extra-financial disclosure. What investors and analysts said (2012)" published by GRI [35].

Most ratings focus on larger companies and not include SMEs in their evaluations. However, in our study, DM_2 evaluations incorporate the large and SMEs preferences. Companies have some problems to express with exact numerical values their social and environmental performance. Currently, companies focus on how they interact with stakeholders and how business activities impact on society [36]. According to Muñoz et al. [37] corporate governance and supplier management were considered by the experts as not being important issues with respect to CSR management in companies, specifically in SMEs. However, nowadays, corporate governance and management issues are more and more important for multinational companies.

The two decision makers/ experts (investor and company) have expressed their opinions on the importance weights of the four criteria and the ratings of each company with respect to the four criteria independently. These opinions can differ from other investor's opinions or from other company's opinions.

Tables 6 and 7 show the original assessment information provided by the two decision makers, where aggregated fuzzy numbers are obtained by averaging the fuzzy opinions of the two DMs. That is $\tilde{v}_j = (\tilde{v}_j^1 + \tilde{v}_j^2)/2$ and $\tilde{x}_{ij} = (\tilde{x}_{ij}^1 + \tilde{x}_{ij}^2)/2$, where $\tilde{v}_j$ and $\tilde{x}_{ij}$ are the relative importance weight and the ratings given by the DMs.

Table 6. The relative importance weights of the four criteria by two DMs

Criterion	DM_1 (investor)	DM_2 (company)	Aggregated fuzzy number
C_1 = SI	M	H	(0.38, 0.63, 0.88)
C_2 = GM	VH	M	(0.50, 0.75, 0.88)
C_3 = E	H	H	(0.50, 0.75, 1.00)
C_4 = OP	H	VH	(0.63, 0.88, 1.00)

Table 7. Ratings of the three first companies with respect to the four criteria by the two DMs

Criteria	Candidates	DM_1 (investor)	DM2 (company)	Aggregated fuzzy number
C_1 = SI	A_1	MS	HS	(0.38, 0.63, 0.88)
	A_2	MS	HS	(0.38, 0.63 , 0.88)
	…	…	…	…
C_2 = GM	A_1	MS	MS	(0.25, 0.5, 0.75)
	A_2	HS	HS	(0.5, 0.75, 1.00)
	…	…	…	…
C_3 = E	A_1	LS	MS	(0.13, 0.38, 0.63)
	A_2	LS	MS	(0.13, 0.38, 0.63)
	…	…	…	…
C_4 = OP	A_1	MS	MS	(0.25, 0.5, 0.75)
	A_2	LS	LS	(0, 0.25, 0.50)
	…	…	…	…

Subsequently, the normalized fuzzy decision matrix has been constructed. However, if all the criteria/attributes, $C_1,\ldots,C_n$, are assessed using the same set of fuzzy linguistic variables, then the fuzzy decision matrix is of the same dimension and therefore needs no normalization [31].

Table 8. The fuzzy normalized decision matrix and fuzzy weights of seven alternatives

Companies	C_1 = SI	C_2 = GM	C_3 = E	C_4 = OP
A_1	(0.38,0.63, 0.88)	(0.25, 0.5, 0.75)	(0.13, 0.38, 0.63)	(0.25,0.5, 0.75)
A_2	(0.38, 0.63, 0.88)	(0.5, 0.75, 1)	(0.13, 0.38, 0.63)	(0, 0.25, 0.5)
A_3	(0.38, 0.63, 0.88)	(0.13,0.38, 0.63)	(0, 0.25, 0.5)	(0.13, 0.38, 0.63)
A_4	(0.50, 0.75, 1)	(0.25,0.5, 0.75)	(0.13,0.38, 0.63)	(0.13, 0.38, 0.63)
…	…	..	…	….
A_{165}	(0.38, 0.63, 0.88)	(0.13,0.38,0.63)	(0, 0.25, 0.5)	(0, 0.25, 0.5)
A_{166}	(0, 0, 0.25)	(0,0, 0.25)	(0,0, 0.25)	(0, 0.25, 0.5)
Weight	**(0.38, 0.63, 0.88)**	**(0.50, 0.75, 0.88)**	**(0.50, 0.75, 1.00)**	**(0.63, 0.88, 1.00)**

Then, the weighted normalized fuzzy decision matrix was constructed using Equation (1).

Table 9. The fuzzy weighted normalized decision matrix

Companies	C_1 = SI	C_2 = GM	C_3 = E	C_4 = OP
A_1	(0.14, 0.39, 0.77)	(0.13, 0.38, 0.66)	(0.06, 0.28, 0.55)	(0.16, 0.44, 0.75)
A_2	(0.14, 0.39, 0.77)	(0.25, 0.56, 0.88)	(0.06, 0.28, 0.55)	(0.00, 0.22, 0.50)
A_3	(0.14, 0.39, 0.77)	(0.06, 0.28, 0.55)	(0, 0.19, 0.44)	(0.08, 0.33, 0.63)
A_4	(0.19, 0.47, 0.88)	(0.13, 0.38, 0.66)	(0.06, 0.28, 0.55)	(0.08, 0.33, 0.63)
…	…	…	…	…
A_{165}	(0.14, 0.39, 0.77)	(0.06, 0.28, 0.55)	(0, 0.19, 0.44)	(0.00, 0.22, 0.50)
A_{166}	(0, 0, 0.22)	(0, 0,0.22)	(0, 0, 0.22)	(0.00, 0.22, 0.50)

Calculated the weighted matrix, the Fuzzy Positive Ideal Solution (FPIS, A^+), which represents the compromise solution, and the Fuzzy Negative Ideal Solution (FNIS, A^-), which represents the worst possible solution, were defined respectively, for triangular set fuzzy, as:

$$A^+ = [(1,1,1), (1,1,1), (1,1,1), (1,1,1)]$$
$$A^- = [(0,0,0), (0,0,0), (0,0,0), (0,0,0)]$$

Finally, we calculated the distance of each company from FPIS and FNIS according equation (3) and the closeness coefficient to the ideal solution as Table 10. A large closeness coefficient of a company indicates good environmental performance.

Table 10. The distance measurement

Companies	d^+	d^-	CC
A_1	2.60327861	1.81888179	0.41131068
A_2	2.64365314	1.79992975	0.40506272
A_3	2.85404127	1.55123512	0.35213117
A_4	2.63571289	1.80023196	0.4058283
…	…	…	
A_{165}	2.94870975	1.45603899	0.33056119
A_{166}	3.58441794	0.69643454	0.16268595

4 Results

Of all firms ranking in the first 50 places; 56% are European, 24% are North American and 20% are Australasian. Moreover, it is worth noting that among the top

20 more than half are European companies, 4 are American and only 2 are Australasian (Table 11). This seems to indicate that the environmental protection issues are more integrated in Europe than in the rest of the analysed world areas.

Table 11. Top 20 Environmental Ranking to Agri-food Industry (Fuzzy TOPSIS)

Fuzzy TOPSIS ranking	Companies	Region	CC
Company_1	A_{65}	Europe	0.52091104
Company_2	A_9	North America	0.50061343
Company_3	A_7	Europe	0.4981034
Company_4	A_{54}	North America	0.4981034
Company_5	A_{22}	Europe	0.49595408
Company_6	A_5	Europe	0.49543631
Company_7	A_8	North America	0.47930895
Company_8	A_{64}	Asia	0.47763862
Company_9	A_6	Europe	0.47401481
Company_10	A_{74}	Europe	0.47401481
Company_11	A_{24}	North America	0.47208818
Company_12	A_{122}	Europe	0.46974765
Company_13	A_{26}	Europe	0.46378209
Company_14	A_{18}	Europe	0.46338196
Company_15	A_{31}	Europe	0.45837799
Company_16	A_{11}	Europe	0.45792493
Company_17	A_{14}	Europe	0.4556082
Company_18	A_{67}	Oceania	0.4556082
Company_19	A_{66}	Europe	0.45059629
Company_20	A_{17}	Europe	0.45057121

5 Conclusion

Assessing corporate sustainability is increasingly practice-relevant and a large number of sustainability ratings, rankings and indices have appeared in the last years. However, the complexity in the development of sustainability rating makes it difficult to design evaluation methodologies.

This paper contributes to the literature by correcting the weaknesses of rating evaluation methodologies in two aspects:

(i) *Incorporation of preferences of investors and companies in the rating evaluation process.* The different criteria, evaluation and scoring systems used by the sustainability rating agencies are not according to the interests of investors and companies [3]. It would be interesting for companies to have tools for evaluating their sustainability at their disposal and it would be interesting for socially responsible investors to have tools for choosing their investments according to SRI criteria.

A Fuzzy TOPSIS Methodology is proposed to incorporate the investors and companies' preferences in the rating agencies evaluation process. We develop an integrated fuzzy technique for order preference by similarity to ideal solution and to improve the quality of decision making for ranking alternatives. In fuzzy TOPSIS all the ratings and weights are defined by means of linguistic variables.

(ii) *Reaching consensus between investors and companies.* Sustainability rating agencies evaluate CSR of companies to offer different SRI alternatives. Rating evaluation process should therefore integrate the different sensibilities and objectives of both stakeholders (investors and companies). In this sense, the methodology proposed allows to integrate the investors and companies preferences and to reach a consensus about what CRS is and how to translate it to investors' decisions.

Finally, the environmental performance in agri-food companies is analized by means on an Environmental Ranking to agri-food industry.

Acknowledgments. This work is supported by P1.1B2010-04 and P1.1B2010-13 research projects, and Sustainability and Corporate Social Responsibility Master Degree (UJI – UNED).

References

1. World Commission on Environment and Development (WCED), Our Common Future. Oxford University Press, Oxford (1987)
2. Freeman, R.E., Harrison, J.S., Wicks, A.C., Parmar, B.L., De Colle, S.: Stakeholder Theory. The State of the Art. Cambridge University Press, Cambridge (2010)
3. Windolph, S.E.: Assessing corporate sustainability through ratings: challenges and their causes. Journal of Environmental Sustainability 1, 61–81 (2011)
4. Graafland, J., Eijffinger, S.C., Smid, H.: Benchmarking of Corporate Social Responsibility: Methodological Problems and Robustness. J. Bus. Ethics 53, 137–152 (2004)
5. Bakucs, L.Z., Fertő, I., Havas, A.: Future impact of new technologies upon food quality and health in Central Eastern European countries. In: Studies on the Agricultural and Food Sector in Central and Eastern Europe, 46 Halle (Saale), IAMO, pp. 82–94 (2008)
6. Nomisma. - Società di Studi Economici: European agriculture of the future the role plant protection products, BOLOGNA. Working paper (2008)
7. DuPuis, E.M.: Not in my body: rBGH and the rise of organic milk. Agric. Human Values 17, 285–295 (2000)
8. Awasthi, A., Chauhan, S.S., Goyal, S.K.: A fuzzy multicriteria approach for evaluating environmental performance of suppliers. Int. J. Production Economics 126, 370–378 (2010)
9. Hartmann, M.: Corporate social responsibility in the food sector. Eur. Rev. Agric. Econ. 38, 297–324 (2011)
10. Maloni, M.J., Brown, M.E.: Corporate social responsibility in the supply chain: an application in the food industry. J. Bus. Ethics 68, 35–52 (2006)
11. Social Investment Forum –SIF-, What is Socially Responsible Investing (SRI)? (2009), http://www.socialinvest.org/resources/sriguide/srifacts.cfm
12. Renneboog, L., Ter Horst, J., Zhang, C.: Socially responsible investments: Institutional aspects, performance, and investor behaviour. J. Bank Financ. 32, 1723–1742 (2008)
13. Muñoz, M.J., Fernández, A., Nieto, L., Rivera, J.M., Escrig, E., León, R.: SMEs and corporate social responsibility: The perspective from Spanish companies. International Journal of Sustainable Economy 1, 270–288 (2009)
14. European Commission -EC-, Communication from the Commission, Concerning Corporate Social Responsibility: A Business Contribution to Sustainable Development (2002), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0347:FIN:EN:PDF
15. Porter, M., Kramer, M.: Strategy & society: The link between competitive advantage and corporate social responsibility. Harv. Bus. Rev. 84, 78–92 (2006)
16. Eurosif: Active share ownership in Europe: 2006 European Handbook. Eurosif (2006), http://www.mvovlaanderen.be/uploads/1244116859-1244116838790-active-share-ownership-in-europe.pdf

17. Stubbs, W., Cocklin, C.: An ecological modernist interpretation of sustainability: the case of Interface Inc. Bus. Strat. Env. 17, 512–523 (2008)
18. Fowler, S.J., Hope, C.: A Critical Review of Sustainable Business Indices and their Impact. J. Bus. Ethics 76, 243–252 (2007)
19. Windolph, S.E.: Assessing corporate sustainability through ratings: challenges and their causes. Journal of Environmental Sustainability 1, 61–81 (2011)
20. Escrig, E., Rivera, J.M., Muñoz, M.J., Fernández, M.A.: Integrating Sustainability through Fuzzy Logic: Fortune Global 100 Sustainability Rating. In: Computational Intelligence in Business and Economics Proceedings of the MS 2010 International Conference, Barcelona, Spain, July 15-17 (2010)
21. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications. Springer, Heidelberg (1981)
22. Santos, F.-J.J., Camargo, H.A.: Fuzzy Systems for Multicriteria Decision Making. CLEI Electronic Journal 13 (2010)
23. Saghafian, S., Hejazi, S.R.: Multi-Criteria Group Decision Making Using A Modified Fuzzy TOPSIS Procedure. In: Proc. of the Int. Conf. on Computational Intelligence for Modeling, Control and Automation. IEEE Computer Society (2005)
24. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Management Sci. 17, 141–164 (1970)
25. Delgado, M., Verdegay, J.L., Vila, M.A.: Linguistic decision making models. Int. J. Intelligent System 7, 479–492 (1992)
26. Herrera, F., Herrera-Viedma, E., Verdegay, J.L.: A model of consensus in group decision making under linguistic assessments. Fuzzy Sets and Systems 78, 73–87 (1996)
27. Bojadziev, G., Bojadziev, M.: Fuzzy Sets, Fuzzy Logic, Applications. Advances in Fuzzy Systems-Applications and Theory, vol. 5. World Scientific, Singapore (1995)
28. Gupta, S., Gupta, A.: A Fuzzy Multi Criteria Decision Making Approach for Vendor Evaluation in a Supply Chain. Interscience Management Review (IMR) 2 (2012)
29. Global Reporting Initiative -GRI-. GRI guidelines (2006), `https://www.globalreporting.org/resourcelibrary/G3-Sustainability-Reporting-Guidelines.pdf`
30. Global Reporting Initiative -GRI-. Sustainability Reporting Guidelines & Food Processing Sector Supplement (FPSS) (2011)
31. Wang, Y.M., Elhag, T.M.S.: Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment. Expert Syst. Appl. 31, 309–319 (2006)
32. Zeleney, M.: Multiple criteria decision making. McGraw Hill, New York (1982)
33. Escrig, E., Muñoz, M.J., Fernández, M.A.: Sustainable Development and the Financial System: Society's Perceptions About Socially Responsible Investing. Bus. Strat. Env. (2012), doi:10.1002/bse.1755
34. Cañal, V., Pérez, B.M., Rodríguez, M.V., Bilbao, A., Arenas, M.: Inversión Socialmente Responsable. Revista de Economía Financiera 19, 8–57 (2009)
35. Global Reporting Initiative -GRI-. The value of extra-financial disclosure. What investors and analysts said (2012), `http://www.accountingforsustainability.org/wp-content/uploads/2012/07/What-investors-and-analysts-said-The-value-of-extra-financial-disclosure.pdf`
36. Silberhorn, D., Richard, C.W.: Defining corporate social responsibility: A view from big companies in Germany and the UK. European Business Review 19, 352–372 (2007)
37. Muñoz, M.J., Fernández, A., Rivera, J.M., León, R., Escrig, E., Ferrero, I.: Materiality Analysis for CSR Reporting in Spanish SMEs. International Journal of Management, Knowledge and Learning 1, 3–22 (2012)

How to Apply Model Driven Engineering to Develop Corporate Social Responsibility Computer Systems

Cristina Campos and Reyes Grangel

Knowledge Engineering Group
Dept. de Llenguatges i Sistemes Informàtics, Universitat Jaume I, Castelló, Spain
{camposc,grangel}@uji.es

Abstract. The research presented in this paper aims to provide a framework to achieve the goal of generating computer systems for Corporate Social Responsibility (CSR) management using the Model Driven Engineering (MDE) software engineering paradigm. CSR has become an asset that companies need to efficiently manage, both at the organizational level and at the computer one. On the other hand, MDE consists in applying top-down transformations so that high-level models are transformed into software models and then from these ones, the code is automatically generated. Based on these principles this proposal considers how MDE can be applied to generate computer systems to manage CSR. The process consists in performing business information models on CSR and applying MDE methods to obtain high quality CSR computer systems in an efficient and economic way. The goal of this paper is to show the steps and methods to guide this process.

Keywords: Business Information Systems, Corporate Social Responsibility, Model Driven Engineering, MDA, Metamodel, UML Profile.

1 Introduction

In the current economic context not only is the final product or service that companies develop valued, but also the process followed to carry it out and the perception that the end user has on transparency and ethics of this product. This value is called Corporate Social Responsibility (CSR) [1, 2] which has become an asset that companies need to manage efficiently, not only at the organizational level but also at computer one. Therefore, there is a growing need to develop models and software to support automated CSR management.

Moreover, efficient software development has become a key factor in today's knowledge society. To develop quality software on time and using available resources to suit the rapidly changing demands and technological advances has become a necessity, both for companies who are end users and for developers. In this context, new paradigms of software development as Scrum Manager or Model Driven Engineering provide a more flexible response to this need.

The Model Driven Engineering (MDE) [3] promotes the development of software from models where the key element is the model and its successive transformations to

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 42–50, 2013.

get the code. There are many benefits of following this paradigm in the current business and technological environment, thus due to the rapid evolution of technology, it is more efficient to develop software from business models that are more stable than technological requirements.

In this context there are available commercial tools for transforming software models performed using computer languages as Java, or .NET. However, these transformations cannot be applied for models in a higher level of abstraction, as business models are [4-7].

There are numerous studies on the advantages of code generation from models following the MDE paradigm, however there is a clear lack of how to link the production of software when you need to start at business level of abstraction. Moreover, in the case of the CSR there are no modeling languages that support the specifications that must be modeled [8].

To provide a joint solution to these needs, a long term project has been defined. The final goal is to generate software for CSR management, from high-level business models, first applying transformations to obtain a software model, and finally getting the code automatically. To achieve this goal, the first step is to provide a framework including a modeling language and a modeling tool to represent business dimension of CSR. This paper shows the steps that should be followed in order to obtain a computer system to manage CSR in any enterprise, considering, as the starting point, a business model to represent CSR concepts and applying transformations among models using the Model Driven Engineering paradigm. The complete framework obtained in this project is called Model Driven CSR Framework.

The use of modeling languages to represent CSR will provide benefits both on business and academic worlds. For enterprises CSR is a new challenge that in some cases is not completely understood. Then, the capability to represent processes, requirements and different aspects concerning CSR allows managers to analyze, evaluate and improve their current CSR processes, the same way as other enterprise modeling languages provide mechanisms to analyze and improve business processes. On the other hand, the definition of a modeling language and the application of MDE to CSR will provide the knowledge about how to specify, manage and measure CSR and will open new research issues in this field.

This paper is organized as follows, after this short introduction. The state of the art on business modeling and CSR is presented in the section 2, in order to define that is the main problem that this research work proposes to solve. In section 3, the proposed solution is explained, that is to say, how to apply MDE to CSR context. Section 4 shows the performed process to generate computer systems to manage CSR starting from business models. Finally, section 5 includes the main conclusions and the further research work.

2 State of the Art

2.1 Enterprise Modeling

Enterprise modeling can be defined as the art of externalizing enterprise knowledge, which adds value to the company and needs to be shared [9]. This type of modeling has been used successfully since its emergence in the 80's in many areas and for different purposes, such as process reengineering and implementation of computer systems. This constant evolution has resulted in a context where there are many languages, methodologies and business modeling tools available and useful for their purpose [10].

Companies use different modeling languages [9, 10] to develop their business models for different business dimensions, such as organization, processes, decisions, etc. UEML [11, 12] or POP* [13, 14] were developed based on this type of languages in order to improve interoperability when it is need to exchange business models using different languages. Some implementations derived from metamodels proposed by UEML and POP* by UML profiles such as 'UML2 Profile For Enterprise Knowledge' [8, 15] has demonstrated the possibility of using UML2 [16] and in particular UML profiles as business modeling languages.

However, a major gap in this context is the weak connection between business models and generation of software, which causes companies to develop few models and use them poorly so models are not updated and become unsuitable for their purpose [10]. Initiatives such as the Model Driven Architecture (MDA) [17] promoted by the Object Management Group (OMG) try to solve these problems.

CSR is a new asset that companies need to manage and, therefore, it is interesting to represent this business dimension using modeling languages. The current models defined for CSR are theoretical and frameworks for the application of CSR but are not useful in the generation of software to manage this enterprise dimension [18]. Modeling languages used in this sense must provide artifacts, and a semantic and syntax to represent the components and the behavior of the CSR business dimension.

2.2 Corporate Social Responsibility

Social responsibility encompasses everything a company contributes voluntarily, worrying not only about economic, but also the social and environmental impact of its activities, and relationships with other companies. This attitude has a favorable impact on their competitive position and becomes an added value because it provides a picture of a company committed to managing responsibly its economic, social and environmental impact [1].

CSR is a broad concept and there have been many definitions. But it seems that there is consensus in considering that CSR endeavors to alienate the three P's (TBL Triple Bottom Line) of a company: Profit, People and Planet [2, 19]. From the point of view of employees it is considered that a company acts in a socially responsible way by taking appropriate decisions and actions carried out within the framework of satisfaction and commitment set by his colleagues.

CSR appears as a tool for enterprises to manage sustainability efficiently. So is tied to the concept of sustainable development. Sustainable development is defined in [20] as 'development that achieves the needs of current generations without compromising the ability of future generations to achieve their own needs'. This concept includes something called the TBL (Triple Bottom Line) which states that the company should have economic, social, and environmental concerns and objectives.

Many agencies have developed guidelines, recommendations and standards on CSR to improve their implementation. The most recognized is 'The Global Reportintg

Initiative (GRI)' which is the institution that provides the framework in which companies report their performance on CSR concepts [21]. GRI defines guidelines about what and how should be reported and how to organize and group reports on economic, social, and environmental performance.

A third generation (G3) guideline is provided in [21] but currently, GRI is working on the development of a fourth generation of Sustainability Reporting Guidelines. These new guidelines will improve previous ones and will consider new aspects concerning requirements for sustainability data, and enable reporters to provide relevant information to various stakeholder groups.

Considering the G3 Guidelines Reporting Principles and Guidance are:

- Principles to define report content: materiality, stakeholder inclusiveness, sustainability context, and completeness.
- Principles to define report quality: balance, comparability, accuracy, timeliness, reliability, and clarity.
- Guidance on how to set the report boundary.
- GRI provides indicators to measure what companies do in terms of CSR and guidance on how and what you have to report on CSR, but it does not provide a model of how to manage the processes and activities of the company to achieve satisfactory results.

As standards in CSR, there are no mandatory rules established, but there are agencies that define widely accepted regulations, such as:

- The network 'Global Compact (GC)' [22] is a USA initiative that includes ten recommendations, basic guidelines and maintains agreements between companies and employees in this field.
- The AA1000 series of standards [23] that aim to ensure that CSR is done correctly, rather than measuring its impact.
- The ISO 26000 [24], in draft form, which is considered to be the standard worldwide RSC, includes concerning stakeholders (standard AA1000SES), concerning investment (Principles for Responsible Investment Nations United) or concerning human rights and the environment (OECD Guidelines for Multinational Enterprises), among others.
- The SGE21 Forética [25] or the project AENOR PNE EJ 165 010 [26] which is under preparation.

Given these recommendations, the implementation of CSR in organizations involves managing many aspects that are not considered in basic business management and that are not supported by business models that enhance the development of computer systems. Therefore, to manage and to have a computer system to support all these aspects is critical to the evolution of a company that wants to achieve, not only economic, but also a prestige and competitive advantages.

In this context, the recommendation of the European Union, presented on the report FInES [27] establishes the vision of business in the future in order to meet the challenges that the current economic situation requires. This paper defines the features required for future business and the challenges that must be met to achieve them. Among these aspects it is considered that companies should be

Community-Oriented Enterprises, and therefore Corporate Social Responsibility must be one of the fundamental values to include in their strategy.

3 Applying MDE to CSR

The Model Driven Engineering approach focuses on models as the primary artifacts in the software development process, with the transformations on these models as the main operation [3]. Model Driven approaches aim to provide powerful solutions to improve software development processes. Such approaches may also be useful for model transformations in the context of enterprise modeling [28].

Among other MDE initiatives proposed by the OMG to promote the generation of code from models it is worth considering: Model Driven Architecture (MDA) [17]. The philosophy of this architecture is based on the idea that the system modeling changes the way in which the software is written. The main problem is that, in many cases, the model on the design of the data or software remains only in the mind of the programmer for a few minutes, and is then forgotten forever with the loss of knowledge. To solve this problem it is intended that MDA will become another step in the evolution of development in the field of software engineering, so that code generation from models is only one level of compilation. Thus software application developers suggest that this is the best method for writing reusable and interoperable software.

This initiative promotes the use of models as a mean for the design and implementation of systems in which the most important activities are modeled. One of its main objectives is to achieve separation between the business and technological domain, i.e. between the business information systems and their models, and the final technological platform in which the computer system to manage CSR is implemented. In this way, computer systems could support the quick technological advances that occur nowadays. To achieve this goal MDA is based on three types of models:

- The CIM (Computation Independent Model) shows the view independently of any computer system. It represents the domain and system requirements from an organizational, business or functional points of view. This model does not change whatever technology is used in the subsequent implementation of the system. It is focused on representing the business information system.
- The PIM (Platform Independent Model) models the functionality of the system, focusing on the aspects without introducing computerized specific features of the technologies that will be used. It focuses on information and computer vision system.
- The PSM (Platform Specific Model) represents the system implementation considering the technological platform and its characteristics. It can be obtained by applying transformations to the PIM model.

The proposal considers the possibility of transforming models from higher levels to lower levels using automated tools to generate code. Thus one can transform a PIM automatically into several PSM's, considering different implementation platforms. What provides this approach is that it achieves greater reusability, portability and interoperability of the software components of the system. The main advantage is that the same model of an organizational behavior and business information system (CIM) is useful for different technologies and therefore no changes are needed if there is no new higher level requirement.

In practice there exist different results that automate transformations of PIM (models including details of implementation although not of the specific computing platform), to PSM models, i.e. the implementation [29]. They have developed software generation tools that are marketed and used by the computer services companies developing quite successful results. But considering transformations from CIM to PIM models, there are only partial results and proposals [4, 30].

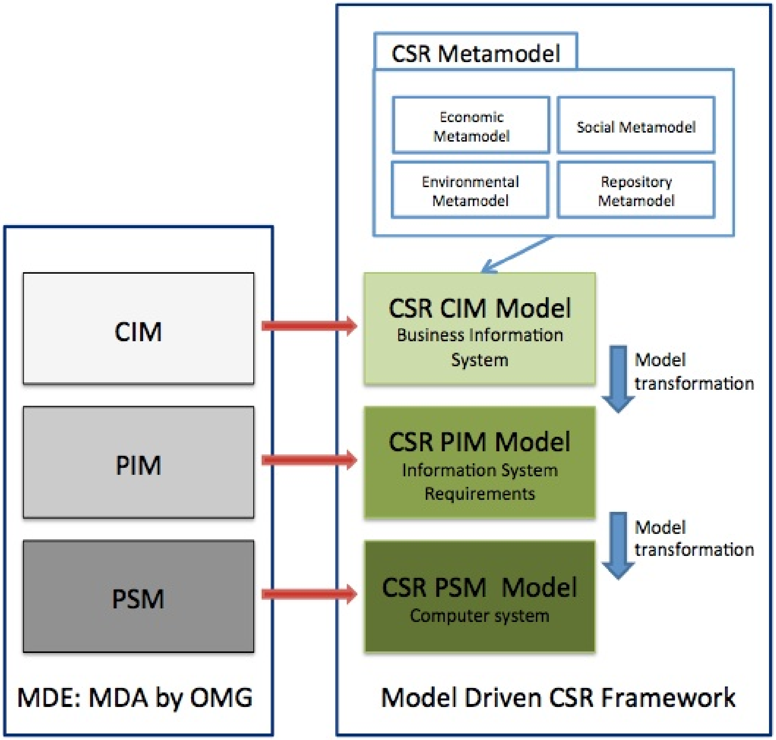


Fig. 1. Big picture for Model Driven CSR Framework

The research presented in this paper is focused on how to apply MDE paradigm, and more specifically MDA, to generate computer systems to manage CSR. Therefore, the idea is to represent the CSR dimension of any enterprise at high level obtaining a CSR CIM Model, which then needs to be transformed into a CSR PIM Model and finally into a CSR PSM Model (see Fig. 1).

4 Process to Generate Computer Systems to Manage CSR

In order to achieve the Model Driven CSR Framework it is necessary to follow a complex process, taking into account that there are no modeling languages that can be used to develop business models on CSR.

This process is based on MDE paradigm, and it will include the following subprocesses or steps:

1. To develop the metamodel to represent the main concepts related to CSR at the CIM level. This first step will require the analysis of the main concepts, concerns and lines involved in CSR, taking in account CSR organization guides and standards, expert opinions as well as research results in this area. CSR is a broad concept and concerns many aspects. In a first approximation of the work, it is proposed to select one of the axes of the Triple Bottom Line as a starting point. Therefore, the proposed metamodel is divided into four packages (see Fig. 1):
   - Economic metamodel: to take into account CSR at this level.
   - Social metamodel: to consider how people and the organization itself is accomplishing CSR requirements.
   - Environmental metamodel: to include environmental aspect of CSR in the model.
   - Repository metamodel: to represent all the documentation linked to CSR and to all measurement performed on it. Therefore, this package is related to the other three.
2. To design a UML profile and implement it on an specific computer tool in order to represent the business CSR model at the CIM level.
3. To establish which elements of the CSR CIM Model need to be transformed into the PIM level. Then, it is necessary to develop the corresponding mapping between CIM and PIM level and implement the transformation in any tool with the aim of transforming the CSR CIM Model into the CSR PIM Model, which represents CSR requirements for the computer system. Taking into account the experiences of this group using these kinds of tools, the selected tool is ATL [31, 32]. Finally, using the ATL tool the process should be repeated in order to obtain CSR PSM Model from CSR PIM Model.
4. To evaluate and improve, both, the metamodel and the profile. The development of these kinds of computer systems requires different iterations based on the evaluation of the first versions or prototypes and the improvements produced by the results evaluation. Therefore, a case study is defined to be used in the previous steps. The evaluation of the metamodel and its corresponding profile has two main benefits. On the one hand, the final result of the transformation, that is to say, the computer system to manage CSR, is improved. On the other hand, the modeling language provided by the implementation of the UML profile is also improved.

5 Conclusions

This is a proposal about what to do in order to apply MDE to develop computer systems to manage CSR. The proposal is the first step of an ambitious project, the results of which will provide advantages in both CSR research and software engineering, and the main result will be the Model Driven CSR Framework. This framework consist of a set of processes, methods and languages that will be useful for enterprises to manage their CSR using computer systems and also to model at high level the different dimensions of CSR: economic, social and environmental.

Moreover, the results are interesting for Computer Engineering as an example of learned lessons on MDE development.

One of the main components of this framework is the metamodel for representing high-level business model for CSR and its implementation as UML Profile, which can be use as a modeling language.

One of the main contributions is the proposed methodology to use all the components of the framework in order to obtain the computer system from CSR CIM Models.

Acknowledgements. This work has been partly funded by the Project 'Definition of a Metamodel and Application of MDA to the Generation of Software to Support Corporate Social Responsibility' (P1.1B2012-35).

References

1. León, R., Muñoz, M.J., Chalmeta, R.: Methodology for Sustainability Strategic Planning and Management. Industrial Management and Data Systems 110(2), 249–268 (2010)
2. Carroll, A.B.: A three-dimensional conceptual model of corporate social performance. Academy of Management Review 4, 497–505 (1979)
3. Bèzivin, J.: In Search of a Basic Principle for Model Driven Engineering. CEPIS, UPGRADE, The European Journal for the Informatics Professional V(2), 21–24 (2004), http://www.upgrade-cepis.org/issues/2004/2/up5-2Bezivin.pdf
4. Ben Salem, R., Grangel, R., Bourey, J.-P.: A comparison of model transformation tools: Application for Transforming GRAI Extended Actigrams into UML Activity Diagrams. Computers in Industry 59(7), 682–693 (2008)
5. Jouault, F., et al.: ATL: a QVT-like transformation language. In: Tarr, P.L., Cook, W.R. (eds.) OOPSLA Companion ACM, pp. 719–720 (2006)
6. Czarnecki, K., Helsen, S.: Classification of Model Transformation Approaches. In: OOPSLA 2003 Workshop on Generative Techniques in the Context of Model-Driven Architecture (2003)
7. Milewski, M., Roberts, G.: The model weaving description language (MWDL)-Towards a formal aspect oriented language for MDA model transformations. In: Proceedings of the 1st Workshop on Models and Aspects - Handling Crosscutting Concerns in MDSD, in conjunction with the 19th European Conference on Object-Oriented Programming (2005)
8. Grangel, R., Chalmeta, R., Campos, C., Palomero, S.: Improving Interoperability using UML Profile for Enterprise Modelling. In: Enterprise Interoperability IV. Making the Internet of the Future for the Future Enterprise (Part I) I-ESA 2010, pp. 35–45 (2010)
9. Vernadat, F.B.: Enterprise Modeling and Integration: Principles and Applications. Chapman and Hall (1996)
10. Grangel, R., Chalmeta, R., Campos, C., Coltell, O.: Enterprise Modelling, an overview focused on software generation. In: Panetto, H. (ed.) Interoperability of ESA Workshops of the INTEROP-ESA International Conference EI2N, WSI, ISIDI and IEHENA 2005, pp. 65–76. Hermes Science Publishing (2005)
11. Berio, G., Opdahl, A., Anaya, V., Dassisti, M.: Deliverable DEM 1: UEML 2.1. Technical report, INTEROP-DEM (2005)
12. INTEROP: Interoperability Research for Networked Enterprises Applications and Software NoE (IST-2003-508011) (2008), http://www.interop-noe.org

13. Grangel, R., Chalmeta, R., Schuster, S., Peña, I.: Exchange of Business Process Models Using the POP* Meta-model. In: Bussler, C.J., Haller, A. (eds.) BPM 2005. LNCS, vol. 3812, pp. 233–244. Springer, Heidelberg (2006)
14. ATHENA: Advanced Technologies for interoperability of Heterogeneous Enterprise Networks and their Applications IP (IST-2001-507849) (2008), http://www.athena-ip.org
15. Grangel, R., Chalmeta, R., Campos, C.: A Modelling Framework for Sharing Knowledge. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS (LNAI), vol. 4693, pp. 1230–1237. Springer, Heidelberg (2007)
16. OMG. OMG Unified Modeling Language (OMG UML), Superstructure. Object Management Group, version 2.4.1 formal/2011-08-06 edition (August 2011)
17. OMG. MDA Guide Version 1.0.1. Object Management Group, document number: omg/2003-06-01 edition (June 2003)
18. Redman, E.: Three Models of Corporate Social Responsibility: Implications in Public Policy, http://files.meetup.com/1325336/Three%20Models%20of%20CSR.pdf
19. Carroll, A.B.: Corporate Social Responsibility: Evolution of a Definitional Construct. Business Society 38, 268 (1999)
20. WCDE, World Commission on Environment and Development: Our Common Future. Oxford University Press (1987)
21. Browna, H.S., de Jongb, M., Levycuilding, D.L.: Building Institutions based on Information Disclosure: Lessons from GRI's Sustainability Reporting. Journal of Cleaner Production 17(6), 571–580 (2009)
22. United Nations Global Compact, http://www.unglobalcompact.org/
23. AccountAbility, http://www.accountability.org/index.html
24. ISO 26000 – Social responsibility, http://www.iso.org/iso/iso_catalogue/management_and_leadership_standards/social_responsibility/sr_iso26000_overview.htm
25. Forética, http://www.foretica.org/
26. PNE 165010 EX de AENOR, http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0042958&PDF=Si
27. FInES Cluster. Future Internet Enterprise Systems: Research Roadmap, Final Report, Version 4.0 (June 2010)
28. Jardim-Gonçalves, R., Grilo, A., Steiger-Garcao, A.: Challenging the interoperability between computers in industry with MDA and SOA. Comput. Ind. 57(8), 679–689 (2006)
29. Benguria, G., Larrucea, X., Elvesater, B., Neple, T., Beardsmore, A., Friess, M.: A Platform Independent Model for Service Oriented Architectures. In: Entrerprise Interoperability. Springer (2007)
30. Grangel, R., Bigand, M., Bourey, J.-P.: Transformation of decisional models into UML: application to GRAI grids. International Journal of Computer Integrated Manufacturing 23(7), 655–672 (2010)
31. AtlanMod INRIA, http://www.inria.fr/equipes/atlanmod (accessed January 2013)
32. Jouault, F., et al.: ATL: a QVT-like transformation language. In: Tarr, P.L., Cook, W.R. (eds.) OOPSLA Companion ACM, pp. 719–720 (2006)

A Model for Estimating Cash Flows in Firms Backed by Venture Capital

Iván Arribas, Elisabeth Bustos, and Gregorio Labatut

Universidad de Valencia, Av. de los Naranjos s/n,
46022 Valencia, Spain
{Ivan.Arribas,Elisabeth.Bustos,Gregorio.Labatut}@uv.es

Abstract. Venture Capital only backs firms for a short period of time. When the time to exit arrives, the firm must inevitably be valued in order to obtain a basis for negotiating the exit price. Discounted cash flow is precisely one of the valuation methods that are used most by Small and Medium-sized Enterprises (SMEs).

This paper presents a model for the estimation of free cash flow in Venture Capital-backed SMEs. Results suggest that profitability and EBITDA determine trends in free cash flow. Our model contributes to the literature by providing a useful tool for both Venture Capital entry and exit decision making. What is more, the model is useful for all stockholders when it comes to deciding how to maximize the worth of their firm.

Keywords: Venture Capital, SMEs, free cash flow, valuation.

1 Introduction

Interest in Venture Capital (VC) has mounted since Gorman and Sahlman (1989) as a result of this sector consolidating in both developed and emerging nations.

There are currently several areas of research devoted to studying the various aspects of VC, either from the perspective of VC-backed firms or public or private venture investors, or in terms of the effect that VC has on the macroeconomic variables of the country it operates in.

More specifically, several authors have provided empirical evidence that VC-backed firms grow more than their non VC-backed counterparts (Hellmann and Puri, 2002; Baum and Silverman, 2004; Balboa et al. 2011).

On a different note, it is also worth highlighting the area of research devoted to obtaining statistical evidence on VC exit strategies. Authors that have contributed to this specific area include Cumming and MacIntosh, (2003); and, more recently, Ozmel et al. (2012).

However, the work on VC exit primarily focuses on exit strategies and the profitability of the mechanisms provided by the market, but not the intrinsic conditions of VC-backed firms.

In this sense, we find the literature clearly lacks contributions regarding when VC should exit firms in order to obtain the desired profit and, more specifically, from the perspective of the expected conditions of the VC-backed firm.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 51–63, 2013.

Bearing this in mind, we believe it would be interesting to provide a model capable of explaining the best time to sell. That moment will coincide with the desired profit, which will be achieved when the shares of a firm attain a certain value.

Consequently, the firm must be assessed at the time of exit by applying a generally accepted method and one of the most widely used by venture capital is discounted cash flows (Pintado et al. 2007). For this reason, we think it is worth proposing a regression model capable of forecasting the future cash flows of VC-backed firms.

The contribution this paper makes is important and is based on a sound theoretical and methodological framework. We selected the economic and financial variables our empirical analysis is based on after carefully reviewing previous research and (from a methodological viewpoint) sample selection is justified in order to avoid bias that could lead to estimation errors and also to validate the results of the model. Furthermore, the regression model has been chosen to relate cash flows to the selected variables.

The paper contributes to the existing literature in several ways. In the first place, it provides a regression model that explains cash flow behavior in response to changes in certain corporate variables. More specifically, our results suggest that the important variables in terms of changes in cash flows are (in order of importance) economic profitability and EBITDA. This result is in keeping with those obtained by the majority of studies on business growth in firms backed by VC.

In the second place, we also provide empirical evidence on the predictive power of the model and therefore contribute to the design of a useful tool in VC exit decision making. This tool allows investors to plan their exit strategies to better fit the market conditions at the time of their exit.

Furthermore, the analysis, applied in European context, focuses on Small and Medium-sized Enterprises (SME), the problems of which are a major concern of all national and international organizations due to the undeniably important role they play in the business fabric of any economically developed country (Watson and Wilson, 2002; Madrid and García Pérez de Lema, 2008; among others).

The study has been organized as follows: first, we define the theoretical framework, review the previous literature and pose the hypotheses to be tested; secondly, we set out the methodology and sample characteristics and explain the variables used; thirdly, we analyze the results and finally, we present the main conclusions reached.

2 Literature Review and Research Hypotheses

Authors such as Amit et al. (1990) have studied VC through the Agency Theory. Using the same approach, Hsu (2004) indicates that there are two main areas of research on VC. The first focuses on the relationships between venture capitalists and the firms they back (Sahlman, 1990).

The second area of research seeks to study the impact of venture capital in both macroeconomic (Belke et al. 2001) and microeconomic terms (Engel, 2002; Alemany and Martí, 2005).

This work corresponds to the latter strand of research. So, we will review the existing literature on the impact that VC has on the growth of VC-backed firms. We will then conclude the review with a summary of the previous studies devoted to the effect of VC on cash flow, making special reference to the valuation methods used by VC. Finally, we will define our hypotheses. By doing so, we will establish the necessary theoretical grounds to base our empirical study on.

2.1 Impact of Venture Capital on Business Growth

Since Gompers and Lerner (2001) indicated that the impact of venture capital was an issue pending research, numerous papers have attempted to fill this gap in the scientific literature, as well as to provide evidence of the positive effect that VC has on business growth.

The literature states that VC stimulates more rapid growth firstly because venture capitalists select firms meticulously and secondly, due to the capital and value added that VC provides the firms it backs, reasons that are acknowledged by authors such as Hellmann and Puri (2002), Davila et al. (2003) and Chemmanur et al. (2007).

More specifically, value added materializes in the shape of prestige in the eyes of stakeholders, which therefore allows firms to enhance their trade relations (Hsu, 2004) while at the same time increasing their capacity for borrowing and, consequently, investment and growth (Ou and Haynes, 2006).

In addition, venture capital investors also contribute value to firms by professionalizing management (Hellmann, 2000), by providing strategic and financial advice (Gorman and Sahlman, 1989) and finally, by fostering the synergies between the firms in their portfolios (Baum and Silverman, 2004).

Whatever the cause, there is currently sufficient literature proving the stimulating effect of VC on growth. In this sense, Balboa et al. (2006) obtain evidence on how VC boosts the growth of firms in the process of expanding, compared to a group of comparable firms not backed by VC. What is more, Balboa et al. (2011) conclude that VC-backed firms gain more growth momentum regardless of their stage of development.

In summary, the literature discussed above confirms the positive effect that VC has on the growth of the firms it backs.

2.2 Cash Flows in the Context of Venture Capital

The literature on VC includes various research papers that use cash flows in their empirical studies. On the one hand, cash flows are widely used in studies aimed at ascertaining the impact of VC on growth (Alemany and Martí, 2005; Engel and Keilbach, 2007; among others).

Furthermore, it is worth highlighting that for Davila et al. (2003), the growth of a firm is positively related to its worth. And for Copeland et al. (2000), cash flow is king in terms of the worth of a firm, although it is also true that this statement has been criticized by other authors (Liu et al. 2006).

Finally, and linking up with the objective of our study, various papers have analyzed the valuation methods used by VC investors. Nevertheless, Scherlis and Sahlman (1989) indicate that no perfect valuation method exists for VC-backed firms, but we can say that in practice, private equity managers show their own preferences by choosing a particular valuation criterion (Rojo et al. 2010).

According to Payne et al. (2009), a firm's stage of development determines the choice of valuation method. In the same vein, Reverte and Sánchez (2012) conclude that the discounted cash flows method is used the most in the case of firms in an early development stage. Barrow et al. (2011) reach the same conclusion regarding the method of choice, although they argue that for mature firms, private equity managers opt for the earnings multiple method.

However, Pintado et al. (2007) found evidence on the basis of a sample of 51 VC investors surveyed that the discounted cash flows method is frequently used more by VC investors for firms in both an early and mature stage of development.

In short, the discounted cash flows method and the earnings multiple method are contesting the title of most frequently used valuation method, but, as indicated by Barrow et al. (2001), experts normally do not make just one valuation, but rather apply a second alternative method in order to confirm the results obtained.

This perspective concludes the theoretical framework of this study and, in keeping with the objective established, we now proceed to define the hypotheses to be tested:

Hypothesis 1: the variations in total assets, EBITDA, sales, profitability and the number of employees are determinants of the increase in worth of VC-backed firms.

Hypothesis 2: the worth of VC-backed firms does not depend on the market price trends of other firms.

3 Data and Methodology

3.1 Sample

We based our study on a sample of 89 Spanish low and medium technology SMEs in the process of expanding that had received VC during the period dating from 2001 to 2007. Table 1 provides a breakdown of the information on the number of firms that have received VC in each of the years of the sample period.

Table 1. Sample characteristics

Year	Number of companies	Percentage
2001	10	11.24
2002	10	11.24
2003	15	16.85
2004	11	12.36
2005	10	11.24
2006	18	20.22
2007	15	16.85
Total	89	100.00

Notice that the data are not Panel data given that firms enter in the analysis in different years (see Table 1) and we have collected financial data for 5 years from their inclusion year. Thus, data are so unbalanced that Panel Data analysis reveals impossible to be applied and standard Least Square regression has been used to model data (see Section 4 below for further explanations).

This sample was obtained following a filtering process, which made it necessary to remove all the firms whose venture capital was not related to a capital expansion operation. Likewise, we also rejected any venture firms due to the fact they lack the historical accounting information necessary for this research. The objective of this study also forced us to reject the firms that recorded losses year after year, as such firms have obviously not demonstrated any growth.

Furthermore, due to the study timeframe falling between 2000 and 2010, every year we can analyze the developments in the variables under study one year before the entry of VC and up to three years after this event. As a result, we will have a study window of four years for the firms backed each year.

The identity of firms and the type of operation was provided by the Spanish Private Equity and Venture Capital Association (ASCRI), which contains the complete population of VC deals in Spain; while the financial reporting data used was obtained from the SABI database (Bureau van Dijk and Informa), which records the official annual accounts of more than 650,000 Spanish companies.

3.2 Explanatory Variables

The basis for choosing the variables is the objective of this research, for which reason the dependent variable is cash flow. More specifically, we have used free cash flow, due to being preferred by 84.3% of investors surveyed when performing VC valuations (Reverte and Sánchez, 2012).

As regards the independent variables, we have tried to combine the experience of other authors with the indicators that in our opinion best gauge trends in free cash flows.

In this sense, we have employed five economic and financial corporate variables and one other variable external to the firm, which represents the market. As regards the corporate variables, we opted for total assets, EBITDA, sales and the number of employees due to being accurate measures of the impact of VC on firm growth, as stated in the empirical research by Balboa et al. (2006) and Martí and Ferrer (2012), among others.

We have also included profitability due to considering it a relevant variable in firm growth. In this sense, Molina and Correa (2002), states that the firms that grow are typically more profitable.

In order to complete the independent variables, we have included the Madrid Stock Exchange Total Index (ITBM after its initials in Spanish) as a variable external to the firm, to correct the effect that economic developments have on the value of the SMEs in the model. This index is a representative indicator of the price of all the publicly listed firms in Spain. Apart from stock prices, the ITBM also includes the distribution of reinvested dividends.

Out of the stock market indices available, we opted for the ITBM because it is better suited to the profit reinvestment and promotion of self-financing policies generally adopted by VC-backed SMEs. It is worth mentioning that the ITBM has been frequently used by Fernández et al. (2011) in their research on profitability and generating worth for stockholders.

Table 2 below provides the definitions of all the variables used in the empirical analysis.

Table 2. Model variables

Variables		
CORPORATE	FCF	Free cash flow (EBITDA less the increase in variable capital necessary, less investment in fixed assets during that same period and less the tax that the firm would hypothetically pay on profit before interest)
	ASSETS	Total assets
	PROF1	Sales
	PROF2	Profitability (EBITDA/ total asset)
	EBITDA	Earnings Before Interest, Tax, Depreciation, and Amortization
	EMP	Number of employees
MARKET	ITBM	Madrid Stock Exchange Total Index (Includes prices plus dividends paid of all the publicly listed firms in Spain)

What we really include in the model is the variation rate (VR) of the variables described, considered between the year prior to the entry of VC and the third year after that event.

We will propose a model that includes time-specific effects as an alternative to the impact of the market on SMEs in order to estimate the unobservable effect of the economic environment on firms for each year and separately.

3.3 Descriptive Statistics

Table 3 provides the variation rate for each of the original variables, taking the year before the entry of VC and three years after that event as periods of reference. As can be observed, all the variables display an increase on average. It is important to note here that the increase is not due to having eliminated SMEs that recorded losses from the sample, but rather to natural growth. The reason is that we did not demand a minimum level of profits when selecting the sample, only that the firms had not recorded losses.

Likewise, it is worth highlighting that the variation rates observed are similar to those obtained by Balboa et al. (2011) in a recent study on the impact of VC on similar explanatory variables.

Table 3. Descriptive analysis of variable VR

	VRFCF	VRASSETS	VREBITDA	VRPROFI1	VRPROFI2	VREMP
N	87	88	87	86	87	81
Mean	4.0800	2.6065	4.8879	52.1052	1.1876	1.5599
Std. deviation	8.7990	4.0797	10.7052	439.4200	4.9625	3.8827
Median	1.5581	1.3513	1.4594	0.8166	0.2149	0.4286
Rank	53.4900	29.2800	71.0300	4,073.2700	40.1100	30.5000

As expected, the variables display correlation significantly different to zero in most cases (see Table 4), although generally speaking it is not very high, so the presence of multicollinearity is therefore discarded. It is worth highlighting the high degree of correlation between the variation rate of free cash flow and EBITDA. On the other hand, the variation rate in profitability is the least related to the rest of variables.

Table 4. Correlation between VR

	VRFCF	VRASSETS	VREBITDA	VRPROFIT1	VRPROFIT2	VREMP
VRFCF	1.000	0.289** (87)	0.743** (87)	0.350** (86)	0.509** (87)	0.254* (81)
VRASSETS		1.000	0.365** (87)	0.285** (86)	0.018 (87)	0.410** (81)
VREBITDA			1.000	0.633** (86)	0.386** (87)	0.340** (81)
VRPROFIT1				1.000	0.002 (86)	0.317** (80)
VRPROFIT2					1.000	0.022 (81)
VREMP						1.000

Correlation and pairwise simple size in parentheses. ** and * indicate significance levels of <1% and < 5%, respectively.

4 Empirical Model and Analysis of Results

We have included the changes in the variables that took place between the year prior to the entry of VC and the third year following that event, so first differences will be used. Murray and Goyal (2003) indicate that this approach can result in a loss of accuracy and also bias coefficients towards zero, but they do not expect this to distort the conclusions.

The following model will test the hypotheses of this research:

VRFCF = f(VRASSETS, VREBITDA VRPROF1, VRFPROF2, VREMP, VRITBM) (1)

As the original data refer to time series, we believed it was also worth including time-specific effects. In this way, we would be able to test whether the timeframe of growth for each of firms, which varies throughout the sample period [2000, 2010], has influenced the trend of free cash flow alongside the rest of co-variables. Therefore, the following model will also be estimated:

VRFCF = f(VRASSETS, VREBITDA VRPROF1, VRFPROF2, VREMP, Time dummies) (2)

whereby the time dummies capture the time-specific effects.

Table 5. Regression results using two different model specifications - Dependent variable VRFCF

Independent variables	Model 1	Model 2
VRASSETS	0.080	-0.091
	(0.871)	(0.232)
VREBITDA	0.406***	0.407***
	(0.106)	(0.106)
VRPROFIT1	0.000	0.000
	(0.002)	(0.002)
VRPROFIT2	0.549***	0.560***
	(0.148)	(0.149)
VREMP	0.062	0.158
	(0.183)	(0.184)
VRITBM	0.105	
	(0.210)	
Time-effect 2001		4.116**
		(1.782)
Time-effect 2002		0.096
		(2.014)
Time-effect 2003		2.895*
		(1.607)
Time-effect 2004		1.224
		(1.780)
Time-effect 2005		3.419
		(2.071)
Time-effect 2006		-0.723
		(1.460)
Time-effect 2007		-0.350
		(1.453)
Observations	81	81
R^2	0.576	0.698
Adjusted R^2	0.549	0.645
F-statistics	16,467	13.123
Probability F	0.000	0.000

This table reports the OLS regressions. The dependent variable is the variation rate of free cash flow. The independent variables are variation rates of total assets, Sales, Profitability, EBITDA, number of employees and the Madrid Stock Exchange Total Index for Model 1; for Model 2, the last independent variable has been removed and seven time effects are included. We report the coefficients and their standard errors in parentheses. ***, ** and * indicate significance levels of <1%, <5% and <10%, respectively.

The two models specified have been estimated using Ordinary Least Squares (OLS) and the results are presented in Table 5.

For the purpose of testing the robustness of the results, we have estimated the same regressions by calculating the variation rates on the basis of the year prior to the entry of VC and the second year after this event. The same results are obtained, which reinforces the conclusions of the proposed models.

The same firm variables are significant in both models: the change in EBITDA and the change in profitability. In addition, the variation rate of the ITBM is not significant in the case of model 1, although time effects are observed in the case of firms that receive VC in 2002 or 2004.

The only coefficients that were significant in both models are associated to the variation rates of EBITDA and profitability. Furthermore, the estimated values of the coefficients are very similar in both models, which is interpreted as a good sign of robustness.

This result partially confirms our first hypothesis, as only two of the five effects considered are found to be significant. In this sense, a 1% increase in EBITDA implies a 0.4% increase in free cash flow, *ceteris paribus*; and a 1% rise in profitability implies a 0.5% increase in free cash flow, *ceteris paribus*.

The variation rate of the ITBM is not significant, which means we must reject the second hypothesis. Apparently, market price trends do not explain the free cash flow behavior of VC-backed firms.

However, there are clearly aspects external to the firm that influence changes in free cash flow, as evidenced by the fact that two of the time-specific effects included in model 2 were significant. That is, the variation rate of free cash flow partly depends on the period considered for calculation purposes.

More specifically, the firms that gained access to VC in 2002 and 2004 (time effects 2001 and 2003, respectively) display significantly larger variation rate in free cash flow than the rest of firms, ranging between three and four percentage points. While not significant, it is still striking that the firms that requested VC in the last years of the sample period (2007 and 2008) display a negative time effect.

Some statistics regarding the quality of the models estimated are shown at the bottom of Table 5. Determination coefficients of 0.58 and 0.70 are obtained for models 1 and 2, respectively, which confers them strong predictive power.

5 Discussion and Conclusions

Before presenting the conclusions of the study, it is worth recalling that this research employs a sample made up of seven different groups of firms, each of which obtained VC in different years. This information adds value to the study, as it incorporates the time-effect, which reflects the economic trend and scenario that affects firms and VC decision making in each specific year, information that will undoubtedly strengthen the conclusions.

This research is based on two premises for which evidence was provided in previous studies. On the one hand, VC-backed firms grow faster than similar firms financed by other means (Davila et al. 2003; Balboa et al. 2011). More specifically, one of the variables that displays this growth is cash flow. On the other hand, the discounted cash flow method is one of the most popular among VC investors when it comes to valuing firms, particularly those in the early stages of development (Barrow et al. 2001; Pintado et al. 2007). The firm must inevitably be valued in order to obtain a basis for negotiating the exit price.

Furthermore, cash flow (and more specifically free cash flow) is a relevant variable in both VC entry and exit decision making. By estimating the behavior of free cash flow, VC investors can make decisions regarding the schedule and the amount to finance by stage, particularly bearing in mind that financing a firm in different stages is the best way for VC investors to control compliance with the business plan (Sahlman, 1990).

As far as exiting is concerned, by estimating future free cash flow, VC investors can predict the best time to exit, regardless of whether the firm is progressing favorably or unfavorably.

For all of the above reasons, we believe it is very interesting to estimate a model that reveals the future behavior of free cash flow in VC-backed firms. In this sense, our model explains that changes in free cash flow depend on the profitability of assets and the EBITDA, as indicated by the significant coefficients obtained by the two models defined.

This result is consistent with a survey on 51 VC-backed firms conducted by Reverte and Sánchez (2012), who conclude that the expected return on investment is the most important factor for VC investors. Hence, it is worth highlighting that our model incorporates the variable that is of most interest to these investors.

The EBITDA is also extremely important for VC investors, because this variable is used to calculate the profitability of the assets.

It is worth highlighting that the model has not considered the rest of corporate variables. That is, total assets, sales and the number of employees are not significant determinants of the change in free cash flow, despite all of them displaying positive variation rates. This result adds robustness to our study due to once again being consistent with the results obtained by Reverte and Sánchez (2012).

Another conclusion that can be drawn is that the trend in the market prices of publicly-listed firms in Spain, represented by the evolution of the ITBM, does not influence the behavior of the free cash flows of VC-backed SMEs. This conclusion is also in keeping with the evidence obtained by Manigart and van Hyfte (1999) in that these firms record stronger growth in free cash flow and, consequently, their value behaves differently.

But surprisingly, an external effect is observed on free cash flow in the case of SMEs that obtained VC in 2002 and 2004. We have attempted to find an economic reason to justify the higher variation rate in those years, but neither GDP nor the CPI registered upturns in those years. In order to be sure, we once again resorted to the ITBM and discovered a slight upturn in 2001 and a visible increase in 2003 in regard to their respective previous years. Therefore, it could be concluded that the ITBM does influence the behavior of free cash flow in those firms in particular.

This might be the explanation, or perhaps the reason is that the SMEs that make up those groups have been more closely monitored by VC investors and, as a result, the value added provided by VC is more evident in those years. This possibility would be in line with indications by Balboa et al. (2011), in the sense that the degree of value added depends on how busy VC investors are with other tasks such as searching for funds, or negotiating the entry into or exit from other firms.

In summary, the model in this research proposes the variations in EBITDA and profitability, with proportions of 0.4% and 0.5% respectively, as the variables that determine the behavior of free cash flow. It is also worth underlining that this model makes it possible to forecast the future behavior of free cash flow on the basis of the variables mentioned. This aspect is important, because it also makes it possible to estimate the free cash flows that would be obtained in different expected EBITDA and profitability scenarios.

Therefore, this research contributes a tool designed to complement and strengthen the conclusions of other procedures used to estimate free cash flows. More specifically, it is aimed at low and medium technology SMEs that are in the process of expanding and are backed by VC.

In addition, the model presented here is useful for both VC and institutional investors. In the case of the former, it allows them to forecast exit values and plan their exit, while in the case of the latter, it helps them to plan their corporate strategy. Moreover, both types of investors can use it in decision making aimed at maximizing the value of the firm.

Finally, we believe it is interesting to point out that while the model is simple, this should not be taken as a weakness, but as a strong point, because this makes it operational for use directly in SMEs.

References

1. Alemany, L., Martí, J.: Unbiased Estimation of Economic Impact of Venture Capital Backed Firms. EFA 2005 Meeting, Moscow (2005)
2. Amit, R., Glosten, L., Muller, E.: Entrepreneurial Ability, Venture Investments, and Risk Sharing. Management Science 36(10), 1232–1245 (1990)
3. Balboa, M., Martí Pellón, J., Zieling, N.: Does venture capital really improve portfolio companies' growth? Evidence from Growth Companies in Continental Europe. EFMA Meeting, Madrid (2006)
4. Balboa, M., Martí, J., Zieling, N.: Impact of funding and value added on Spanish venture capital-backed firms. Innovation: The European Journal of Social Science Research 24(4), 449–466 (2011)
5. Barrow, C., Richardson, A., Copin, G., Paliard, R., Lange, J., Leleux, B., St-Cyr Hec, L.: Valuing High Growth Potential Companies: An International Comparison of Practices by Leading Venture Capitalist and Underwriters. Management International 6(1), 55–73 (2001)

6. Baum, J., Silverman, B.: Picking Winners or Building Them? Alliance, Intellectual, and Human Capital as Selection Criteria in Venture Financing and Performance of Biotechnology Start-ups. Journal of Business Venturing 19(3), 411–436 (2004)
7. Belke, A., Fehn, R., Foster, N.: Venture Capital Investment and Labor Market Performance: A Panel Data Analysis. University of Vienna, Working Paper No. 0112 (2001)
8. Bertoni, F., Ferrer, M.A., Martí, J.: The Different Role Played by Venture Capital and Private Equity Investors on the Investment Activity of Their Portfolio Firms. Small Business Economics (2011), `http://dx.doi.org/10.1007/s11187-011-9384-x` (forthcoming)
9. Chemmanur, T., Krishnan, K., Nandy, D.: How does Venture Capital Financing Improve Efficiency in Private Firms? A Look Beneath the Surface. Boston College, Working Paper (2007)
10. Copeland, T., Koller, T., Murrin, J.: Valuation: Measuring and Managing the Value of Companies, 3rd edn. Wiley, New York (2000)
11. Cumming, D., MacIntosh, J.: A Cross-Country Comparison of Full and Partial Venture Capital Exits. Journal of Banking and Finance 27, 511–548 (2003)
12. Davila, A., Foster, G., Gupta, M.: Venture Capital Financing and the Growth of Start-up Firms. Journal of Business Venturing 18(6), 689–708 (2003)
13. Engel, D.: The Impact of Venture Capital on Firm Growth: An Empirical Investigation. ZEW, Discussion Paper No. 02 (2002)
14. Engel, D., Keilbach, M.: Firm Level Implications of Early Stage Venture Capital Investment: An Empirical Investigation. Journal of Empirical Finance 14(2), 150–167 (2007)
15. Fernández, P., Aguirreamalloa, J., Corres, L.: IBEX 35: 1991-2010. Rentabilidad y creación de valor. IESE Business School, Universidad de Navarra, Research Paper No. DI-890 (2011)
16. Gompers, P., Lerner, J.: The Venture Capital Revolution. Journal of Economic Perspectives 15(2), 145–168 (2001)
17. Gorman, M., Sahlman, W.: What do Venture Capitalists Do? Journal of Business Venturing 4(4), 231–248 (1989)
18. Hellmann, T.: Venture Capitalists: The Coaches of Silicon Valley. In: Lee, C., Miller, W., Gong Hancock, M., Rowen, H. (eds.) The Silicon Valley Edge: A Habitat for Innovation and Entrepreneurship, pp. 276–294. Stanford University Press, Stanford (2000)
19. Hellmann, T., Puri, M.: On the Fundamental Role of Venture Capital. Federal Reserve Bank of Atlanta Economic Review, Fourth Quarter (2002)
20. Hsu, D.: What do Entrepreneurs Pay for Venture Capital Affiliation? Journal of Finance 59(4), 1805–1844 (2004)
21. Liu, J., Thomas, J.K., Nissim, D.: Is Cash Flow King in Valuations? Financial Analysts Journal 63(2), 56–68 (2007)
22. Madrid, A., García Pérez De Lema, D.: Las ayudas financieras a la innovación a la pyme: Sesgo de motivación y de selección administrativa. Revista Internacional de la Pequeña y Mediana Empresa 1(1), 1–19 (2008)
23. Manigart, S., van Hyfte, W.: Post-investment Evolution of Belgian Venture Capital Backed Companies: An Empirical Study. In: Babson Entrepreneurship Conference, Working Paper (1999)
24. Martí, J., Ferrer, M.A.: Investment-Cash Flow Sensitivity in Small and Medium-Sized Enterprises at the Expansion Stage. Revista Internacional de la Pequeña y Mediana Empresa 4(1), 68–89 (2012)

25. Molina, M., Correa, A., González, A.L.: Crecimiento y pyme: un estudio empírico. Revista Europea de Dirección y Economía de la Empresa 11(3), 77–102 (2002)
26. Murray, F., Goyal, V.: Testing the Pecking Order Theory of Capital Structure. Journal of Financial Economics 67(2), 217–248 (2003)
27. Ou, C., Haynes, G.W.: Acquisition of Additional Equity Capital by Small Firms- Findings from the National Survey of Small Business Finances. Small Business Economics 27, 157–168 (2006)
28. Ozmel, U., Robinson, T., Stuart, T.: Strategic Alliances, Venture Capital, and Exit Decisions in Early Stage High-tech Firms. Journal of Financial Economics (2012) (forthcoming)
29. Payne, G.T., Davis, J.L., Moore, C.B., Bell, R.G.: The Deal Structuring Stage of the Venture Capitalist Decision-Making Process: Exploring Confidence and Control. Journal of Small Business Management 47(2), 154–179 (2009)
30. Pintado, T., García Pérez De Lema, D., van Auken, H.: Venture Capital in Spain by Stage of Development. Journal of Small Business Management 45(1), 68–88 (2007)
31. Reverte, C., Sánchez, M.: Empirical Evidence on Investment Valuation Models Used by Spanish Venture Capital Companies: Special Reference to the Discounted Cash Flow Method's Implementation. International Journal of Entrepreneurship and Small Business 17(4), 495–510 (2012)
32. Rojo, A., Gálvez, M., Alonso, J.: La valoración por las Sociedades de Capital Riesgo en la práctica. Revista Española de Capital Riesgo 3/20110, 3–15 (2010)
33. Sahlman, W.: The Structure and Governance of Venture-capital Organizations. Journal of Financial Economics 27, 473–521 (1990)
34. Scherlis, D.R., Sahlman, W.A.: A Method for Valuing High-Risk, Long-Term Investments. Note Harvard Business School (1989)
35. Watson, R., Wilson, N.: Small and Medium Size Enterprise Financing: A Note on Some of the Empirical Implications of a Pecking Order. Journal of Business Finance and Accounting 29(3), 557–578 (2002)

Non-linear Tradeoff between Risk and Return: A Regime-switching Multi-factor Framework

Vicent Aragó and Enrique Salvador

Finance and Accounting Department, Universitat Jaume I,
Avda. Sos Baynat s/n, E-12071, Castellon de l Planan, Spain
{arago,esalvado}@uji.es

Abstract. This study develops a multi-factor framework where not only the market risk is considered but also potential changes in the investment opportunity set. Although previous studies find no clear evidence about a positive and significant relation between return and risk, favourable evidence can be obtained if a non-linear relation is established. The positive and significant tradeoff between return and risk is essentially observed during low volatility periods suggesting a procyclical risk aversion of investors. Different patterns for the risk premium dynamics in low and high volatility periods are obtained, both in risk prices and risk (conditional second moments) patterns.

Keywords: Risk premium, ICAPM, pro-cyclical risk aversion, non-linear multivariate GARCH, intertemporal risk.

1 Introduction

The relation between expected return and risk has motivated many studies in the financial literature. Most of the recent asset pricing models are based in this fundamental trade-off, so understanding the dynamics of this relation is a key issue in finance. One of the first studies establishing a theoretical relation between expected return and risk is the [1] CAPM model. These authors proposed a positive linear relationship between the expected return of any asset and its covariance with the market portfolio; in other words, the expected return of the market portfolio is proportional to its conditional variance. This static model has been analyzed empirically in several studies obtaining no clear evidence about the sign and significance of this relationship [2]. [3] proposed an extension of this model adding a second risk factor in the relationship that may improve the static CAPM model. The market risk premium in the Merton's model is proportional to its conditional variance and the conditional covariance with the investment opportunity set (hedging component). This framework established in a time-continuous economy is an extension of the static CAPM model assuming a stochastic set of investment opportunities. The expected market risk premium in equilibrium is:

$$E_t\left(R_{W,t}\right)=\left[\frac{-J_{WW}W}{J_W}\right]\sigma_{W,t}^2+\left[\frac{-J_{WB}}{J_W}\right]\sigma_{WB,t} \quad (1)$$

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 64–74, 2013.

Where *J(W(t), B(t), t)* is the utility function (subscripts representing partial derivatives), *W(t)* is the wealth level, *B(t)* is a variable that describes the state of investment opportunities in the economy, $E_t\left(R_{W,t}\right)$ is the expected excess return on aggregate wealth, $\sigma^2_{W,t}$ and $\sigma_{WB,t}$ are, respectively, the conditional variance and the conditional covariance of the excess returns with the investment opportunity set, and $\left[\frac{J_{WW}W}{J_W}\right]$, $\left[\frac{J_{WB}}{J_W}\right]$ could be viewed as the risk prices of the sources of risk.

Assuming risk-averse investors $J_W > 0$ and $J_{WW} < 0$, the model establishes a positive relation between risk premium and market volatility. However, the relation between the risk premium with the second risk factor (σ_{WB}) depends on the sign of J_{WB} and σ_{WB}. If J_{WB} and σ_{WB} share the same sign the investors demand a lower risk premium, but if the sign is different a higher risk premium is demanded. Assuming that Equation 1 is the proper model for the empirical study of the risk-return trade-off, the omission of this risk factor could lead to misspecifications of the empirical models and misleading evidence about the risk-return relationship.

Despite the important role of this trade-off in the financial literature, there is no clear consensus about its empirical evidence. In the theoretical framework, the parameters (the risk prices in brackets) are considered constant over time[1] and the variables (the sources of risk) are allowed to be time-varying. However, to make this model empirically tractable one must make several assumptions; the most common is considering constant risk prices [4]. Another common assumption made in the empirical analysis of the risk-return trade-off is considering a set of investment opportunities constant over time, remaining the market risk as the only source of risk [2]. This assumption leads to the validation of the static CAPM model. It is also necessary to assume specific dynamics for the conditional second moments. The most common are the GARCH models [5] [2]. Finally, the empirical model is established in a discrete time economy instead of the continuous time economy used in the equilibrium model of the theoretical approach. Many of the empirical papers studying the risk-return use one or more of the assumptions explained above.

The great controversy in the empirical validation of the risk-return trade-off is motivated by the disappointing results obtained about the sign and significance of this relation. There is no consensus about whether these results are due to: (a) wrong specifications of conditional second moments [6]; (b) misspecifications of the empirical models caused by the omission of the hedge component,[7]; (c) both causes.

However, another potential problem related with the empirical validation of the risk-return trade-off is the assumption of a linear relationship between return and risk. Some authors ([8], [9]) are concerned with this point and develop alternative

---

1 There are other general equilibrium models where time-varying risk aversion coefficients are obtained in models with habit persistence such as [10] or other theoretical frameworks where a non-linear and time-varying relation between risk and return is considered ([8])).

2 [11] proposes an alternative specification, the MIDAS regression, for modelling conditional second moments against GARCH models.

theoretical models for the risk-return trade-off where non-linear patterns are included through regime-switching models. The equilibrium model in [8] is slightly different from Merton's approach. A more complex, non-linear and time-varying relation between expected return and volatility is obtained. [8] also remarks the importance of the hedge component in the determination of the risk-return trade-off in his non-linear framework.

This study tries to shed light on the empirical validation of the risk-return trade-off. Although there is a large literature focused on this empirical validation, there are only few studies using multi-factor models that consider the hedge component[3]. The main empirical approach used in the literature is the GARCH-M framework, which assumes a linear relation between return and risk. However, there are other empirical approaches to analyze empirically the risk-return trade-off. Most of them use different econometric techniques to validate a linear relationship between return and risk based on the Merton's ICAPM model (i.e. [11] using the MIDAS regression or [12] using a temporal and cross-sectional analysis of a wide range of portfolios comprising the whole market). However, in this paper we use another econometric approach based on the equilibrium model of [8] in which we do not consider a linear relationship between return and risk but non-linear. It is showed in this paper that for shorter span empirical analysis, the relationship between expected return and volatility follows non-linear rather than linear patterns as suggested the ICAPM model. The RS-GARCH approach proposed in this study let obtain favorable evidence for a positive and significant risk–return tradeoff.

The main contributions of this paper are the followings: Firstly, this study analyzes the risk-return tradeoff in Spain during the last few years. Secondly, according to the papers of [8] and [9] it proposes a multi-factor model (considering a stochastic set of investment opportunities) where both the risk prices and sources of risk are state-dependent, allowing us to consider non-linear relationships between return and risk. Thirdly, it shows differences in the patterns followed by risk prices and conditional volatilities in different states (defined as low and high volatilities), being the risk price values lower during high volatility states and the conditional volatility more persistent during low volatility states. Fourth, it shows that a significant risk-return tradeoff can be obtained when it is assumed a non-linear relationship between return and risk. This evidence is essentially observed during low volatility states but not during high volatility states or when a linear relationship between return and risk is analyzed suggesting a procyclical risk-aversion of investors. Fifth, it seems that the relevant aspect for this evidence is the assumption of a non-linear relation between return and risk although the hedge component is important overall in the non-linear framework.

This paper is organized as follows. Section 2 provides a description of the data. Section 3 develops the empirical framework used in the paper. Section 4 gives the main empirical results and Section 5 concludes.

---

[3] One of the most common assumptions in the literature is the consideration of a constant set of investment opportunities, or, alternatively, independent and identically distributed rates of return. This assumption implies that the market risk premium only depends on its conditional variance and could be validated using univariate rather than multi-factor models.

2 Data Description

This study uses 720 weekly [13] excess market returns from the Spanish market, including observations from 1 January 1996 to 15 October 2009. Even though there are slight differences in the parameter estimations using different data frequency, there is no particular reason that the conclusions in this study should be affected by the selection of data frequency. Some authors remark this point in their studies [14].

The excess market returns are computed using the quotations of the IBEX-35 index, first obtaining logarithmic returns[4] and then subtracting from these returns the risk-free rate. Following [6] the money market rate suitably compounded at weekly frequency is used as the proxy for the risk-free rate. The choices for the proxy used as the hedging component against changes in the investment opportunity set are the followings rates for the Spanish market. ([12] use similar proxies for the American case): 1-year Treasury bill, 3-year Treasury bond, 5-year Treasury bond, 10-year Treasury bond, an equally averaged portfolio with the previous 3 bonds and the difference between the yields on the 10-year and the 3-year Treasury bond. Thomson Datastream is used to obtain the data about the stock index, International Financial Statistics for the data corresponding to the risk-free rate and the AFI (*Analistas Financieros Internacionales*) database[5] for the data about the proxies used as the intertemporal hedging component. For brevity we do not show the main summary statistics in this version of the paper but they are the typical for financial series (non-normal distribution, fat tails, heterokedasticity). Besides these stylized facts, the correlation matrix for the different proxies shows a low correlation between the excess returns of the market portfolio and the potential alternative investments. This result shows that the last series could be considered as proxies reflecting the alternative investment set available to the investors. Due to the lack of consensus in the literature about the best proxy representing the alternative investment set, this study uses the different assets shown above which present different characteristics (in their terms and maturity) and add robustness to the study.

3 Empirical Methodology

This section presents the empirical models used in the study. The main contribution of this paper is the assumption of a state-dependent risk price and state-dependent conditional volatilities, which implies a non-linear relationship between return and risk, following the equilibrium model in [8]. So, assuming bivariate GARCH dynamics for conditional volatilities, (more specifically, the BEKK model of [15], state-independent multi-factor models that establish a linear relation between return and risk are presented in Section 3.1, followed by state-dependent multi-factor models

[4] We use logarithmic returns multiplied by 100 to facilitate the convergence of the empirical models.

[5] AFI is a Spanish private consulting company.

that establish a non-linear risk-return trade-off through regime-switching both in the risk premium and conditional volatilities, in Section 3.2.[6]

3.1 State-Independent Multi-factor Model

This section presents a multi-factor model derived from the [3] ICAPM model. The 'general' model allows time-varying conditional second moments, but the risk aversion (risk price) coefficients for market risk $\left[\frac{J_{WW}W}{J_W}\right]$ and intertemporal component risk $\left[\frac{J_{WB}}{J_W}\right]$ are constant over time

$$\begin{aligned} r_{m,t} &= \lambda_{10} + \lambda_{11}\sigma^2_{m,t} + \lambda_{12}\sigma_{mb,t} + \varepsilon_{m,t} \\ r_{b,t} &= \lambda_{20} + \lambda_{21}\sigma_{bm,t} + \lambda_{22}\sigma^2_{b,t} + \varepsilon_{b,t} \end{aligned} \tag{2}$$

where λ_{ij} for *i=1,2* and *j=0,1,2* are the parameters to estimate and represent the different risk prices and $\sigma^2_{m,t}$, $\sigma^2_{b,t}$, $\sigma_{mb,t}$ represent the conditional second moments (market variance, intertemporal hedging component variance and covariance between market portfolio and hedging component). A restricted model is also estimated, where the alternative investment set is time invariant $(\lambda_{21} = \lambda_{22} = 0)$ [7].

As we explained above, it is necessary to make an assumption about the dynamics of the volatilities in order to empirically validate the theoretical ICAPM model. To analyze bivariate relationships, one of the most used models in the literature is the BEKK model of [15]. This model sets the following variance equation:

$$H_t = \begin{pmatrix} \sigma^2_{m,t} & \sigma_{mb,t} \\ \sigma_{mb,t} & \sigma^2_{b,t} \end{pmatrix} = CC' + A\varepsilon_{t-1}\varepsilon'_{t-1}A' + BH_{t-1}B' \tag{3}$$

where C is a lower triangular *2x2* matrix of constants, A and B are *2x2* matrices of parameters, ε_{t-1} is a Tx2 vector of innovations and H_{t-1} is the lagged covariance matrix.

The model is estimated by the maximization of the Quasi-Maximum Likelihood function of [16], assuming that the innovations follow a normal bivariate distribution $\varepsilon_t \sim N(0, H_t)$, which allows us to obtain robust estimates of standard errors.

[6] The asymmetric response of volatility to news of different signs (leverage effect) is not considered for several reasons: (1) there is no improvement about the significance of the risk-return trade-off in previous studies; (2) the convergence of the proposed models is harder to achieve due to the inclusion of the new parameters. These reasons lead to the consideration of a more parsimonious model. Moreover[14] states that the choice of volatility specification in the GARCH-M context is of second-order importance providing different specifications similar results.

$$L(\theta)=\sum_{t=1}^{T}\ln\left[f\left(r_t,\Omega_t;\theta\right)\right] \quad where \quad f\left(r_t,\Omega_t;\theta\right)=\left(2\pi\right)^{-1}\left|H_t\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\varepsilon_t^{'}H_t^{-1}\varepsilon_t\right) \qquad (4)$$

where I H_t I represents the determinant of the covariance matrix and the remaining terms have been defined above.

3.2 Regime-switching Multi-factor Model

This section introduces a new multi-factor model where both the risk prices and the conditional second moments are dependent of the state in the economy. In this case, we propose two states[7]. The consideration of regime-switching in the empirical relation allows us to obtain state-dependent estimations for the risk prices and conditional second moments. This implies a non-linear and state-dependent relation between expected return and risk following the general equilibrium model developed in [8].

The mean equation specification in this model is

$$\begin{aligned} r_{m,t,s_t} &= \lambda_{10,s_t}+\lambda_{11,s_t}\sigma^2_{m,t,s_t}+\lambda_{12,s_t}\sigma_{mb,t,s_t}+\varepsilon_{m,t,s_t} \\ r_{b,t,s_t} &= \lambda_{20,s_t}+\lambda_{21,s_t}\sigma_{bm,t,s_t}+\lambda_{22,s_t}\sigma^2_{b,t,s_t}+\varepsilon_{b,t,s_t} \end{aligned} \qquad (5)$$

for $s_t=1,2$ where λ_{ij,s_t} for *i=1,2* and *j=0,1,2* are state-dependent parameters, r_{m,t,s_t} and r_{b,t,s_t} are the state-dependent excess market and hedging component returns, σ^2_{m,t,s_t}, σ^2_{b,t,s_t} and σ_{bm,t,s_t} are the state-dependent conditional second moments, and ε_{m,t,s_t} and ε_{b,t,s_t} are the state-dependent innovations[8].

It is assumed that the state-dependent conditional second moments follow a GARCH bivariate dynamics (more specifically, a BEKK model). That is, there are as many covariance matrices as states. The state-dependent covariance matrices are

$$H_{t,s_t}=\begin{pmatrix}\sigma^2_{m,t,s_t} & \sigma_{mb,t,s_t}\\ \sigma_{mb,t,s_t} & \sigma^2_{b,t,s_t}\end{pmatrix}=C_{s_t}C_{s_t}{'}+A_{s_t}\varepsilon_{t-1}\varepsilon_{t-1}^{'}A_{s_t}{'}+B_{s_t}H_{t-1}B_{s_t}{'} \qquad (6)$$

The consideration of several states leads to a noteworthy rise in the number of parameters to estimate. In order to reduce this over-parameterization we only let parameters accompanying lagged innovations and lagged variances to be regime-switching[9]. Furthermore, the difference among states is defined by two new parameters *sa* and *sb* that properly weight the estimations obtained in one state for the other state. Therefore, the state-dependent covariance matrices in our model are:

---

[7] Previous studies considering three states (e.g., [17]) show that the third state only reflects odd jumps in the return series. The explanatory power of the third state is low and it is worthless in light of the difficulties of the estimation process that it produces.

[8] We also estimate a restricted model where $\lambda_{21}=\lambda_{22}=0$.

[9] [13] make a similar assumption to avoid potential convergence problems.

$$H_{t,s_t=i} = \begin{pmatrix} \sigma^2_{m,t,i} & \sigma_{mb,t,i} \\ \sigma_{mb,t,i} & \sigma^2_{b,t,i} \end{pmatrix} = CC' + A_i \varepsilon_{t-1} \varepsilon'_{t-1} A'_i + B_i H_{t-1} B'_i \quad for \quad s_t = i = 1,2 \tag{6.1}$$

where $A_2 = sa \cdot A_1$ and $B_2 = sa \cdot B_1$, A_1 and B_1 are *2x2* matrices of parameters, and *C* is a *2x2* lower triangular matrix of constants (the same for the 2 states).

The shifts from one regime to another are governed by a hidden variable following a first-order Markov process with transition matrix

$$\hat{P} = \begin{pmatrix} \Pr\left(s_t = 1 \middle| s_{t-1} = 1\right) = p & \Pr\left(s_t = 1 \middle| s_{t-1} = 2\right) = (1-q) \\ \Pr\left(s_t = 2 \middle| s_{t-1} = 1\right) = (1-p) & \Pr\left(s_t = 2 \middle| s_{t-1} = 2\right) = q \end{pmatrix} \tag{7}$$

where p and q are the probability of being in state 1 and 2 if in the previous period the process was in state 1 and 2 respectively.

The ex-ante probabilities (the probabilities of being in each state in the period t using the information set at period (t-1) are (10.1) and (10.2):

$$P\left(s_t = 1 \middle| \Omega_{t-1};\theta\right) = p * P\left(s_{t-1} = 1 \middle| \Omega_{t-1};\theta\right) + (1-q) P\left(s_{t-1} = 2 \middle| \Omega_{t-1};\theta\right) \tag{8.1}$$

$$P\left(s_t = 2 \middle| \Omega_{t-1};\theta\right) = 1 - P\left(s_t = 1 \middle| \Omega_{t-1};\theta\right) , \tag{8.2}$$

where

$$P\left(s_t = k \middle| \Omega_t;\theta\right) = \frac{P\left(s_t = k \middle| \Omega_{t-1};\theta\right) f\left(r_t \middle| s_t = k, \Omega_t;\theta\right)}{\sum_{k=1}^{2} P\left(s_t = k \middle| \Omega_{t-1};\theta\right) f\left(r_t \middle| s_t = k, \Omega_t;\theta\right)} \tag{9}$$

for k=1,2 are the filtered probabilities (the probabilities of being in each state in the period t with the information set up to t).

Assuming state-dependent innovations following a normal bivariate distribution $\varepsilon_{t,s_t} \sim N\left(0, H_{t,s_t}\right)$, the vector of unknown parameters θ is estimated by maximizing the following maximum-likelihood function:

$$L(\theta) = \sum_{t=1}^{T} \ln\left[\sum_{k=1}^{2} P\left(s_t = k \middle| \Omega_t;\theta\right) f\left(r_t, \Omega_t;\theta\right)\right] \quad where \quad f\left(r_t, \Omega_t;\theta\right) = (2\pi)^{-1} \left|H_t\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \varepsilon'_t H_t^{-1} \varepsilon_t\right) \tag{10}$$

where the state-dependent likelihood function is weighted by the ex-ante probability of being in each state.

4 Empirical Results

This section presents the empirical results for the models proposed. We estimate the models explained in the previous section for the different proxies used as the intertemporal hedging component; models using the 1-year T-bill, the 3-year T-bond, the 5-year T-bond, the 10-year T-bond, the equally-weighted bond portfolio and the term spread are named .a, .b, .c, .d, .e, .f for brevity. Section 4.1 shows the results for the linear models (without regime-switching) in the two cases mentioned: general and restricted version. Section 4.2 explains the results for the non-linear multi-factor models (general and restricted), including regime switching.

4.1 Multi-factor Models Estimations

The estimated models in this section are those introduced in section 3.1. The case without restrictions is named general model and the restricted version are the models where we assume constant risk premiums for the hedge component $\lambda_{21} = \lambda_{22} = 0$. The estimated parameters for the mean equation are presented in Table 1.

Table 1. Mean equation estimations for linear multi-factor model. Estimated parameters for mean equation. ***,** and * represents significance at 1%, 5% and 10% levels.

		$r_{m,t} = \lambda_{10} + \lambda_{11}\sigma^2_{m,t} + \lambda_{12}\sigma_{mb,t} + \varepsilon_{m,t}$ $r_{b,t} = \lambda_{20} + \lambda_{21}\sigma_{bm,t} + \lambda_{22}\sigma^2_{b,t} + \varepsilon_{b,t}$					
		Model 2.a	Model 2.b	Model 2.c	Model 2.d	Model 2.e	Model 2.f
λ_{10}	R	0.1639	0.1192	0.1388	0.1851	0.1487	0.1297
	G	0.1722	0.1858	0.1706	0.1992	0.1782	0.1626
λ_{11}	R	0.0125	0.0221	0.0151	0.0126	0.0190	0.0221
	G	0.0107	0.0162	0.0163	0.0139	0.0176	0.0205
λ_{12}	R	-0.2247	0.4998	0.2283	-0.0171	0.2779	0.4398
	G	-0.0433	0.7222	0.3634	0.0127	0.4165	0.5212
λ_{20}	R	0.0008	0.0243**	0.0387**	0.0631**	-0.0505***	0.0173**
	G	-0.0033	-0.0103	-0.0187	-0.0040	-0.1169*	-0.0091
λ_{21}		0.2417	0.0701	0.0864	0.0402	0.0688	0.0495
λ_{22}		0.2597	0.6618	0.4032*	0.1720	0.3700	0.6155

It is clear that most of the parameters in this multi-factor model are non-significant for the mean equation. The coefficients that reflect the market risk price (λ_{11}) are positive but non-significant in all cases considered. Similar results are obtained for the hedging component risk factor (λ_{12}).

4.2 Regime-switching Multi-factor Models Estimation

This section shows the estimations for the state-dependent models presented in Section 3.2. These models exhibit state-dependent risk prices and conditional second moments. Table 2 displays the estimation for the state-dependent mean equation in all cases considered. We can associate states 1 and 2 with low and high volatility periods respectively.

Positive and significant estimations for the market risk price in low volatility states ($\lambda_{11,s=1}$) are obtained in all cases considered (for all proxies used as the intertemporal hedging component in the general and restricted version of the model)[10]. A positive and significant influence over the market risk premium of the risk price is also observed, representing the covariance between risk premium and hedging component

[10] The results for the intercept are also significant. Some authors ([6], [11]) interpret this fact as market imperfections.

($\lambda_{12,s=1}$) in low volatility states. Generally, this covariance exhibits a negative influence in the total risk premium demanded (see Figure 1).

So, the product of the risk price times the covariance between excess market return and hedging component $\left(\lambda_{12}\sigma_{mb,t,s_t=1}\right)$ shows that the total risk premium required by the investor $\left(\lambda_{11}\sigma^2_{m,t,s_t=1} + \lambda_{12}\sigma_{mb,t,s_t=1}\right)$ is slightly lower than the market risk premium.

Only when the covariance is positive does the premium associated with the hedging component lead to higher values of the total risk premium regarding the market risk premium.

Table 2. Mean equation estimations for non-linear multi-factor model. Estimated parameters for mean equation. ***,** and * represents significance at 1%, 5% and 10% levels.

$$r_{m,t,s_t} = \lambda_{10,s_t} + \lambda_{11,s_t}\sigma^2_{m,t,s_t} + \lambda_{12,s_t}\sigma_{mb,t,s_t} + \varepsilon_{m,t,s_t}$$
$$r_{b,t,s_t} = \lambda_{20,s_t} + \lambda_{21,s_t}\sigma_{bm,t,s_t} + \lambda_{22,s_t}\sigma^2_{b,t,s_t} + \varepsilon_{b,t,s_t}$$

		Model 2.a	Model 2.b	Model 2.c	Model 2.d	Model 2.e	Model 2.f
Panel A. Low volatility state (s_t =1)							
$\lambda_{10,s_t=1}$	R	-1.1540**	-0.8954	-2.3156***	-2.3689***	-2.7943***	-2.4819**
	G	-0.8077	-1.5745***	-2.6614***	-2.5942***	-2.2375***	-4.0322***
$\lambda_{11,s_t=1}$	R	0.4044**	0.1169**	0.1311**	0.1867***	0.2980**	0.2270**
	G	0.3415**	0.1682**	0.1144***	0.1982***	0.0758**	0.3011**
$\lambda_{12,s_t=1}$	R	2.7521	4.6265	3.7124**	2.0452***	1.6691**	1.3646***
	G	7.6601***	1.8731**	-0.3294	2.5449***	0.0169**	1.6573***
$\lambda_{20,s_t=1}$	R	0.0099	0.0406**	0.0742**	0.0785	-0.0745**	0.0524**
	G	-0.0009	0.0096	0.4503	-0.0172	-0.2907**	0.1715***
$\lambda_{21,s_t=1}$	G	-0.1518	0.1997***	0.4438**	-0.0479	-0.8829***	-0.4219**
$\lambda_{22,s_t=1}$	G	1.8914	0.4396	0.9609	0.3025	0.2076	-1.3879**
Panel B. High volatility state (s_t =2)							
$\lambda_{10,s_t=2}$	R	-1.4062**	0.3597***	0.3436*	0.3502**	0.2323***	0.2662*
	G	-1.2897**	0.1993**	0.1891	0.3652**	0.3583***	0.2908**
$\lambda_{11,s_t=2}$	R	0.0733	0.0198	0.0137	0.0043	0.0191	0.0153
	G	0.0662	0.0302***	0.0337	0.0111	2.8310	0.0169
$\lambda_{12,s_t=2}$	R	-1.2216	0.3280	-0.3172	-0.3739	-0.0507	-0.7152
	G	-1.4985	0.3132	0.7456*	-0.4473	-0.0323	-0.2767
$\lambda_{20,s_t=2}$	R	-0.0027	0.0210	0.0354	0.0605	-0.0404**	0.0183**
	G	-0.0117***	-0.0355	-0.1469*	0.03480	-0.0768	-0.0184
$\lambda_{21,s_t=2}$	G	-0.1059	-0.1202*	0.1151	0.0637	0.0540	-0.0215
$\lambda_{22,s_t=2}$	G	0.9297	1.0195	1.1358**	0.0635	0.1869	0.7740*

Panel B of Table 2 shows the results obtained for state 2. Generally, a significant relation is not observed between expected return and risk in high volatility states. A positive but no significant estimation is obtained for the risk price (market risk ($\lambda_{11,s=2}$), and covariance between market risk and hedge component ($\lambda_{12,s=2}$)). Moreover, the risk aversion coefficients in state 1 (corresponding to low volatility states) are higher than those corresponding to state 2 (high volatility states). This result suggests that there is less risk aversion in high volatility states. This finding is not consistent with the spirit of the theoretical models that suggests that higher volatility should be compensated with higher returns. However, [9] and [14] found the same evidence: during high volatility states there is a decreasing level of risk aversion. One possible explanation could be the different risk aversion profile for investors in each state. During calm (low volatility) periods more risk-averse investors are trading in markets, but in high volatility periods only the less risk-averse investors remain in the market because they are the only investors interested in assuming such risk levels, decreasing the risk premium demanded during these periods. Moreover, recent papers such as [18] have reported similar evidence obtaining a significant risk-return trade-off during boom periods. In this study we do not define the states of the market depending on the business cycle (boom/crisis) but we use regime volatilities. However, the evolution for regime volatilities is very close to those of business cycles and very often low volatility states corresponds with calm periods while the less common high volatility states are associated with crisis periods [14]. The procyclical risk-aversion (investors show more risk-aversion during boom periods than during crises periods) documented in the paper of [18] is also supported in this approach using volatility regimes where investors show more risk-aversion during low volatility periods than during high volatility periods.

5 Summary

This paper analyzes empirically the risk-return trade-off for the Spanish market using several proxies for the alternative investment set. We propose two multi-factor models considering conditional second moments according a bivariate GARCH specification based on theoretical frameworks which develop linear and non-linear relationships between return and risk. The results show that only a positive and significant risk-return trade-off is obtained in the non-linear case and only in the states governed by low volatility process (State 1). However, it is found no favorable evidence either in the linear framework or in high volatility states. These results support the findings of previous papers which present a procyclical risk aversion behaviour of investors. During low volatility states (associated with boom periods) investors are more risk averse than during high volatility periods (associated with crises). The investor profile in each context may also have influence in this lower risk aversion coefficient. The weight of the hedging component in the risk premium is less important than the market risk factor although the former has also a significant impact in low volatility periods. Strong assumptions of a linear relation between return and risk could lead to model misspecification and an inability of the empirical model to capture a significant risk-return relationship since the existence of periods where a risk-return trade-off is not observed could lead to non-significant estimation of this relation for the entire sample.

References

1. Sharpe, W.: Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. Journal of Finance 19(3), 425–442 (1964)
2. Glosten, L., Jagannathan, R., Runkle, D.: On the relation between the expected value and the variance of the nominal excess return on stocks. Journal of Finance 48(5), 1779–1801 (1993)

3. Merton, R.: An intertemporal asset pricing model. Econometrica 41(5), 867–888 (1973)
4. Goyal, A., Santa-Clara, P.: Idiosyncratic risk matters! Journal of Finance 58(3), 975–1008 (2003)
5. Bollerslev, T.: Generalised autoregressive conditional heteroskedasticity. Journal of Econometrics 31(3), 307–327 (1986)
6. Leon, A., Nave, J., Rubio, G.: The relationship between risk and expected return in Europe. Journal of Banking and Finance 31(2), 495–512 (2007)
7. Scruggs, J.: Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach. Journal of Finance 53(2), 575–603 (1998)
8. Whitelaw, R.: Stock market risk and return: an equilibrium approach. Review of Financial Studies 13(3), 521–547 (2000)
9. Mayfield, S.: Estimating the market risk premium. Journal of Financial Economics 73(3), 867–887 (2004)
10. Campbell, J.Y., Cochrane, J.: Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behaviour. Journal of Political Economy 107(2), 205–251 (1999)
11. Ghysels, E., Santa-Clara, P., Valkanov, R.: There is a risk–return trade-off after all. Journal of Financial Economics 76(3), 509–548 (2005)
12. Bali, T., Engle, R.: The ICAPM with Dynamic Conditional Correlations. Journal of Monetary Economics 57(4), 377–390 (2010)
13. Capiello, L., Fearnley, T.: International CAPM with Regime-Switching Parameters. In: FAME Research Paper Series rp17, International Center for Financial Asset Management and Engineering, Geneva, Switzerland. (2000)
14. Lundblad, C.: The risk-return trade-off in the long run: 1836-2003. Journal of Financial Economics 85(1), 123–150 (2007)
15. Baba, Y., Engle, R., Kraft, D., Kroner, K.: Multivariate simultaneous generalized ARCH, Unpublished Working Paper (University of California at San Diego) (1990)
16. Bollerslev, T., Wooldridge, J.: Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11(2), 143–172 (1992)
17. Sarno, L., Valente, G.: The cost of carry model and regime shifts in stock index futures markets: An empirical investigation. The Journal of Futures Markets 20(7), 603–624 (2000)
18. Kim, S.W., Lee, B.S.: Stock Returns, Asymmetric Volatility, Risk Aversion, and Business Cycle: Some New Evidence. Economic Inquiry 46(2), 131–148 (2008)

Generalizing Some Usual Risk Measures in Financial and Insurance Applications

Jaume Belles-Sampera, Montserrat Guillén, and Miguel Santolino

Department of Econometrics, Riskcenter-IREA, University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain
jbellesa8@alumnes.ub.edu, {mguillen,msantolino}@ub.edu

Abstract. We illustrate a family of risk measures called GlueVaR that combine Value-at-Risk and Tail Value-at-Risk at different tolerance levels and have analytical closed-form expressions for the most frequently used distribution functions in financial and insurance applications, i.e. Normal, Log-normal, Student t and Generalized Pareto distributions. Tail-subadditivity is a remarkable property of a subfamily of GlueVaR risk measures. An implementation to the analysis of risk in an insurance portfolio is investigated.

Keywords: quantiles, subadditivity, tails, risk management.

1 Introduction and Background

A challenging task that risk managers must face on a daily basis is to assess and control the level of risk in financial and insurance institutions. For that purpose they usually quantify risk using the Value-at-Risk (a quantile in terms of a statistical distribution model) or the Tail Value-at-Risk which is the expected loss beyond Value-at-Risk at a given tolerance level. Choosing the distribution, the tolerance level and the risk measure is not an easy task. Regulators sometimes establish the risk measure and the tolerance level to be implemented in practice, but freedom is given to chose the distribution or to change the former in order to evaluate risk from alternative perspectives.

Our aim is to illustrate the use of a new family of risk measures, which we name GlueVaR and which were proposed recently by Belles-Sampera et al. [4]. This family combines the most popular risk measures and considers more than just one parameter to capture managerial attitudes towards risk. In addition, closed-form expressions of GlueVaR for frequently used statistical distributions can be easily deduced.

Value-at-Risk (VaR) at level α is the α-quantile of a random variable *X*, i.e. $\mathrm{VaR}_\alpha(X) = \inf\{x \mid F_X(x) \geq \alpha\} = F_X^{-1}(\alpha)$, where F_X is the cumulative distribution function (cdf) of *X* and α is the confidence or the tolerance level $0 \leq \alpha \leq 1$. Value-at-Risk has been adopted as a standard tool to assess the risk and to calculate capital requirements in the financial industry. VaR may fail to satisfy the

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 75–82, 2013.

subadditivity property[1]. A risk measure is subadditive when the aggregated risk is less than or equal to the sum of individual risks. Subadditvity is an appealing property when aggregating risks in order to preserve the benefits of diversification.

Tail Value-at-Risk (TVaR) measures average losses in the most adverse cases rather than just the minimum loss, as the VaR does. TVaR at level α of a random variable X is defined as, $\mathrm{TVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\lambda^1 \mathrm{VaR}_\lambda(X)\delta\lambda$. Therefore, risk assessment based on the TVaR have to be considerably higher than those based on VaR and significant differences arise depending on which risk measure is adopted. Notwithstanding, TVaR is subadditive unlike VaR..

2 GlueVaR Risk Measures

Consider a probability space and the set of all random variables defined on this space. Any risk measure[2] ρ is a mapping from the set of random variables to the real line $\Re$, $X \mapsto \rho(X) \in \Re$. Distortion risk measures were introduced by Wang ([31, 32]) and are closely related to the distortion expectation theory [36]. There are two key elements to define a distortion risk measure: first, the associated distortion function; and, second, the concept of the Choquet Integral[3]. A detailed literature review of distortion risk measures is available in [13] and [3]. The distortion function and the distortion risk measure can be defined as follows:

Definition 1. [Distortion function] Let $g:[0,1] \to [0,1]$ be a function such that *g(0)=0*, *g(1)=1* and *g* is non-decreasing. Then *g* is called a distortion function.

Definition 2. [Distortion risk measure] Let *g* be a distortion function. Consider a random variable *X* and its survival function $S_X(x) = P(X > x)$. Function ρ_g defined by $\rho_g(X) = \int_{-\infty}^{0}\left[g(S_X(x)) - 1\right]dx + \int_0^{+\infty} g(S_X(x))dx$ is called a distortion risk measure.

The mathematical expectation is a distortion risk measure whose distortion function is the identity function, $\rho_{id}(X) = E(X)$ (see, for instance, [13]). Therefore, a straightforward way to interpret a distortion risk measure is as follows: first, the survival function of the random variable is distorted ($g \circ S_X$); second, the mathematical expectation of the distorted random variable is computed. VaR and TVaR measures are distortion risk measures as shown in Table 1.

---

[1] VaR is subadditive for elliptically distributed losses (see, for example, [27]). However, the subadditivity of VaR cannot be generalized as has been shown, for instance, in [2] and [1].

[2] See [30], for instance, for an extensive introduction to risk measures.

[3] In honour of Gustave Choquet who introduced the concept of the integral for non-additive measures [9].

Table 1. VaR and TvaR distortion functions

Risk measure	**Distortion function**
VaR	$\psi_\alpha(u)=\begin{cases}0 & if \quad 0\le u<1-\alpha\\ 1 & if \quad 1-\alpha\le u\le 1\end{cases}$
TVaR	$\gamma_\alpha(u)=\begin{cases}\dfrac{u}{1-\alpha} & if \quad 0\le u<1-\alpha\\ 1 & if \quad 1-\alpha\le u\le 1\end{cases}$
For a confidence level $\alpha\in(0,1)$	

A new family of risk measures, named GlueVaR, was introduced by Belles-Sampera et al. [4]. A GlueVaR risk measure is described by its distortion function:

$$\kappa_{\beta,\alpha}^{h_1,h_2}(u)=\begin{cases}\dfrac{h_1}{1-\beta}\cdot u & if \quad 0\le u<1-\beta\\ h_1+\dfrac{h_2-h_1}{\beta-\alpha}\cdot\left[u-(1-\beta)\right] & if \quad 1-\beta\le u<1-\alpha\\ 1 & if \quad 1-\alpha\le u\le 1\end{cases} \qquad (1)$$

where $\alpha,\beta\in[0,1]$ so that $\alpha\le\beta$, $h_1\in[0,1]$ and $h_2\in[h_1,1]$. Parameter β is the additional confidence level besides α. The shape of the GlueVaR distortion function is determined by the distorted survival probabilities h_1 and h_2 at levels $1-\beta$ and $1-\alpha$, respectively. We call parameters h_1 and h_2 the heights of the distortion function.

A wide range of risk measures may be defined under this framework. Note that VaR_α and TVaR_α are particular cases of this new family of risk measures[4]. Parameters α,β, h_1 and h_2 can be selected so that GlueVaR is located in between VaR_α and TVaR_α, simply by making inequalities $\psi_\alpha(u)\le\kappa_{\beta,\alpha}^{h_1,h_2}(u)\le\gamma_\alpha(u)$ hold for any $u\in[0,1]$. Parameters can also be chosen so that $\mathrm{GlueVaR}_{\beta,\alpha}^{h_1,h_2}$ is located between TVaR_α and TVaR_β.

---

[4] Namely, for a random variable *X*, $\mathrm{VaR}_\alpha(X)$ and $\mathrm{TVaR}_\alpha(X)$ correspond to the risk given by the GlueVaR with distortion functions $\kappa_{\alpha,\alpha}^{0,0}$ and $\kappa_{\alpha,\alpha}^{1,1}$, respectively.

3 Linear Combination of Risk Measures

Given a random variable *X* and for fixed tolerance levels α and β such that $\alpha<\beta$, $\mathrm{GlueVaR}_{\beta,\alpha}^{h_1,h_2}(X)$ can be expressed as a linear combination[5] of $\mathrm{TVaR}_{\beta}(X)$, $\mathrm{TVaR}_{\alpha}(X)$ and $\mathrm{VaR}_{\alpha}(X)$ using the following transformation:

$$\begin{cases} \omega_1 = h_1 - \dfrac{(h_2-h_1)\cdot(1-\beta)}{\beta-\alpha} \\ \omega_2 = \dfrac{h_2-h_1}{\beta-\alpha}\cdot(1-\alpha) \\ \omega_3 = 1-\omega_1-\omega_2 = 1-h_2 \quad , \end{cases} \tag{2}$$

because then

$$\mathrm{GlueVaR}_{\beta,\alpha}^{h_1,h_2}(X) = \omega_1\cdot\mathrm{TVaR}_{\beta}(X) + \\ +\omega_2\cdot\mathrm{TVaR}_{\alpha}(X)+\omega_3\cdot\mathrm{VaR}_{\alpha}(X), \tag{3}$$

as it is shown in [4].

A useful consequence of (3) is that when analytical closed-form expressions of $\mathrm{VaR}_{\alpha}(X)$ and $\mathrm{TVaR}_{\alpha}(X)$ are known for a random variable *X*, we can automatically derive the closed-form expression of $\mathrm{GlueVaR}_{\beta,\alpha}^{h_1,h_2}(X)$ without further complications. For example, Normal, Log-normal, Student-*t* and Generalized Pareto distributions have simple closed-form expressions of GlueVaR. All of these distributions are investigated and the closed-form expressions of GlueVaR are provided in Belles-Sampera et al. [4]

Many research studies analyzing the properties of risk measures can be found in the literature ([2, 19, 20, 14, 29, 7, 16]). Artzner et al. [2] established a set of axioms that a risk measure should satisfy: positive homogeneity, translation invariance, monotonicity and subadditivity. Subadditivity is only guaranteed when the distortion function is concave (see [12, 15, 34, 35]). Fat right-tails have been

---

[5] An interpretation of GlueVaR risk measures as aggregation operators can be undertaken. An aggregation operator is a function that combines inputs into a single value, where inputs may be degrees of preference, membership or likelihood, or support of a hypothesis. Therefore, a linear combination of risk measures may be understood as an aggregation operator. A complete state of the art on aggregation operators can be found in [21, 22]. Additionally, VaR and TVaR may be understood as aggregation operators for discrete distributed random variables, as has been shown in [5].

extensively studied in insurance and finance (see, for example, [33, 17, 18, 11, 28, 8]), but studies about subadditivity of risk measures in the tail region are scarce ([10, 25]). Belles-Sampera et al. [4] already proved that right tail-subadditivity for a pair of risks is satisfied by GlueVaR if its associated distortion function $\kappa_{\beta,\alpha}^{h_1,h_2}$ is concave in $[0,1-\alpha)$.

4 Illustration and Risk Attitudes in GlueVaR

Let us fix two levels of severity, namely α and β, with $\alpha < \beta$. Then, the risk can be measured in a highly conservative scenario with TVaR at level β; in a conservative scenario with TVaR at level α; and in a less conservative scenario with VaR at level α. So GlueVaR is a combination of these risk scenarios and projects a specific risk aversion attitude. We select, for instance, $\alpha = 95\%$ and $\beta = 99.5\%$, which is equivalent to one bad event every twenty years or one bad event every two hundred, respectively. The other two parameters are directly related to the weights given to these scenarios. For instance, we could say that the three components of GlueVaR in (3) are equally important, i.e. $\omega_1 = \omega_2 = \omega_3 = 1/3$ or, in other words, that we have a balanced attitude with respect to the three scenarios of risk assessment.

Data for the cost of claims from a major Spanish motor insurer are used to illustrate the applicability of these results. These data contain n=519 observations of the cost of individual claims (the aggregated cost of bodily injury and property damage) in thousands of euros, and were analyzed in [6] and [23].

Table 2 presents the values of some risk measures in this application. In the first row the results when the empirical distribution is used are presented. In subsequent rows Normal, Log-normal, Student-*t* with 4 d.f. and Generalized Pareto distributions are fitted and their respective risk values are shown. In addition to VaR and TVaR at the 95% confidence level and TVaR at the 99.5% level, three different GlueVaR risk measures linked to those confidence levels have been computed. These GlueVaR match three different attitudes considered. The first one, the attitude associated to the GlueVaR with weights $(1/3, 1/3, 1/3)$ in expression (3) or, in terms of heights the one with $(h_1, h_2) = (11/30, 2/3)$. The other two attitudes are linked to GlueVaR risk measures with weights $(1/6, 1/4, 7/12)$ and $(-1/9, 10/9, 0)$ in (3), respectively. Or, in terms of heights, the allocated height values are $(23/120, 5/12)$ for the former and $(0, 1)$ for the later. The reasons behind these selections are as follows. The first GlueVaR risk measure is a potential candidate to satisfy tail-subadditivity while the other two are not[6]. Regarding the second and the third GlueVaR measures that are considered in the current application, the third has a negative ω_1 weight in (3) while in the second one all weights are positive. In fact, because of the heights of the third GlueVaR considered here, it holds that this risk measure lies between $VaR_{95\%}$ and $TVaR_{95\%}$. Hence, each GlueVaR presented in Table 2 represents a different risk attitude.

Calculations have been made in R and R programmes are available from the authors.

Table 2. Several risk figures for random variable 'cost of claims in thousands of €'

				$GlueVaR^{h_1,h_2}_{99.5\%,95\%}$		
	$VaR_{95\%}$	$TVaR_{95\%}$	$TVaR_{99.5\%}$	$\left(\frac{11}{30},\frac{2}{3}\right)$	$\left(\frac{23}{120},\frac{5}{12}\right)$	$(0,1)$
Empirical	47.6	125.5	479.0	217.4	139.0	86.3
Normal	87.0	105.9	143.4	112.1	101.1	101.8
Log-normal	48.9	119.1	397.8	188.6	124.6	88.2
Student-*t* (4 d.f.)	109.0	157.5	298.6	188.3	152.7	141.8
Pareto	44.2	94.6	301.4	146.7	99.7	71.7

The first comment to be made regarding the figures shown in Table 2 is that all the GlueVaR values computed here are considerably smaller than those given by $TVaR_{99.5\%}$. Secondly, as a decision maker that wants to preserve tail-subadditivity is supposed to be more risk averse than another decision maker who is not worried by this issue, it is natural to obtain higher risk values in the first GlueVaR column compared to the last two columns. On the other hand, we have obtained numerical evidence that the third GlueVaR risk measure lies between $VaR_{95\%}$ and $TVaR_{95\%}$. Finally, the values obtained with the second GlueVaR suggest that this risk measure represents an attitude less conservative than the attitude linked with the first GlueVaR but, at the same time, more conservative than the one linked to the last GlueVaR.

Last but not least, we also note that the assumption on the statistical model is crucial. The results differ a lot depending on which distribution is considered. However, the conclusions about the GlueVaR measures are the same no matter which distribution is considered. So, we argue that the choice of the statistical model and the choice of the risk measure do not interfere. In our application we have not addressed how practitioners should tackle the choice of the statistical distribution.

[6] As shown in Belles-Sampera et al. [4], a necessary condition for a GlueVaR risk measure to satisfy tail-subadditivity is that $h_2 \le h_1 \cdot \frac{1-\alpha}{1-\beta}$. With $\alpha = 95\%$ and $\beta = 99.5\%$, only the first set of weights fulfil that condition.

5 Conclusions

GlueVaR measures can easily be used to evaluate the risk when the risk manager has a particular risk attitude and wants to combine standard risk measures. Analytical expressions of these measures are effortlessly implemented in practice.

The two levels of qualitative information that GlueVaR risk measures incorporate (one related to the confidence levels of bad and worst-case scenarios; the other related to the plausibility of those scenarios) can help to achieve an old challenge about the need to include a risk attitude in risk assessment. We have also shown that the applicability of GlueVaR is straightforward.

Aknowledgements. The authors thank the Spanish Ministry of Science and the Spanish Ministry of Economy and Competitiveness for supports ECO21787-C03-01 and ECO2012-35584. Montserrat Guillén thanks ICREA Academia.

References

1. Acerbi, C., Tasche, D.: On the coherence of expected shortfall. Journal of Banking & Finance 26(7), 1487–1503 (2002)
2. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Mathematical Finance 9(3), 203–228 (1999)
3. Balbás, A., Garrido, J., Mayoral, S.: Properties of distortion risk measures. Methodology and Computing in Applied Probability 11(3, SI), 385–399 (2009)
4. Belles-Sampera, J., Guillén, M., Santolino, M.: Beyond Value-at-Risk: GlueVaR Distortion Risk Meaures. IREA Working Papers 201302, University of Barcelona, Research Institute of Applied Economics (2013)
5. Belles-Sampera, J., Merigó, J.M., Guillén, M., Santolino, M.: The connection between distortion risk measures and ordered weighted averaging operators. Insurance, Mathematics & Economics (2013), `http://dx.doi.org/10.1016/j.insmatheco.2013.02.008`
6. Bolancé, C., Guillén, M., Pelican, E., Vernic, R.: Skewed bivariate models and nonparametric estimation for the CTE risk measure. Insurance: Mathematics and Economics 43(3), 386–393 (2008)
7. Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., Montrucchio, L.: Risk measures: Rationality and diversification. Mathematical Finance 21(4), 743–774 (2011)
8. Chen, D., Mao, T., Pan, X., Hu, T.: Extreme value behavior of aggregate dependent risks. Insurance: Mathematics and Economics 50(1), 99–108 (2012)
9. Choquet, G.: Theory of Capacities. Annales de l'Institute Fourier 5, 131–295 (1954)
10. Danielsson, J., Jorgensen, B.J., Sarma, M., de Vries, C.G.: Sub-additivity re-examined: the case for Value-at-Risk. Tech. Rep., CiteSeerx (2005)
11. Degen, M., Lambrigger, D.D., Segers, J.: Risk concentration and diversification: Second-order properties. Insurance: Mathematics and Economics 46(3), 541–546 (2010)
12. Denneberg, D.: Non-Additive Measure and Integral. Kluwer Academic Publishers, Dordrecht (1994)
13. Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R.: Actuarial Theory for Dependent Risks. Measures, Orders and Models. John Wiley & Sons Ltd., Chichester (2005)

14. Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R., Laeven, R.: Risk measurement with equivalent utility principles. Statistics & Decisions 24(1), 1–25 (2006)
15. Dhaene, J., Laeven, R.J.A., Vanduffel, S., Darkiewicz, G., Goovaerts, M.J.: Can a coherent risk measure be too subadditive? Journal of Risk and Insurance 75(2), 365–386 (2008)
16. Ekeland, I., Galichon, A., Henry, M.: Comonotonic measures of multivariate risks. Mathematical Finance 22(1), 109–132 (2012)
17. Embrechts, P., Lambrigger, D.D., Wuethrich, M.V.: Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. Extremes 12(2), 107–127 (2009)
18. Embrechts, P., Neslehova, J., Wuethrich, M.V.: Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness. Insurance: Mathematics and Economics 44(2, SI), 164–169 (2009)
19. Föllmer, H., Schied, A.: Convex measures of risk and trading constraints. Finance and Stochastics 6(4), 429–447 (2002)
20. Frittelli, M., Rosazza Gianin, E.: Putting order in risk measures. Journal of Banking & Finance 26(7), 1473–1486 (2002)
21. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions: Construction methods, conjunctive, disjunctive and mixed classes. Information Sciences 181(1), 23–43 (2011)
22. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions: Means. Information Sciences 181(1), 1–22 (2011)
23. Guillén, M., Prieto, F., Sarabia, J.M.: Modelling losses and locating the tail with the Pareto Positive Stable distribution. Insurance: Mathematics and Economics 49(3), 454–461 (2011)
24. Hosking, J., Wallis, J.: Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3), 339–349 (1987)
25. Hua, L., Joe, H.: Tail comonotonicity: Properties, constructions, and asymptotic additivity of risk measures. Insurance: Mathematics and Economics 51(2), 492–503 (2012)
26. Jiang, L.: Convexity, translation invariance and subadditivity for g-expectations and related risk measures. Annals of Applied Probability 18(1), 245–258 (2008)
27. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton Series in Finance. Princeton University Press, New York (2005)
28. Nam, H.S., Tang, Q., Yang, F.: Characterization of upper comonotonicity via tail convex order. Insurance: Mathematics and Economics 48(3), 368–373 (2011)
29. Song, Y., Yan, J.-A.: Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. Insurance: Mathematics and Economics 45(3), 459–465 (2009)
30. Szëgo, G.: Measures of risk. Journal of Banking and Finance 26(7), 1253–1272 (2002)
31. Wang, S.S.: Insurance pricing and increased limits ratemaking by proportional hazard transforms. Insurance: Mathematics and Economics 17(1), 43–54 (1995)
32. Wang, S.S.: Premium calculation by transforming the layer premium density. ASTIN Bulletin 26(1), 71–92 (1996)
33. Wang, S.S.: An actuarial index of the right-tail risk. North American Actuarial Journal 2(2), 88–101 (1998)
34. Wang, S.S., Dhaene, J.: Comonotonicity, correlation order and premium principles. Insurance: Mathematics and Economics 22(3), 235–242 (1998)
35. Wirch, J.L., Hardy, M.R.: Distortion risk measures: Coherence and stochastic dominance. In: IME Conference, Lisbon (2002)
36. Yaari, M.E.: The dual theory of choice under risk. Econometrica 55(1), 95–115 (1987)

Risk Behavior of Stock Markets Before and After the Subprime Crisis

Esteban Otto Thomasz[1] and Aurelio Fernández Bariviera[2]

[1] Center of Quantitative Research (CMA), University of Buenos Aires, Córdoba 2122 1120AAQ Ciudad Autónoma de Buenos Aires República Argentina
[2] Department of Business, Universitat Rovira i Virgili, Av. Universitat 1 43204 Reus, Spain
Centro de Estudios Científicos y Técnicos, FACPCE, Argentina

Abstract. The aim of this paper is to present a simple classification of traditional risk indicators of stock markets. The last 10 years provides extensive evidence of changes of behavior of markets around the globe, combining a period of continuous growth with low volatility, an extreme crisis and a recuperation period. We present a simple analysis of risk indicators, such as implied volatility, value at risk, measuring of extreme events, etc. of 30 stock markets around the globe. Additionally we make a comparative analysis of such risk measures before and after the subprime crisis.

1 Introduction

The aim of this paper is to study changes in traditional risk indicators across markets during the period January 2003 to October 2012. The sub-prime crisis provides a unique opportunity to measure the performance of traditional risk estimators and the different impact across worldwide markets and gives a new track record as to incorporate the influence of extreme events in the behavior of stock markets.

Stock markets´ risk analysis can be divided into three approaches: extrapolation of past behavior, stochastic simulations and generative simulations. Extrapolation of past behavior is related to all the time series models applied to finance, starting from the Box and Jenkins methodology, GARCH models (Engle, 1982) and the application of other non-linear models such as the Fractional Brownian Motion (Mandelbrot, 1997).

Stochastic simulations approximate the behavior of financial prices by using computer simulations to generate random prices paths (Jorion, 2007). One traditional example is the Monte Carlo methods developed by Ulam in 1940 (Eckhardt, 1987). The generative approach (Epstein, 2006) consists of a creation of a group of agents with different rules of behavior that generates a simulation when interaction occurs within a finite space. This approach is related to programming of agent based models (Borril and Testefasion, 2011) with rules of behavior inspired on behavioral finance (Ritter, 2003). Some examples are Lux (1998) and LeBaron (2011), among many others.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 83–90, 2013.

However, in this paper we do not yet construct any model. We present a preliminary descriptive classification, with and hybrid approach: we try to measure the change of risk indicators over past history over two different non-overlapping periods of time. It must be point out that we are interested in the differences across markets and not within a single one, having a macro-financial view. The working hypothesis is that one vision of market risk might be defined to sudden change and reaction, more than average behavior. This paper contributes to the literature by analyzing the behavior of different risk proxies before and after the subprime crisis in 30 countries.

The remainder of the paper is organized as follows. The second section summarizes the risk proxies and the data used in this paper. The third section evaluates the performance of the selected risk indicators for the whole period and makes a comparative analysis in two non-overlapping periods. Finally the fourth section presents the main conclusions.

2 Methodology and Data

We have selected the following risk indicators:

▫Implied volatility, measured through the standard deviation

$$\sigma = \left[\sum_{i=1}^{n}(x_i - \bar{X})^2\right]^{1/2}$$

▫Extreme events, measured as the cases than exceed minus three standard deviations. This can be seen as a pragmatic way of gathering evidence to prove the existence of fat tails.

$$EE = \sum_{i=1}^{n} x_i \ \forall x_i < -3\sigma$$

▫Non Parametric value at risk, define as the quantil of the sample that will not be exceeded with a 99% confidence level.

$$RAR = x_0 \frac{\square}{\square} F(x_i < x_0) = 0{,}01$$

▫Extreme changes in volatility, measured as the range of 60 days moving average standard deviation.

$$Range_{\sigma_t} = max\left[\sum_{t=1}^{60}\frac{\sigma_t}{60}; \sum_{t=2}^{61}\frac{\sigma_t}{60}; \ldots; \sum_{t=n-120}^{n-60}\frac{\sigma_t}{60};\right] - min\left[\sum_{t=1}^{60}\frac{\sigma_t}{60}; \sum_{t=2}^{61}\frac{\sigma_t}{60}; \ldots; \sum_{t=n-120}^{n-60}\frac{\sigma_t}{60};\right]$$

In all cases, indicators are calculated on daily logarithmic returns of the series, in US dollars.

$$x_t = ln\left(P_t^{u\$s} / P_{t-1}^{u\$s} \right)$$

The series consists of 2569 observations, dating from December 2nd 2002 to October 4th 2012.

Regarding the sample, we have selected 30 stock markets around the globe, including developed and emerging economies of the most important economic regions. The countries than conform the sample are the following: Argentina, Austria, Belgium, Brazil, Canada, Chile, Colombia, Cyprus, Estonia, Finland, France, Germany, Greece, India, Indonesia, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, Peru, Portugal, Russia, Slovakia, Slovenia, Spain, the United Kingdom (UK) and the United States (US). It must be point out that this is a highly representative sample of the world's stock market, considering that only Canada, the US, Japan, France, Germany, Italy, Spain, and the UK concentrates 64,5% of aggregate capitalization.

3 Results

3.1 Risk Ranking during the Whole Period

In this section we present values of the indicators previously described, performing an ordinal ranking, classifying markets into three boxes with the same number of elements divided in high, medium and low risk.

Regarding the implied volatility of the historical series, it can be seen that Cyprus, Brazil and Russia were the most volatile markets. In the other extreme we can mention Italy, the US and Slovenia, achieving the lowest volatility record of the sample.

Table 1. Implied Volatility

High		Medium		Low	
Cyprus	2.643	Irland	1.733	India	1.562
Brasil	2.548	France	1.731	Korea	1.525
Russia	2.197	Indonesia	1.730	UK	1.466
Greece	1.974	Germany	1.721	Portugal	1.462
Argentina	1.925	Peru	1.703	Estonia	1.415
Austria	1.890	Netherlands	1.699	Chile	1.398
Spain	1.775	Luxembourg	1.592	Slovakia	1.339
Mexico	1.774	Canada	1.580	Italy	1.313
Colombia	1.756	Belgium	1.572	US	1.298
Finland	1.749	Japan	1.563	Slovenia	1.293

In the case of extreme events (Table 2), Canada, Peru and Austria concentrate the highest probability (thus highest risk), whereas Chile, Luxembourg and India presents the lowest cases that exceed three standard deviation in the left tail (lowest risk).

Table 2. Cases of extreme events, in percentage

High		Medium		Low	
Canada	1.29%	Germany	0.97%	Belgium	0.86%
Peru	1.25%	Argentina	0.97%	Portugal	0.86%
Austria	1.13%	Cyprus	0.97%	Slovakia	0.86%
US	1.09%	Finland	0.97%	Slovenia	0.86%
France	1.09%	Estonia	0.93%	Japan	0.86%
Irland	1.09%	Russia	0.93%	Brasil	0.82%
Netherlands	1.05%	Greece	0.93%	Korea	0.82%
Spain	1.05%	UK	0.90%	Chile	0.82%
Colombia	1.05%	Mexico	0.90%	Luxembourg	0.78%
Italy	1.01%	Indonesia	0.90%	India	0.74%

Comparing Table 1 and Table 2, it seems that extreme eents are more likely in developed countries, in spite of the fact that in those countries, the implied volatility is lower.

In the case of non-parametric return at risk (Table 3), Cyprus, Brasil and Russia had the higher values, thus performing higher risk. In the other hand, Italy, Solovakia and Slovenia show the lower values, thus being the less risky markets of the sample according to this indicator.

Table 3. Non parametric return at risk

High RAR		Medium RAR		Low RAR	
Cyprus	7.646	France	5.276	Japan	4.329
Brasil	7.034	Irland	5.240	UK	4.277
Russia	6.553	Mexico	5.151	Portugal	4.188
Austria	5.842	Netherlands	5.125	Korea	4.141
Greece	5.735	Germany	5.115	Estonia	4.107
Argentina	5.588	Finland	5.114	US	3.944
Peru	5.538	Indonesia	4.993	Chile	3.922
Spain	5.413	Belgium	4.540	Italy	3.906
Canada	5.298	India	4.445	Slovakia	3.650
Colombia	5.288	Luxembourg	4.380	Slovenia	3.605

In the case of the volatility range, Brazil, Russia and Austria present the highest values, meaning that changes in volatility is high, and thus performing higher risk in this sense. In the other extreme there is Korea, Estonia and Slovakia, with the smallest changes in volatility (less risk).

Table 4. Volatility range

High		Medium		Low	
Brasil	7.018	Cyprus	4.474	Indonesia	3.946
Russia	6.797	UK	4.390	Belgium	3.940
Austria	5.228	Germany	4.369	Greece	3.807
Netherlands	4.921	Irland	4.333	Portugal	3.749
Mexico	4.888	Spain	4.283	Finland	3.673
Canada	4.846	US	4.229	Chile	3.636
Peru	4.786	Argentina	4.211	India	3.184
France	4.576	Japan	4.179	Korea	3.148
Italy	4.537	Luxembourg	4.077	Estonia	2.947
Colombia	4.499	Slovenia	3.963	Slovakia	2.158

Up to this point, we have classified stock markets in an ordinary scale according to four different risk estimators. Following the previous classification, it can be said the Colombia was the most risky market, followed by Brazil, Russia and Spain. Meanwhile, the less risky markets were Korea, Chile, Slovakia and India.

3.2 Risk Change Before and After Crisis Inception

In this section we split the sample into two non-overlapping subperiods, recalculate all risk proxies and compute their change. Following Bariviera *et al.* (2012) we consider that the financial crisis starts with the bankruptcy of Lehman Brothers on September 15th 2008. Consequently, period 1 goes from December 2nd 2002 to September 12th 2008, totalizing 1.510 observations. Period 2 goes from September 15th 2008 to October 4th 2012, totalizing 1.059 observations.

The largest increase in implied volatility can be seen in Greece, Spain and Austria. The lowest increases correspond to Argentina, Indonesia and Colombia (decrease). In fact, the subprime crisis was harder in countries with expanded financial systems. Developing countries were less affected since the lack of well established financial markets acted as a protection shield.

Table 5. Increase in standard deviation between periods

High		Medium		Low	
Greece	111%	UK	75%	Italy	54%
Spain	110%	Germany	72%	Peru	50%
Austria	104%	Netherlands	72%	Slovenia	48%
Portugal	99%	Estonia	68%	Chile	42%
Canada	97%	Irland	66%	Slovakia	36%
Cyprus	94%	India	66%	Japan	34%
France	91%	Russia	64%	Brasil	33%
US	89%	Mexico	61%	Argentina	26%
Luxembourg	79%	Korea	60%	Indonesia	24%
Belgium	75%	Finland	55%	Colombia	-3%

The largest increase in RAR can be seen in Portugal, Canada and Italy, and the lowest in Argentina, Indonesia and Colombia.

Table 6. Increase in RAR between periods

High		Medium		Low	
Portugal	129%	Belgium	80%	Germany	64%
Canada	126%	Mexico	80%	Irland	60%
Italy	124%	Chile	79%	Russia	57%
Spain	115%	Slovenia	72%	Finland	56%
US	107%	Korea	70%	Peru	54%
France	106%	Japan	66%	Slovakia	53%
Greece	100%	Luxembourg	65%	India	44%
Cyprus	94%	Estonia	65%	Argentina	41%
UK	86%	Netherlands	65%	Indonesia	15%
Austria	84%	Brasil	64%	Colombia	-8%

Regarding extreme events, the largest increase correspond to the US, Canada and Austria, and the lowest to Chile, Indonesia and Colombia.

Table 7. Increase in extreme events between periods

High		Medium		Low	
US	0.0264	Mexico	0.0169	Japan	0.0127
Canada	0.0263	Irland	0.0168	Peru	0.0125
Austria	0.0241	Italy	0.0165	India	0.0115
France	0.0232	Estonia	0.0162	Russia	0.0114
Spain	0.0223	Luxembourg	0.0157	Slovenia	0.0111
Greece	0.0210	UK	0.0153	Argentina	0.0107
Netherlands	0.0207	Brasil	0.0150	Finland	0.0107
Germany	0.0204	Korea	0.0150	Chile	0.0102
Belgium	0.0191	Portugal	0.0143	Indonesia	0.0056
Cyprus	0.0172	Slovakia	0.0127	Colombia	(0.0002)

As for volatility range, the highest increase can be seen in Brazil, the US and Canada, and the lowest in Cyprus, Slovakia and Colombia.

Table 8. Increase in volatility range

High		Medium		Low	
Brasil	259%	Korea	127%	Netherlands	87%
US	249%	UK	123%	Chile	84%
Canada	231%	Argentina	113%	Greece	76%
Austria	193%	Portugal	109%	Finland	71%
Italy	192%	Belgium	101%	Irland	63%
Japan	142%	Germany	100%	Slovenia	47%
Luxembourg	131%	Spain	98%	India	42%
Russia	130%	Estonia	89%	Cyprus	24%
France	129%	Peru	89%	Slovakia	21%
Mexico	128%	Indonesia	88%	Colombia	-20%

In all, following the previous classification, Austria, Canada, France and the US appear to be the most risky markets in terms of past extrapolation. It has to be mention that the four of them are central markets that traditionally are associate with less risk measures, but according to this study they might seem as the most risky. Meanwhile, Finland, Slovakia, Indonesia and Colombia appear to be the less risky.

4 Summary

There are markets that are systemically more risky, maintaining high volatility over time. In these cases, extrapolation of the past –to quantify risk- seems to be an acceptable choice. This is for instance the case of Colombia.

On the other hand, developed markets tend to be more stable, and traditional risk measures based on past behavior tend to underestimate risk when a systemic crisis occurs. The group shows deeper changes in volatility in the event of crisis, such as Austria, Canada, France and the US.

Then, we see some countries with persisting problems, relating to deeper economic unbalances which are not yet been solved out, such as Spain and Greece. Last, we observe some markets with low risk measures during time and also between periods, such as Slovakia, the least risky market according to the classification presented.

This primary classification will serve as the base to move to a deeper analysis of stock market risk. The leading idea is to prove that under some circumstances consolidated markets are more risky in the sense of sudden change. On the other hand, emerging markets can be seen as more risky on average, but less risky in the sense of sudden change. Also, we cannot ignore special cases where endogenous macroeconomic imbalances generate shock amplification processes (Thomasz and Casparri, 2010) that increase market risk more rapidly than the average, like the cases of Spain and Greece.

In all, the contribution of this work is related to the concept of risk defined as sudden change, which can easily change the allocation of resources among markets, in some cases in favor of a group emerging markets. However, this study is still very limited, in the sense that algorithms used are very simple and we have chosen to evaluate the effect of one singular event as the sub-prime crisis. Nevertheless, this is a starting point to evaluate the application of a methodology that better quantify the unexpected changes among markets, improving the definition and estimation of market risk. Besides in future works, we would like to explore the possibility of anticipating structural breaks in risk measures.

References

1. Bariviera, A.F., Guercio, M.B., Martinez, L.B.: A comparative analysis of the informational efficiency of the fixed income market in seven European countries. Economics Letters 116(3), 426–428

2. Borrill, P., Tesfatsion, L.: Agent-Based Modeling: The Right Mathematics for the Social Sciences? Working Paper No. 10023, Iowa State University (2011)
3. Eckhardt, R.: Stan Ulam, John von Neumann, and the Monte Carlo method. Los Alamos Science, Special Issue 15, 131–137 (1987)
4. Engle, R.: Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. Econometrica 50, 987–1008 (1982)
5. Epstien, J.: Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton University Press (2007)
6. Jorion, P.: Financial Risk Manager Handbook. Wiley & Sons. Inc., New York (2007)
7. LeBaron, B.: Active and passive learning in agent-based financial markets. Eastern Economic Journal 37, 35–43 (2011)
8. Lux, T.: The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. Journal of Economic Behavior & Organization 33, 43–165 (1998)
9. Mandelbrot, B.: Fractals and Scaling in Finance. Springer, New York (1997)
10. Ritter, J.: Behavioral Finance. Pacific-Basin Finance Journal 11(4), 429–437 (2003)
11. Thomasz, E., Casparri, M.: Chaotic dynamics and macroeconomic shock amplification. Computational Intelligence in Business and Economics, pp. 507–514. World Scientific (2010)

Estimating Risk with Sarmanov Copula and Nonparametric Marginal Distributions

Zuhair Bahraoui, Catalina Bolancé, and Ramon Alemany

Dept. Econometrics, Riskcenter-IREA
University of Barcelona Av. Diagonal,
690, 08034 Barcelona, Spain
{zuhair,bolance,ralemany}@ub.edu

Abstract. We show that Sarmanov copula and kernel estimation can be mixed to estimate the risk of an economic loss. We use a bivariate sample from a real data base. We show that the estimation of the dependence parameter of the copula using double transformed kernel estimation to estimate marginal cumulative distribution functions provides balanced risk estimates.

Keywords: Copula, kernel estimation, value at risk, conditional tail expectation.

1 Introduction

Estimating the risk of loss is a major challenge in finance and in insurance, which has been extensively studied in the literature (see, for instance, the books by [1], [2] and [3] or articles such as [4], [5] and [6], among many others). In this work, we propose to use the Sarmanov copula (see [7]) with non-parametric marginals to estimate the risk of a loss that is obtained as the aggregation of two dependent losses. Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) are the selected risk measures. We show that estimating marginals using double transformed kernel estimation (DTKE) as proposed by [6] is the method that best fits our purpose. We apply our proposal to a real insurance database corresponding to a random bivariate sample of the cost of claims.

The aim of this work is to show as the transformed kernel estimator of cumulative distribution function allows us to obtain a good fit of the Sarmanov copula. Unlike the rest, in this copula the dependence structure is not separated strictly of the marginal distributions, i.e. the marginals are incorporated into the dependence structure; therefore, the estimation of these marginal distributions is essential to estimate the parameter of the copula.

2 Sarmanov Copula

Let (X,Y) be a bivariate random vector with marginal probability distribution functions (pdfs) f_X and f_Y. Also, let ϕ_1 and ϕ_2 be two bounded non constant function such that:

$$\int_{-\infty}^{+\infty} f_X(t)\phi_1(t)dt = 0, \quad \int_{-\infty}^{+\infty} f_y(t)\phi_2(t)dt = 0 .$$

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 91–98, 2013.

Then the bivariate pdf introduced by [7] is defined as:

$$h(x,y) = f_X(x) f_y(y)\left(1+\omega\phi_1(x)\phi_2(y)\right).$$

From Sklar's theorem, we deduce that the associated copula can be expressed as:

$$C(u,v) = uv + \omega \int_0^u \int_0^v \phi_1\left(F_X^{-1}(t)\right)\phi_2\left(F_y^{-1}(s)\right) dt ds. \qquad (1)$$

and its density is:

$$c(u,v) = 1 + \omega\phi_1\left(F_X^{-1}(u)\right)\phi_2\left(F_y^{-1}(v)\right). \qquad (2)$$

where F_X and F_Y are cumulative distribution functions (cdfs) of X and Y, respectively. Parameter ω is a real number that satisfies the condition $1+\omega\phi_1(x)\phi_2(y) \geq 0$ for all x and y. This parameter is related to the correlation between X and Y (if it exists), ω is called the dependence parameter. As [8] shows, the dependence between X and Y is:

$$\rho = \frac{v_1 v_2}{\sigma_1 \sigma_2},$$

where $v_1 = E(X\phi_1(X))$, $v_2 = E(X\phi_2(X))$, $\sigma_1^2 = Var(X)$ and $\sigma_2^2 = Var(Y)$ When we take $\phi_1(x) = 1-2F_X(x)$ and $\phi_2(y) = 1-2F_Y(y)$, we have the classical Farlie-Gumbel–Morgenstern (FGM) copula. In this case the dependence parameter has the range $-1/3 \leq \omega \leq 1/3$.

Another special case is when we consider functions of the type:

$$\phi_1(x) = x - \mu_X \text{ and } \phi_2(y) = y - \mu_Y \qquad (3)$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$. The author in [8] shows that, if the support of f_X and f_Y is contained in $[0,1]$, then the range of the dependence parameter is:

$$\max\left(\frac{-1}{\mu_X\mu_Y}, \frac{-1}{(1-\mu_X)(1-\mu_Y)}\right) \leq \omega \leq \min\left(\frac{1}{\mu_X(1-\mu_Y)}, \frac{1}{(1-\mu_X)\mu_Y}\right).$$

If the support of f_X is contained in $[a,b]$ and f_Y is contained in $[c,d]$, we can easy prove that:

$$\max\left(\frac{-1}{(b-\mu_X)(d-\mu_Y)}, \frac{-1}{(\mu_X-a)(\mu_Y-c)}\right) \leq \omega \leq \min\left(\frac{1}{(b-\mu_X)(\mu_Y-c)}, \frac{1}{(\mu_X-a)(d-\mu_Y)}\right).$$

.

2.1 Simulating from the Sarmanov Copula

To generate a bivariate random variable from Sarmanov's copula in (1), we use the procedure described by [9] which is based on the conditional distribution of a random vector (U,V)

$$P(V \leq v \mid U = u) = C_u(v) ,$$

where:

$$C_u(v) = \lim_{\delta \to 0^{\leftarrow}} \frac{C(u+\delta, v) - C(u,v)}{\delta} = \frac{\partial C(u,v)}{\partial u} .$$

The algorithm is implemented as follows:

1. Generate two independent random variables u and t from an Uniform$(0,1)$ distribution.
2. Set $v = C_u^{-1}(t)$, where $C_u^{-1}(t)$ denotes a quasi-inverse of C_u.
3. The desired pair is (u,v).

For our case, when we consider functions ϕ_1 and ϕ_2 as defined in (3), we have:

$$C_u(v) = v + \omega\left(F_X^{-1}(u) - \mu_X\right)\int_0^v \left(F_Y^{-1}(s) - \mu_Y\right)ds . \tag{4}$$

This result can easily be shown if the derivative of (1) is calculated or, alternatively, using the following relationship (see Lee, 1996):

$$P(Y \leq y \mid X = x) = F_Y(y) - \omega \int_y^{+\infty} f_Y(t)\phi_2(t)dt , \tag{4}$$

Taking $v = F_Y(y)$ and $u = F_X(x)$. Expression (4) can be calculated using the change of variable $t = F_Y^{-1}(s)$ in the integral.

3 Nonparametric Approximation of cdf

In this work, we propose to use different ways of obtaining a kernel estimation of the cdf in order to estimate marginal cdfs F_X and F_Y, respectively. We consider classical kernel estimation, transformed kernel estimation and double transformed kernel estimation. We describe these three methods to the specific case when marginals F_X and F_Y are the same.

Classical kernel estimation (CKE) of cdf F_X is obtained by integration of the classical kernel estimation of its pdf f_X. By means of a simple change of variable it follows that:

$$\hat{F}_X(x) = \int_{-\infty}^{x} \hat{f}_X(t)dt = \int_{-\infty}^{x} \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{t-X_i}{b}\right)dt \\ \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-X_i}{b}} k(t)dt = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right), \quad (5)$$

where $k(\cdot)$ is a pdf, which is known as kernel function. Very common kernels are Gaussian or Epanenchnikov kernels (see [10]). Function $K(\cdot)$ is the cdf of $k(\cdot)$. Parameter b is the bandwidth; it controls the smoothness of the function estimate. Silverman in [11] analyzes the statistical properties of (5).

Classical kernel estimation is not a good alternative when data are right skewed (see [6]). An alternative is transformed kernel estimation, that consists of transforming the data so that the transformed observations are symmetric. Authors in [12] propose to use a shifted power transformation family:

$$T_{(\lambda_1,\lambda_2)}(x) = \begin{cases} (x+\lambda_1)^{\lambda_2}\, sign(\lambda_2) \, if \; \lambda_2 \neq 0 \\ \ln(x+\lambda_1)^{\lambda_2} \quad if \; \lambda_2 = 0 \end{cases}, \quad (5)$$

where $\lambda_1 \geq -\min(X_1,\ldots,X_n)$ and $\lambda_2 \leq 1$ for right skewed data. Transformed kernel estimation (TKE) of a cdf is:

$$\hat{F}_X(x) = \sum_{i=1}^{n}\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{T_{(\lambda_1,\lambda_2)}(x) - T_{(\lambda_1,\lambda_2)}(X_i)}{b}\right). \quad (6)$$

Then, the transformed kernel estimation is a classical kernel estimation with transformed data ([12] describe a method to choose transformation parameters λ_1 and λ_2).

Double transformed kernel estimation (DTKE) needs two steps. First, a transformation of the data $T(X_i) = Z_i, i = 1,\ldots,n$, is chosen, where the transformed data have a distribution that is close to the Uniform$(0,1)$ distribution. Second, data are transformed again using the inverse of the Beta$(3,3)$ distribution with pdf and cdf:

$$m(x) = \frac{15}{16}(1-x^2)^2, -1 \leq x \leq 1$$

and

$$M(x) = \frac{3}{16}x^5 - \frac{5}{8}x^3 + \frac{15}{16}x + \frac{1}{2}.,$$

The resulting transformed data have a distribution that is close to the $\text{Beta}(3,3)$ distribution. This distribution can be estimated optimally using classical kernel estimation (see the discussion in [6]).

The double transformation kernel estimation (DTKE) is:

$$\hat{F}_X(x) = \sum_{i=1}^{n} \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{M^{-1}(T(x)) - M^{-1}(T(X_i))}{b} \right). \tag{6}$$

where T(x) is the generalized Champernowne cdf:

$$T(x) = \frac{(x+c)^{\gamma} - c^{\gamma}}{(x+c)^{\gamma} + (M+c)^{\gamma} - 2c^{\gamma}}$$

with parameters $\gamma, M > 0$ and $c \geq 0$ (authors in [13] describe a maximum pseudo-likelihood method to estimate parameters of the Champernowne distribution).

4 Value-at-Risk and Tail Value-at-Risk

Let $S = X + Y$ be the sum of two possibly dependent random variables X and Y. The Value-at-Risk of S with a confidence level α is:

$$VaR_{\alpha}(S) = \inf\{s, F_S(s) \geq \alpha\}.$$

and the Tail Value-at-Risk of S with a confidence level α is:

$$TVaR_{\alpha}(S) = E(S \mid S \geq VaR_{\alpha}(s)) = \frac{1}{1-\alpha} \int_{\alpha}^{1} VaR_{\alpha}(u) du .$$

Our goal in this section is to calculate the VaR and TVaR using Sarmanov copula to model dependency and the nonparametric approach to estimate marginal cdfs using the Monte Carlo method. The procedure is described here:

1. Estimate with non-parametric method the cdfs $\hat{F}_X$ and $\hat{F}_Y$.
2. Replace the cdf estimates in (2) in order to obtain the parameters of the copula that maximize likelihood.
3. Generate pairs (u_i, v_i) for $i = 1 \ldots r$ using (4), where r is the number of simulated pairs.
4. Solve $\hat{F}_X(X_i) = u_i$ and $\hat{F}_Y(Y_i) = v_i$ and obtain the simulated losses (X_i, Y_i).
5. Calculate $S_i = X_i + Y_i$ and estimate $VaR_{\alpha}(S)$ and $TVaR_{\alpha}(S)$ empirically once r repetitions are available.

5 Results

The data corresponds to a random sample of claims that were obtained from motor insurance accidents. An insurance company kindly gave us access these data. We have two costs: Cost1, contains the amount paid to the insured person to compensate for own damages to the vehicle and all other losses to third-parties damages, and Cost2, corresponds to the expenses related to medical treatments and hospitalization as a result of the accident (see [14] for more information on these data).

To estimate marginal cdfs with classical kernel estimation (CKE), transformed kernel estimation (TKE) and double transformed kernel estimation (DTKE) we use the Epanechnikov kernel and the bandwidth based on the asymptotic minimization of weighted integrated mean squared error (WISE, see [6]).

In Table 1 we summarize the values of the dependence parameters of the Sarmanov copula for the different kernel estimations for the marginal cdf. We note that there are significant differences between the estimated dependence parameters when we use CKE, TKE or DTKE to estimate the marginal cdfs. The log-likelihood column refers only to the dependence parameter estimation, that is not a full likelihood, which is why positive values are obtained. The log-likelihood shows that the best fit is obtained using DTKE to estimate marginal cdfs.

Table 1. Estimated parameters of Sarmanov copula for bivariate claims cost data

	w	Log-Likelihood*
CKE	9.53×10^{-9}	4.96227
TKE	109.05	29.68099
DTKE	0.99	87.86507

* This is not full likelihood, so values can be larger than zero

Tables 2 and 3 summarize the $VaR_\alpha(S)$ and $TVaR_\alpha(S)$ calculated for different confidence levels α, using a Monte Carlo simulation method. We use $r = 2000$ simulated samples. In Fig. 1 we plot the estimated $VaR_\alpha(S)$ together with the confidence interval at 95% for the empirical estimation.

To estimate the confidence intervals, we use the Bootstrap method. We generate samples with replacement with the same size of the original sample. This methodology allows us to obtain the intervals with different confidence levels.

Table 2. Estimated VaR with Sarmanov copula for bivariate claims cost data at tolerance levels 95%, 99% and 99.5%

VaR	0.95	0.99	0.995
CKE	7639.09	20582.40	24781.87
TKE	8622.01	48179.00	100784.01
DTKE	9610.28	24638.11	26338.82

Table 3. Estimate TVaR with Sarmanov copula for bivariate claims cost data at tolerance levels 95%, 99% and 99.5%

$TVaR$	0.95	0.99	0.995
CKE	47745.92	61114.87	81381.10
TKE	246990.90	326447.20	465581.20
DTKE	36907.52	46006.60	56690.85

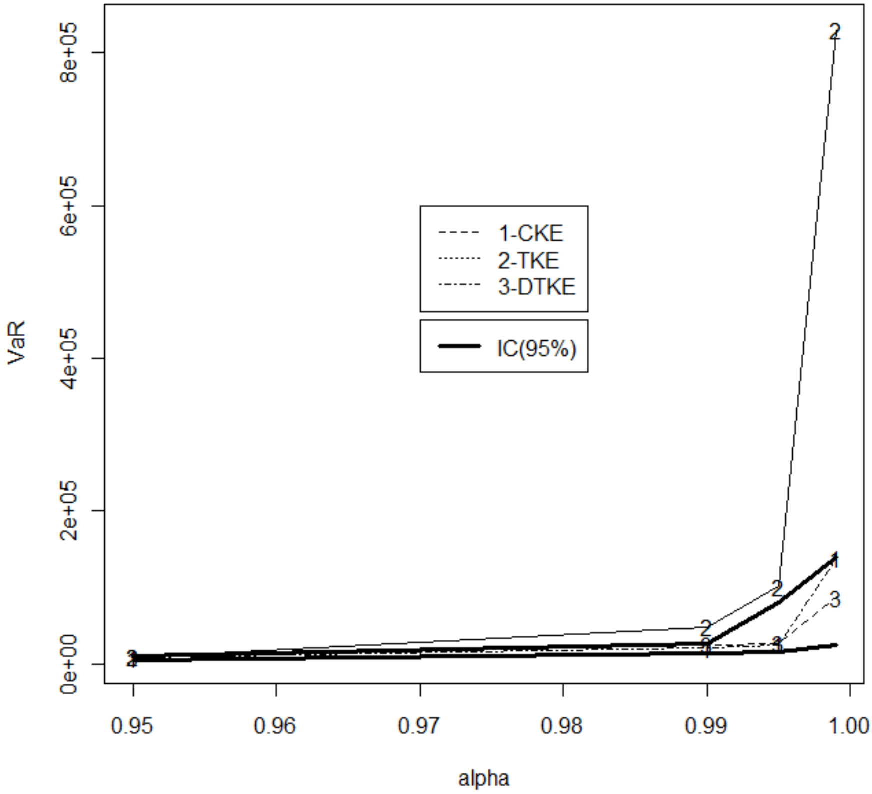


Fig. 1. Estimated VaR

The results show that using the DTKE we obtain the best results, inside confidence levels and ensuring balanced risk estimation, neither overestimated nor underestimated.

6 Conclusions

In this paper we present an example, using a random sample of claims from a real database, where, to estimate the risk of an aggregated loss, we mix copulas with kernel estimations. We show that double transformed kernel estimation (DTKE) of a cdf can be a useful tool combined with copulas, which allows to estimate the VaR and the TVaR using Monte Carlo simulation method.

We propose to use the Sarmanov copula, which so far has not been used in the context of the quantifying risk. As we say in the introduction, in this copula the dependence structure is not separated strictly of the marginal distributions. We show how this fact can affect the risk estimate since the estimation of the marginal distribution affects significantly the goodness of fit of this copula. As principal future lines for research we want to analyze how kernel estimation can improve the goodness of fit of alternative and well known copulas.

References

[1] McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concept, Techniques and Tools. Princeton University Press (2005)
[2] Jorion, P.: Value at risk. The new Benchmark for managing Financial Risk. The McGraw-Hill Companies (2007)
[3] Bolancé, C., Guillén, M., Nielsen, J.P., Gustafsson, J.: Quantitative Operational Risk Models. CRC finance series. Chapman and Hall, New York (2012)
[4] Dhaene, J., Denuit, M., Goovarts, M.J., Kaas, R., Vyncke, D.: Risk measure and comonotonicity: A Review. Stochastic Models 22, 573–606 (2006)
[5] Dowd, K., Blake, D.: After VaR: The theory, estimation, and insurance applications of quantile-based risk measures. The Journal of Risk and Insurance 73, 193–229 (2006)
[6] Alemany, R., Bolancé, C., Guillén, M.: Non-parametric estimation of Value-at-Risk. Insurance: Mathematics and Economics 52, 255–262 (2013)
[7] Sarmanov, O.V.: Generalized normal correlation and two-dimensional Fréchet. Soviet Mathematics. Doklady. 25, 1207–1222 (1996)
[8] Lee, M.L.T.: Properties and aplications of the Sarmanov family of bivariate distributions. J. Theory and Methods 25, 1207–1222 (1996)
[9] Nelson, R.B.: An Introduction to Copulas, 2nd edn. Springer, Portland (2006)
[10] Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
[11] Azzalini, A.: A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328 (1981)
[12] Bolancé, C., Guillén, M., Nielsen, J.P.: Kernel density of actuarial loss functions. Insurance: Mathematics and Economics 32, 19–36 (2003)
[13] Buch-Larsen, T., Guillén, M., Nielsen, J., Bolancé, C.: Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. Statistics 39, 503–518 (2005)
[14] Bolancé, C., Guillén, M., Pelican, E., Vernic, R.: Skewed bivariate models and nonparametric estimation for CTE risk measure. Insurance: Mathematics and Economics 43, 386–393 (2008)

Implications of Unisex Assumptions in the Analysis of Longevity for Insurance Portfolios

Arelly Ornelas, Montserrat Guillén, and Manuela Alcañiz

Dept. Econometrics, Riskcenter-IREA
University of Barcelona
Av. Diagonal, 690, 08034 Barcelona, Spain
{arelly.ornelas,mguillen,malcaniz}@ub.edu

Abstract. Since the Court of Justice of the European Union (EU) imposed a rule that insurance prices cannot be different for men and women, it is relevant to look into insights of mortality rate calculation and its implications. We already know that mortality curves are distinct between men and women, mainly in youth ages. However we are interested in the longevity risk, therefore we studied elderly ages. We start the analysis describing mortality rates for each sex and age. The data used come from INEGI and CONAPO in Mexico. For each 5-year period, the number of deaths and the number of living people by age and sex are available. We assume possible scenarios of gender proportion in mortality tables. Using survival analysis we estimated the parameters of a Weibull distribution, moreover we calculated the value at risk at different confidence levels. We emphasize the limitations of using unisex mortality tables to estimate longevity risk if the proportion of male and female is not equilibrated in the insurance portfolio. The actuarial modeling is necessary even when regulatory circumstances impose rules on the use of risk factors such as gender for tarification.

Keywords: Longevity risk, Value at risk, Unisex life tables.

1 Introduction

In this article, we model the mortality behavior in the Mexican general population using data from 1990 to 2009 for each 5-year period and then we compare the results with those assumed for a subpopulation of citizens where the proportion of men and women is not equal. This is what happens in insurance portfolios. For those who are insured, the proportion of men and women does not need to be balanced as it is in the general population.

It is widely known that insured policyholders generally have more wealth than the average citizen does, as they can afford insurance. It is also believed that for the same reason, members of the insured subpopulation invest a larger amount of money to prevention, thus leading to lower rates of mortality compared to the general population of the same age and gender. A conflict occurs when the regulations, as it is the case in Europe, or current commercial practices, dictate that only unisex life tables can be used to price life insurance products. Several questions arise: should insurers use unisex life tables? What are the implications of banning the use of different mortality assumptions for men and women?

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 99–107, 2013.

Due to the obligation to abandon gender discrimination in pricing life insurance in Europe, we were motivated to start this analysis and to contribute to the understanding of mortality model in this context.

On March 1, 2011, the Court of Justice of the European Union (EU) ruled that the differentiation in price and service provision depending on gender is discriminatory and therefore sex discrimination cannot be used when pricing new insurance contracts. So far, men and women could get different prices when buying an insurance contract as an exception within the European directive against sex discrimination in access to goods, services and supply.

The decision on eliminating price differences took place in spite of the statistical evidence that all over the world women have more longevity than men and that women's annual mortality rate is lower than men's annual mortality rate of the same age up to very old age groups.

Life insurance is a particular field where a non-discriminate is expected to have a large impact because it is very difficult to replace gender information by objective behavioral explanatory factors. Thus, the elimination of gender in pricing is a fundamental change when addressing longevity risk.

Women are mostly affected by the payment of higher life insurance premiums than they did in the past, because the non-discrimination implies that women's mortality rates should compensate men's mortality rates, which are higher. Men are badly affected in products that imply some longevity risk, such as annuities at retirement.

The probability of death of an individual at a given age depends on his current specific age and sex, among other factors such as occupation, lifestyle and daily habits. An important measure for an actuary is life expectancy, which shows the average time in years that a person is expected to survive. Life expectancy is the fundamental part of pension calculation and it is estimated based on an analysis of the number information on living and deceased people. Life expectancy can be calculated at birth or at any age, the latter case is of interest in the actuarial field. According Vaupel [18] to increase the life expectancy of a population means to reduce premature deaths by age, something that is achieved by prevention and appropriate health policies. However, we will analyze longevity risk, i.e. the number of years that only a small fraction of the population survives. For this reason, we will apply the concept of value at risk to lifetime duration.

2 Data Description

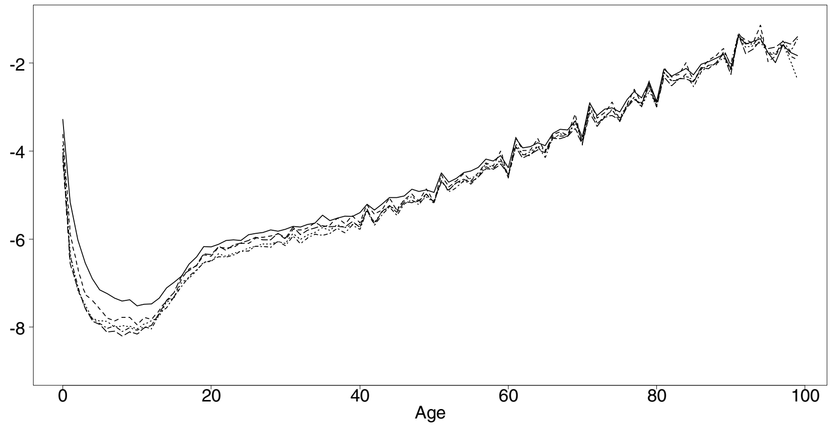
Data for this study were obtained from the website of the National Institute of Statistics, Geography and Informatics (INEGI) [21] in Mexico and the National Population Council (CONAPO) [22]. Mortality rates were computed as the ratio between the number of people alive and the number of registered deaths with age x in a year t. The former were extracted directly from censuses and population counts conducted by the INEGI. The latter were extracted from the CONAPO website. The data analyzed here reflect the demographic behavior of the 31 states of Mexico and one federal district. We have information for some specific years 1990, 1995, 2000, 2005 and 2009; we have separated males and females with age between 0 and 99 years.

In order to avoid biased information, data without description of sex and/or age were excluded. The omission of such information could be caused by deaths at a young age, when gender may not be reported, or in the case of old age, where the exact death-age is not known. The percentage of unclassified data is very small (less than 0.1%), so we are not losing a lot of information.

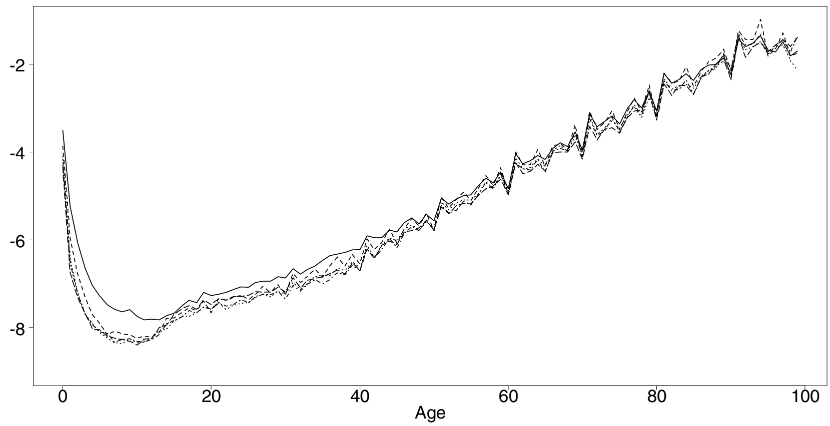
We calculate the crude rate of mortality m(x, t) as:

$$m(x, t) = d(x, t)/l(x, t), \quad (1)$$

where l(x, t) is the number of living individuals and d(x, t) the number of deaths for each year t and age x.



(a)



(b)

Fig. 1. Log crude mortality rates for Mexican population: (a) Males, (b) Females. From age 0 to 99. Years 1990, 1995, 2000, 2005 and 2009. Solid line represents 1990.

We performed an exploratory analysis of the evolution of mortality rates over the study years t and at different ages x, separated by sex. Figure 1 shows that the curves of log mortality rates in men and women are very similar in childhood and adulthood; the largest differences can be found in the youth that is when men are more likely to have accidents. This phenomenon related to young man fatalities is known as the accident hump and has not changed much over the two decades. Figure 1 also shows that mortality has improved in the last two decades, mostly before age 20. The solid line represents the year 1990, which was the year with bigger mortality.

In order to explore the effect of gender proportion we vary the percentage of females in the population. The mixed rate at age x ant t years is defined as follows:

$$m_w(x, t) = w\, m_f(x, t) + (1-w)m_m(x, t), \tag{2}$$

where $w \in \{0\%, 25\%, 50\%, 75\%, 100\%\}$, $m_w(x, t)$ is the mixture of mortality rates, $m_f(x, t)$ is the mortality rate for females and $m_m(x, t)$ is the mortality rate for men. As we are interested in the mortality for elderly ages we subset the data. The final database contained ages from 65 to 99.

3 Methodology

3.1 Review of Survival Analysis

Survival analysis refers to a methodology that focuses on the study of mortality behavior. Since the definition of mortality can be applied in many other situations, survival analysis has been widely used to study termination or any event related to the end of a process. In biostatistics, survival analysis refers to the duration of a disease that does not necessarily end up in a fatality. In engineering, survival usually refers to the duration of devices until there is a failure. More generally, survival analysis estimates the time to the occurrence of an event of interest.

We denote by T the duration of life. The survival function is the probability of an individual surviving beyond time t, $S(t) = P[T > t] = 1 - F(t)$, where $t > 0$. S corresponds to the probability that an individual does not die before t time units, i.e. the probability that event E occurs after t years. Some properties of the survival function: $S(0) = 1$, $S(\infty) = 0$, $S(t)$ is a decreasing monotonic function.

There are two approaches of fitting survival function: non-parametric and parametric. In the first case the most known is the Kaplan and Meier method. Specifically for older ages Fledelius et al (2004) [7] made a comparison between the two approaches.

Parametric methods are very useful to fit the mortality distribution, or simply the statistical distribution of T. The most common distributions are the Exponential, the Weibull, the Lognormal and the Gamma distributions. All these statistical distributions are supported on the interval $[0, \infty)$, and have non-symmetric density functions.

In this particular case, we just fitted the three first distributions mentioned above. We obtained better results with the Weibull case than for the other distributions. We assume T is distributed as a Weibull (λ, ρ) with probability density function.

$$f(t) = \lambda\rho t^{\lambda-1} \exp(-\rho t^{\lambda}), \qquad \text{where } \rho > 0 \text{ and } \lambda > 0. \tag{3}$$

We call ρ and λ scale and shape, respectively. More properties of the Weibull distribution are shown in [20]. The corresponding survival function is $S(t) = \exp(-\rho t^{\lambda})$. From this expression, it is possible to estimate the Weibull parameters using least squares and following the linear relationship:

$$log(-log(S(t))) = log(\rho) + \lambda\, log(t). \tag{4}$$

More details about Survival analysis theory can be found in Cox and Oakes (1984) [2].

3.2 Value at Risk

In some situations we are interested in quantify the maximum possible losses. In finance, a useful risk measure is called VaR (Value al risk) [4]. Some firms use VaR in order to measure and control the level of risk that can absorb. Value at risk is defined as the "Maximum loss that will not be overcome with some probability already fixed". See [4] and [6] for further information.

Let us call α a confidence level (α in the range from 90% to 100%). VaR is the value *v so* that the probability that loss *L* is higher than *v* is less than (1 - α). It is possible to find the VaR when the distribution of a random variable and its parameters are already know, that it is because the VaR is the quantile of the distribution function.

$$\mathrm{VaR}_{\alpha} = \inf\{\, v \in \mathrm{R} : \mathrm{P}(L > v) \leq (1 - \alpha)\} \tag{5}$$

In our specific case, we are going to use the Value at risk as a measure of longevity risk. We assume that insured individuals who live longer than the corresponding Value at risk exceed the age when most individuals have already died. We will compare VaR for different scenarios of men and women mixture, as in insurance portfolios there is no guarantee that the proportion of men and women is exactly equilibrated as it is in the general population.

4 Results

With the crude mortality rate $m_w(x, t)$ weighted by the proportion of females in the population we proceed to calculate the empirical survival function for each year *t* and for the weighted composition of females. Once we have the empirical survival function we find the Weibull survival fit and the VaR calculation. We used the `demography` package in R version 2.13.

In the first place, we adjust a parametric model following the approach explained in the methodology section. We assumed that the survival behavior follows a Weibull distribution. By using the linear relation described in equation (4), we obtained the parameter estimates, where ρ is the scale parameter and λ is the shape parameter. Table (1) shows their values.

The shape parameter λ describes the failure rate form. Indeed if λ is less than one, this indicates that the mortality rate decreases over time. If λ is greater than one indicates that the mortality rate increases over time. Finally, when λ is equal to one, this indicates a constant behavior of the mortality rate.

In our case, λ is greater than one for all the data collection years. Besides, it is possible to identify a slightly increase of the shape parameter when female proportion is larger than male proportion, which is what is expected.

Table 1. Parameter estimates for the Weibull distribution. $log(-log(S(t))) = log(\rho) + \lambda\, log(t)$

	1990		1995		2000		2005		2009	
	ρ	λ	ρ	λ	ρ	λ	ρ	λ	ρ	λ
Male	18.09	1.62	18.24	1.66	19.59	1.61	19.14	1.67	19.69	1.65
25% F	18.52	1.64	18.55	1.68	20.05	1.62	19.59	1.69	20.22	1.67
50% F	18.98	1.66	18.88	1.71	20.54	1.64	20.06	1.71	20.78	1.69
75% F	19.46	1.68	19.22	1.73	21.04	1.67	20.56	1.73	21.37	1.71
Female	19.97	1.71	19.57	1.76	21.58	1.69	21.09	1.75	22.00	1.74

Afterwards we calculated the life expectancy (Table 2) and the VaR Since we had the Weibull parameters we could directly calculated the quantile given that the confidence level α is equal to 90% and 95%. Results are shown in Table 3 and Table 4.

Life expectancy from 65 years old has grown in the last two decades. The difference between the life expectancy from 1990 and 2009 is 1.42 years for females and 1.22 years for males. Comparing life expectancy from women and men older than 65, the main difference arise in 2009 when women are expected to live 18.51 years and men 16.85 years after age 65. It is easy to see that women live longer than men.

Table 2. Life expectancy from 65 years in Mexico

	1990	1995	2000	2005	2009
Male	15.63	15.79	16.75	16.47	16.85
25% F	15.98	16.06	17.09	16.82	17.24
50% F	16.33	16.34	17.44	17.19	17.65
75% F	16.70	16.62	17.81	17.56	18.07
Female	17.09	16.91	18.19	17.95	18.51

Table 3. Value at risk confidence level 90%

	1990	1995	2000	2005	2009
Male	30.23	30.14	32.93	31.50	32.66
25% F	30.78	30.46	33.51	32.07	33.36
50% F	31.36	30.79	34.11	32.67	34.08
75% F	31.95	31.12	34.72	33.29	34.81
Female	32.56	31.46	35.35	33.92	35.57

Table 4. Value at risk confidence level 95%

	1990	1995	2000	2005	2009
Male	35.55	35.32	38.80	36.86	38.32
25% F	36.14	35.62	39.41	37.47	39.07
50% F	36.74	35.92	40.03	38.11	39.83
75% F	37.36	36.23	40.67	38.75	40.61
Female	37.99	36.53	41.31	39.41	41.40

Considering a confidence level equal to 90%, for example in year 2009, VaR means that 10% women are expected to live 35.57 beyond 65 years, whereas 10% men are expected to live 32.66 years beyond 65 years. Table 3 confirms the fact that women live longer than men. The VaR grows as the female proportion does. This situation is observed for all the years of study and it is even more remarkable for 2005 and 2009.

Taking a confidence level more strict than before, like 95%, we can say that in 2009 five in one hundred women live 41.4 years more after 65 years and, correspondingly, five in one hundred men live more than 38.32 years beyond 65 years.

The longevity risk is larger when the female proportion increases compared to the entirely male population. An insurance company should be taking into account this fact in order to prevent future losses. All our analyses were done with the whole population, but it is well known that insurance policyholders have more concern about prevention than non-insured people. Consequently, an insured person usually lives longer than a non-insured having similar age and gender. Therefore, we should note that our results could be underestimating the value-at –risk if we were examining an insurance portfolio instead of the general population.

5 Conclusions

The main recommendation is to completely untie risk management from insurance pricing. Sex differentiation is not banned for internal risk management and, therefore, risk measurement must be made using life table separately for each sex in the insurance companies. Thus, the widespread use of unisex mortality table, constructed as a linear combination between the mortality table male and female, can be useful for pricing but should not replace the tables and projections made by separating the two gender groups.

Insurance underwriting based on alternative risk factors is pointed out as the solution to traditional gender and date of birth, which are the classic starting point when buying a life insurance policy. This involves the construction of systematic questionnaires, thus reflecting the risk of the insurance contract. The systematic application of these data collection systems can lead to operating expenses that could increase the cost of insurance provision.

Therefore, after the implementation of the Directive to insurance in December 2012, the danger is that the first reaction of insurers is raising prices until they have enough experience to know the statistical and actual consequences faced by the use of unisex price rates. As already mentioned, the request for detailed information about the habits of the insured can be inaccurate and easily manipulated. Moreover, some questionnaires can omit relevant information. Factors that are not easily verifiable by the insurer a priori should be excluded from the tariff. In [13] the authors explain implication of different factors in the mortality.

It is often argued that eliminating risk factors in pricing can cause adverse selection. This means that low risk people, when they see that prices increase decide to cancel their policy and, therefore the insurance portfolio suffers an imbalance, because only high-risk customers renew their policies. For example in the case of life insurance, men now benefit from lower prices and that would be attractive to new men customers, while women who see premium increases compared to the current premium paid, may decide not to buy insurance. In the limit, the insurance company may only have men customers and should adjust up the price again.

Some studies indicate that disregarding certain risk factors can increase loss coverage, i.e. the aggregate amount of losses covered. Detailed examples and counter examples can be found in Ornelas et al (2013) [16].

Unisex premiums should reflect the composition of the portfolio, as the percentage of men and women who are insured. Following Guillén (2012) [8] the insured databases require constant evaluation of its quality and monitoring. The actuarial modeling is necessary even when regulatory circumstances impose rules on the use of risk factors such as gender.

We have analyzed the use of unisex tables in the insurance market; mainly we focused in the differences between sexes for elderly ages. It is also interesting to evaluate the risk in the complete age range, from 0 to 100 or more. However, it is difficult to find a parametric distribution that fits the whole age range. For that reason, we are exploring some alternatives, such as nonparametric models.

Acknowledgments. Guillén thanks ICREA Academia. All authors thank support received from The Spanish Ministry and FEDER grants: ECO2012-35584 and ECO2010-21787-C03-01.

References

1. Brockett, P.L., Cox, S.H.: Statistical adjustment of mortality tables to reflect known information. Transactions of the Society of Actuaries 36, 63–75 (1984)
2. Cox, D.R., Oakes, D.: Analysis of Survival Data. Monographs on Statistics and Applied Probability 21 (1984)
3. Curry, C., O'Connell, A.: An analysis of unisex annuity rates. EOC working paper. Pensions Policy Institute 16 (2004), http://www.eoc.org.uk/research
4. Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R.: Actuarial theory for dependent risks: measures, orders and models. John Wiley & Sons, England (2005)

5. Engelman, M., Canudas-Romo, V., Agree, E.M.: The implications of increased survivorship for mortality variation in aging populations. Population and Development Review 36(3), 511–539 (2010)
6. Fler, B., Höfler, B.: Risk Measures - Value at Risk and Beyond. GRIN Verlag, Germany (2008)
7. Fledelius, P., Guillén, M., Nielsen, J.P., Petersen, K.S.: A comparative study of parametric and non-parametric estimators of old-age mortality in Sweden. Journal of Actuarial Practice 11, 101–126 (2004)
8. Guillén, M.: Sexless and beautiful data: from quantity to quality. Annals of Actuarial Science 6(2), 231–234 (2012)
9. Guillén, M., Nielsen, J.P., Pérez-Marín, A.M.: Multiplicative hazard models for studying the evolution of mortality. The Annals of Actuarial Science 1(1), 165–177 (2006)
10. Guy Thomas, R.: Some Novel Perspectives on Risk Classification. The Geneva Papers on Risk and Insurance-Issues and Practice 32(1), 105–132 (2007)
11. Hudson, R.: Mortality projections and unisex pricing of annuities in the UK. Journal of Financial Regulation and Compliance 15(2), 166–179 (2007)
12. Kalben, B.B.: Why men die younger: causes of mortality differences by sex. North American Actuarial Journal 4(4), 83–116 (2000)
13. Kuh, D., Hardy, R., Hotopf, M., Lawlor, D.A., Maughan, B., Westendorp, R., Cooper, R., Black, S., Mishra, G.D.: A review of lifetime risk factors for mortality. British Actuarial Journal 15, 17–64 (2009)
14. McDowell, M.: An examination of the implications of implementing the principle of equal treatment between men and women in the access to and supply of goods and services in the insurance and related financial services. Report of Working Group on Insurance-related issues in the Gender Non-Employment Directive (2004/113/EC) (2006)
15. O'Connell, A.: Longevity Bulletin. Institute and Faculty of Actuaries 3 (2012)
16. Ornelas, A., Guillén, M., Romero, M.J.: La igualdad una exigencia para el seguro: problemas derivados de las tablas unisex. Fundación de Estudios Financieros (2013)
17. Oxera report: The use of gender in insurance pricing (Analysing the impact of a potential ban on the use of gender as a rating factor). Association of British Insurers. Research Paper. 24 (2010)
18. Vaupel, J.W., Zhang, Z., van Raalt, A.A.: Life expectancy and disparity: an international comparison of life table data. BMJ Open 1, 1 (2011)
19. Rinne, H.: The Weibull Distribution: A Handbook. CRC Press Inc. (2009)
20. Warth, J.: Productos de Riesgos Preferentes en un Entorno Unisex. Genre: Risk Insights 15(3) (2011)
21. National Institute of Statistics, Geography and Informatics (INEGI), `http://www.inegi.org.mx/`
22. National Population Council (CONAPO), `http://www.conapo.gob.mx/`

Stock Market Simulation: Heavy Tails through Normal Perturbation*

Esteban Otto Thomasz[1] and Maria Teresa Casparri[2]

[1] Faculty of Economic Sciences, University of Buenos Aires, Córdoba 2122
Ciudad Autónoma de Buenos Aires, ZIP: 1120 AAQ, República Argentina
[2] Centro de Investigación en Métodos Cuantitativos aplicados a la Economía y la Gestión,
Córdoba 2122, Ciudad Autónoma de Buenos Aires,
ZIP: 1120 AAQ, República Argentina

Abstract. The aim of this paper is to present a simple simulation model which can generate price paths with heavy tails returns. The model consists of two agents represented by two excess demand functions and a normal stochastic perturbation. Depending of the value of the parameters, the model can generate a wide range of simulations, from a pure stochastic with normal distribution to a heavy tail process. The main achievement is the simplicity of the functional form and the parameter setting as to change simulations. In that sense it can be used as a complement to Monte Carlo simulation.

1 Introduction

The subprime crisis showed that traditional risk measurement techniques, based on past behavior of the time series, are irrelevant when amplification mechanisms rule the market. On the other hand provides great evidence and an opportunity to improve risk estimators and risk modeling.

Thus, the objective of this work is to provide a simple simulation model that emulates the behavior of the world market index before and during the subprime crisis.

In the first section we present the model. In the second section we present some simulations generated by the model. Then we compare the simulation results with other theoretical models and the empirical series. Finally we present some concluding remarks.

2 The Model

The model follows a simple agent based model (Borrill, 2010) specification. The approach is to generate a simulated trajectory from the interaction of two rules of behavior, following the generative methodology (Epstein, 2007).

---

* This work is supported by PICT 2011 N° 0919 of the National Agency of Scientific and Technological Promotion (*Agencia Nacional de Promoción Científica y Tecnológica*).

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 108–114, 2013.

The model specification has elements that can be found in Day & Huang (1990) NS Weihong & Day (1993). There are two types of agents: passive investors (De Long, 1990) and chartists. These agents are represented by an excess demand function. The price moves according to the aggregate excess demand function of the market. Passive investors demand follows one simple deterministic rule: it depends on the deviation of the observed price from the theoretical fundamental value. Chartist demand is based on the extrapolation of the past: they buy or sell according to the difference between the observed price and the average of the last three periods. In addition, to prevent the model from being totally deterministic, a stochastic perturbation is added to the aggregate excess demand function of the market. In this case, we use a random variable normally distributed with mean 0 and variance equal to 1.

The model equations are presented below:

Passive investors demand:

$$D_{fd}(P_t) = \propto (P^* - P_t) f(P_t)$$

Where $\propto$ is the strength of passive investors demand, $P^*$is the fundamental value and $f(p)$ is a loss function. One traditional specification of the loss function is the following (Fernandez Diaz, 1995):

$$f(P_t) = \left((P_s - P_t)(P_t - P_i)\right)^{-1}$$

Where Ps is the minimum price and Pt the maximum price within passive investors participate in the market.

In this case, we take the same values of Ps and Pt as in Fernandez Diaz (1999). Taking 0 and 2 as maximum and minimum value, the function is reduced to the following expression:

$$D_{fd}(P_t) = \propto (P^* - P_t)\left(\frac{1}{2P_t - P_t^2}\right) \quad with \quad 0 < P_t < 2$$

Chartists demand:

$$D_{cd}(P_t) = \beta(P_t - \bar{P})$$

$$\bar{P} = \frac{P_{t-1} + P_{t-2} + P_{t-3}}{3}$$

Price movement equation: $P_{t+1} = P_t + \gamma D_T(P_t) + \varepsilon_t$

with $\varepsilon_t \sim N(0,1)$

3 Simulations

In this section some simulations are presented. To do this, we begin by increasing the influence of the passive investors' demand, while the chartists are fixed. In all cases the simulations were compute with MS-Excel with an horizon of 4.078 periods.

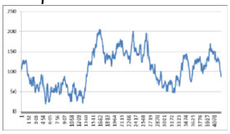
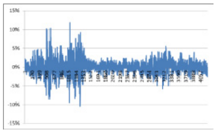
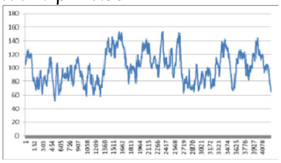
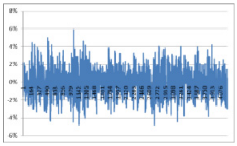
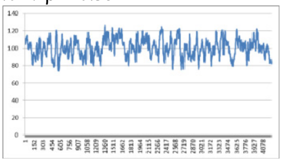
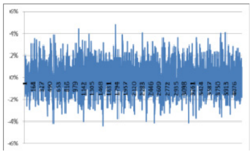
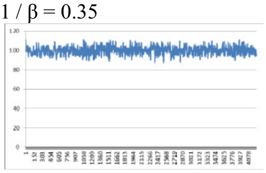
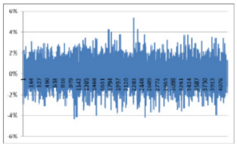
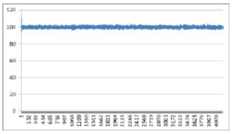
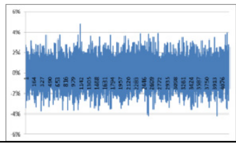
Price	Returns
$\alpha = 0.0001$ / $\beta = 0.35$ 	
$\alpha = 0.001$ / $\beta = 0.35$ 	
$\alpha = 0.01$ / $\beta = 0.35$ 	
$\alpha = 0.1$ / $\beta = 0.35$ 	
$\alpha = 1$ / $\beta = 0.35$ 	

Fig. 1. Increase in the passive investors influence

It can be observed that with the increase of the passive investors influence, the model converges to the fundamental value. In that extreme case, variations are only explained by the random variable.

In the following charts another set of simulations are presented. In this case, we increase the influence of the chartists leaving fixed the influence of passive investors.

In this case, the increase in the chartists influence generates deeper volatility in the process. In the last case, the trajectory generates a limit cycle.

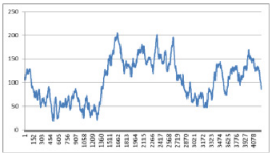
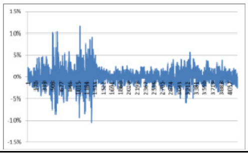
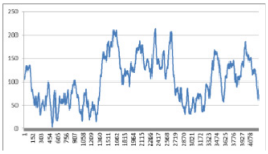
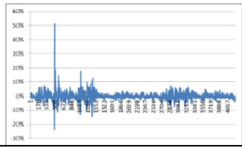
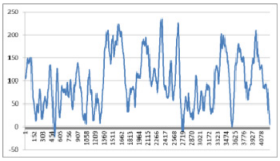
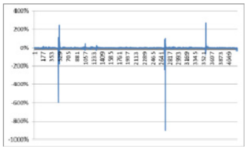
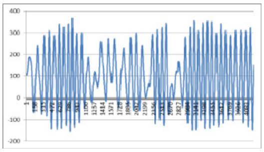
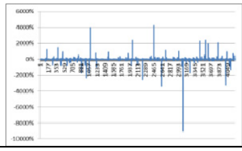
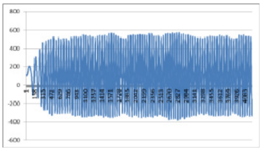
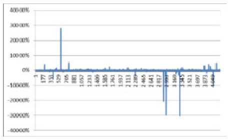
Price	Returns
β = 0.35 / α = 0.0001 	
β = 0.4 / α = 0.0001 	
β = 0.45 / α = 0.0001 	
β = 0.5 / α = 0.0001 	
β = 0.52 / α = 0.0001 	

Fig. 2. Increase in the chartists influence

4 Results

Results on both sets of simulations shows extreme values of the parameters perform two kinds of results: convergence to the fundamental value in one case and limit cycle in the other.

However, the most interesting trajectories are those generated below extreme values of the control parameters, where some similarities to empirical series can be seen.

This is mainly generated by the interaction of two different rules of behavior, which is generally accepted as a necessary configuration to perform bubbles and crashes (Vogel, 2010).

In this sense, Karino and Kawagoe (2009) conclude that for the simulation of price bubbles is necessary for agents to follow the prospective theory – within the framework of a Ball and Holt (1998) model. If only the expected utility theory is applied the price will always converge to the fundamental value.

Beside the two different rules of behavior, the other important element of the model is the random variable. It must be clear that without this stochastic perturbation the model does not generate any interesting behavior. It will be seen that these normally distributed random perturbations can generate clustering and fat tails because of model structure.

From this perspective, even when information shocks follow a normal distribution, the market process can change the Gaussian form of fundamental changes into a heavy tail distribution in returns (Lux, 1998).

5 Model Selection

In this section some theoretical models are presented, including one simulation of the model presented, as to compare them with the behavior of the empirical series.

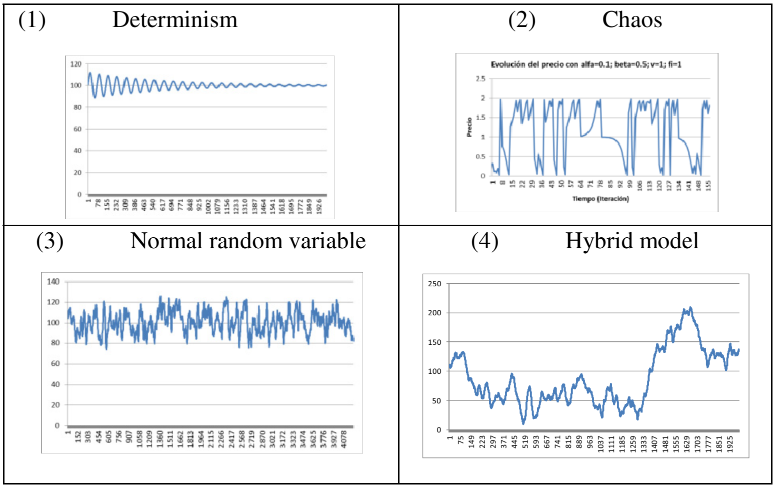


Fig. 3. Theoretical models and empirical series

The first model is purely deterministic and performs a case of purely endogenous and convergent dynamics. The second trajectory represents a path from a chaotic model. The third trajectory represents a random variable with normal distribution. The fourth trajectory was generated by the hybrid model presented in this work. Last, the fifth trajectory corresponds to the empirical evolution of the world market. Clearly, the simulation of the model presented in this work (4) looks closer to empirical series, far beyond the other theoretical models. In the next group of graphs we present the basic statistical properties of each simulation.

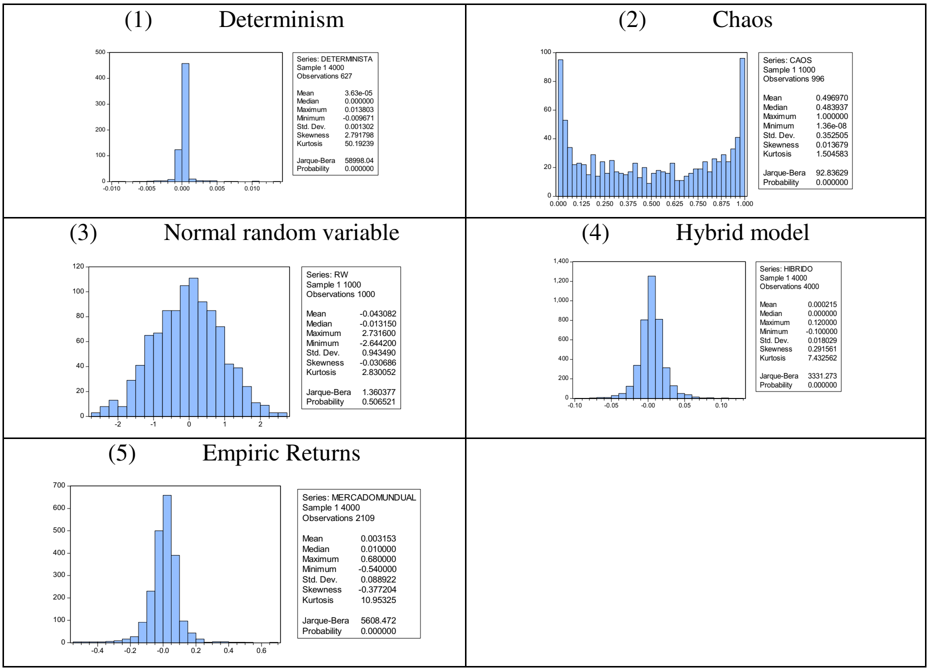


Fig. 4. Distribution of theoretical models and empirical series

Analyzing the statistical properties, it is clear that the hybrid approach generates a series with fat tails, with some similarities to the empirical distribution. In this case, figure (5) perform the daily returns of the world market in American dollars. The world market is an equation of the following markets: U.S.A., Germany, France, England, Italy, Canada, South Korea, Japan, Indonesia, Russia, India, Ireland, Spain, Greece, Brazil, Peru, México, Colombia, Venezuela, Chile and Argentina weighted by GDP. The set represents over the 70% of total world stock market capitalization.

6 Summary

We have presented an agent based model with two results: internal and empirical consistency:

i. Internal consistency refers to the correspondence between changes in the simulation due to changes in control parameters.
ii. Empirical consistency refers to the fact that under certain values of the control parameters the model generates a trajectory with some statistical properties as the empirical one.

One important property of the model is the hybrid approach (Ellner, Nychka and Gallart, 1992). In this sense, the model is neither purely deterministic nor purely stochastic. Moreover, the heavy tails were produce by a mild perturbation process in the sense of Mandelbrot (2008), avoiding the need to use other theoretical distributions different of the normal one.

The simplicity of the model –it consists of two simple rules of behavior and one stochastic perturbation normally distributed- converts it into a friendly tool to provide simulations of stock prices under different scenarios (deeper or lower propagation market mechanism) that can be used to complement other techniques, such as the traditional *Monte Carlo* approach.

A future area of research is to provide a stronger microfundamentation of the rules of behavior of the model.

References

1. Ball, S., Holt, C.: Classroom Games: Speculation and Bubbles in an Asset Market. Journal of Economic Perspectives 12(1), 207–218 (1998)
2. BorrillP., TesfatsionL., Agent-BasedModeling: The Right Mathematics for the Social Sciences? Working Paper No. 10023, Iowa State University (2011)
3. Day, R., Huang, W.: Bulls, Bears and Market Sheep. Journal of Economic Behavior and Organization 14, 299–329 (1990)
4. De Long, J., Shleifer, A., Summers, L., Waldmann, R.: Positive Feedback Investment Strategies and Destabilizing Rational Speculation. The Journal of Finance 45(2), 379–395 (1990)
5. Ellner, Nychka, Gallant: LENNS, a program to estimate the dominant Lyapunov exponent of noisy nonlinear systems from time series data. Institute of Statistics Mimeo Series N 2235 (1992)
6. Epstein, J.: Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton University Press (2007)
7. Fernández Díaz, A.: Dinámica Caótica en Economía. McGraw Hill, Madrid (1999)
8. Karino, Kawagoe: Bubble and Crash in the Artificial Financial Market. Lecture notes in economics and mathematical systems, vol. 631. Springer (2009)
9. Lux, T.: The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. Journal of Economic Behavior & Organization 33, 143–165 (1998)
10. Ritter, J.: Behavioral Finance. Pacific-Basin Finance Journal 11(4), 429–437 (2003)
11. Vogel, H.: Financial Market Bubbles and Crashes. Cambridge University Press (2010)
12. Weihong, H., Day, R.: Chaotically Switching Bear and Bull Markets: The Derivation of Stock Price Distributions from Behavioural Rules. Evolutionary Dynamics and Nonlinear Economics, 169–182 (1993)

Simulation-Optimization Methods in Vehicle Routing Problems: A Literature Review and an Example

Angel A. Juan[1], Javier Faulin[2], Elena Pérez-Bernabeu[3], and Oscar Domínguez[4]

[1] Computer Science Dept., IN3-Open University of Catalonia, c/Roc Boronat, 117, 08018 Barcelona, Spain
ajuanp@uoc.edu
[2] Dept. of Statistics and OR, Public University of Navarre, Los Magnolios Blg. Campus Arrosadia, 31006 Pamplona, Spain
javier.faulin@unavarra.es
[3] Dept. of Applied Statistics, OR and Quality, Universitat Politècnica de València, Plaza Ferrándiz y Carbonell, s/n 03801 Alcoi, Spain
elenapb@eio.upv.es
[4] SIANI Institute, University of Las Palmas, Las Palmas de Gran Canaria, Spain
oscar@opein.com

Abstract. One of the application fields of the simulation tools is Transportation and Logistics, but inside the decision making process in Transportation, the use of simulation tools to calculate optimal routes for Vehicle Routing Problems (VRPs) has been scarce and infrequent. We analyze the role of simulation in the calculation of routes for delivery problems in applied vehicle routing decisions along with a description of the performance of simulation algorithms to solve the Stochastic Vehicle Routing Problem-SVRP. In the last part of the paper, a biased-randomized algorithm for solving the capacitated vehicle routing problem with two-dimensional loading (2LCVRP) is introduced, dealing with the combination of two important aspects: vehicle routing and vehicle packing. Some experimental results contribute to validate our approach as a promising one, both in terms of the quality of the solutions as in terms of the computational time needed to obtain them.

Keywords: VRP, CVRP, simulation, optimization, 2LCVRP.

1 Introduction

Simulation is a very well-known mathematical tool applied in many research areas and used to solve many real problems in industry and services. One of those application areas for Simulation is Transportation, Logistics and Supply Chain Management. There have been many research reports and papers which have used simulation to solve real problems in the Transportation field [22], Logistics field [17], or Supply Chain Management field [41]. A classical Operations Research problem inside the Transportation and Logistics arena is the Vehicle Routing Problem which has been studied to solve pickup and delivery problems since its first mention in the Dantzig and Ramser's paper [9]. Nevertheless, the use of simulation in the resolution

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 115–124, 2013.

of routing problems has been scarce and with limited results. The purpose of the current paper is making a short description of the state-of-the-art of the use of simulation in the routing problems field, focusing on the Stochastic Vehicle Routing Problem (SVRP) because of its random nature. We also propose an example of how simulation-based algorithms employing random variants can be used for solving the capacitated vehicle routing problem with two-dimensional loading (2LCVRP). This methodology is based on a real case implemented at Opein (http://www.opein.es/index.php/en/), a company for machinery rental in Spain.

1.1 The Vehicle Routing Problem

The Capacitated Vehicle Routing Problem (CVRP) is the most popular routing problem in the Combinatorial Optimization arena. Vehicle Routing Problems (VRPs) constitute a relevant topic for current researchers and practitioners. The amount of related articles published in refereed journals has grown exponentially in the last 50 years [10]. Known as NP-hard problem [34], it is still attracting attention worldwide due to its potential applications, both to real-life scenarios and also to the development of new meta-heuristics for solving other combinatorial problems [21], [40]. The basic goal is to find an optimal set of routes for a fleet of homogeneous vehicles so that a set of customers' demands are satisfied. All routes begin and end at one or several depots, where all resources are initially located. CVRP constraints can be described as follows [40]: (i) Each non-depot node is supplied by a single vehicle; (ii) all vehicles begin and end their routes at the depot (node 0); (iii) a vehicle cannot stop twice at the same non-depot node; and (iv) no vehicle can be loaded exceeding its maximum capacity. The characteristics of the CVRP are shown in Table 1. The objective function of this problem is defined by the sum of the (usually Euclidean) distances or costs between the nodes visited by each vehicle, taking all the vehicles into account.

Table 1. Characteristics of CVRP

Characteristics of VRP	Options for the CVRP	Characteristics of VRP	Options for the CVRP
Size of fleet	Multiple vehicles with capacity constraints	Kind of network	Non-oriented. Symmetrical in distances
Type of fleet	Homogeneous	Max. time on route	No constraint
Origin of vehicles	Single depot	Activities	Only deliveries
Kind of demand	Known deterministic demand	Constraints	(i), (ii), (iii) and (iv)
Location of demand	At each node	Objective	Minimize distances (costs)

1.2 The Stochastic Vehicle Routing Problem

The Stochastic Vehicle Routing Problem-SVRP is a variant of the VRP, which can be described as a CVRP in which one or more of the problem parameters are described as random variables (Table 2). One of the most interesting classifications of the SVRP is done by Cordeau *et al.* [8]. This routing problems seem the most suitable to be analyzed by simulation because of its stochastic nature, as these are the model of real applications such as the money collection in bank branches [30], among others.

Table 2. Types of Stochastic Vehicle Routing Problems

Acronyms	Stochastic parameter	Acronyms	Stochastic parameter
VRPSC	Customers	VRPSCD	Customers and Demands
VRPSD	Demands	VRPSTT	Travel Times

Among the different types of SVRP problems presented in Table 2, the most popular one is the VRPSD, introduced by Tillman [39], who built a heuristic approach using the saving concept developed by Clarke and Wright [6],[7],. The classical goal here consists of determining the optimal set of routes that minimizes the tangible costs subject to the following constraints: (i) all routes begin and end at the depot; (ii) each vehicle has a maximum load capacity, which is considered to be the same for all vehicles; (iii) all (stochastic) customers' demands must be satisfied; (iv) each customer is supplied by a single vehicle; and (v) a vehicle cannot stop twice at the same customer without incurring in a penalty cost. The main difference between the Capacitated Vehicle Routing Problem (CVRP) and the VRPSD is that in the former each customer's demand is known beforehand, while in the latter the actual demand of each customer has a stochastic nature, i.e.: its statistical distribution is known beforehand, but its exact value is revealed only when the vehicle reaches the customer. The VRPSC is analyzed with a two stages procedure: during the first stage a solution is obtained providing a set of routes visiting the depot and each customer exactly once; while in the second stage the list of the absent customers is revealed and the initial solution has to be adapted skipping the absent vertices. The VRPSCD includes the stochastic nature of the two parameters (customers and demands). Good studies for this problem can be found at [3] and [19]. Finally, the VRPSTT, in which there is a penalty cost if the route duration exceeds a given deadline, was firstly studied by Laporte *et al.* [31] and other studies [29], [30].

1.3 Use of Monte Carlo Simulation in the VRP

Monte Carlo Simulation (MCS) can be defined as a set of techniques that make use of random number generation to solve certain stochastic or deterministic problems [33]. MCS has proved to be extremely useful for obtaining numerical solutions to complex problems which cannot be efficiently solved by using analytical approaches. Fernández de Córdoba *et al.* [14] applied MCS along with Clarke and Wright [6] algorithm to solve the Rural Postman Problem (RPP), and the CVRP [15].

MCS is still in an early stage of implementation in the VRP arena, but there are some recent studies Faulin et al. [13], Juan et al. [27] and Juan et al. [28].

2 More Recent Applications of Simulation in the VRP and SVRP

Usually, the type of methods employed to solve the routing problems are mostly related to different versions of heuristics and metaheuristics (tabu search [35], genetic algorithms, sweep-based algorithms [36] and others), adding constraints such as time window ones [38]. Nevertheless, there has been a continuous progress in the use of simulation in routing and we can consider in our study three different perspectives:

1. *Simulation as instances for routing problems for VRP*. The use of simulation as a way of providing new VRP instances or VRP cases to test new algorithms is well-known in routing in the last years. Thus, Bard et al. [2] use Monte Carlo simulation to replicate vehicle routing zone construction, or Archetti et al. [1] have designed a Monte Carlo simulation experiment to obtain different demands for new instances for the Split Vehicle Routing Problem. Similar cases have been developed by Juan et al. [25] and Faulin et al. [12] where the instances where generated by simulation.
2. *Design of probabilistic routing procedures for VRP*. This is the case of the Greedy Randomized Adaptive Search Procedures or GRASP techniques which uses adapted randomized algorithms to locate better solutions to the routing problems. The GRASP procedures allow a guided search in big spaces of solutions [37]. A similar technique with different specifications in its control is the ALGACEA algorithm developed by Faulin and Juan [11].
3. *Simulation as routing algorithm designer for SVRP*. Considering the traditional description of the SVRP [20], the design of algorithms for that problem using simulation has only carried out by Juan et al. [26], who defined a hybrid procedure of simulation and reliability techniques to solve some instances of the VRPSD. Nevertheless, the most popular procedures to solve SVRP are hybrid methods of probabilistic structures and metaheuristic procedures, as in Bianchi et al. [4].

3 An Example: Two-Dimensional Loading Capacitated VRP

In this example we are considering a relatively new variant of the CVRP, the so-called Two-dimensional Loading Capacitated Vehicle Routing Problem, or 2LCVRP [24]. In this variant, the customers' demands consist of lots of items which cannot be stacked because of their fragility, weight, or dimensions. Notice that the way these items are assigned to vehicles and packed into them might have a significant influence over the distribution costs. Thus, in addition to the routing issue, decision-makers might face also a packing problem. Some real-life examples of the 2LCVRP can be found, for instance, in the transportation of large items –such as kitchen appliances, furniture, or heavy machinery in edification. It is necessary to consider not only the weight of the load but also the shapes of the products so that the corresponding items are packed in the assigned vehicle without overlapping. In order to satisfy the CVRP assumptions, all items requested by a single customer have to be loaded in the same vehicle. Otherwise customers would be visited more than once by different vehicles. More variants of the 2-dimensional packing problem are described in [42]. As a generalization of the CVRP, the 2LCVRP is NP-complete and, in fact, extremely difficult to solve in practice. Even when exact approaches based on branch-and-bound techniques are able to solve some instances with up to 100 items, there are much smaller instances which remain unsolved as seen in [23] and [24].

A natural way to classify 2LCVRP instances is based on their packing constraints [16]. Some authors consider a truck loading policy based on the order in which customers will be served, i.e. a Last-In-First-Out loading policy ([18], [24]), while others do not consider these loading policies at all ([18]). One of the main goals of this example is to provide a parameter-less yet efficient algorithm to support decision-making in the context of the 2LCVRP, introducing a multi-start algorithm which makes use of biased randomization of classical heuristics. In particular, for the routing part of the algorithm, it integrates a biased randomization of CWS heuristic [6]. The randomized version makes use of a pseudo-geometric distribution [28], and it also uses some memory-based strategies [27]. During the solution-construction process, the algorithm verifies two feasibility conditions: (a) that the total items weight is lower than the vehicle maximum capacity (vehicle-capacity constraint); and (b) that the items can be loaded into the assigned vehicle without overlapping (packing constraint). For (b), the algorithm employs an adapted version of the Best Fit packing heuristic [5]. The algorithm is also encapsulated into a multi-start process, which allows it to escape from local minima.

3.1 Problem Description

Consider a set of $n+1$ nodes representing the central depot (node 0) and the n customers to be supplied (nodes 1 to n) and E is the set of edges $e = (i, j,)$ connecting nodes i and j, while c_e represents the traveling cost associated with the edge connecting both nodes. All the demands must be served by a fleet of K identical vehicles. These vehicles are initially located at the depot, and each vehicle has a maximum weight (i.e. capacity) and surface-dimensions constraints, say W and H for width and height, respectively. Given these assumptions, the main goal is to find a minimal-cost feasible solution. Then, this solution is given as a set of routes that minimize the total distribution costs while satisfying all customers' demands, i.e. all vehicle load capacity constraints in terms of weight, width, and height. A mathematical model for this problem was developed in [24].

3.2 Methodology

While other approaches use a two-stage schema –one for solving the vehicle packing problem and then another for solving the vehicle routing problem–, our approach integrates both decisions into one single constructive stage. The proposed algorithm will construct routing solutions implicitly taking into account the loading constraints. The CWS heuristic [6] proposes an initial dummy solution with as many routes as customers. Then, it follows a constructive process in which routes are iteratively merged, as far as the capacity constraint is not violated, according to a savings-based criterion. This heuristic is a fast method to provide relatively good solutions to medium-size CVRP instances. It is possible to significantly improve the quality of the solutions by just applying a biased and parameter-less randomization process to the CWS heuristic [28], and using this previous work as a basis, we propose to include also a packing-checking condition during the route merging process. Merging routes will only be allowed if the feasibility of the emerging solution –both in terms of capacity and in terms of loading– is not compromised. We use a biased randomized version of the Best Fit heuristic [5] as an efficient packing algorithm. Also, in our approach we consider the so-called 2|UR|L problem as described in Fuellerer et al. [16], which makes use of the following realistic assumptions: (1) *unrestricted order*, in the sense that any item is allowed to be moved inside the truck while other items are being downloaded; and (2) *non-oriented loading*, meaning that items can be rotated while being loaded in order to look for a better packing solution. We use the cache technique is a fast local search process that adds 'memory' to the algorithm, allowing it to achieve a faster convergence to the pseudo-optimal solutions [27]. Also, the use of splitting (divide-and-conquer) techniques aims at reducing the original problem size and, therefore, can be also seen as another local search process. In order to avoid getting trapped into a local minimum, the entire randomized method is encapsulated into a multi-start approach.

The algorithm requires the probability distribution which will be used to perform the biased randomization processes (e.g. a geometric distribution), as well as all the instance inputs: nodes coordinates, nodes demands, vehicle maximum capacity, vehicle dimensions, and shape of each different product. An initial dummy solution, in which one different vehicle is assigned to each different customer, and an efficiency or savings list are constructed. Then, a multi-start process is started. This multi-start process is particularly useful for two reasons: (a) it allows the algorithm to escape from local minima; and (b) it makes very easy to parallelize the randomized algorithm, by just running multiple instances of it, each one using a different seed for the pseudo-random number generator, that can be executed over several threads, cores, or computers. After each iteration of the multi-start process, a new feasible solution is obtained. This new solution is generated using our biased randomized version of the CWS heuristic, which also integrates a packing algorithm to guarantee solution feasibility. We try to improve the solution being constructed in each iteration by applying a memory-based technique (fast local search), i.e. updating new routes by better routes obtained in previous iterations. Moreover, if the solution can be considered a 'promising' one, we apply another local search process called splitting. If the new solution improves the best solution so far, then the former is updated. Eventually, the algorithm will return the best solution produced in the entire multi-round process. One of the key steps in the main procedure is the generation of new random solutions, which aim at minimizing total costs while satisfying the capacity and packing constraints.

Using the initial dummy solution as starting point, a route-merging constructive process starts. In this process, new edges are randomly selected from the sorted savings list and their corresponding routes are merged if possible, i.e. if no constraint is violated, including the packing constraint. Notice that any efficient packing algorithm could be used to solve the packing problem in order to test loading constrains feasibility. As a result, at the end of each constructive process, a random feasible solution is obtained. The quality of the random solution will depend upon the way edges have been randomly selected from the list. There is where the biased probability distribution can be very helpful. In other words, if a uniform probability distribution were used instead of a biased one, the resulting solutions would not be competitive at all, since the 'common sense' inherent to the heuristic would be lost.

3.3 Numerical Experiments

The algorithm described in this paper has been implemented as a Java application. At the core of this implementation, we used the SSJ library provided in L'Ecuyer, Meliani, and Vaucher [32], and in particular the LFSR113, which develops a fast random number generator. An Intel Xeon at 2.0 GHz and 4 GB RAM was used to perform all tests, which were run on the Eclipse platform for Java over Windows 7. In order to compare the efficiency of our approach, some classical benchmark instances for the 2LCVRP were selected from the web site www.or.deis.unibo.it/research.html. This site contains detailed information on a large number of benchmark instances for different CVRP problems. We chose arbitrarily some instances, but representing small, medium and large size instances in terms of the number of nodes. Also, these instances are provided with a fleet of unlimited vehicles which have always a loading surface of 40x20 squared units, and a capacity of 8,000 cubic units. Moreover, 5 different classes were developed for each classical instance, differing only by the general shape of the products [18]. For each of the chosen benchmark instances, our algorithm was tested in all 5 classes. For each instance-class combination, 10 independent iterations (replicas) were run using a different seed for the pseudo-random number generator. Each replica was run for a maximum time of 500 seconds where both, the best solution found (BEST10) as well as the average value of the generated solutions (AVG10), were registered. These BEST10 and AVG10 values were compared with the results obtained by Fuellerer et al. [16]. The Fuellerer's Ant Colony Optimization (ACO) algorithm is run 3 hours for each instance-class combination. The web site http://prolog.univie.ac.at/research/VRPandBPP/ contains 10 different solutions (replicas) generated by Fuellerer's ACO metaheuristic to all five classes of the classical 2LCVRP instances. The comparisons between Fuellerer's ACO and our Multi-start algorithm are summarized in Table 3.

The comparison in Table 3 shows that our approach seems to perform quite well showing negative, or under 3%, average gaps with respect to Fuellerer's results. Moreover, notice that computational times were extremely different. While ACO was running for 3 hours any instance-class, our biased randomization algorithm runs 500 seconds each replica, i.e. 83 minutes each instance-class.

We observe that, for any instance in which the loading and packing issues are relaxed and the routing issues become predominant (class 1), our algorithm seems to be always very competitive. In many other instance-class combinations, our algorithm has been able to obtain negative gaps too, that is Fuellerer's objective function was often improved, but the vehicles needed are less than in our solution in many cases.

Table 3. Gaps between Fuellerer (2009) ACO and our simulation-based algorithm

Instance	Class	BEST10			AVG10			*Average*	
		Fuellerer	OBS	Gap	Fuellerer	OBS	Gap	*BEST10*	*AVG10*
E016-05m	class 1	334.96	332.30	-0.80%	334.96	332.30	-0.80%		*-1.56%*
	class 2	334.96	332.30	-0.80%	334.96	332.30	-0.80%		
	class 3	352.16	343.75	-2.39%	352.16	343.75	-2.45%	*-1.55%*	
	class 4	342.00	332.20	-2.95%	342.00	332.20	-2.95%		
	class 5	334.96	332.30	-0.80%	334.96	332.30	-0.80%		
E031-09h	class 1	610.00	601.61	-1.39%	611.22	601.61	-1.60%		*0.53%*
	class 2	610.00	640.32	4.74%	611.23	640.32	4.54%		
	class 3	610.00	607.67	-0.38%	613.05	607.67	-0.89%	*0.83%*	
	class 4	614.24	623.72	1.52%	614.73	623.72	1.44%		
	class 5	610.23	608.22	-0.33%	613.43	608.22	-0.86%		
E041-14h	class 1	861.79	856.61	-0.60%	862.37	856.61	-0.67%		*-0.40%*
	class 2	863.27	867.08	0.44%	863.68	867.08	0.39%		
	class 3	862.62	856.61	-0.70%	862.62	856.61	-0.70%	*-0.36%*	
	class 4	861.79	859.00	-0.32%	862.45	859.00	-0.40%		
	class 5	861.79	856.61	-0.60%	862.02	856.61	-0.63%		
E016-03m	class 1	278.73	278.73	0.00%	278.73	278.73	0.00%		*2.88%*
	class 2	278.73	289.23	3.63%	278.73	289.23	3.63%		
	class 3	284.23	294.19	3.39%	284.46	294.19	3.31%	*2.91%*	
	class 4	282.95	297.80	4.99%	282.95	297.80	4.99%		
	class 5	278.73	285.93	2.52%	278.91	285.93	2.46%		
Average				0.46%			0.36%		

4 Conclusions

This paper analyzes the role of simulation-based approaches in solving vehicle routing problems, either with or without stochastic characteristics. As discussed in the paper, simulation is becoming an interesting alternative to the use of other classical optimization methods. An example of application illustrates these ideas. The example shows how it is possible to use a simulation-based algorithm for solving a complex problem combining vehicle routing and vehicle packing issues. Instead of offering a sequential two-stage approach to solve these problems, we develop an integrated approach which is able to consider both routing and packing costs simultaneously to better support the decision-making process. The computational results obtained so far are promising and they also show some ways in which our approach could probably be improved. This methodology has been successfully tested on a machinery rental company.

References

1. Archetti, C., Savelsbergh, M.W.P., Speranza, M.G.: To split or not to split: That is the question. Transportation Research Part E 44, 114–123 (2008)
2. Bard, J.F., Jarrah, A.I., Zan, J.: Validating vehicle routing zone construction using Monte Carlo simulation. European Journal of Operational Research 206, 73–85 (2010)
3. Bent, R.W., Van Hentenryck, P.: Scenario-Based Planning for Partially Dynamic Vehicle Routing with Stochastic Customers. Operations Research 52, 1526–1563 (2004)
4. Bianchi, L., Dorigo, M., Gambardella, L.M., Gutjahr, W.J.: A survey on metaheuristics for stochastic combinatorial optimization. Natural Computing 8, 239–287 (2009)
5. Burke, E.K., Kendall, G., Whitwell, G.: A new placement heuristic for the orthogonal stock-cutting problem. Operations Research 52, 655–671 (2004)
6. Clarke, G., Wright, J.: Scheduling of Vehicles from a central Depot to a Number of Delivering Points. Operations Research 12, 568–581 (1964)
7. Cordeau, J.F., Gendreau, M., Laporte, G., Potvin, J.-Y., Semet, F.: A guide to vehicle routing heuristics. Journal of the Operational Research Society 53, 512–522 (2002)
8. Cordeau, F., Laporte, G., Saverlsbergh, M.W.P., Vigo, D.: Vehicle Routing. In: Barnhart, C., Laporte, G. (eds.) Transportation. Handbooks in OR&MS, vol. 14, pp. 367–428 (2007)
9. Dantzig, G.B., Ramser, J.M.: The truck dispatching problem. Management Science 6, 81–91 (1959)
10. Eksioglu, B., Volkan, A., Reisman, A.: The vehicle routing problem: A taxonomic review. Computers & Industrial Engineering 57, 1472–1483 (2009)
11. Faulin, J., Juan, A.: The ALGACEA-1 method for the capacitated vehicle routing problem. International Transactions in Operational Research 15(5), 23, 599–621 (2008)
12. Faulin, J., Gilibert, M., Juan, A., Ruiz, R., Vilajosana, X. SR-1: A Simulation-based Algorithm for the Capacitated Vehicle Routing Problem In: Proceedings of the 2008 Winter Simulation Conference. pp. 2708–2716. (2008)
13. Faulin, J., Juan, A., Jorba, J., Grasman, S., Barrios, B.: SR-2: A Hybrid Intelligent Algorithm for the Vehicle Routing Problem. In: Xhafa, F., Herrera, F., Abraham, A., Köppen, M., Benitez, J.M. (eds.) Proceedings of the 8th International Conference on Hybrid Intelligent Systems. IEEE Computer Society (2008)
14. Fernández de Córdoba, P., García Raffi, L.M., Mayado, A., Sanchis, J.M.: A Real Delivery Problem Dealt with Monte Carlo Techniques. TOP 8, 57–71 (2000)
15. Fernández de Córdoba, P., García Raffi, A., Sanchis, J.M.: A Heuristic Algorithm Based on Monte Carlo Methods for the Rural Postman Problem. Computers and Operations Research 25, 1097–1106 (1998)
16. Fuellerer, G., Doerner, K., Hartl, R., Iori, M.: Ant colony optimization for the two-dimensional loading vehicle routing problem. Computers and Operations Research 36, 655–673 (2009)
17. Gagliardi, J.P., Renaud, J., Ruiz, A.: A Simulation Model to Improve Warehouse Operations. In: Proceedings of the 2007 Winter Simulation Conference, Piscataway, New Jersey, pp. 2012–2018 (2007)
18. Gendreau, M., Iori, M., Laporte, G., Martello, S.: A tabu search heuristic for the vehicle routing problem with two-dimensional loading constraints. Networks 51, 4–18 (2008)
19. Gendreau, M., Laporte, G., Séguin, R.: A tabu search algorithm for the vehicle routing problem with stochastic demands and customers. Operations Research 43, 330–335 (1996)
20. Gendreau, M., Laporte, G., Séguin, R.: Stochastic Vehicle Routing. European Journal of Operational Research 88, 3–12 (1996)

21. Golden, B., Raghavan, S., Wasil, E. (eds.): The Vehicle Routing Problem: Latest Advances and New Challenges. Springer, New York (2008)
22. Goldsman, D., Pernet, S., Kang, K.: Simulation of Transportation Logistics. In: Proceedings of the 2002 Winter Simulation Conference, pp. 901–904 (2002)
23. Iori, M., Martello, S.: Routing problems with loading constraints. TOP 18, 4–27 (2010)
24. Iori, M., Salazar, J.J., Vigo, D.: An exact approach for the vehicle routing problem with two-dimensional loading constraints. Transportation Science 41(2), 253–264 (2007)
25. Juan, A., Faulin, J., Ruiz, R., Barrios, B., Gilibert, M., Vilajosana, X.: Using oriented random search to provide a set of alternative solutions to the capacitated vehicle routing problem. In: Chinneck, J., Kristjansson, B., Saltzman, M. (eds.) Operations Research and Cyber-Infrastructure. OR/CS Interfaces Series, vol. 47, pp. 331–346. Springer, New York (2009)
26. Juan, A., Grasman, S., Faulin, J., Riera, D., Méndez, C., Ruiz, B.: Applying Simulation and Reliability to Vehicle Routing Problem with Stochastic Demands. In: Proceedings of the XI Conference of the AIIA 2009, pp. 201–214. Reggio-Emilia, Italy (2009)
27. Juan, A., Faulin, J., Jorba, J., Riera, D., Masip, D., Barrios, B.: On the use of Monte Carlo simulation, cache and splitting techniques to improve the Clarke and Wright saving heuristics. Journal of the Operational Research Society 62, 1085–1097 (2011)
28. Juan, A., Faulin, J., Ruiz, R., Barrios, B., Caballe, S.: The SR-GCWS hybrid algorithm for solving the capacitated vehicle routing. Applied Soft Computing 10, 215–224 (2010)
29. Kenyon, A.S., Morton, D.P.: Stochastic vehicle routing problem with random travel times. Transportation Science 37, 69–82 (2003)
30. Lambert, V., Laporte, G., Louveaux, F.V.: Designing collection routes through bank branches. Computers & Operations Research 20, 783–791 (1993)
31. Laporte, G., Louveaux, F.V., Mercure, H.: The vehicle routing problem with stochastic travel times. Transportation Science 26, 161–170 (1992)
32. L'Ecuyer, P., Meliani, L., Vaucher, J.: SSJ: A Framework for Stochastic Simulation in Java. In: Proceedings of the 2002 Winter Simulation Conference, pp. 234–242. Institute of Electrical and Electronics Engineers, Inc., Piscataway (2002)
33. Law, A.: Simulation Modeling & Analysis. McGraw-Hill, New York (2007)
34. Prins, C.: A Simple and Effective Evolutionary Algorithm for the Vehicle Routing Problem. Computers and Operations Research 31, 1985–2002 (2004)
35. Rego, C.: A subpath ejection method for the vehicle routing problem. Management Science 44(10), 1447–1459 (1998)
36. Renaud, J., Boctor, F.F.: A sweep-based algorithm for the fleet size and mix vehicle routing problem. European Journal of Operational Research 140(3), 618–628 (2002)
37. Resende, M.G.C.: Metaheuristic hybridization with Greedy Randomized Adaptive Search Procedures. In: Tutorials in Operations Research. State-of-the-Art Decision-Making Tools in the Information-Intensive Age, pp. 295–319. INFORMS, Hanover (2008)
38. Solomon, M.M.: Algorithms for the vehicle-routing and scheduling problems with time window constraints. Operations Research 35(2), 254–265 (1987)
39. Tillman, F.: The multiple terminal delivery problem with probabilistic demands. Transportation Science 3, 192–204 (1969)
40. Toth, P., Vigo, D.: The Vehicle Routing Problem. SIAM Monographs on Discrete Mathematics and Applications (2002)
41. Yoshizumi, T., Okano, H.: A Simulation-Based Algorithm for Supply Chain Optimization. In: Proceedings of the 2007 Winter Simulation Conference, Piscataway, New Jersey, pp. 1924–1931 (2007)
42. Wäscher, G., Haussner, H., Schumann, H.: An improved typology of cutting and packing problems. European Journal of Operational Research 183, 1109–1130 (2007)

Technical Staff Motivation in Nigeria: A Strategic Imperative

Ann I. Ogbo[1], Christopher Chukwudi Orga[2], and Wilfred I. Ukpere[3]

[1] Department of Management, University of Nigeria Enugu Campus, Enugu Nigeria
[2] Department of Business Administration,
Enugu State University of Science and Technology, Enugu, Nigeria
[3] Department of Industrial Psychology & People Management,
Faculty of Management University of Johannesburg, South Africa
wiukpere@uj.ac.za, wilfredisiomaukpere@gmail.com

Abstract. The paper is focused on the strategies to motivating technical staff in Nigeria. Nigeria is an emerging economy with various challenges imposed by globalisation and climatic change agents. Strategies adopted and their efficacy in motivating technical staff are expected to result in better appreciating and handling of environmental/climatic changes, satisfaction and improved productivity of the staff. The study investigated the peculiar characteristics of technical staff, the effects of monetary rewards on the technical staff, the effects of 'private practice' and non-provision of work on technical staff. A sample of 110 technical staff was selected. The findings indicate that junior technical staff (artisans and foremen) preferred monetary rewards, while senior technical staff (technical officer cadre) preferred non-monetary rewards. More of the technical staff preferred provision of work and 'private practice' as motivational strategies/tools. It is concluded that a technical staff who is provided with work and who is allowed 'private practice' is more likely to be committed to his/her job and would avoid strike. It is recommended that technical staff should be given time of their own to engage in 'private practice' and should always be provided with job to avoid loss of skill.

Keywords: Emerging Economy, Monetary Reward, Private Practice, Climatic Change.

1 Introduction

Technical staff are odd bunches. [15] posits that their mentality, view of life and history are uniquely different from most other career people. [17] believes that one of the major differences is that technical people are into technology because they love it, and many have been doing it all their life, being able to get paid to do it is just a huge bonus. The above comments notwithstanding, getting technical people to link their desire to focus on projects related to the goals of the business can be a challenging task. Knowing how to properly motivate employees often does not come easy for technical staff.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 125–137, 2013.

The employee has his/her aspirations just like the firm or organization has its objectives. There may be conflict, where organizational goals and personal goals are incongruent, and therefore, makes employee motivation an uphill task. The productivity of an average worker in Nigeria has dropped [4], as most workers hardly find meaning in their jobs. Despite the incursions made by globalization and information and communication technology across the globe, the average Nigerian worker still lags behind in productivity. Researchers have argued that the Nigerian worker is inherently lazy [19]. Others have compared an average Nigerian working in Nigeria and another working abroad. The productivity tilts in favour of the Nigerian working abroad. This has been attributed to environmental factors [5]. In realization of the impact of environment on productivity of the workforce, many programmes have been developed by administrators and business managers to motivate workers. There has been piece rate system of payment, profit sharing arrangement and share or stock allotment to improve the productivity of workers.

However, a group of workers- the technical staff- seems to exhibit the same trait irrespective of the country of operation. Weinstein [17] in aggregating motivational strategies for technical staff isolated change management, creation of acceptance and excitement across the organization for funded projects, inspiring laggards as well as tech superstars. The implication is that abilities are not evenly spread and the manager should understand this and carry everybody along. Weinstein [17] finally contends that understanding what makes your people 'tick', such as encouraging downtime, considering formal motivational programmes, applying the same standards to everyone, and creating ideal leadership career paths for those that understand both their work and the company's business interests, are panacea for motivating technical staff, which leads ultimately to improved productivity. It is clear here that understanding the workers interests and the business interests is primarily necessary in the motivation this rare breed of the workforce namely, the technical staff.

The crux of entrepreneurship development plans has been the acquisition of basic skills. The technical aid corps scheme, the industrial training fund, the establishment of colleges, polytechnics and universities of technology accede to not only the importance of technical staff, but the urgency in coping with liberalization and internationalization swooping across nations for which Nigeria joined in 1986 when she accepted IMF conditionality. Motivating technical staff seems not an option, but an imperative. Today, nations develop nuclear power, launch rockets on the moon and grapple with harsh environmental and climatic changes across the globe. Only a trained and well-motivated technical staff can lead Nigeria to a higher level of progress. The study therefore has become imperative to find the best ways of motivating technical staff and to determine the pitfalls from such motivational strategies with particular reference to technical staff in Nigeria.

2 Statement of the Problem

There have been incessant strikes by workers in Nigeria especially at the state university of science and technology, Enugu. The Project Development Institute and other research institutes were also at one time or the other on strike; the argument and agitation being terms and conditions of employment. No doubt productivity cannot be at its peak with the spate of work stoppages ravaging the nation's industrial sector and institutions of learning. With frequent interruptions, the technical man, the scientists can hardly perform. The grand challenge for scientists is to focus on discoveries that reduce poverty rather than on winning prizes. Discoveries and inventions that increase wealth and reduce poverty are the 'heroes' of science and technology.

The poverty level has attained unprecedented height in the last ten years. Flood and other natural disasters remind man of climatic change oozing from man's industrial activities, which also cause disequilibrium in the ecosystem. There are intense heat, intense rainfall, melting of ice and ravaging disease, which seem intractable. The scenario points to the failure of the technical staff to either make environmentally friendly inventions or foresee the adverse effects of their inventions. The middle class has literarily collapsed owing to non-motivational efforts by industries and even universities of technology to make technical jobs attractive. The establishment of universities of technology aimed at arresting the collapse of the middle class and technical education does not seem to have yielded the desired result. There seems to be little efforts presently at the university level to lead in research on these technical issues. No doubt a hungry man cannot effectively defend his master.

Motivating technical staff to operate at their maximum level to make inventions and counteract adverse environmental effects that have resulted in poverty, diseases and deprivation of essential things of life will definitely be a right step in joining in improving the standard of living of the people. There have been incidents of collapsed buildings, unprecedented floods, dams overflowing and epileptic power supply. Motivating technical staff therefore seems not an option but an imperative; a study in that area has become necessary to help restore the middle class in Nigeria and counteract the environmental challenges. The problem of the study therefore is the increasing environmental challenges and the paucity of technical staff in Nigeria.

3 Objectives of the Study

The study had the main objective of determining the tools for motivating technical staff in Nigeria. The sub-objectives were:

1. To ascertain the peculiar characteristics of technical staff.
2. To find out the impact of monetary reward on technical staff motivation
3. To find out the effect of 'private practice' on technical staff motivation
4. To ascertain the effect of non-provision of work on the skill of technical staff.

4 Research Questions

To effectively achieve the above objectives, the following questions were asked:

1. Are there peculiar characteristics of technical staff?
2. Has monetary reward an impact of technical staff?

3. Has ‘private practice’ an effect on technical staff?
4. Does non-provision of work affect skill of technical staff?

5 Literature Review and Conceptual Framework of the Study

The directing function of management is made possible using three tools, namely leadership, communication and motivation. It is well known that no two human begins are the same; the enormity of the directing function therefore cannot be overstated. [12] asserts that aspirations differ, temperaments are not alike; they differ intellectually and physically. The job of confidence building is important and can be achieved through leadership, communication and motivation. [12] further defines motivation as the energizing force that induces or compels and maintains behaviour. Behaviour, of course, is motivated and is also goal directed. [12] opines that motivated behaviour is sustained, goal directed and results from felt deprivation (need). The urge to act or move is usually propelled. To invent a drug for HIV, to invent machine for a by-pass operation, etc., result from the scourge of AIDs and death from pains due to professional mistakes of doctors.

Motivation may be define the drive or readiness of an organization or man to work hard towards greater performance needs; the desire of an individual at a point in time is to achieve objectives, the broad intention as well as defined growth. The above indicates that both the individual and the organization can be motivated. The need however, may not always be the same. This is the reason for integration of interests in the workplace. The objective to be pursued and the growth to attain put pressure (tension) on either the organization or the individual. [13] see motivation as the conscious efforts to consider the needs and wants of workers by management geared toward eliciting compliance through the satisfaction of these needs, wants and desires. [13] further define need as a state of felt deprivation. No doubt, a satisfied worker is a fulfilled worker and is a productive worker. The need makes the worker to look for want-satisfiers. These want-satisfiers put tension on the worker until the needs are satisfied. A process of need – want-satisfaction chain exists which is repeated if the feedback makes it necessary. Why we do things cannot be observed directly; it can only be inferred from goal-directed behaviour.

Motivation can be extrinsic or intrinsic. [10] sees extrinsic motivation as the motivation to encourage an activity as a means to an end while intrinsic motivation is the motivation to be involved in an activity for its own sake. It should be noted however, that a learner can be high in both intrinsic and extrinsic motivation and high in one and low in the other. [8] informs that the type of motivation learners or workers experience depends on the context they are in. This means that motivation can change overtime. [8] also suggests that challenge, control, curiosity, fantasy, and aesthetic value are all sources of intrinsic motivation. It is obvious that people do things, such as go to school, in order to improve themselves to avoid being illiterate people. Why they improve themselves and don’t want to be illiterate people is a mystery. It is a black box and it has not been fully uncovered. The perspective on motivation could be as below:

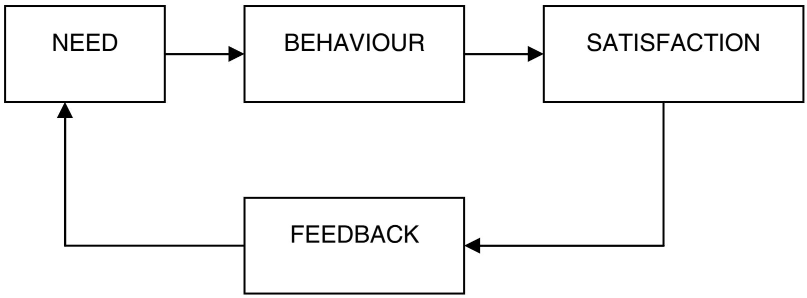


Fig. 1. Perspective on Motivation

Certain needs and wants cause one to do certain things (behaviour) which satisfy these needs, and which can change thereby determining which needs and wants are primary.

The subject of motivation is favoured with many theories and we could only analyse a few to select the suitable one for the study. Hierarchy of Needs Theory in which the needs at the bottom are the most urgent and need to be satisfied before certain attention is paid to the others. The most famous example [7] assures, is the Maslow's hierarchy of needs, which accepts that needs are hierarchical and only after satisfying the lower levels of needs that a person is free to progress to the ultimate need of self-actualization. The needs are arranged from bottom as physiological needs, safety needs, belonging and love needs, esteem and the self-actualization needs. A compact classification sees the first two as fundamental needs, the next two as psychological needs and the last one as self-actualization need. Arguments still persist as to whether sex is a basic or psychological need.

Alderfer's ERG Theory [cited in 16] classifies needs into three categories as against Maslow's five categories. Alderfer's needs theory also is hierarchical. Its bottom up approach shows needs as existence needs (physical well-being), relatedness needs (satisfactory relations with others) and growth needs (development of competence and realization of potential. This is similar to Maslow's hierarchy except that Alderfer believes that sex should not be in the bottom category. Alderfer [cited in 16] believes that as you start satisfying high needs they become more intense (e.g. the more power you get the more you want power) like an addiction. The Maslow and Alderfer theories of hierarchy of needs suggest that not everyone is motivated by the same things. It depends where you are in the hierarchy. The needs hierarchy mirrors the organizational hierarchy to a certain extent because some people are still motivated by needs much lower than their position.

Another type of need theory is the Acquired needs Theory. This was made popular by David McCelland and states that some needs are acquired as a result of life experiences. He identifies need for achievement, need for affiliation and need for power to control others. [10] outlines the McCelland theory and suggests managerial behaviour or implications to managers as below.

Table 1. McClelland's Theory

Style	More effective	Less effective
n_{arch}	Seek: to excel, may avoid both low and high risks as a result, in order to pursue meaningful success.	Work alone or with other high achievers
n_{power}	Seek: either personal or institutional power; either way they want to direct others, but the constituent power is in service to the institute's success, of those with that focus tends to make better managers.	Direct orders
n_{aff}.	Seek: harmonious work relationship to accept, to be accepted and to include others. They can be more comfortable conforming to group norms.	Work in settings with significant personal interaction.

Source: [10].

He finally summits that, you can motivate people. You can provide environment where people motivate themselves. Apply what you know about people's styles to strengthen their individual work environment. Add along the way is focus; focus in inelastic motivation factors by building strong work relationships and expend those relationships so that more is possible try and build healthy vibrant work environment and remember that work is as valuable as building any other relationship or one's life.

Cognitive Evaluation Theory is yet another theory of motivation. The theory suggests that there are actually two motivation systems: intrinsic and extrinsic which correspond to two kinds of motivations. The intrinsic motivations are achievement, responsibility and competence. They come from actual performance of the task or work. The extrinsic factors of motivating according to the theory include pay, promotion, feedback, working conditions and they come from a person's environment, controlled by others. The intrinsic factors are the higher order needs of Maslow's hierarchy.

The Two-Factor Theory is also called the motivator-hygiene factor theory. The hygiene factors are factors whose absence have neutral impact on motivation and whose presence has no perceived motivation but simply placates subordinate. The motivators are factors whose presence motivates and whose absence does not cause any particular dissatisfaction; it just fails to motivate. Again the hygiene factors are the bottom of Maslow's hierarchy while the motivators are at the top of the hierarchy. The two scales of hygiene factors (dissatisfiers) and motivators (satisfiers) are independent and you can be high on both [10].

Other theories exist and include Equity Theory which states that it is not the actual reward that motivates but the perception, and the perception is based not on the reward in isolation, but in comparison with the efforts that went into getting it and the rewards and efforts of others. The Reinforcement Theory of B. F. Skinner indicates the effects of the consequences of a particular behaviour on the future occurrences of the behaviour which are positive reinforcement, negative reinforcement, punishment and extraction. Vroom's Expectancy Theory brought quantitative analysis into motivation theory. It brought together many of the elements of the previous theories. It combines perceptual aspects of equity theory with the behavioural aspects of the other theories; it gives the equation:

$$M = E \times I \times V.$$

This shows that motivational force is related to expectancy, instrumentality and valence in a multiplicative manner.

The summary is that people tend to sustain behaviour that results in the satisfaction of their needs. The cognitive theory of motivation is adopted for the study because it best explains the innate and instinctive attributes of the technical staff.

6 Recognition and Reward in Organization

[3] states that the 21st century requires a holistic approach to recognizing and rewarding employees. It is possible to place too much emphasis on pay and other extrinsic rewards but the changing nature of the relationship between employers and employees requires a new kind of 'currency'. Effective systems of recognition and reward engage an individual's entire being. They also as, [20] put, encourage employees to unleash scores of productive energy while exhibiting regenerative qualities that foster creativity, emotional reserves that translate into passion and even spiritual attributes that result in the inspired performance needed to achieve a larger vision.

The implication of the above findings of [3] is that successful managers should have respect for both people and process and that relationship should be emphasized against command and control management. Such managers regard employees as part of their customer base, contiguously looking for ways to satisfy and retain employee commitment while ultimately inspiring them to peak performance. Peak performance of course is the desire of mangers, even not- for- profit organizations.

However, the nature of the workplace may make for the pacing of workers and low relationship may be present without jeopardizing performance. The workplace diversity which has resulted in multiracial and multicultural organizations simply suggests that one size approach to human resource will not be it at all. Also the workplace diversity suggests the presence of different cadres of workers, namely academic, administrative, technical and operative. The rewards such as carrot, stick and motivational tools as [3] puts it, have left experts with one mistake or the other. Enlightened managers sue for total reward system that links direct and indirect payments to performance requirements tied to the organization's success; such an approach seems more effective than simpler more restrictive linear systems that function on a quid pro quo: produce this and you get that syndrome.

Successful companies in USA adopted holistic systems of reward and recognition [3]. The holistic system reward and recognition usually adopted included job design, decision – making process, pay equity, performance planning and management systems, self-direction communication, leadership styles and professional development. This incorporates anything that influences employees to unleash their motivation and passion, and has resulted in the making of free agents in the workforce. The lean size of today's organizations owing to downsizing to improve profit and sometimes to sustain itself has led to workers working to improve their skills and enhance their marketability, impedances and reputation and thus justify their continued employment [9]. This simply means that these free agents require and need flexibility to move through organizational systems without being locked into one department. The technical staff hardly is this flexible because of the special nature of his function. [18] posits that smart managers understand the importance of respecting people's intelligence and telling it like it is. They work collaboratively with employees. They make conscious decisions to join force instead of subordinating or dominating. As nerve systems such as skill-based pay, total reward programmes, and management become more main stream, the challenge for management will be to avoid any suggestions that they are manipulative or disrespectful. In fact, placing too much emphasis on pay and pay system will distract for the intrinsic value of work "itself". [18] concludes by stating that skilled and technical workers are motivated more by the intrinsic value of work.

Clearly, having meaning in the work leads to non-monetary motivation. Some researchers however state that the intrinsic value becomes necessary only when the extrinsic or short term rewards have been reasonably achieved. McGregor insists that commitment to objectives is a function of the rewards associated with the achievement. [14] emphasizes that recognizing the short term nature of employment and the need to influence peak performance, organizations have generated elaborate programmes to motivate employees including informal and formal awards and providing recognition by means of factors external to the work itself. [14] further lists these awards as base-pay packages, variable pay, incentives, cash and cash equivalent benefits, gain-sharing plans, commission, stock options and alternative pay programmes. This indicates that intrinsic rewards are inherent in the nature of the work itself and context or environment in which it is performed. Such workers are intimately reenergized and satisfied either because it fulfils individual's desire to support the organization's mission or value system or their own relationship with co-workers.

[11] identifies seven ways to motivate a technical staff which include:

1. Getting the tools that are needed to do the job.
2. Making sure that they have the right skills and providing opportunities to learn.
3. Getting a viable work environment
4. Giving people as much information as they need to do their job.
5. Shielding the team from office policies, and
6. Making sure each member remembers he is part of a team
7. Meing there when needed and responding to problems and concerns.

Quite a litany of options or rewards! While all the options are necessary in every organization, the stage in the life cycle of the staff determines the degree and application. An employee may need little relationship and little supervision as he ages in the organization and to do the opposite may be counterproductive.

7 Characteristics of Technical and Professional Staff

The appreciation of the power of organizational members has led to the separation and development of the management function of direction. The existence of good plans no longer assures of undertaking of assigned tasks. The manager is not only left with the responsibility of controlling the activity that develops from plans but also tries to get the organization members to go to work willingly and enthusiastically.

This, as [6] posits, is because of such factors as the increasing educational level of employees, greater utilization of professional personnel, advancing technology and the power of labour organizations. The manager in charge of human resource in realization of the above, and in clear understanding of the needs of the employee, designs a programme to elicit the right type of behaviour from the employee. [6] further sees motivation as the skill of aligning employee and organizational interest so that behaviour results in achievement of employee wants simultaneously with attainment of organizational objectives.

The wants of employees range from pay, security of job, congenial associates, credit for work done, a meaningful job opportunity to advance, comfortable, safe and attractive working conditions, competent and fair leadership, reasonable orders and directions to a socially relevant organization. The technical staff characteristics, [11] reiterates "are clear from their tendencies or inclinations, which include: tend to be introverts, tend to think more logically than emotionally, tend to be problem solvers and tend to be technical creative".

The nature of technical work, which most often confines the worker to solitary conditions, no doubt will make them introverts. The logic involved in the execution of their job, the sequence of operation and the time-centeredness also make an average technical staff think logically and avoid emotions as much as possible. The job of many technicians lies in the actualization of the objective or work at hand. He/she therefore is a problem solver. To say that he/she is technically creative simply states the obvious. The creativity of the technical staff at both the design and implementation stages is an expected outcome. He/she is therefore a techno-determinist who believes that technological progress is inevitable and avoids technology laggards, who are slow to adopting new technology.

In summary, the technical staff is hardly singled out for rewards or motivation. However, the scanty literature reveals that rewards that are intense could motivate the technical staff. Avenues for creativity, self-expression and logical thinking seem the best for the technical staff. It is however to be realized that basic needs should be fairly satisfied before any staff, the technical staff inclusive, gives his/her best. The available literature however did to show the effect of private jobs, a method which allows the technician or technical staff to get a job, cost it and produce it without accounting to the organization.

8 Research Methodology

The survey design was adopted. The area of the study was the Enugu State University of Science and Technology, Enugu, South East, Nigeria. Both primary and secondary sources of data were adopted. The population of study was 151, consisting of all engineering technical staff of the university. A sample of 110 was adopted using the Taro Yamane formula. The stratified random sampling technique was employed in selecting subjects for the study. A semi-structured questionnaire was utilized for the study. Also personal interview and personal observations were utilized as methods for the collection of data. The reliability of the instrument was tested using cronbach's alpha coefficient, which yielded reliability co-efficient of 0.88. The content validity was tested using product moment coefficient of correlation which yielded a coefficient of 0.85, which was good. The data were presented and analysed using frequency tables and simple percentages.

9 Results and Findings

Table 2. Distribution of Respondents on the Personal Characteristics of Technical Staff

Characteristics	Staff		Total	%
	Artisans/ Foremen	Technical Officers		
Introversive	53	37	90	91
Logical thinking	51	39	97	98
Creativity	47	38	85	86
Problem solving	56	39	93	94
Fun seeking	2	0	2	2
Philanthropy	1	0	1	1
Emotional	0	0	0	0

Source: Field Survey, 2011

From table 2 above, 90 (91%) indicated that technical staff are introversive 97(98%) indicated they are capable of logical thinking, 85 (86%) indicated they have creativity, 93 (94%) indicated the personal characteristics included problem solving. However, 2 (2%) indicated that technical staff are not fun-living. None indicated that they are emotional. Only 1 (1%) indicates philanthropy as a characteristics of technical people.

Table 3. Distribution of Respondents on the Reward Preferred

Reward	Staff		Total	%
	Artisan & foremen	Technical officers		
Monetary	49	8	57	58
Non-monetary	11	31	42	42
Total	60	39	99	100

Source: Authors Fieldwork.

Table 3 shows that 57 (58%) indicated that monetary rewards were necessary, while 42 (42%) indicated that non-monetary rewards were necessary. Majority of the technical officers preferred non-monetary reward while majority of the artisans and foremen preferred monetary reward.

Table 4. Distribution of Respondents on the Provision of Work

Opinion	Staff		Total	%
	Artisans & foremenn	Technical officers		
Provision of work	51	39	90	91
Non- provision of work	9	0	9	9
Total	60	39	99	100

Source: Authors Fieldwork.

Table 4 shows that 90 (91%) of the respondents preferred provision of work, while 9 (9%) preferred non-provision of work.

Table 5. Distribution of Respondents on their Preference to Private Jobs

Opinion	Staff		Total	%
	Artisan & foremen	Technical officers		
Private job	57	34	91	92
No private job	2	5	7	7
Indifferent	1	0	1	1
Total	60	39	99	100

Source: Authors Fieldwork.

Table 5 shows that 91 (92%) preferred private jobs as an alternative reward, 7 (7 percent) preferred no private job, while 1 (1%) was indifferent.

The study revealed the following:

- That technical staff of Enugu State University of Science and Technology, Enugu, Nigeria exhibit peculiar characteristics. The characteristics included introversiveness, logical thinking, problem solving attitude, and creative tendencies.
- The artisans and foremen among the technical workforce prefer monetary reward, while the technical officers prefer non-monetary rewards.
- The technical staff members prefer provision of work.
- Majority of the workforce (technical) prefer private jobs they popularly called (PP) to other rewards.
- The staff will avoid strike, if given the chance.

10 Discussion of Findings

The findings of logical thinking, problem solving, creativity and introversiveness are in line with [11]. It was not exhaustive as people could be philanthropic and fun-seekers. However, the findings revealed that although it could all apply in other work situations the technical people have 3 out of 99 chances for philanthropy and fun-seeking. The discovery that non-monetary rewards are preferred by technical officers, while artisans and foremen (also technical people) is in line with both Abraham Maslow's hierarchy of needs and Chris Agyris's mature and immature people. The earlier the lower needs of food, shelter and security are satisfied, the more the emergence of higher ones like social and self-actualization needs. The higher technical people – technical officers – present the characters of the mature people and theory of Y of both Agyris and McGregor respectively.

Subject of recent litigations has been whether or not the employer must provide work. The study here clearly shows that the worker who is technically inclined insists on provision of work. The reasons of loss of learning effect, loss of dexterity, and in cases of piece rate system as is common among technical people make provision of work a preferred option among the technical people studied.

The latest finding, and perhaps the addition to motivational literature, is the finding on private jobs. That was variously called "private practice", 'pp', etc. An average technical staff in the area studied prefers private jobs where he collects a job, costs it, produces it, gets the money and keeps it. For employers afraid of the free agents, this reward is an option. Where 'pp' is in practice, the average technical staff hates strikes.

11 Conclusion and Recommendations

In conclusion, it could be inferred that all the technical staff in the area studied have peculiar characteristics which include introversiveness, problem-solving, logical thinking and urge for creativity. Artisans and foremen are more likely to be motivated by monetary rewards, while technical officers are more likely to be motivated by non-monetary rewards. A technical staff provided with a job, is more likely to stay and even improve on his performance than one without jobs provided. Where private practice is allowed, the staff is more likely to improve performance. Technical staff members dislike strike. From the foregoing the following recommendation are worth proposing:

1. Technical staff should be identified and employed based on possession of certain characteristics.
2. Artisans and foremen should be motivated using monetary inducements.
3. Technical officers should be motivated by non-monetary inducements.
4. To motivate technical staff, jobs should always be provided.
5. Technical staff should be allowed some elements of private practice in which the organization's facilities could be used solely for the benefit of the staff. This should however, be carefully done to avoid it being the rule rather than the exception.
6. A national, industry-wide, and university wide study should be made to further authenticate the findings in the study.

References

1. Abanobi, C.: My Encounter with the Father of the Internet. Daily Sun 6(1907) (2010)
2. Aziagba, P.C.: Staff Development as Motivation for Paraprofessionals: A case Study in Southern Nigeria (2008), http://www.google.com (retrieved October 29, 2010)
3. Bowen, R.B.: Making Recognition and Rewards a “Whole-Person” Experience. Houghton Muffin, Boston (1999)
4. Chukwu, M.I.: Managing Poverty Eradication through Public Relations. ESUT Journal of Management Sciences 1(1) (2005)
5. Fashoyin, T.: Industrial Relations in Ngieria, 2nd edn. Longman Nigeria Plc, Ikeja (1992)
6. Flippo, E.: Personnel Management, 5th edn. McGraw Hill, Arizona (1982)
7. Ile, N.N.: Comparative and International Management. Ochumba Press Limited, Enugu (2000)
8. Iwu, E.E.A.: Administrative management theory: A synoptic comparative viewpoint. Veamaks Publishers, Enugu (2000)
9. Jeffrey, P.: The Human Equation. Haward Business School Press, Cambridge (1998)
10. Jennings, S.: Employee Motivation in the Workplace: Different Types of Motivation Theories (2009), http://www.google.com (retrieved on October 29, 2010)
11. Mochal, T.: Four things you should know about your technical staff (2007), http://www.techpublic.com/blog/project-management/four-things-you-should-know-about-your-technical-staff/163 (retrieved on July 26 2012)
12. Nwachukwu, C.C.: Management Theory and Practice. Ontisha African Fep Pub. (1992)
13. Orga, C.C., Ugwu, J.N.: Management: An Introductory Analysis. Veamaks, Enugu (2006)
14. Saleni, R.: Leading After a Layoff, Avon, MA: Adams Media. Adams Media, Avon (2005)
15. Sidious, D.: Motivating technical employees (2007), http://sithsigma.wordpress.com/200711/20/motivatingtechnical-employees/
16. Robbins, S.P., Judge, T.: Organisational Behaviour. Pearson Prentice Hall, New Jersey (2009)
17. Weinstein, A.: Redefining technology market strategic insight for competitive intelligence professionals. In: Society of Competitive Intelligence Professionals Annual International Conference Proceedings, Arlington, pp. 580–581 (1996)
18. Welch, J.: Jack: Straight from the gut. Werner Book, New York (2001)
19. Yesufu, T.M.: Dynamics of Industrial Relations in Nigeria. The Nigerian Experience, Ibadan: OUP (1982)
20. Zinghein, P., Schuster, J.: Pay People Right. Jossey Bass, San Francisco (2000)

A Hybrid Algorithm for Solving a Bilevel Production-Distribution Planning Problem

Herminia I. Calvete[1,*], Carmen Galé[2], and María-José Oliveros[3]

[1] Dpto. de Métodos Estadísticos, IUMA, Universidad de Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain
[2] Dpto. de Métodos Estadísticos, IUMA, Universidad de Zaragoza, María de Luna 3, 50018 Zaragoza, Spain
[3] Dpto. de Ingeniería y Diseño de Fabricación, Universidad de Zaragoza, María de Luna 3, 50018 Zaragoza, Spain
{herminia,cgale,mjoliver}@unizar.es

Abstract. This paper addresses a hierarchical production-distribution system with two levels of decision making. At the top of the hierarchy, a distribution company faces a multi-depot vehicle routing problem, i.e. it decides on the allocation of retailers to depots and on the routes of vehicles from each depot. At the bottom of the hierarchy, a manufacturing company, after receiving the order from the distribution company, decides which manufacturing plants will produce the required items. We present a hybrid algorithm to solve this problem. This algorithm uses the genetic algorithm structure to control the allocation of retailers to depots. It determines the routes to serve retailers by using an ant colony optimization algorithm and addresses the production issue by solving a linear programming problem. The efficiency of the algorithm is shown by applying it to a realistic instance.

Keywords: Production, distribution, bilevel, ant colony, evolutionary.

1 Introduction

The need for integrating the production and distribution systems of a supply chain has been increasingly addressed in recent years [1,2]. Calvete, Galé and Oliveros [3] have proposed a bilevel optimization model to deal with a hierarchical production-distribution system which integrates the problem faced by a distribution company which aims to minimize the cost of serving items from a set of depots to its retailers and the problem faced by a manufacturing company which serves the depots. The distribution company decides on the allocation of retailers to depots and on the routes of vehicles which serve them. The distribution company is supplied by a manufacturing company which, after receiving the order from the distribution company, decides which manufacturing plants will produce the required items. In terms of the hierarchical decision process, the manufacturing company is at the top of the hierarchy and is therefore the leader while the manufacturing company is at the bottom of the hierarchy and is therefore the follower. This problem is formulated as a

* Corresponding author.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 138–144, 2013.

bilevel optimization problem and solved by using ant colony optimization. In this paper, we present a hybrid algorithm for solving the problem. This algorithm combines the ideas of evolutionary algorithms and ant colony optimization.

In order to make this paper self-contained, in section 2 we formulate the production-distribution planning problem. Section 3 goes on to develop the algorithm. Section 4 presents an initial study of the performance of the algorithm by showing the results of applying it to a realistic instance. Finally, some concluding remarks are presented in Section 5.

2 The Hierarchical Production – Distribution Planning Problem

A distribution company owns a set of depots from which it serves a set of retailers. This company decides on the depot that serves each retailer and designs the routes which serve them. Moreover, it is responsible for having sufficient quantities of product available at each depot to satisfy the demand of the retailers assigned to that depot. The distribution company aims to minimize its overall costs which include the costs of delivering the product from the depots to the retailers and the costs of acquiring the product and unloading at the depots.

The product is supplied by a manufacturing company which owns a set of manufacturing plants. After receiving an order from the distribution company, it allocates the production to its plants aiming to minimize its operation costs.

In order to formulate the model for the production-distribution planning problem, the following notations will be used:

Parameters:

K, k set of plants, plant index
L, l set of depots, depot index
R, r set of retailers, retailer index
S, s set of vehicles used, vehicle index
E set of existing arcs connecting depots and retailers
b_r demand of retailer r
R_l set of retailers served from depot l
R_s set of retailers served by vehicle s
S_l set of vehicles used at depot l
A_k production availability of plan k
c_{ij}^{11} cost of connecting i to j, $(i,j) \in E$
c_{kl}^{12} cost of providing an item manufactured in plant k to depot l
c_{kl}^{22} operation cost for manufacturing an item at plant k to depot l

Decision variables:

x_{ij}^{s} distribution company variable
this takes the value 1 if vehicle s uses arc $(i,j) \in E$ and 0 otherwise
y_{kl} manufacturing company variable
this represents the amount of items manufactured at plant k for depot l

The distribution company variables must verify the constraints of a multi-depot vehicle routing problem (MDVRP) [4,5], which is an extension of the vehicle routing problem. The constraints of the manufacturing problem reflect the production process at manufacturing plants.

$$\min_{x_{ij}^s, y_{kl}} \sum_{s \in S} \sum_{(i,j) \in E} c_{ij}^{11} x_{ij}^s + \sum_{k \in K} \sum_{l \in L} c_{kl}^{12} y_{kl}$$

s.t.

$\{x_{ij}^s\}$ verifies MDVRP constraints

$x_{ij}^s \in \{0,1\}, \ (i,j) \in E, s \in S$

where, for given $\{x_{ij}^s\}$, the variables $\{y_{kl}\}$ solve:

$$\min_{y_{kl}} \sum_{k \in K} \sum_{l \in L} c_{kl}^{22} y_{kl}$$

s.t.

$$\sum_{l \in L} y_{kl} \le A_k, \quad k \in K$$

$$\sum_{k \in K} y_{kl} \ge \sum_{s \in S_l} \sum_{r \in R_s} b_r, \quad l \in L$$

$$y_{kl} \ge 0, k \in K, l \in L$$

This is a mixed-integer bilevel programming problem [3,6]. For $\{x_{ij}^s\}, \{y_{kl}\}$ to be a feasible solution of the problem, $\{y_{kl}\}$ has to be an optimal solution of the manufacturing company problem, which involves variables $\{x_{ij}^s\}$ in its constraints. Note that $\sum_{s \in S_l} \sum_{r \in R_s} b_r$ depends on the variables x_{ij}^s, since that quantity is not known until the retailers are assigned to depots.

3 The Algorithm

The algorithm constructs feasible solutions of the bilevel problem. For this purpose, the allocation of retailers to depots is controlled by a genetic algorithm (GA). Once the retailers have been assigned, the problem of designing the routes for delivering from each depot to its retailers is solved using an ant colony optimization algorithm (ACOA). Besides, once the retailers have been assigned, the manufacturing company problem can be solved. This is a linear programming problem which is solved optimally.

To start the algorithm, an initial population of chromosomes is generated. Each chromosome is a string whose size is the number of retailers. Each allele of the chromosome contains the depot assigned to the corresponding retailer. At each iteration of the algorithm, new chromosomes are generated by using crossover and mutation operators. The size p of the population, the probability of crossover and the probability of mutation are parameters of the algorithm. The elitist strategy is used to select at each iteration the best p chromosomes. The algorithm proceeds by constructing populations until the stopping condition is met.

The fitness of a chromosome is defined by the value of the objective function: $\sum_{s\in S}\sum_{(i,j)\in E} c_{ij}^{11}\, x_{ij}^{s} + \sum_{k\in K}\sum_{l\in L} c_{kl}^{12}\, y_{kl}$. In order to compute this expression, we must compute the variables $\{x_{ij}^{s}\}$. For this purpose, we solve L vehicle route problems (one for each depot) in order to design the routes that serve the retailers by using an ACOA. The variables $\{y_{kl}\}$ are computed by solving a linear program.

4 Illustrative Instance

The instance was proposed in [3]. The distribution system consists of 69 retailers and 4 depots. The time-distance between any two of them is 525 minutes or less. The cost c_{ij}^{11} is proportional to the time-distance. A fleet of homogeneous vehicles with a capacity of 40 load units is considered. The driving time is set to 480 minutes and the working time is set to 600 minutes. For each retailer, service time is 15 minutes. Retailer demand ranges from 2 to 15 units. Total demand is 447, so at least 11 vehicles are needed to deliver. The manufacturing company owns 4 plants with a production availability of 2000 item units each. The operation cost c_{kl}^{22} ranges from 1 to 4. The cost c_{kl}^{12}ranges from 0.36 to 5.36. The algorithm was coded using Borland C and the numerical experiments have been performed on a PC Intel(R) Core (TM) i7 at 2.93GHz having 8 GB of RAM under Windows 7.

In the experiment we have studied the influence of the number N of iterations of the algorithm. With this aim in mind, we have applied the algorithm with $N = 50$, 100, 200, 300, 400, 500, 600 and 700 iterations, keeping the remaining parameters fixed throughout the experiment. For each value of N, 5 runs of the instance were done. Concerning the GA, the population size was set to 8, the crossover probability to 0.9 and the mutation probability to 0.5. Regarding the ACOA, after pilot testing and taking into account that it is applied to each new chromosome, 10 iterations each with 4 ants and local search 2-opt were applied.

Table 1. Algorithm results of the illustrative instance

N	$\bar{f}_1$	f_1^b	f_1^w	$\%f_1$	$\bar{T}$
50	4969.48	4863.24	5074.76	4.50	1.43
100	4905.25	4840.60	4962.14	3.15	2.88
200	4908.26	4767.04	5022.66	3.21	5.48
300	4825.75	4789.44	4868.24	1.48	8.20
400	4790.42	4757.84	4860.14	0.74	10.69
500	4812.37	4771.64	4863.64	1.20	13.50
600	4809.75	4755.38	4849.44	1.14	16.77
700	4796.26	4756.56	4818.64	0.86	19.62

The results of the experiments are displayed in Table 1. The first column identifies the value of the number of iterations. Columns two to four refer to the distribution company objective function value $f_1$and show the average $\overline{f_1}$, the best value f_1^b and the worst value f_1^w found among the 5 runs of the instance. The sixth column presents the percentage gap between the average and the best value of f_1 obtained throughout the experiment, $f_1^{best} = 4755.38$, i.e. $\%f_1 = \frac{\overline{f_1}-f_1^{best}}{f_1^{best}}$. Finally, the last column shows the average of the computing time in minutes.

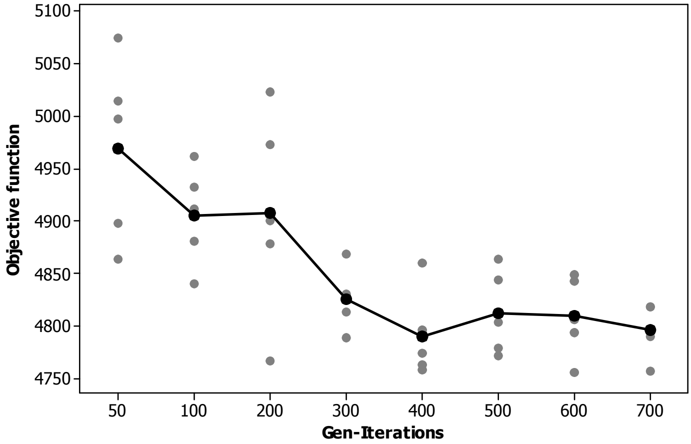


Fig. 1. Individual value plot of the distribution company objective function

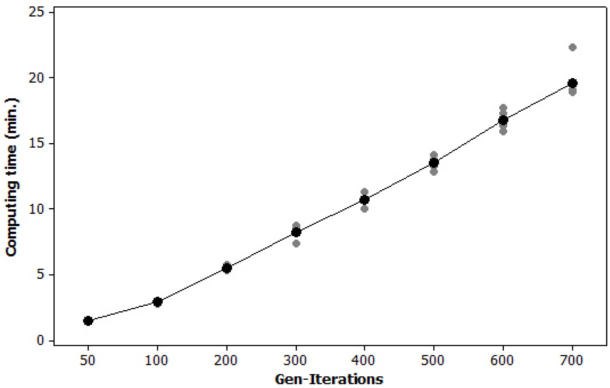


Fig. 2. Individual value plot of the computing time

Figure 1 shows the value of f_1 obtained throughout the computational experiment. The grey dots are associated with the values of f_1 obtained in each run of the experiment. The black dots display the corresponding mean value. As expected, the value of f_1 decreases when the number of iterations increases, but it seems to become stabilized from $N = 400$. The best value of f_1 obtained is 4755.38. This represents a 5.1% improvement over the current best value of 5011.44 provided in [3]. This new solution involves 14 routes. Depots 1 to 4 deliver 46, 61, 124 and 216 units, respectively. Depot 1 is supplied from plant 4, depot 2 from plant 1 and depots 3 and 4 are supplied from plant 3.

Figure 2 displays the computing time invested in minutes. A grey dot represents the computing time invested in a run of the experiment. A black dot displays the corresponding mean value. As can be clearly observed, the computing time largely increases when the number of iterations increases. The average of the computing time ranges from 1.42 to 19.62.

Figure 3 displays the relationship between the value of the distribution company objective function provided by the algorithm and the computing time invested by the algorithm.

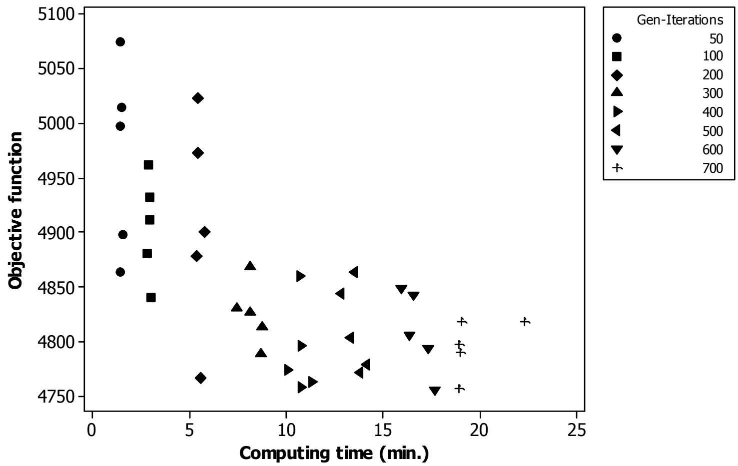


Fig. 3. Objective function versus computing time

5 Conclusions

In this paper we have proposed a metaheuristic algorithm for solving a production-distribution problem. The algorithm combines GA for assigning customers to depots, ACO for solving the associated routing problems and classical optimization for solving the production problem. The algorithm has been applied to solve the realistic instance presented in [3]. In general, the algorithm provides better values of the objective function than those existing in the literature. As expected, the best values are obtained when more computing time is spent. Hence, more research is needed in applying the algorithm to benchmark problems in order to confirm its performance and to improve the computing time. It would also be interesting to investigate the influence of other factors such as crossover or mutation probabilities, population size and the number of ants in the performance of the algorithm.

Acknowledgments. This research work has been funded by the Gobierno de Aragón under grant E58 (FSE), by Ibercaja under grant UZ2011-CIE-01 and by UZ-Santander under grant UZ2012-CIE-07.

References

1. Erengüc, S.S., Simpson, N.C., Vakharia, A.J.: Integrated Production/Distribution Planning in Supply Chains: an Invited Review. Eur. J. Oper. Res. 115, 219–236 (1999)
2. Vidal, C.J., Goetschalckx, M.: Strategic Production-Distribution Models: a Critical Review with Emphasis on Global Supply Chain Models. Eur. J. Oper. Res. 98, 1–18 (1997)
3. Calvete, H.I., Galé, C., Oliveros, M.J.: Bilevel Model for Production-Distribution Planning Solved by Using Ant Colony Optimization. Comput. Oper. Res. 38, 320–327 (2011)
4. Renaud, J., Laporte, G., Boctor, F.: A Tabu-Search Heuristics for the Multi-Depot Vehicle Routing Problem. Comput. Oper. Res. 23, 229–235 (1996)
5. Cordeau, J., Gendreau, M., Laporte, G.: A Tabu Search Heuristic for Periodic and Multi-Depot Vehicle Routing Problems. Networks 30, 105–119 (1997)
6. Bard, J.F.: Practical Bilevel Optimization. Kluwer Academic Publishers, Dordrecht (1998)

A Simulation Study Regarding Different Aircraft Boarding Strategies

Silvia Mas[1], Angel A. Juan[2], Pol Arias[1,*], and Pau Fonseca[3]

[1] Dept. Telecom. and Systems Eng., Autonomous University of Barcelona, Sabadell, Spain
[2] Computer Science Dept., IN3-Open University of Catalonia, Barcelona, Spain
[3] Operations Research Dept., Technical University of Catalonia, Barcelona, Spain
silvia.masm@campus.uab.cat, {ajuanp,pol.arias5}@gmail.com, pau@fib.upc.edu

Abstract. The airline industry is constantly subject to the search of new methods in order to increase efficiency, profitability, and customer satisfaction. Since airlines only generate revenue when their airplanes are on the air, the time they spend at the airports should be the shortest possible. Hence, the airplane turnaround time becomes a process which airlines pay special attention on. The boarding process has a very important role, since it is one of the significant elements of the turnaround time, and a slow boarding process might lead to many kinds of problems to the airline, from financial issues to customer complaints. This paper analyzes the major interferences among the passengers that cause delays in boarding times, and after comparing the different aircraft boarding strategies, it proposes the most efficient strategy.

Keywords: Aircraft Boarding, Simulation, Delay Times.

1 Introduction

Ground handling operations include all services that are carried out during the aircraft turnaround. The aircraft turnaround comprises the time from which the captain sets the airplane parking breaks, until he releases its breaks again. In other words, the turnaround begins when the ramp staff blocks the airplane (chocks on), and finishes when the chocks are off and the airplane starts the pushback. Most of the activities are independent and can take place simultaneously, such as catering, cleaning, and fueling; however, other activities such as the passenger boarding, cannot start until other processes have been finished.

It has been found that many previous works refer to the boarding process as a problem, since it is an activity that cannot start until processes such as fueling, cleaning or catering are ready. Likewise, Figure 1 shows that the passenger boarding process constitutes a critical path. Of course, the main priority during a boarding process is always safety, rather than carrying out a fast boarding. This explains why sometimes the boarding process does not start until fueling is finished, even when it could be done.

* Corresponding author.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 145–152, 2013.

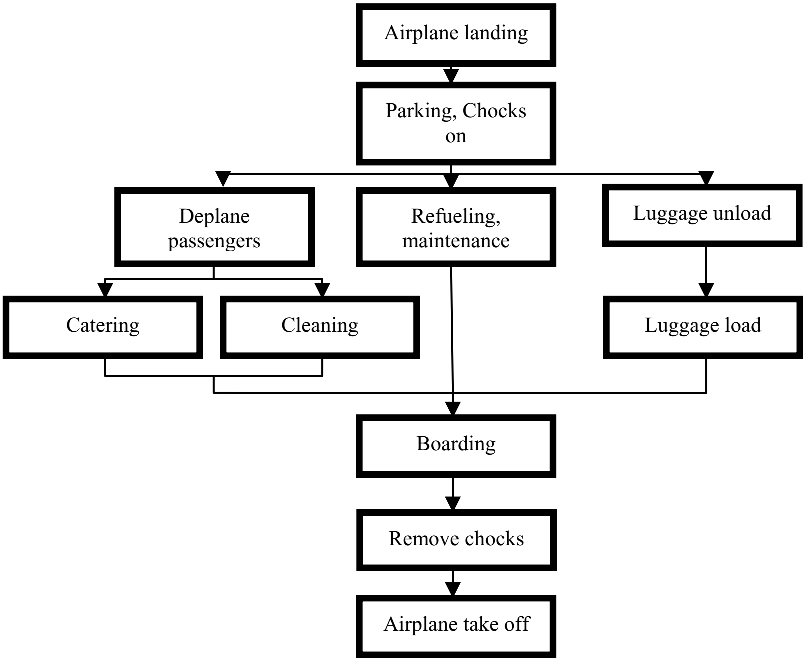


Fig. 1. Turnaround Scheme, Albert Steiner and Michel Philipp [1]

In addition to this, airline managers have to be aware that the most cost-effective boarding process will have to maintain quality and customer satisfaction.

One of the purposes in reducing the boarding time refers to reduce the number of interferences between passengers inside the airplane. A boarding interference is defined as an instance of a passenger blocking the access of another passenger to his seat. Therefore, the minimization of the total boarding time is related to the minimization of passenger interferences. Also, the total boarding time is related to the number of carry-on luggage that passengers have.

This paper is mainly focused on the boarding process while using a bridge. The most popular boarding strategies adopted by many of the airlines are the following:

- Back-to-front (BF) boarding policy (Figure 2) is the traditional strategy, adopted by most airlines for both narrow and wide-body aircraft. This strategy consists in boarding first class firstly (block 1). Then, passengers are called in groups to board the aircraft, following the sequence from back to front –i.e. blocks 2, 3, 4, 5, and 6.
- Outside-inside boarding strategy (Figure 3), also known as window-middle-aisle boarding. First class passengers are boarded first (block 1). Then, passengers in window seats are boarded (block 2), then middle seats (block 3), and finally aisle seats (block 4).

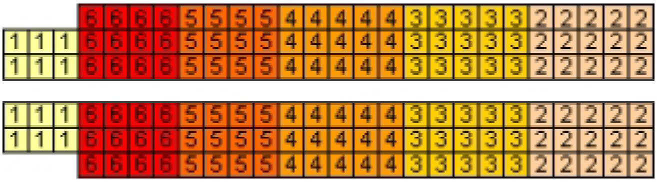


Fig. 2. Back-to-front boarding policy

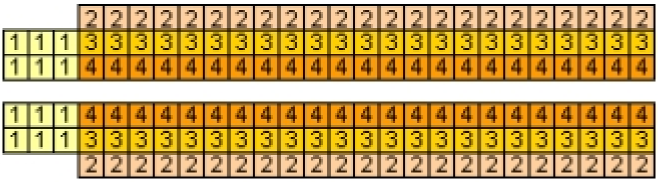


Fig. 3. Outside-inside boarding policy

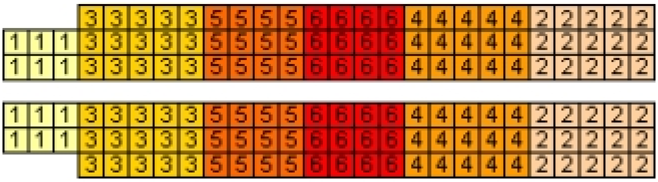


Fig. 4. Rotating zone boarding policy

- Rotating zone boarding strategy consists in boarding passengers sitting in the middle of the aircraft last (Figure 4). Thus, passengers are grouped into zones and board the aircraft first in the front (block 1), then in the back (block 2), then front again (block 3), then back (block 4), and so on.
- Random boarding strategy does not specify any condition while boarding passengers and the aircraft is boarded in one zone randomly (Figure 5). First class passengers are also boarded firstly (block 1). Then, passengers board the airplane in a first-come first-serve basis (block 2); or in other words, following a FIFO process (first-in first-out).

In this paper, simulation is used to analyze some of the aforementioned policies in order to search for the most efficient boarding strategy. The paper is structured as follows: Section 2 provides a brief overview of related work. Section 3 describes our approach. Section 4 includes some numerical experiments. Finally, Section 5 summarizes the main results.

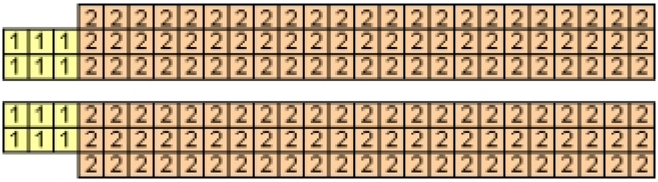


Fig. 5. Random boarding policy

2 Related Work

The aircraft passenger boarding problem has been previously studied mostly through simulation-based solutions for analyzing and improving passenger airplane boarding. **Marelli et al. [2]** conducted a simulation-based analysis performed for Boeing. Boeing Corporation created a computer simulation model called *Boeing Passenger Enplane/Deplane Simulation* (PEDS). The result of its study was that the outside-inside boarding strategy reduced boarding times significantly. **Van Landeghem and Beuselinck [3]** carried out a simulation study based on airplane boarding. According to this study, the fastest way to board the passengers on an airplane was to do it individually by their row and seat number. **Van den Briel et al. [4]** did not take into account airplane design parameters, and the study showed how strategies based on reducing interferences are better than the traditional back-to-front policy. These authors designed the so-called reverse-pyramid method, with the aim of boarding passengers while utilizing as much as possible the aircraft. The model was mainly developed to minimize passenger boarding interferences, and it was used MINLP, a mixed-integer nonlinearly constrained optimization solver.

Bauer et al. [5] developed a computer simulation to model the boarding process. These authors considered different boarding strategies and individual variations of passengers. Additionally, they treated the boarding problem as a stochastic process, and specifically, they used queuing theory to reach a better understanding of bottlenecks and their effects.

3 An Overview of Our Approach

By implementing a simulator in Visual Basic for Applications (VBA), using Excel, all the possible scenarios are tested; Figure 6 shows the interface of the simulator created. In order to obtain a reliable conclusion about which boarding policy performs better, we have considered 18 different scenarios. For the three different boarding strategies studied (random, back-to-front, and front-to-back), three specific aircraft models are determined:

- Medium-small airplanes with capacity for up to 152 passengers.
- Medium airplanes with capacity for up to 178 passengers.
- Medium-large airplanes with capacity for up to 212 passengers.

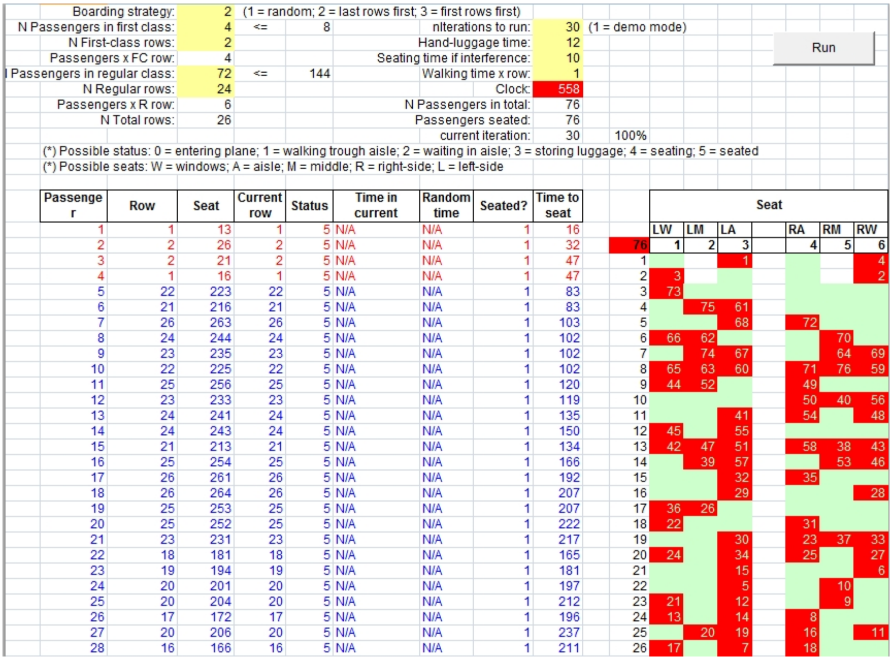


Fig. 6. Simulator overview

Moreover, for each model of airplane considered, an occupancy parameter is applied. Obviously, the more occupancy level, the more passenger interference, and therefore, more time will be required to board all passengers. In addition to this, we also want to model the occupancy level in order to find out how this parameter affects to a certain boarding procedure. Thus, in our model we consider two different occupancy levels: medium (50% occupancy), and high (100% occupancy).

In this context, we will try to figure out which policy works better according to the specific airplane size, and the current occupancy level of passengers. Notice that we have not considered the outside-inside boarding strategy since this policy cannot be easily employed in a real-life situation –if applied, passengers traveling together such as families or friends would have to board into the plane separately.

Firstly, according to the boarding strategy being analyzed, the model makes a pre-assignation of several parameters, and keeps them in memory. In order to determine the parameters, real times during 30 boarding processes at the airport of Barcelona were observed, and average estimates were derived from those observations. The current model assumes that these parameters are constant values. Thus, for our simulation trials the following values are used:

- Walking time per row = 1 time unit (0.5 seconds)
- Seating time interference = 10 time unit (5 seconds)
- Baggage stowage time = 12 time unit (6 seconds)

Secondly, there is a need to specify which boarding strategy is going to be executed. For instance, we can indicate the random approach, the back-to-front method, or the front-to-back strategy. Except the boarding approach where no seats are assigned to passengers, in the rest of strategies each passenger has a certain pre-assigned seat. Then, according to the aircraft type, there is a need to describe the number of rows and seats located in first class and in economy class, as well as the number of passengers that will occupy each class. Also, it is interested to model the time required for a passenger to walk and to stow the luggage, as well as the time provided by a seating interference. As explained before, in the current version of our model these three parameters have been considered as constants. Finally, the number of iterations to run the simulation has to be chosen. This collection of data is thrown in a grid that represents the airplane, where the different seats and the aisle are shown. The model divides the airplane in four blocks where passengers are assigned to, so they will be seated according to these blocks. Within a certain block, passengers are randomly sorted. Finally, simulation parameters such as the clock time or the total number of passengers being boarded are being tracked during the simulation.

4 Numerical Experiments

In this article, three boarding strategies are tested (Random, Back-to-Front, and Front-To-Back). Each strategy is used in every plane also having two different occupancy rates, so 18 different scenarios are tested. In each scenario 30 iterations have been computed to test its robustness. Because of the quantity of information extracted from each scenario, a sample has been chosen. Figure 7 shows an Analysis of Variance (ANOVA) for the scenario Medium–Large aircrafts with Medium Occupancy level. The results from this ANOVA tests with a p-value = 0.000 and non-overlapping confidence intervals, allow us to conclude that there exist significant differences in average boarding times among the different boarding policies considered.

```
                                 Individual 95% CIs For Mean Based on
                                 Pooled StDev
Level    N   Mean       StDev    -----+---------+---------+---------+----
Random  30  663,13      41,77    (*-)
BtF     30  754,53      38,57             (*-)
FtB     30  957,57      43,96                                   (-*)
                                 -----+---------+---------+---------+----
                                     700       800       900      1000


Pooled StDev = 41,49
```

Fig. 7. ANOVA Results for Medium – Large Aircraft with Medium Occupancy

Similar results can be observed in Figure 8, which illustrates an ANOVA test for the scenario Medium-Large aircraft with High Occupancy level. Notice that, again, the associated p-value is really low (p = 0.000) and, as in the previous experiment, confidence intervals do not overlap.

```
                                   Individual 95% CIs For Mean Based on
                                   Pooled StDev
Level    N   Mean        StDev   -----+---------+---------+---------+----
Random  30  1314,3        53,7   (*)
BtF     30  1707,8        48,8                     (*)
FtB     30  1934,1        68,3                                (*)
                                 -----+---------+---------+---------+----
                                   1400      1600      1800      2000


Pooled StDev = 57,6
```

Fig. 8. ANOVA Results for Medium – Large Aircraft with High Occupancy

Finally, Table 1 summarizes the results associated to the 18 different scenarios considered. It is worthy to highlight that, according to these results, in all scenarios the Random method results to be the one employing less time to complete the boarding. Also, notice that the higher the occupancy level, the larger the difference among the different strategies.

Table 1. Summary Results (times in time-based units)

	Medium-Small Aircraft			
	Random	BtF	FtB	nIterations
Medium Occupancy Level	495.8	565.1	747.1	30
High Occupancy Level	1008.5	1330.9	1522.7	30
	Medium Aircraft			
	Random	BtF	FtB	
Medium Occupancy Level	574.1	654.4	855.7	30
High Occupancy Level	1140.8	1517.9	1746.2	30
	Medium-Large Aircraft			
	Random	BtF	FtB	
Medium Occupancy Level	663.1	754.5	957.6	30
High Occupancy Level	1314.3	1707.8	1934.1	30

5 Conclusions

This paper has analyzed, using simulation, different boarding strategies in a set of common scenarios. Our results seem to confirm that the traditional and most common boarding method, which corresponds to the back-to-front approach, is not the most efficient one. In contrast, the random boarding strategy seems to perform the best in all scenarios. This conclusion is coherent with some previous studies, which suggested that airline managers should apply a certain boarding method according to the airplane size, or the occupancy of the flight. Also, as observed in our simulation study, the passenger occupancy level becomes a very important issue. In fact, our results help to quantify how the time difference among boarding methods increases as the occupancy level raises. For future work, we plan to consider planes with two boarding doors and two aisles.

References

1. Steiner, A., Philipp, M.: Speeding up the airplane boarding process by using pre-boarding areas. In: 9th Swiss Transport Research Conference (2009)
2. Marelli, S., Mattocks, G., Merry, R.: The role of computer simulation in reducing airplane turn time. AERO Magazine 1 (1998)
3. Van Landeghem, H., Beuselinck, A.: Reducing passenger boarding time in airplanes: A simulation based approach. European Journal of Operational Research 142(2), 294–308 (2002)
4. Van den Briel, M.H.L., Villalobos, J.R., Hogg, G.L., Lindemann, T., Mule, V.A.: America West airlines develops efficient boarding strategies. Interfaces 35, 191–201 (2005)
5. Bauer, M., Bhawalkar, K., Edwards, M.: Boarding at the Speed of Flight. UMAP Journal 237 (2007)

Routing Simulation Model with Allocation of Burdens Multimodal

Oscar Javier Herrera

Educational Researcher, ECCI -UCC Bogotá Colombia
ojavierho@gmail.com

Abstract. In the current work, we present the design of an allocation model of process plants to different customers considering the use of a fleet of vehicles with certain capabilities for the transport of goods, which is framed within theoretical model of mixed integer linear programming, developed from routing model with cargo allocation model and combined with multimodal transport. Furthermore, this model was applied using a simulation process on a real system, specifically the supply of recycled products Recycling Association ARB Bogotá Colombia, making a projection of their operation according to their medium-term strategic plan.

This model is focused for businesses developing logistics supply or distribution especially within a supply chain using a fleet of vehicles with different capacities and in which a destination can only be served by a single source within the corresponding nodal network.

Keywords: Supply Chain, Logistics, Procurement, mixed integer programming, Multimodal Transport Model, competitiveness.

1 Introduction

It is remarkable how the management of the supply chain (SCM) and specifically procurement activities is one of the best ways to improve the performance of organizations [1]. The SCM is a management model that pursues synergies through the integration of key business processes across the supply chain. The main objective is to serve the final consumer and the other owners of resources as effectively and efficiently as possible, that is, by products and / or services of greater value perceived by end customers and achieved at the lowest possible cost. "[2] therefore, the planning, scheduling and control activities are an important part routing to achieve high levels of productivity for any organization and if harmonize all activities associated with product flow in both directions, from up and down within the supply chain [3] and are actually responsible for the procurement, production and delivery of a product and / or service to the end customer. To do this we need the management of different types of physical and financial flows from the stage of procurement of materials to final product suppliers, manufacturers, distributors and customers, but should be done so that the input supply and output is done with effective distribution [4]. In this way, organizations gain competitive advantages and that transactions are processed faster, with more flexibility and lower cost [5].

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 153–162, 2013.

2 Problem Description

As problem at hand, must be the gathering center called Pennsylvania recycled products belonging to the Association of Recyclers of Bogotá ARB, has established a transport route that ensures the efficient use of available resources to perform the operation of collection material in each of the sources, in addition to complying with set delivery times with customers.

The planning method handled by the collection center has flaws because the routing is determined daily and empirically, and it is evident that there is an organization in terms of documentation handled in this process that allows the human resource plan for collection time so that it can meet the logistical objectives workday.

Due to poor planning, delays occur when you come to any source collection, difficulties also arise as to the readiness of the material with some suppliers, not because the material is classified or not supplying the personnel of thus not met the required amount of material, and the distances between points are high which leads directly increase the cost of fuel, also do not have close communication with collection points in terms of quantity, on the other hand, not controlled undelivered orders generating a poor level of service. All this represents unproductiveness that cause increased costs that were not budgeted in the provisioning operation.

Furthermore, when determining the transport route should take into account the distance between sources to chart a path that achieves better collect more material in less time, also choose routes that are not as busy in hours Reviews in the city of Bogota, in order to ensure the reduction of time and distance traveled and increase operational efficiency. Also keep in mind within the route, the check engine tanking and avoiding possible breakdowns of vehicles are there in the company; truck Hyundai HD65 and DFM Pick up truck. Thus, to solve this problem, we examined the causes related to different areas of the collection center, concluding through inquiry and observation, that how to design and manage the current routing had the highest level of impact directly affecting productivity Bogotá Association of Recyclers ARB.

3 Methodology

As a first instance, we evaluated different logistics applications in mixed integer programming bibliographic references given in Operations Research to establish parametric aspects and / or restrictive in various models according to the distribution center where one carries out the implementation of the model and subsequent simulation, in this way to establish the best course of action (optimal) decision system with the constraint of limited resources found [6].

It is found that transport-related models have been studied with various published results which are modeled as classical problems of combinatorial optimization, integer programming, and in many cases are solved exactly in [7], [8], [9], [10], [11], so taking the problem of design and optimization of routes and frequencies has been less studied [12], finding the following difficulties:

a. Formulation of the problem: to define the decision variables (in particular the choice of line by the carrying) and the objective function.
b. Non-linearity and non convexity of the problem.

c. Combinatorial nature of the problem, with discrete variables.
d. Multiple objectives: there is a trade-off between the objectives mainly system users and operators, which means it can not be single optimal solution, but several non-dominated solutions. A solution is not dominated when no other solution that improves on some objective function without worsening the rest.
e. Spatial arrangement of routes: a willingness formalize them.

The first tools of optimal design of routes and frequencies arise in the 70's, based on intuitive ideas, without a model formulation and its objective function, in some cases without exploration of the solution space. In the 80's makes some objective functions, and incorporates new parameters such as coverage of demand, load factor (ratio of passengers standing on the number of seats) and transfer buses [13]. In the 90's are other approaches such as the use of meta-heuristics and exploration of the solution space. The ease of integrating existing modules to incorporate graphical interfaces, stimulate the development of new methods, which are differentiated by their:

a. Adaptability in the data available, mainly those related to the topology of the transit network and travel demand (origin-destination matrices);
b. Interactivity with the user, in order to allow the incorporation of human knowledge (human technician) in the decision making process;
c. Efficiency: quality results and reasonable processing times;
d. Flexibility regarding the planning horizon, the first methods referred to planning short and medium term.

Within the broader managed techniques as they have evolved: Model of the Traveler, better known as Traveling Salesman Problem (TSP) is one of the most studied problems in the field of optimization because of its unique features. The TSP is a problem of NP-Hard. When the number of points of the network is high, it becomes unfeasible the use of exact algorithms for solving such problems due to the great amount of time required for resolution. In these cases when other algorithms require that an approximate solution to provide optimum whose computation time required is less, such as the heuristic algorithms, and more specifically, metaheuristics. For the specific case of the TSP problem genetic algorithm was used for resolution. In order to analyze the results found for the same, have developed local search heuristic algorithms used by other authors in solving the TSP problem [14]. These algorithms are: Savings and Integration.

Algorithms are also metaheuristics, in the particular case where the number of variables involved in the model is very high, the classic solution algorithms is not efficient, due to the time needed to find the optimal solution of the model.

3.1 Models Used as the Basis

- Multimodal Transport Model

In first instance and in the highlights of logistics, will address the issue of transportation of goods, supplies, materials etc., Both internally and externally of the company [15], defined as taking the initial course, the origins or local sources and destinations and demand or which establishes a fleet of vehicles for transport with their respective capacity constraints [16].

The objective of this model is transported from the source "i" to the destination "j" using the vehicle "k" at the lowest possible cost.
Model data are:

a. Level of supply at each source and the amount of demand at each destination.
b. The unit cost of transportation of goods to each destination.
c. The number of vehicles and their availability and capacity restrictions

Please check that the lines in line drawings are not interrupted and have a constant width. Grids and details within the figures must be clearly legible and may not be written one on top of the other. Line drawings should have a resolution of at least 800 dpi (preferably 1200 dpi). The lettering in figures should have a height of 2 mm (10-point type). Figures should be numbered and should have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table. Please center the captions between the margins and set them in 9-point type (Fig. 1 shows an example). The distance between text and figure should be about 8 mm, the distance between figure and caption about 6 mm.

To ensure that the reproduction of your illustrations is of a reasonable quality, we advise against the use of shading. The contrast should be as pronounced as possible.

If screenshots are necessary, please make sure that you are happy with the print quality before you send the files.

- Mixed model with multiple sources Routing with Load Allocation
The goal of this model is to design a delivery routes several clients of a system using multiple sources, taking into account the capacity demand conditions, with the conditionality, specifically routing connectivity model from various sources [17] where one takes into account that each client can be served only by a distribution center [18]. The model is supplied with parameters fixed costs and variable costs Transportation Distribution by product from each of the distribution centers or logistics operators to different clients of the system.

In this model, one must first define a maximum travel distance on a unit that must perform logistics operators or processing centers. This parameter is used as a restriction in the functional development of a whole system within the predetermined nodal network.

3.2 Establishment of the Model

To develop the model of this paper, the mathematical structure was established for each of the models which are based on the model developed logistic supply. These activities were carried out by establishing the theoretical exposition.

Based on these structures, the integration model was determined according to the characteristics of each of these models, thereby obtaining generic building the model in question: Routing Model with Multimodal Freight Assignment and subsequently developed an application to a typical case in the procurement system projected Bogotá Association of Recyclers, framed within a logistics consulting project that takes them by the proposing institution of this work.

- General model

Decision Variables

Xijk = Quantity to transport the production center "i" to destination "j" in the means of transport "k". i=1,2……m; j=1,2……n; k=1,2……l

Vijk = Number of trips to be made from the production plant "i" to the consumer center "j" in transportation "k"

Yij = binary decision variable where; 1 if i j serves, or 0 if i does not attend aj

Parameters

Cij = Unit cost of transporting the product, source or process plant i, the consumer or customer center j within nodal network supply system.

CFij = Fixed cost transport process plant i, the consumer or customer center j within nodal network supply system.

Cmaxi = maximum allowable cost for transportation from the process plant i, the consumer or customer center j within nodal network supply system.

ai = available capacity to supply the process plant i

bj = requirement demand by the consumer or customer center j

CCK = Vehicle Payload k

Nk = number of vehicles type k

Vmaxk = Maximum travel by vehicle type k

Mathematical structure:

$$\text{Min Z: } \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{l} CijXijk + \sum_{i=1}^{m}\sum_{j=1}^{n} CFijYij \quad \text{Objective Function} \tag{1}$$

s.a.

$$\sum_{i=1}^{m} Yij = 1 \quad \forall \quad j = 1,\ldots,n \qquad \text{Connectivity Constraint} \tag{2}$$

$$\sum_{j=1}^{n} CFijYij = C\max i \quad \forall \quad i = 1, \ldots, m \quad \text{Cost Restriction max. permitted} \tag{3}$$

$$\sum_{j=1}^{n}\sum_{k=1}^{l} XijkYij \leq \text{ai} \quad \forall \quad i = 1, \ldots, m \quad \text{Ability Restriction Offer} \tag{4}$$

$$\sum_{k=1}^{l}\sum_{i=1}^{m} YijXijk \geq \text{bj} \quad \forall \quad j = 1,\ldots,n \quad \text{Restriction or Market Demand} \tag{5}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} Vijk \leq V_{maxk} \quad \forall \quad k = 1, \ldots, l \quad \text{Restricting Number of trips} \tag{6}$$

$$Xijk \leq CCkNkVijk \quad \forall \quad i = 1,..,m; \ j = 1,..,n; \ k = 1,..,l \quad \text{Vehicle capacity constraint} \tag{7}$$

$$Xij \geq 0 \quad \text{Nonnegativity constraint} \tag{8}$$

$$Vijk = \text{Positive integer variable} \tag{9}$$

$$Y\text{ij} = \text{binary variable} \tag{10}$$

Established a language appropriate computational cumshot note modelares features of this integration, it was determined to run in Excel Solver for its simplicity and because it allows linear models to work together as binary variables as established this model.

3.3 Application Development

We begin by setting the nodal pattern of interaction between the various components of the system, process plants or logistics operators (origins or sources) and consumption centers or system customers (destinations).

For this, application development a nodal network initial case in the procurement system projected Recycling Association ARB Bogotá Colombia, taking as an example typical case and three collection centers which collect recycled products seven intermediate customers or PR collection points, which can be visited by geography and whose assignments and input parameters are presented in figure 1. Otherwise, it is of note that there are two types of cars or trucks, the first 5.2-tonne capacity and the second of 8.5 tons, for the first type has 4 vehicles and the second 3 besides that no time constraints presented by the collection point for provisioning.

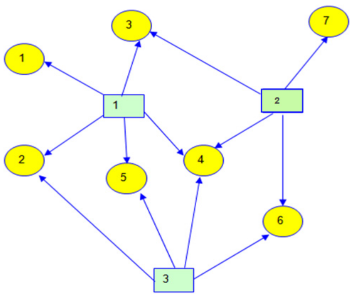


Fig. 1. Assignments Nodal Network Provisioning

Table 1. Data Entry (Fixed Costs, Variable Costs, Demand and Supply)

Fixed costs of transport from the storage facilities "i" to the collection points "j"							
Collection Centers	PR1	PR2	PR3	PR4	PR5	PR6	PR7
1	$ 25.000	$ 28.500	$ 30.000	$ 26.000	$ 25.000		
2		$ 27.000		$ 27.000	$ 24.000	$ 25.000	
3			$ 28.000	$ 26.500		$ 26.000	$ 27.500

Unit transportation costs from the distribution center "i" collection point "j"								
Collection Centers	PR1	PR2	PR3	PR4	PR5	PR6	PR7	offer
1	$ 50	$ 45	$ 48	$ 55	$ 52			35000
2		$ 56		$ 49	$ 51	$ 53		40000
3			$ 46	$ 47		$ 49	$ 50	40000
Demand	10500	9500	12000	13500	11000	11500	14000	

Based on this information the model set is fed, so as to obtain the appropriate allocation of clients to each of the hubs fulfilling the condition that each customer can only be served by a single hub as Table 2 presents the results.

Table 2. Allocations for each distribution center to the respective network clients

Solution Matrix Var. Amount to transport (Vehicle 1)

Collection Centers	PR1	PR2	PR3	PR4	PR5	PR6	PR7	Total
1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0
2	0,0	0,0	0,0	0,0	11.000,0	0,0	0,0	11000
3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0
Total	0	0	0	0	11.000	0	0	

Solution Matrix Var. Amount to transport (Vehicle 2)

Collection Centers	PR1	PR2	PR3	PR4	PR5	PR6	PR7	Total
1	10.500,0	0,0	0,0	0,0	0,0	0,0	0,0	10500
2	0,0	9.500,0	0,0	0,0	0,0	11.500,0	0,0	21000
3	0,0	0,0	12.000,0	13.500,0	0,0	0,0	14.000,0	39500
Total	10.500	9.500	12.000	13.500	0	11.500	14.000	

4 Simulation of the Routes Found Promodel

Within the constraints of the model is also necessary to take into account the ability of each of the vehicles, for reference Hyundai HD65 is 4.2 tons, and 900 kg for DFM, according to the technical specifications.

Furthermore, each was determined distances between the centers of collection and collection points via a system of georeferencing. From the distance data obtained relates employee travel time to get from one point to another, and it is calculated travel speed in order to verify that exceed allowable limits for this type of transport vehicles in Bogota, according to current regulations set up 35km/hr.

Management Indicators Used

a. Used Capacity: This is determined from the total registered route Kilograms of total capacity used on the same truck.
b. Average Time per Source: establishing the relationship as the total time spent on each tour and visited many sources.
c. Cost Mileage: Determine the costs that are incurred in the transport business, from the cost structure.

To perform the run is necessary to apply different types of analysis data obtained from the system [19], as shown below:

Analysis of Independence: When you start modeling in Promodel® should study the relationship of the variables to determine any dependency between them [20], [21], for this particular time - Mileage. Given the amount of data required is obtained by grouping sets of data [22], performing two divisions ranges from distance every two-three kilometers, will determine the correlation coefficient thereof in order to establish the connection and thus the dependence of the data. Made these analyzes, it is possible to conclude that there was correlation between variables median, however not enough to determine dependence since none of the values obtained is greater than 0.85.

It is clear that homogeneity analysis is not possible due to the aforementioned data division, which means that as the distance interval increases, in the same way it affects the time involved for each course, so the times for each step are also different [23].

Analysis is performed goodness of fit for each of the ranges set forth above by the Stat:Fit application to determine the probability distribution which best represents the variables studied [19]. From the list of the best distributions by Stat: Fit, and acceptance of the same for those that represent a data more closely, and given their application is selected own lognormal distribution for each interval. Otherwise, as there is insufficient data to perform the analysis of the time spent on the recruitment and delivery of material, set the time as defined by a normal distribution under the central limit theorem [24], [25], which ensures that the sampling distribution for any population approaches a normal type for independent random variables.

Simulation Results

In evaluating each of the routes Promodel ® is possible to have a more realistic view of system behavior, since the variability seen in the same characteristic of the different aspects influencing the time spent on each run.To have a better use of mobile resource, capacity was determined as the most significant variable when choosing the appropriate path for each distribution center, the main objective is to use the maximum capacity of each truck to meet the requirements and material delivery times to customers. Furthermore simplify matters and the results in Table 3 and 4 show the capacity utilization, time and travel distance in the material collection performed in the simulation for the different sources, compared with the behavior current. It is noteworthy that the proposed model handles time variability that includes stops at traffic lights and late delivery of material sources.

Table 3. Comparison in capacity utilization, time and travel distance, current vs. simulated. V1

Vehicle 1. Hyundai

	Utilization		Total time		Distance		Total cost		
	Proposal	Present	Proposal	Present	Proposal	Present	Proposal	Present	saving
Monday	98,8%	42,5%	22,00 Hr	3,28 Hr	26,30 Km	25,60 Km	$ 218.714,17	$ 49.576,38	$ (169.137,79)
Tuesday	80,3%	32,4%	21,00 Hr	12,07 Hr	93,90 Km	105,70 Km	$ 262.618,82	$ 191.432,25	$ (71.186,57)
Wednesday	93,1%	66,8%	15,00 Hr	5,37 Hr	44,41 Km	91,30 Km	$ 169.847,77	$ 119.822,04	$ (50.025,73)
Thursday	91,1%	36,1%	0,00 Hr	13,38 Hr	63,70 Km	105,70 Km	$ 49.857,99	$ 203.229,94	$ 153.371,95
Friday	96,2%	74,0%	0,00 Hr	5,92 Hr	71,55 Km	52,25 Km	$ 56.002,19	$ 94.210,83	$ 38.208,65
								Total	$ (98.769,49)

Table 4. Comparison in capacity utilization, time and travel distance, current vs. simulated. V2

Vehicle 2. DFM

	Utilization		Total time		Distance		Total cost		
	Proposal	Present	Proposa	Present	Proposa	Present	Proposal	Present	saving
Monday	95,9%	78,8%	5,37 Hr	8,8	56,50 Km	55,50 Km	$ 92.584,08	$ 122.691,51	$ 30.107,44
Tuesday	97,9%	94,7%	5,87 Hr	7,13	21,40 Km	70,70 Km	$ 69.614,24	$ 119.548,75	$ 49.934,51
Wednesday	89,1%	73,6%	4,91 Hr	5,12	42,40 Km	45,80 Km	$ 77.405,31	$ 81.957,72	$ 4.552,41
Thursday	96,9%	78,4%	6,22 Hr	3,97	51,10 Km	75,30 Km	$ 96.012,49	$ 94.690,62	$ (1.321,87)
Friday	90,5%	74,5%	5,40 Hr	2,98	48,90 Km	102,90 Km	$ 86.905,73	$ 107.377,33	$ 20.471,59
								Total	$ 103.744,08

The monetary value is calculated in Colombian pesos, about $56 a week per vehicle. The ratio of cost within the routing system is linked to the total travel mileage, and the cost of Labor, in the analysis of the model takes into account the parameter mileage between sources, we calculate savings in money for each one of the days, the transportation cost incurred on each route.

5 Conclusions

- The routing model meets current logistical parameters of collection centers, but you want to have an approach to organizing and improving the flow of material, for it must implement the use of tools to optimize the resources have currently, the established routing model and simulation for vehicles is estimated reduction in operating costs $113 dollars a week, generating significant long-term benefit to the company.

- The simulation model applied to multimodal transportation allocation allows better visualization of the routing system for the collection centers, reducing travel time and movements, compared to current empirical model proposed routes reflect a decrease in mileage and time equivalent to 7.48 hours and 210.59 miles unnecessary operation also ran the route on Saturdays was distributed during the week, therefore the space on this day can be leveraged for other operating activities of the distribution center.

- To establish the routes with the restrictions given by the supplier, the company requires you to have a better planning of the logistics operation in addition to keep track of the route performed by each vehicle's storage facility.

- By having a routing system established collection centers as proposed, allows the company logistics operator better control and management of mobile resources in order to evaluate its correct operation.

Overall this model allows for a fairly satisfactory-as the case where they applied for decision-making in logistics activities which have various possibilities for the distribution and allocation posterior nodal routing in a network, in addition to the ability to carry out assignments logistics operators with the demands of customers in the system at the same time takes into account the capabilities and availability to be taken from each of these operators or network distribution, obtaining at last, values optimal with respect to reducing the costs associated with this operation.

Note that this model includes fixed costs of transportation, which sets basic parameters as input to facilitate and expedite the calculation of these costs by the distances between each customer and distribution centers in the system network, eg including drivers' wages, fuel consumption of vehicles and other items of machinery inherent in spending by the mobilization of goods, making it easier and gives greater accuracy in the calculation of these costs and then be used as information input for the development of the mathematical model. Moreover, also provides transportation variable costs driving unit value that implies the mobilization of products from each source processing center or to different consumption centers or destination; taking thus integrating elements of cost involved in the development of the transport along the supply chain for a company engaged in this work logistics.

References

1. Kaihara, T.: Supply chain management with market economics. International Journal of Production Economics 73(1), 5–14 (2001)
2. Pires, Silvio, R.I., Díaz, C., Luis, E.: Gestión de la Cadena de Suministros, p. 25. McGraw-Hill/Interamericana de España, S.A.U., Madrid (2007)

3. Handfield, R.B., Nichols, E.L.: Supply Chain Redesign: Transforming Supply Chains into Integrated Value Systems. Prentice Hall (2002)
4. Villa, A.: Introducing some supply chain management problems. International Journal of Production Economics 73(1), 1–4 (2001)
5. Al-Mudimigh, A., Zairi, M., Ahmed, A.: Extending the concept of supply chain: the effective management of value chains. International Journal of Production Economics 87(3), 309–320 (2004)
6. Méndez, G.A.: Guía temática Programación Lineal y Entera. Editorial U.D.F.J.C., Bogotá (2000)
7. Moskowitz, W.: Operations Research. Editorial Prentice Hall (1991)
8. Prawda Witenberg, J.: Métodos y modelos de investigación de operaciones, vol. 1. Modelos determinísticos Limusa Noriega editores, México D.F. (1999)
9. Taha, H.: Investigación de Operaciones una Introducción 6° Edición. Prentice Hall, México D.F (1998)
10. Winston Wayne, L.: Investigación de operaciones, aplicaciones y algoritmos. Segunda edición. Grupo editorial Iberoamérica (1994)
11. González Ariza, A.: Manual Práctico de Investigación de Operaciones 3° Edición Ediciones Uninorte Colombia (2007)
12. Baaj, M.H.: An AI-based approach for transit route system planning and design. Journal of Advanced Transportation 25(2), 187–210 (1991)
13. Newell, G.F.: Some issues related to the optimal design of bus routes. Transportation Science 13, 20–35 (1979)
14. Charkroborty, P.: Optimal route network design for transit systems using genetic algorithms. Engineering Optimization 34(1), 83–100 (2002)
15. Ballou, H.R.: Logística administración de la cadena de suministro 5° Edición, Editorial Pearson, México D.F (2004)
16. Lopez Cesar, A.: Memorias de Clase "Modelos Cuantitativos". Universidad de la Sabana, Bogotá (2011)
17. Jairo, T.: Aspectos generales sobre Logística. Bogota: Cátedra Gestión de Sistemas Logísticos. Maestría en Ingeniería Industrial, Universidad Distrital Francisco José de Caldas (2008)
18. Herrera Oscar, J.: Modelo de Ruteo Mixto con varias fuentes con Asignación de Cargas. Fusión Industrial, 54–62 (2008)
19. Blanco Rivero, L., Fajardo Piedrahita, I.: Simulación con Promodel: Casos de Producción y Logística. Jireh Impresiones, Colombia (2001)
20. Garcia Dunna, E., Azarang Esfandiari, M.: Simulación y análisis de modelos estocásticos. McGraw Hill, México (1996)
21. Llinás, H., Rojas, C.: Estadística Descriptiva y Distribuciones de Probabilidad, 1ª Edición. Ediciones Uninorte, Colombia (2006)
22. Mendenhall, W., Beaver, R.: Introducción a la probabilidad y estadística, 12ª Edición. Cengage Learning Editores S.A., México (2006)
23. Vásquez, M.G.: Probabilidad y estadística. Editorial McGraw Hill, cuarta edición (1992)
24. Santaló, L.: Probabilidad e inferencia estadística, pp. 81–82. Editorial McGraw Hill Buenos Aires, Argentina (1970)
25. Alvarado, H., Batanero, C.: Articulo: Significado del Teorema Central del Límite, en Textos Universitarios de Probabilidad y Estadística (2008)

A Simulation-Based Approach for Solving the Aircraft Turnaround Problem

Iris Silverio[1], Angel A. Juan[2], and Pol Arias[1,*]

[1] Dept. of Telecommunications and Systems Engineering,
Autonomous University of Barcelona, Sabadell, Spain
[2] Computer Science Dept., IN3-Open University of Catalonia, Barcelona, Spain
{iris.silverio.pons,ajuanp,pol.arias5}@gmail.com

Abstract. This paper describes the Aircraft Turnaround Problem (ATP) and then proposes a simulation-based method for solving it. The ATP is receiving an increasing interest from airline companies due to the importance of reducing delay costs in today's competitive environment. Our approach uses a simulation model to analyze different scenarios and thus estimate the optimal level of buffer times during turnaround operations. Real-life data has been used to complete a set of experimental results that contribute to illustrate the potential applications of the proposed methodology.

Keywords: Aircraft Turnaround Problem, Simulation, Delay Times.

1 Introduction

Nowadays, commercial airlines are deeply interested in reducing total scale times, since longer scale times generate higher opportunity costs for the airlines, as well as other costs associated with the use of airport facilities. The problem of reducing scale times is related to a scheduling optimization problem, where the total time required to process a set of tasks (makespan) has to be minimized. However, these processing times are stochastic variables and, consequently, the associated makespan has a random nature as well. This has implications over flight departure times when these are scheduled right after expected makespans. In effect, if a departure time is scheduled right after the completion of a task and a delay occurs, passengers will have to wait for their flight to be ready even when the scheduled departure time has already been surpassed. Of course, waiting times represent also a cost to the airline companies in terms of bad reputation and unsatisfied clients who might decide to suit the company, to use a different airline in the future, etc. Therefore, potential delay times in tasks completion must be accounted for when scheduling flight departure times if waiting costs are to be considered. According to a survey carried out by the Gatwick airport [1], the air traffic control is responsible for 30% of flight delays, while 25% of delays are due to land services (ground handling) and problems related to flight

* Corresponding author.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 163–170, 2013.

stopovers. Most airlines consider that one of the keys to success in the European air transport market is punctuality during stopovers. As a consequence, it is a usual practice in airline companies to establish a buffer time between the expected makespan and the scheduled departing time, so that both scale-related costs and delay or waiting costs are balanced.

Different approaches have been used so far for solving the so-called "Aircraft Turnaround Problem" (ATP). On the one hand, some authors consider a sequential view of the airport operations-handling process, i.e. some tasks cannot be started until some others have been finished [2]. This approach leads to a version of the ATP known as Resource Constraint Project Scheduling Problem (RCPSP). On the other hand, a different approach is based on the assumption that different tasks can be simultaneously processed by different services, which is related to the well-known Job-Shop Scheduling Problem (JSSP) [3]. Despite a few scientific approaches exist for solving the ATP, they are rarely used by flight coordinators in real-life scenarios to schedule the ‘optimal’ departure times. The goal of this article is to propose a relatively simple simulation-based methodology that can be used to support decision-making processes in balancing scale and delay costs.

The paper is structured as follows: Section 2 performs a literature review on the ATP. Section 3 introduces and describes the proposed stochastic model and the associated simulation-based method to solve it. Section 4 analyzes some numerical experiments that make use of real-life data to illustrate our approach. Finally, Section 5 summarizes the main results obtained in this work and proposes some future research lines.

2 The Aircraft Turnaround Problem and Related Work

As advanced before, two main approaches are used to solve the TAP. The first one distinguishes each operation to be carried out over the turnaround, and the objective is to find the optimal order to minimize the total time stopover by reducing costs. Kuster [4] proposes a variant of the Resource-Constrained Project Scheduling Problem (RCPSP), the x-RCPSP, defining alternative activities that can replace the basic operations to modify the model. The method Project Evaluation and Review Technique (PERT) has also been used to evaluate and try to improve the turnaround process by planning a set of activities, some of which are interdependent. Some activities can be initiated only when the previous has finished, this relationship is indicated by nodes, referring to transactions connected together with arcs indicate dependencies. In its work, Wu [5] presents a methodology using the PERT approach. The main objective is to find the critical path that implies a delay on the overall system. The paper discusses various scenarios and concludes that the critical path is composed of operations on board of the aircraft: arrival of the aircraft, landing passengers, catering refill, etc.

In the second approach, the turnaround is treated as a single block. In this case the goal is to minimize costs considering timeliness of flights from adding a buffer of extra time, so more time is allowed to complete the entire scale –thus reducing the risk of suffering delays. A study by Fricke [6] in collaboration with Lufthansa City Line shows that the main cause of delays is the turnaround of the aircraft. Among these, some common sub-causes are related to handling errors, slow shipping, etc. As a solution to the problem of delays affecting turnaround processes, this author proposes adding a time buffer which can partially absorb potential time delays. Wu [7] studies the relationship between the timeliness and efficiency of flight operations for the stopover, also considering the option of adding a buffer of time to minimize the total costs. To analyze the problem, a mathematical model of stochastic data timeliness is designed, by distinguishing the operational costs of the aircraft, those generated by the delays, and the opportunity cost of the aircraft. Another paper of Wu [8] analyzes the possibility of adding an extra amount of time to the turnaround. The main advantage is that flight delays are absorbed by the buffer, but it affects to the opportunity costs and costs of ground operations on land.

3 An Overview of Our Approach

Figure 1 shows a visual representation of the different variables that must be considered in the ATP. Scheduled arrival time is indicated by (t_{sa}), and the real time by (T_a). The difference between them both is the delay due to late arrival ($delay_A$). If the aircraft had arrived punctual, it would have been ready to depart once the maintenance process is finished (t_{sr}). However, being behind schedule implies that it is ready a few minutes later (t_r). The aircraft has an scheduled time (t_{sd}) at the end of extra time margin (t_b), but actually departs later (t_d). The difference between the scheduled time of departure and the actual total delay is ($delay_T$). The additional time buffer is added to absorb delays caused during the previous flight or during ground operations, in order to reduce costs associated with schedule delay. Some of the variables and symbols used in this section are summarized next:

C_t: Turnaround Total Cost.
C_d: Delay Cost.
C_o : Opportunity Cost.
α: Binary variable that indicates if there is delay or not.
$delay_A$: Arrival Delay.
$delay_M$: Maintenance Delay.
$delay_T$: Total Delay.
t_{sa} : Scheduled Arrival Time.
t_{sd} : Scheduled Departure Time.
T_a : Arrival Time, real arrival time.
t_d : Departure Time, real departure time.
t_{sm} : Scheduled Maintenance Time.
T_m : Maintenance Time, turnaround real time.
t_{sr} : Scheduled Readiness Time, scheduled time when the aircraft is ready to depart.
t_r : Readiness Time, real time when the aircraft is ready to depart.
t_{sb} : Buffer Time.

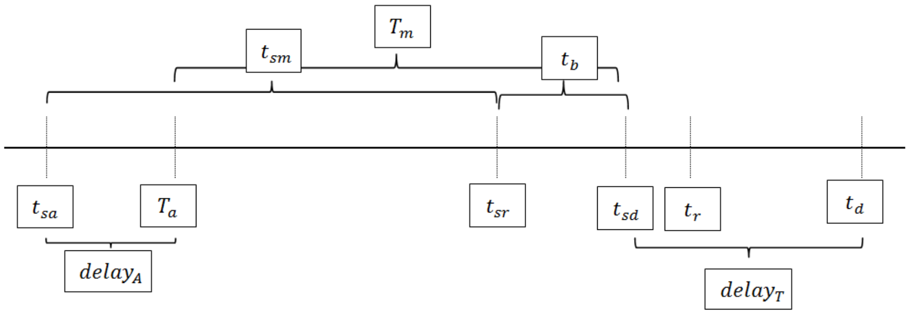


Fig. 1. Turnaround scheme

The objective function of the aircraft turnaround problem can be described as:

$$\text{Min } E[C_t(t_{delay})] = \alpha \cdot E[C_d(t_{delay})] + (1 - \alpha) \cdot E[C_o(t_{delay})] \quad (1)$$

where:

$$\alpha \text{ in } \{0,1\} \quad (2)$$

$$t_{sa} = E[T_a] \quad (3)$$

$$t_{sm} = E[T_m] \quad (4)$$

$$0 \leq t_b \leq t_{max} \quad (5)$$

$$T_a \sim We(\mu, \sigma) \quad (6)$$

$$T_m \sim We(\mu', \sigma') \text{ .} \quad (7)$$

$$t_d = \text{Max}\{t_{sa} + t_{sm} + t_b, t_{sa} + t_{sm} + t_b + ((t_a - t_{sa}) + (t_m - t_{sm}) - t_b \quad (8)$$

$$delay_T = t_d - t_{sd} \quad (9)$$

The objective is to minimize the total expected cost of the scale (C_t), which will consist of a weighted combination of delay costs (C_d) and opportunity costs (C_o).

4 Numerical Experiments

Microsoft Visual Basic for Applications and Excel were used to perform the simulation experiments. Real-life data on daily arrival and maintenance delays was obtained from http://www.flightstats.com. We focused in data provided for the Barcelona-El Prat airport. Also, in order to model the costs, we used a study developed by the University of Westminster [9]. According to this study, the average waiting costs of delay is about 25€ per minute, while the average opportunity costs of a landed and ready-to-go aircraft is about 19€ per minute.

Regarding daily operations in June 2012 at the Barcelona airport, 329 flights were analyzed. From this analysis, it was obtained that the average delay caused by arrival tardiness of an aircraft is about 5 minutes and 47 seconds. Also, the average delay caused by ground operations is about 2 minutes and 20 seconds.

Using bootstrapping techniques and the real-life data, we simulated the aircraft turnaround process. **Table 1** shows the average cost associated with each minute of delay that is not absorbed by the buffer *(Cost_Delay)* as well as the opportunity cost per minute *(Cost_Extra)*. Notice that, according to **Table 1**, in the analyzed data no time buffer is needed, being the average costs about 202.81€.

Table 1. Average Results

	Buffer time (b) in minutes					
	b = 0	**b = 5**	**b = 10**	**b = 15**	**b = 20**	**b = 25**
Cost_Delay	202,81 €	152,81 €	117,10 €	90,65 €	72,34 €	57,45 €
Cost_Extra	0,00 €	57,00 €	124,86 €	199,76 €	280,84 €	364,52 €
TOTAL_COST	202,81 €	209,81 €	241,95 €	290,41 €	353,18 €	421,97 €

Next, an Analysis of Variance (ANOVA) test was run in order to determine if there exists a significant difference among the average costs associated with different levels of buffer. **Figure 2** and **Table 2** show the results of the test.

The ANOVA shows a low p-value (p = 0.000), this indicating that average costs associated with different buffer levels are significantly different. In fact, it can be observed that the first group –the one in which buffer is set to zero– shows the lower costs. So far, the analysis has been made using the average delay of the daily operations. However, delays are not uniform throughout the entire day. In effect, **Figures 3 and 4** show daily variability in arrival delays and maintenance delays. Therefore, we decided to complete a further analysis by splitting the day schedule into five parts and then to determine the best buffer size for each of these parts.

Table 2. ANOVA Table

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	5	14658177	2931635	30.49	0.000
Error	1718	165195114	96155		
Total	1723	179853290			

```
                                    Individual 95% CIs For Mean Based on
                                    Pooled StDev
Level    N    Mean   StDev   --+---------+---------+---------+-------
b=0     79    44,9   127,0   (-----*----)
b=5    329   209,8   380,8                    (-*--)
b=10   329   242,0   337,6                      (--*--)
b=15   329   290,4   305,1                          (--*--)
b=20   329   353,2   279,2                                (-*--)
b=25   329   422,0   264,1                                     (--*--)
                             --+---------+---------+---------+-------
                               0        120       240       360

Pooled StDev = 310,1
```

Fig. 2. ANOVA Results

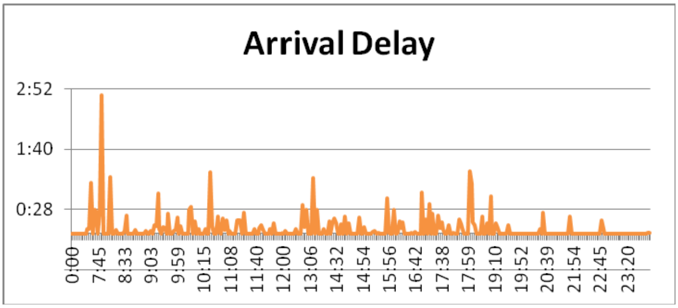


Fig. 3. Arrival Delays

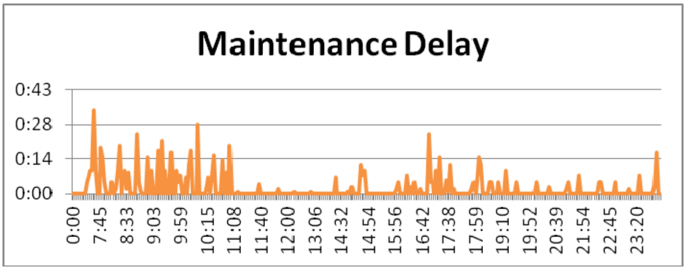


Fig. 4. Maintenance Delays

Table 3 shows the costs obtained throughout simulation for each of the five segments in which we have divided the daily schedule. Notice that in each case, the best estimated buffer can vary between 0 and 10 minutes.

In effect, according to **Table 3**, flights attended between 0:00 and 7:00 hours, between 12:00 and 16:00 hours, and between 20:00 and 0:00 hours, do not need a buffer of time. On the contrary, flights between 7:00 and 12:00 hours and between 16:00 and 20:00 hours would benefit from a time buffer of about 5 minutes. This segmentation shows that the daily operation can be divided in time slots, thus increasing the accuracy of the solution given, i.e. by selecting the appropriate buffer size in each segment, total costs are reduced and the efficiency of the turnaround process is enhanced.

Table 3. Segmented Results

Hours	Buffer time (b) in minutes			
0:01 - 7:00	**b = 0**	**b = 5**	**b = 10**	**b = 15**
Cost_Delay	75,00 €	40,91 €	18,18 €	0,00 €
Cost_Extra	0,00 €	69,09 €	146,82 €	228,00 €
TOTAL_COST	75,00 €	110,00 €	165,00 €	228,00 €
Hours				
7:01 - 12:00	**b = 0**	**b = 5**	**b = 10**	**b = 15**
Cost_Delay	328,16 €	259,71 €	206,55 €	167,23 €
Cost_Extra	0,00 €	42,98 €	97,58 €	162,70 €
TOTAL_COST	328,16 €	302,69 €	304,14 €	329,93 €
Hours				
12:01 - 16:00	**b = 0**	**b = 5**	**b = 10**	**b = 15**
Cost_Delay	160,71 €	113,57 €	80,71 €	58,21 €
Cost_Extra	0,00 €	59,17 €	129,20 €	207,10 €
TOTAL_COST	160,71 €	172,74 €	209,91 €	265,31 €
Hours				
16:01 - 20:00	**b = 0**	**b = 5**	**b = 10**	**b = 15**
Cost_Delay	233,78 €	179,73 €	138,51 €	106,42 €
Cost_Extra	0,00 €	53,92 €	117,59 €	188,20 €
TOTAL_COST	233,78 €	233,65 €	256,11 €	294,62 €
Hours				
20:00 - 0:00	**b = 0**	**b = 5**	**b = 10**	**b = 15**
Cost_Delay	50,00 €	25,70 €	16,20 €	9,15 €
Cost_Extra	0,00 €	78,02 €	166,68 €	257,36 €
TOTAL_COST	50,00 €	103,72 €	182,88 €	266,52 €

5 Conclusions

In this paper, we have presented a simulation-based approach for solving the Aircraft Turnaround Problem (ATP). The ATP is a challenging research area because it introduces random behavior into the problem. We deal with these stochastic variables by employing simulation to analyze different scenarios, each of them characterized by a different time-buffer size. The analysis is performed using real-life data and segmenting the daily activity schedule into five sections. The tests performed show the convenience of assigning a different buffer time to each segment, according to the variations in arrival and maintenance delays. For future work, we plan to adjust the real-life data representing uncertainty in the system by using some probability distributions, e.g. Weibull or LogNormal.

References

1. ACI-Europe, Official News Bulletin for Members, Airport Council International – Europe: Brussels (2000)
2. Wu, C.L., Caves, R.E.: Modeling and Simulation of Aircraft Turnaround Operations at Airports. Routledge (2007)
3. Piotr Lukaszewicz, P.: Metaheuristics for Job Shop Scheduling Problem. Comparison of Effective Methods, 11–12 (2005)
4. Kuster, J., Jannach, D.: Handling Airport Ground Processes Based on Resource-Constrained Project Scheduling. Advances in Applied Artificial, 166–167 (2006)
5. Wu, C.L.: Airline Operations and Delay Management: Insights from Airline Economics, Networks and Strategic Schedule Planning. Journal of Airport Management 4(4), 398–399 (2010)
6. Fricke, H., Schultz, M.: Delay Impacts onto Turnaround Performance. ATM Seminar (2009)
7. Wu, C.L., Caves, R.E.: Aircraft Operational Costs and Turnaround Efficiency at Airports. Journal of Air Transport Management 6(4), 201–208 (2000)
8. Wu, C.L., Caves, R.E.: Modeling and optimization of aircraft turnaround time at an airport. Transportation Planning and Technology 27(1), 25–46 (2004)
9. Cook, A., Tanner, G., Anderson, S.: Evaluating the True Cost to Airlines of One Minute of Airborne or Ground Delay (2004), http://westminsterresearch.wmin.ac.uk

Sovereign Bond Spreads Evolution in Latin American Countries

Antonio Terceño, M. Teresa Sorrosal, Lisana B. Martinez, and M. Glòria Barberà

Universitat Rovira i Virgili, Faculty of Business and Economics
Department of Business Management
Avda Universitat 1, Reus 43204, Spain
{antonio.terceno,mariateresa.sorrosal,lisanabelen.martinez,
gloria.barbera}@urv.cat

Abstract. This paper presents an empirical analysis related to sovereign bond spreads variations of seven Latin American countries during 2000 to 2012. The self-organizing maps methodology allows identifying the similarities among the financial markets evolutions. Moreover, we detect the impact of the Argentina financial crisis and the recent contagion effect of the last financial crisis over the emerging region during the year 2009.

Keywords: Self-organizing maps, sovereign bond spreads, financial crisis, Latin American countries.

1 Introduction

Since the start of the recent financial crisis, in September 2008, most of countries have been affected in many economic and financial aspects. The sovereign bond spreads have increased markedly.

In particular, we are interested in studying the impact of the financial crisis on the spread dynamics. To fulfill the purpose of the article, self-organizing maps (SOM) are applied in order to identify different groups of countries since the year 2000 to 2012, according to spreads similarities.

SOM are suited for clustering tasks, since this kind of networks gathers elements according to their homogeneity considering all the attributes or variables defined by the researcher. One key advantage of this network is its unsupervised learning process that makes not necessary to define a priori groups. Groups are determined according to the similarities and differences of the elements considering all variables that form the patterns. SOM reduce the dimension of the input information (patterns are n-dimensional) to a bi-dimensional map. This fact reduces the complexity of dealing with n variables to a graphical interpretation. The planar location of the elements in the map keeps relation with the values of the n variables considered. In other words, the closer elements in the plane, the more homogeneous are the variables of these elements.

Given that the main objective is to analyze country spreads movements through the years, patterns are defined as a 12-dimensional 'country-year' vector where its

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 171–180, 2013.

elements represent the annualized spread and the monthly spread variations. The correlation among monthly sovereign bond spreads changes is low. Consequently, in order to perform our analysis, it is essential to apply a methodology that considers jointly all spreads values that define a pattern and group those with greater similarities.

The structure of the article is the following: section 2 presents a short review of sovereign bond spreads. Section 3 introduces self-organizing Kohonen maps. Section 4 shows the methodology and data used to the analysis. Section 5 presents the empirical results obtained. Finally, section 6 exposes the conclusions of the study.

2 Sovereign Bond Spreads

Since mid ´90 the new forms of globalization made the emerging financial markets more open to external shocks and investors pay less attention to country fundamentals. The rise in capital flows to emerging markets, driven by structural changes and the switch of macroeconomic policies make them more vulnerable to US interest changes, so interest rate of emerging markets tended to move in the same direction as US interest rate variations [1].

[2] state that many emerging markets have implemented some changes, in particular, into the corporate financing sector. The main purpose is to avoid the problem related to the original sin, which makes markets more vulnerable to external shocks since their debt are usually denominated in external currency, floated rate and short-term. These risky characteristics related to sovereign debt raise the economic vulnerability and the government probability to not being able to meet its obligations behind changes in external and internal conditions [3]. Securities of the emerging markets were characterized by high country risk premiums and high domestic interest rates in detriment of growth and income distribution. The changes of the country specific fundamentals and of market sentiments drove fluctuations in spreads as a consequence of the financial contagion effect during financial crisis.

The literature related to sovereign bond spreads of emerging markets is considerable. Numerous studies focus on its determinants and also on the spread evolution during the nineties financial crisis. However, there is a scarcity of works with respect to the last financial crisis effects over Latin American countries.

Several empirical works have analysed the effect of factors on the sovereign bond spreads of emerging countries. On the one hand, some empirical works emphasize the importance of domestic factors such as variations in fundamentals, liquidity and solvency variables and the importance of policy and fiscal variables. On the other hand, several works emphasize the influence of global factors, such as global liquidity, risk appetite and contagion effects.

[2] found that liquidity and solvency variables explain most of the variations of bond spread in 11 emerging markets[1] during 1990s. Furthermore, those authors found that the US interest rate and some macroeconomic fundamentals play a significant role as determinants of bond spreads.

[1] Five of them are Latin American countries, considered in our work, and the rest ones are from Asia.

Recently, [4] analyze the key drivers of country risk premiums measured by sovereign bond spreads of 46 emerging countries during 1997 to 2008. Those authors find that lower levels of political risk are associated with reduced spreads, in particular during financial turmoil when markets are more vulnerable to institutional instability. Furthermore, those authors find that fiscal factors matter for credit risk in emerging markets since countries with both high deficits and debt have greater risk of default. Consequently, they conclude that despite the important role of global financial conditions and the investor's risk appetite, domestic fundamentals are important to reduce the borrowing cost of this kind of economies.

Otherwise, many authors emphasize the role of global factors to establish the main determinants of sovereign bond spreads. [5] explain that global factors, such as risk appetite, global liquidity and contagion effects explain a large variation of country sovereign risk.

[1] and [6] consider changes in market sentiments as important drivers of sovereign bond spreads related to global factors determinants. [7] probed some spreads changes over time and found that they are explained mainly by shifts in market sentiments than by variations in fundamentals. Their results show that higher credit quality leads to higher probability of issue and to lower spreads, which could be interpreted like the market discrimination among issues according to risk. Furthermore, it is significant that both factors are important drivers of the EMBIG (Emerging Markets Bond Index Global) spreads. [8] find that fall in sovereign spreads between 1995 and 1997 could not be totally explained by an improvement in fundamentals. Similarly, [9] confirm that sovereign bond index rates may be explained by global market factors, local risk factors and other country specific factors.

Furthermore, [10] analyse the impact of global financial conditions, US macroeconomics news and domestic fundamentals on the evolution of the EMBI spreads between 1997 and 2006 for 18 emerging countries. Their results suggest that the long-run evolution of EMBI spread depends on external factors related to global liquidity conditions, risk appetite, crisis contagion and the effect of US macroeconomics news over emerging markets.

Another strand in the literature stresses the importance of the contagion effect of financial and economic crisis originated among emerging countries sharing economic linkages. This effect was an important channel to propagate financial crisis during the nineties. For instance, [11] probes that the contagion effect during the currency crisis in emerging markets during 1990 has been a key channel of transmission, since the high degree of real and financial interdependence among the economies.

The first great financial crisis of the 21st century is producing considerable global impact, not only because of its scale but because of the fact that it has been originated in a developed economy such as US and consequently the financial effects spillover to many different countries.

The financial impact of the global crisis on Latin America has been in some aspects less severe than previous ones. [12] analyze the transmission of the US subprime crisis to emerging markets since early 2007 to February 2009. They find that emerging countries have not been affected at the beginning of the turmoil, but also that during this brief period emerging markets had undertaken reforms (lower public debt and large international reserves) to insulate from adverse shocks from the rest of the world (called decoupled effect). However, after September 2008 emerging markets started to experience a strong deterioration, since the credit crunch and the decline in the international trade spillover (recoupling effect).

[13] examines the sovereign emerging market bond spreads during 1997-2009 and identifies that the volatility of stock markets, the central bank transparency and the onset of the last global financial crisis are the main determinants of the bond spreads. Moreover, he finds that the global financial crisis did not raise bond spread in Asia and Africa. So, this last result should be interpreted as these markets are not synchronized with other parts of the world. In the same way, [14] also finds that the pattern reaction of these two regions in front of the aftermath of the financial crisis is different, being very fast and cyclical for Latin America and regular and slower in East Asia.

3 Self-Organizing Kohonen Maps

As we have mentioned, we try to analyze the spreads in Latin American countries, focusing our study on the homogeneity of its evolution. In order to answer this question, we propose a methodology that is appropriate for clustering problems, grouping different elements (or patterns) according to the similarity between all the characteristics that define them. This methodology is a particular kind of artificial neural network, named Self-Organizing Maps (SOM).

SOM were developed by Kohonen and they are inspired in the way that the brain stores information, which is, placing similar information in closer areas of the brain. Analogously, the output of a SOM is a bi-dimensional map where the input patterns are located (and grouped) through a distance function that measures the degree of similarity between them. In this way, the closer two patterns are in the map, the more similar they are.

Regarding SOM architecture, they are formed by two layers of neurons: an input layer with as many neurons (or units) as variables are used to form the patterns, and an output layer consisting of a bi-dimensional map. Between the two layers, there are feed-forward connections and, moreover, in the output layer we find lateral and auto-recurrent connections. These two last types of connections allow the competition process that is characteristic of the unsupervised learning in SOM. After the learning process, only one neuron (the winner neuron) remains active in the output layer, indicating the place in the map that occupies the pattern whose information has been introduced into the system.

The number of groups can be established by an optimization problem (which minimizes the number of groups and maximizes the homogeneity of the elements within the group) or defining a concrete number a priori.

We can summarize the behavior of a SOM network in the following steps:

1. First of all, we define the patterns (input data) as vectors of n components: $x^p = (x^p_1, x^p_2, \ldots x^p_i, \ldots x^p_n)$, where p refers to the specific pattern and the subscript $i=\{1, 2, \ldots n\}$ corresponds to each one of the variables that are used to describe the data.

2. The input data are propagated to the output layer through feed-forward connections with an associated weight $w_{k,i}$, where the subscript ki indicates the connection between the neuron i of the input layer and the neuron $k=\{1, 2, \ldots m\}$ of the output layer. Initially, weights are random values.
3. Using a measure of distance (the Euclidean distance, $d^p_{k},=[\sum(x^p_{i,} - w_{k,i})^2]^{1/2}$, in our case) between each of the input vectors and the vector that are formed by the weights that arrives to each one of the neurons of the output layer, we obtain the winner neuron for this pattern (it is the output neuron k^* that satisfies $d^p_{k*} = min\ [d^p_k]_{k=1,\ldots m}$).
4. The weights of the winner neuron are modified (also the ones which belong to its neighborhood area, formed by adjacent units in a rectangular or hexagonal area that decreases with time). Weights in time t+1 are obtained as $w_{k,i}\ (t+1)= w_{k,i}\ (t)+ \alpha(t)\cdot\ [x^p_{i,} - w_{k*,i}\ (t)]$, where $\alpha(t)$ is the learning coefficient and takes a value between 0 and 1, decreasing with time in order to ensure the convergence to the equilibrium of the network.
5. These steps are repeated with all the patterns of the input data and with the necessary number of iterations until we get the stability of the association between the different patterns and the same unit of the output layer.

After this process, we obtain a bi-dimensional map where the patters are distributed in function of the similarity between all the characteristics that have been used to describe them. Its relative position respect the other patterns shows the degree of similarity between them, and its absolute position have to be understood for the specific values of each component in the map. This information is obtained from the SOM in the form of components maps.

The unsupervised learning process that is proper to this kind of networks allows self-organizing capabilities that are used in the main applications of the SOM, that are, among others, pattern recognition, resolution of optimization problems and clustering (or grouping).

SOM have been applied to problems of grouping in the financial context [15], [16], [17], [18] and particularly to financial crises analysis [19], [20], [21], [22]. In these works, SOM are used to examine visual predictions of financial crises, for analyzing the bankruptcy risk of companies, etc. In all cases, SOM show great descriptive and predictive ability.

4 Methodology and Data

Our database comprises 7 Latin American countries: Argentina, Brazil, Chile, Colombia, Mexico, Venezuela and Peru. United States is our benchmark to estimate bond spreads. The selection of the countries is based on the most important markets of the region and on the available information of the variables required.

We work with a total of 91 patterns corresponding to data of 7 countries during 13 years. The patterns are defined as vectors of 12 components or variables. The first one is the annual spread of each year of the sample and the rest 11 variables are the spreads variation between two consecutive months, *i.e* between January and February (variable 2), February and March (variable 3) and so on to November and December (variable 12).

The emerging market sovereign bond spreads are measured by the EMBIG elaborated by JPMorgan. The EMBIG is the most comprehensive emerging markets benchmark which considers US-dollar denominated Brady Bonds, Eurobonds, traded loans and local market debt instruments issued by sovereign and quasi-sovereign organism.

The spread is a standard measure of sovereign default risk, commonly known as country risk. This benchmark for emerging countries has been broadly used by many previous works [4], [5], [8], [23], [24], among others.

Data is obtained from DataStream, for the period 2000 to 2012 measured at the end of each month.

SOM network is implemented in Matlab using the toolbox developed by the Laboratory of Information and Computer Science in the Helsinki University of Technology.

The input layer of the SOM consists of 12 units. Each of them contains the value of one of the variables that define the patterns.

When the network is implemented for all the countries selected, the resulting output is of 6x8 dimensions, forming 6 groups. This is the result of a dual criteria objective function, which minimizes the number of groups and maximizes the homogeneity of the patterns within each group.

5 Empirical Results

In this section we present the results considering all countries presented applying SOM. Figure 1 shows the Kohonen map and the grouping related to this analysis.

Arg02 Arg06, Bra06 Ven06, Per06 Col06, Mex06 Ven12	Chi06 Arg12	Arg05 **1**	Chi12 **2**	Ven10 Col12, Bra12, Per12, Mex12	Bra10, Col10, Per10, Mex10
	Arg00, Ven00, Per00, Mex00	Col00 Ven05			Arg10, Chi10
Chi05, Col05, Bra05, Mex05 Per05,		Bra00	Bra04, Mex04 Per04	Col04 Ven04	
	Chi04	Chi09	Per01	**3**	Chi02, Bra02, Ven02, Per02, Mex02
Bra09, Col09 Mex09, Per09	Chi03, Mex03 Bra03, Per03	Col01 Col03	Chi01		Col02
Ven09	Ven03 Arg04	**4**	Chi11, Bra11, Col11, Mex11, Ven11, Per11	Arg11	Arg07
Ven01 Arg03 Arg09				**5**	Chi07
Arg01, Mex01, Bra01	Chi00 Ven08	Arg08, Chi08	Bra08, Per08 Col08, Mex08	**6**	Bra07, Col07 Ven07, Per07 Mex07

Fig. 1. Self-organizing Kohonen map

There are 6 different groups of patterns. As could be observed, inside these groups there is an interesting year grouping among countries. The first group is the one which includes the wider number of patterns of different years. The fourth group also presents different chronological data and the rest of groups contain at most three diverse years. This dispersion is related with the similarity between monthly spreads variations.

To analyze each pattern position in the map, it is necessary to evaluate the value of all the variables (normalized) in the corresponding area. This information is represented in Figure 2 through the scale values represented by colors, which results in numeric scale to the right of each map. In all cases, the dark blue color indicates the minimum values of each variable, while red indicates the highest values.

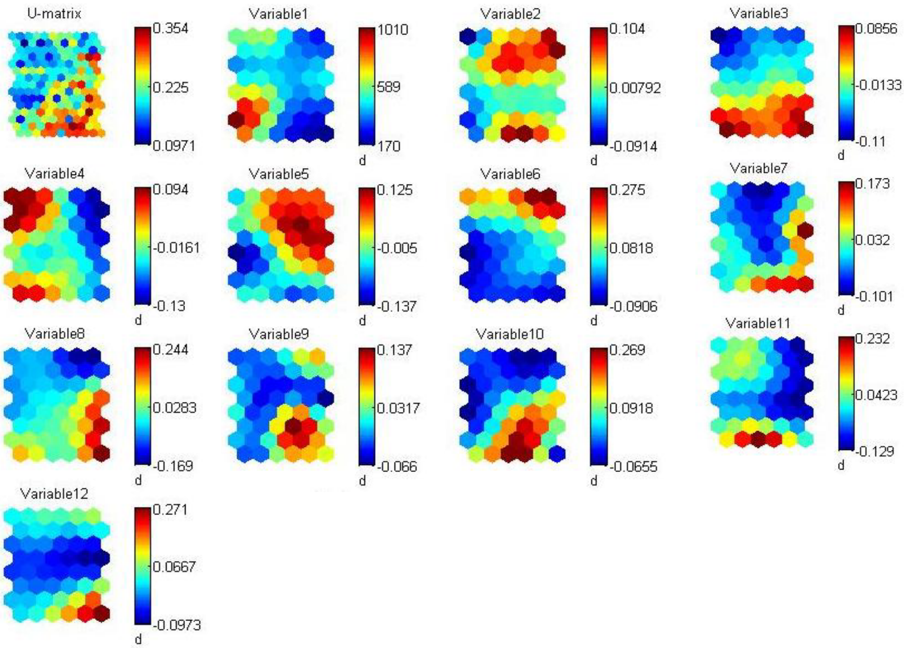


Fig. 2. Distance and components maps

Considering the distribution presented in variable 1 (annual sovereign bond spreads) in the characteristics map, we observe that on the right bottom are located those variables with the lower spreads variation, meanwhile, on the left are those with show the higher values.

According to these results there are some interesting aspects to highlight. As could be appreciate in Figure 1, Argentina is a particular country since 2001 as a consequence of the default crisis impact, which produced high bond spreads variations during the next year which lasted until 2005.

Moreover, Chile presents some discrepancies respect to the block of countries during the years 2000 and 2009. An interesting aspect of this county is the small size of the public debt market related to the private bonds market volume. This characteristic may affect sovereign bond spreads evolution in addition to the economic and financial stability of this country related to the rest of the sample.

In 2001, Peru is located in another country-group. So, according to these results we observe that besides some specific variables, sovereign bond spreads of these emerging markets behaves similarly.

Another important aspect to highlight is the financial crisis contagion effect over the emerging markets considered.

As mention [25] in mid-2007 the financial crisis took off and had a climax in September 2008 with the collapse of Lehman Brothers, which marks a point of inflection in the crisis.

During the first two years since the start of the financial crisis in US, there are not important variations in the spreads levels in any of the emerging markets analyzed; besides some of them suffered some economics problems related to the global impact of the crisis. According to [26], Latin American countries have been better prepared during previous crisis and it will serve to moderate the impact of the greatest financial crisis of the world.

Nevertheless, as could be appreciated in Figure 1, the year 2009 is represented by high values according to the variable 1 of the Figure 2 of the Kohonen map. That means that besides the financial crisis effects were slowly and sometime later over these countries, the emerging financial markets were affected anyway.

The following years do not present in general great sovereign bond spreads variations, except the year 2012 for Argentina and Venezuela which show high spreads oscillations during this last year.

An important aspect to highlight is that in the present analysis we work with sovereign bond spreads variations in order to obtain, according to the SOM, a country-year classification based on the similarities between its spreads evolution. So, it is pertinent to say that these last countries mentioned, Argentina and Venezuela, present during all the period considered in the sample very high spreads levels related to the rest of the countries included.

6 Conclusions

In this paper we analyze sovereign bond spreads evolution for seven Latin American countries from 2000 to 2012. During this period there are two important events in the financial environment. The first one was in Argentina at the end of 2001 when the government finished the convertibility period and was declared in default.

The second is the great financial and economic crisis originated in US and propagated to the rest of the developed economies. So, we focus on the behavior of sovereign bond spreads of some emerging markets during the whole period applying an interesting methodology denominated self-organizing Kohonen maps in order to appreciate the possible existence of a homogeneous movement in the zone.

This methodology allows clustering elements according to the overall similarity and makes unnecessary to define a priori groups. As we could observe, it is possible to find a chronological association by years. In general, all countries present similar spreads variations during the whole period of analysis. This last result is not appropriated to Argentina since 2002 until 2005 as a consequence of the financial and economic crisis that took place during that period in that country. Moreover, during the last year of analysis, Venezuela and Argentina do not present similar behavior to the rest of the countries, given their high volatility in their financial markets as a consequence of the economics problems.

Moreover, as it was expected, the last financial crisis affected some time latter this markets, being affected during the year 2009 as a consequence of the financial contagion effect from the rest of the world. However, this result affected partially these economies during a brief period of time. This is an interesting and contradictory result respect to the impact over European financial markets.

Finally, we need to consider the limitation of the methodology used in the work in the sense that it is able to make groups of patterns but the interpretation is left to the author. The helpfulness of the results will depend on the definition of the patterns and on the knowledge of the context.

Further research might include the appraisal of using more detailed information about the components of the bond spreads as variables of study. This modification could widen the analysis about the characteristics and evolution of the debt in Latin American countries. Additionally, it would be also interesting to compare the results with the ones obtained in other geographical areas, as for instance in the Asian countries.

References

1. Arora, V., Cerisola, M.: How does U.S. Monetary Policy Influence Economics Conditions Emerging Markets? International Monetary Fund. Working Paper /00/48 (2000)
2. Min, H.G., Lee, D.H., Nam, C., Park, M.C., Nam, S.: Determinants of Emerging-Market Bond Spreads: Cross-Country Evidence. Global Finance Journal 14, 271–286 (2003)
3. Guercio, M.B., Terceño, A.: Nuevas Emisiones, Nuevos Riesgos: La Experiencia de las Economías Latinoamericanas. Análisis Financiero 116, 34–46 (2011)
4. Baldacci, E., Gupta, S., Mati, A.: Political and Fiscal Risk Determinants of Sovereign Spreads in Emerging Markets. Review of Development Economic 15, 251–263 (2011)
5. González-Rozada, M., Levy-Yeyati, E.: Global Factors and Emerging Market Spreads. The Economic Journal 118, 1917–1936 (2006)
6. Eichengreen, B., Mody, A.: What Explains Changing Spreads on Emerging Market Debt? National Bureau of Economic Research Working Paper 6408, 1–48 (1998)
7. Kamin, S.B., von Kleist, K.: The Evolution and Determinants of Emerging Market Credit Spreads in the 1990s. BIS Working Papers 68, 1–33 (1999)
8. Ferrucci, G.: Empirical Determinants of Emerging Market Economies Sovereign Bond Spreads. Bank of England. Working Paper 205, 1–42 (2003)
9. Jüttner, D., Chung, D., Leung, W.: Emerging Market Bond Returns – An Investor Perspective. Journal of Multinational Financial Management 16, 105–121 (2006)
10. Özatay, F., Özmen, E., Sahinbeyoglu, G.: Emerging Market Sovereign Spreads, Global Financial Conditions and U.S. Macroeconomic News. Economic Modelling 26, 526–531 (2009)

11. Fratzscher, M.: On Currency Crises and Contagion. International Journal of Finance and Economics 8, 109–129 (2003)
12. Dooley, M., Hutchison, M.: Transmission of the U.S. Subprime Crisis to Emerging Markets: Evidence on the Decoupling–Recoupling Hypothesis. Journal of International Money and Finance 28, 1331–1349 (2009)
13. Siklos, P.: Emerging Market Yield Spreads: Domestic, External Determinants, and Volatility Spillovers. Global Finance Journal 22, 83–100 (2011)
14. Edwards, S.: The International Transmission of Interest Rate Shocks: The Federal Reserve and Emerging Markets in Latin America and Asia. Journal of International Money and Finance 29, 685–703 (2010)
15. Serrano-Cinca, C.: Self-organizing neural networks for financial diagnosis. Decisions Suport Systems 17, 227–238 (1996)
16. DeBoeck, G., Kohonen, T.: Visual Explorations in Finance: with Self-Organizing Maps. Springer, London (2000)
17. Moreno, D., Marco, P., Olmeda, I.: Self-Organizing Maps could improve the classification of Spanish mutual funds. European Journal of Operational Research 174(2), 1039–1054 (2006)
18. Sarlin, Marghescu: Visual predictions of currency crisis using self-organizing maps. Intelligent Systems in Accounting, Finance and Management 18, 15–38 (2011)
19. Sarlin, P., Peltonen, T.A.: Mapping the State of Financial Stability. Working Paper Series 1382. European Central Bank (2011)
20. Fioramanti, M.: Predicting Debt Crises Using Artificial Neural Netwoks: A comparative approach. Journal of Financial Stability 4, 149–164 (2008)
21. Du Jardin, P., Séverin, E.: Predicting Corporate Bankruptcy Using a Self-Organizing Map: An Empirical study to Improve the Forecasting Horizon of a Financial Failure Model. Decision Support Systems 51, 701–711 (2011)
22. Chen, N., Ribeiro, B., Vieira, A., Chen, A.: Clustering and Visualization of Bankruptcy Trajectory using Self-Organizing map. Expert Systems with Applications 40, 385–393 (2013)
23. Ciarlone, A., Piselli, P., Trebeschi, G.: Emerging Markets Spreads and Global Financial Conditions. International Financial Markets, Institutions and Money 19, 222–239 (2009)
24. Gumus, I.: Exchange Rate Policy and Sovereign Spreads in Emerging Market Economies. Review of International Economic 19, 649–663 (2009)
25. Aβmann, C., Boysen-Hogrefe, J.: Determinants of Government Bond Spreads in the Euro Area in Good Times as in Bad- Kiel Institute for the World Economy. Working Paper 1548 (2009)
26. Ocampo, J.: The Impact of the Global Financial Crisis on Latin America. CEPAL Review 97, 9–32 (2009)

A Model of Currency Crisis under the Possibility of Government Resignation*

Javier I. Garcia-Fronti

Facultad de Ciencias Económicas
Universidad de Buenos Aires, Argentina
javier.garciafronti@economicas.uba.ar

Abstract. In Argentina, a peg with the dollar -established in 1991- produced good economic results during the first half of that decade. However, from 1997 onwards, the monetary policy ('*Convertibilidad*') has been accompanied by a deep economic recession. Despite governmental attempts to keep the fixed exchange regime, external factors and high level of local interest rates forced the resignation of the policymaker in December 2001 and the new government immediately devalued the currency. This discussion is topical now, regarding the present Greek crisis. This paper extends the Ozkan and Sutherland (1998) model to contemplate the collapse of the fixed exchange regime that is determined by the resignation of a completely committed policymaker to a fixed exchange regime.

Keywords: Currency crisis, devaluation.

Introduction

The currency crises models could be classified into three main types: first, second and third generation. Following Agénor and Krugman, first generation models focus on certain 'inconsistent fundamentals' as explanation of currency crises [1] [2], while the second generation ones concentrate on the behaviour of rational policymakers who decide when to leave the fixed regime [3]. Recently, the so called 'third generation' models have emerged, focusing in the role played by the private sector [4].

The Argentinean economic situation in 2001, which combined economic recession and fixed exchange regime, has drawn a particular scenario that challenges previous models and explanations of currency crises. Moreover, this national case has particular international relevance for two main reasons. On the one hand, during the 1990s, the International Monetary Fund (IMF) has presented Argentina's monetary regime (the 'Convertibilidad') as an example for other emergent markets. On the other hand, during the year 2000, the Argentinean bond market represented the 25% of the total of worldwide emergent bond markets.

* This work was written as part of the research Project PICT 2011-0919, "Gobernanza financiera: Las propuestas de regulación y sus impactos socioeconómicos. El caso de Argentina".

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 181–189, 2013.

During the 1980s, Argentina presented a complex and changing economic situation, suffering from peaks of hyperinflation to lack of investors' confidence. At the beginning of the 1990s, the legal establishment of a new exchange rate regime ('Convertibilidad') inaugurated a decade where one peso was equal to one dollar. This regime immediately produced positive macro-economic effects such as the end of the hyperinflation process and the steadily economic growth for a period of 6 years. However, the 'miracle' of the 'Convertibilidad' was accompanied by negative long term economic and social consequences that were apparent from 1997 onwards. Finally, during 2001, the high local interest rates have produced a deep economic recession and high and increasing levels of unemployment. In addition, Argentina had a completely committed policymaker to the fixed exchange regime who seemed to disregard the economic and social costs of the recession.

This work extends the Ozkan and Sutherland [3] model to understand a currency crisis under the possibility of government resignation. Formally this resignation is model adding a Poisson process to the external shock. To do so, it is organised into two sections. The following one describes the model proposed by Ozkan and Sutherland in 1998 [3]. The next chapter depicts an extension that contemplates the case of a policymaker 'completely committed' to the fixed exchange regime and her/his resignation.

1 The Ozkan and Sutherland (OS) Model

This chapter presents the main assumptions, principles, variables and algorithms proposed by Ozkan and Sutherland (1998) to explain the ERM crisis in Great Britain.

According to the authors, the policymaker's main problem is the maximisation of the utility function (W), which is represented by the following equation [3]:

$$W(t) = E_t\left[\int_t^{\infty}\left(Z - y^2(\tau)\right)e^{-\delta(\tau-t)}d\tau\right] \tag{1}$$

where E_t is the expectation conditional on information at time "t", the variable δ is the policymaker's discount factor, "y" is the index of demand pressure and "Z" is the benefit of having a fixed exchange system.

In this interpretation, the demand pressure index (y) depends on the interest rate, the exchange rate and the external demand shock as indicated by the following expression:

$$y = -\gamma.i_\tau + \eta.s - \beta.\varepsilon \tag{2}$$

Where i_τ is the interest rate of a zero coupon bond with maturity τ, the variable s is the log of the nominal exchange rate and the variable ε is an external shock with $d\varepsilon = \sigma.dw$ and w is a standard Brownian motion. In other words ε is a driftless Brownian motion with variance σ.

The external demand shock (ε) causes deviations of the index of demand pressure from its optimal value of zero in two ways. Directly, by factor β and indirectly through the value of the interest rate. The policymaker aims to stabilise the demand pressure around a target level of zero. The model assumes that the policymaker has only two possible exchange rate options: a fixed exchange regime or a floating one. In this context, he/she has to select the trigger point to switch [3].

To solve the model it is crucial to distinguish between the private sector's expected trigger points and the policymaker's optimal choice. The expected trigger points[1] will be designated by $\varepsilon_e^u = -\varepsilon_e^l$, while the policymaker's optimal choice of trigger points will be designated by $\varepsilon^u = -\varepsilon^l$. In this way, 'rational expectations equilibrium' requires that the policymaker's optimal triggers coincide with the expected ones by the agents in the foreign exchange market ($\varepsilon^u = \varepsilon_e^u$, $\varepsilon^l = \varepsilon_e^l$).

In this model, the interest rate follows the 'uncovered interest rate parity'. However, the authors incorporate different maturities (τ) (considering pure-discount bonds). They define the expression for the interest rate for each maturity, assuming that the variable 's' is initially fixed, i.e. ($s(t) = \bar{s} = 0$), and the foreign interest rate is zero. So i_τ is given by:

$$i_\tau = \frac{E_t\left[s\left(t+\tau\right)\right]}{\tau} \tag{3}$$

In this model, two main factors affect the interest rate (i_τ): the size of the step change in the exchange rate when the country leaves the fixed regime and the probability of hitting the trigger point. As equation (3) shows, the interest rate reflects the expected future spot rate. The expected spot rate is a function of the external shock, the maturity and the trigger points expected by the investors [the higher the value of external shock (ε), the longer the maturity or the smaller the trigger points, the higher the probability to hit the trigger]. This is summarized in the following equation:

$$i_\tau = \frac{E_t\left[s\left(t+\tau\right)\right]}{\tau} = \frac{h\left(\varepsilon, \tau, \varepsilon_e^l, \varepsilon_e^u\right)}{\tau} \tag{4}$$

where $\varepsilon_e^l, \varepsilon_e^u$ are the expected devaluation trigger points. Following Guenther and Lee [6], the solution for $h(\varepsilon, \tau)$ is:

$$h\left(\varepsilon, \tau\right) = \frac{\beta}{\eta}\varepsilon + \sum_{n=1}^{\infty}\phi_n.\exp\left(\frac{\sigma^2 n^2 \pi^2 \tau}{2.\left(\varepsilon_e^u - \varepsilon_e^l\right)^2}\right)\sin\left(\frac{n\pi\left(\varepsilon - \varepsilon_e^l\right)}{\left(\varepsilon_e^u - \varepsilon_e^l\right)}\right) \tag{5}$$

[1] The model uses a symmetry condition, that is the upper limit is the opposite to the lower limit.

where $$\varphi_n = -\frac{2\beta}{\eta n\pi}\left(\frac{\left(\varepsilon_e^u - \varepsilon_e^l\right)}{n\pi}\sin(n\pi) - \varepsilon_e^u.\cos(n\pi) + \varepsilon_e^l\right)$$

To compute this interest rate function ($i\left(\varepsilon, \tau, \varepsilon_L^e, \varepsilon_u^e\right)$) a Matlab programme is presented in Appendix.

While the previous analysis has taken the trigger points as exogenous, this section includes them as part of the solution of the policymaker to find the "time consistent" triggers. The policymaker must choose the optimal trigger points maximising the utility function W(t) and taking private sector expected trigger points as given. For each expected trigger, the policymaker chooses an optimal trigger and a consistent solution that requires a chosen trigger to be equal to that expected by the investors.

To maximise W(t), the authors make a transformation to a function $V(\varepsilon)$ that it is only function of the shock. This allows us to solve the model as a function of a state variable. In mathematical terms:

$$\textbf{MAX } W(t) = V(\varepsilon) \tag{6}$$

Now the problem is to solve the ordinary differential equation:

$$\frac{\sigma^2}{2}V^{"}\left(\varepsilon\right) = \delta.V\left(\varepsilon\right) + y^2(\varepsilon) - Z \tag{7}$$

Because the value of the utility function is zero in the float regime, there are two boundary conditions for guarantee the "Value matching condition" at trigger points, i.e.

$$V(\varepsilon^l) = 0 \qquad\qquad V(\varepsilon^u) = 0 \tag{8}$$

And a "smooth pasting condition" as an analogy of the "one time investment" [6].

$$V^{'}(\varepsilon^l) = 0 \qquad\qquad V^{'}(\varepsilon^u) = 0 \tag{9}$$

2 Exogenous Probability of Switching of Policy Makers

To explain the Argentina 2001 currency crisis, it is necessary to make an extension that incorporates certain particularities not included in the OS model: the presence of a policymaker who is completely committed to the fixed exchange rate regime and his/her replacement by another policymaker with standard OS behaviour.

This part analyses how to model the optimising problem in the case of a country where the agents of the foreign exchange market make decisions based on the assumption that both the strong commitment fixed exchange policymaker and the standard one have equal probabilities to be in office.

This model takes for granted that a "completely committed' policymaker to a fixed exchange regime (Bigger Z value) is in office and his/her resignation could be modelled by a Poisson process. Additionally, this model assumes that this resignation will be followed by the collapse of the fixed exchange regime and its replacement by a floating one. Finally, this variation also supposes that the policymaker and the agents in the foreign exchange market are aware of the Poisson process parameter which is exogenous to the model.

In mathematics terms, the driver of the model (the external shock) is a more complex stochastic process which mix Brownian motion with a Poisson process (i.e.: $d\xi = \sigma.dz + \lambda.dq$). After the jump event of the Poisson process, the policymaker will resign and the utility function will fall to zero. Consequently, the fixed exchange regime will collapse and a floating exchange regime will be established. To calculate this solution, it is necessary to use the 'modified Ito lemma' [7], tool that offers a framework to solve this kind of mathematical optimising problem including Poisson processes.

In the original model:

$$E_t\left[dV\right] = (y^2 - Z).dt + \delta.V(\varepsilon).dt \tag{10}$$

Assume now that the shock is a brownian motion plus a poisson proccess. That is,

$$d\xi = \sigma.dz + \lambda.dq \tag{11}$$

Nothing changes in (10) if change the definition of the shock,

$$E_t\left[dV\right] = (y^2 - Z).dt + \delta.V(\xi).dt \tag{12}$$

Applying the modified ito's lemma:

$$E_t\left[dV\right] = \frac{\sigma^2}{2}V''\left(\xi\right).dt + \lambda.(-V\left(\xi\right)).dt \tag{13}$$

Equating (12)=(13),

$$E_t\left[dV\right] = (y^2 - Z).dt + \delta.V(\xi).dt = \frac{\sigma^2}{2}V''\left(\xi\right).dt + \lambda.(-V\left(\xi\right)).dt$$

Finnally,:

$$\frac{\sigma^2}{2}V''\left(\xi\right) = \left(\lambda + \delta\right).V\left(\xi\right) + y^2 - Z \tag{14}$$

where the variable λ is the parameter of the Poisson process. Moreover, this work develops a Matlab code for solving this numerical problem based on the book by Smith [8].

This final equation (14) could help to understand the Argentinean devaluation of 2001. The Poisson process, representing the external probability of new government, anticipate the devaluation because it increase the policymaker's discount factor of the original second generation model.

3 Conclusion

Summing up, this work presented a portray of the Ozkan and Sutherland (1998) model. To understand the Argentinean 2001 currency crisis, this paper developed an extension of the OS framework introducing a policymaker 'completely committed' to the fixed exchange rate regime who would be forced to resign and replaced by another standard policymaker.

This model assume an exogenous probability of new government, clearly this is a limitation of the model. In further research project will be important to endogenise it and calibrate to the Greek case [9].

References

1. Agénor, P.R.: The economics of adjustment and growth. La Editorial, UPR, London (2004)
2. Krugman, P.: A model of balance-of-payments crises. Journal of Money, Credit and Banking, 311–325 (1979)
3. Ozkan, F.G., Sutherland, A.: A currency crisis model with an optimising policymaker. Journal of International Economics 44(2), 339–364 (1998)
4. Krugman, P.: Balance sheets, the transfer problem, and financial crises. International Tax and Public Finance 6(4), 459–472 (1999)
5. Guenther, R.B., Lee, J.W.: Partial differential equations of mathematical physics and integral equations. Dover Publications (1996)
6. Dixit, A.: Art of Smooth Pasting, vol. 55. Routledge (1993)
7. Dixit, A.K., Pindyck, R.S.: Investment under uncertainty. Princeton UP, Princeton (1994)
8. Smith, G.D.: Numerical solution of partial differential equations: finite difference methods. Oxford University Press, USA (1986)
9. De Grauwe, P.: The Greek crisis and the future of the eurozone. Intereconomics 45(2), 89–93 (2010)

Appendix (Based on [8])

Program in Matlab 5.3 for solving the original OS model. This code need the presence of solve_ode.m in the computer

```
Main.m
global beta delta eta gamma sigma limit
beta=0.5;
delta=0.05;
eta=0.5;
gamma=0.5;
sigma=1.0;
` The precision level in the optimal trigger is 0.1
for limit=1.5:.1 :2
   limit
      h=(2*limit)/1000;
```

```
  global T1
     T1=-limit:h:limit+1; T1=T1';
  'Call a function to solve the ODE A(t)y'' + B(t)y' +
C(t)y = f(t)
  ' the solve_ode.m is from Smith (2001).
     [T,Y]=Solve_ode('a','b','c','f',-
limit,limit,0,0,1000);
     global T1
     T1=-limit:h:limit+1; T1=T1';
     s=zeros(size(T1));
     ' Defining the exchange rate function after the
collapse.
     for J = 1001:size(T1)
        s(J)=beta/eta*T1(J);
     end
     subplot(5,1,1),plot(T,Y,T1,zeros(size(T1)))
     legend('Utility function')
     xlabel('e')
     ylabel('V(e)')
subplot(5,1,2),plot(T,interest(T,1,limit),T1,zeros(size(T
1)))
     legend('Interest rate')
     xlabel('e')
     ylabel('i(e)')
     subplot(5,1,3),plot(T1,s)
     legend('Exchange rate')
     xlabel('s')
     ylabel('s(e)')
     r=input('Press to continue')
     ' Check the smooth pasting condition
     if (abs(Y(1000)-Y(999)))/(2*s/1000)<=0.05
        break;
     end
  end
```

```
  function y=a(t)
  global sigma
  y=sigma*sigma/2*ones(size(t));
```

```
  function y=b(t)
  y=zeros(size(t));
```

```
  function y=c(t)
  global delta
  y=-delta*ones(size(t));
```

```
  function y=f(t,ee)
  zeta=1*ones(size(t));
  global beta gamma limit T1
```

```
  D=+(gamma*interest(t,1,ee)+beta*t.^1).^2;
  y=+(gamma*interest(t,1,ee)+beta*t.^1).^2-zeta;
  subplot(5,1,4),plot(t,+(gamma*interest(t,1,ee)+beta*t.^
1),T1,zeros(size(T1)))
     legend('Demand preasure index')
     xlabel('e')
     ylabel('Y(e)')
     subplot(5,1,5),plot(t,zeta,T1,zeros(size(T1)))
     legend('Z Benefits of fixed regime ')
     xlabel('e')
     ylabel('Z(e)')
```

```
  function [y]=interest(t,tau,ee)
  global beta eta gamma sigma
  sumatoria=0.;
  for n=1: 1 : 100,
     phi=-2*beta/(eta*n*pi)*(2*ee/(n*pi)*sin(n*pi)-
ee*cos(n*pi)-ee);
     sumatoria=sumatoria+phi*exp(-
sigma*sigma*n*n*pi*pi*tau/(8*ee*ee))*sin(n*pi*(t+ee)/(2*e
e));
  end
  y=beta/eta*t/tau+sumatoria/tau;
```

Program in Matlab 5.3 for solving the OS model for Z=10
All the function all the same except for the definition of the parameter delta in the main function f()

```
Main.m
global beta delta eta gamma sigma limit
beta=0.5;
delta=0.05+0.5;
eta=0.5;
gamma=0.5;
sigma=1.0;
‘ The precision level in the optimal trigger is 0.1
for limit=1.5:.1 :2
      h=(2*limit)/1000;
global T1
   T1=-limit:h:limit+1; T1=T1';
‘ Call a function to solve the ODE A(t)y'' + B(t)y' +
C(t)y = f(t)
   [T,Y]=Solve_ode('a','b','c','f',-
limit,limit,0,0,1000);
   global T1
   T1=-limit:h:limit+1; T1=T1';
   s=zeros(size(T1));
```

```
   ` Defining the exchange rate function after the
collapse.
   for J = 1001:size(T1)
      s(J)=beta/eta*T1(J);
   end
   subplot(5,1,1),plot(T,Y,T1,zeros(size(T1)))
   legend('Utility function')
   xlabel('e')
   ylabel('V(e)')
subplot(5,1,2),plot(T,interest(T,1,limit),T1,zeros(size(T
1)))
   legend('Interest rate')
   xlabel('e')
   ylabel('i(e)')
   subplot(5,1,3),plot(T1,s)
   legend('Exchange rate')
   xlabel('s')
   ylabel('s(e)')
   r=input('Press to continue')
   ` Check the smooth pasting condition
   if (abs(Y(1000)-Y(999)))/(2*s/1000)<=0.05
      break;
   end
end
```

Utilizing Neuronal Calculus in Predicting Inflation

Stelian Stancu, Alexandra Maria Constantin, and Oana Madalina Predescu(Popescu)

The Bucharest University of Economic Studies, Romania
{stancu_stelian_ss,constantin_alexandra_maria}@yahoo.com

Abstract. The neuronal network, veritable instruments of optimal solution generating and diagnosis utilized in the field of artificial intelligence, tends to increase its spectrum of applicability reaching financial-banking predictions. This paper aims to achieve a study regarding the efficiency of applying neuronal networks, with different architectures, in the process of predicting inflation rates in Romania. Also, we will compare results estimated by applying neuronal networks with results obtained through predictions gained by applying classic econometric methods.

Keywords: neuronal calculus, GMDH, prediction, inflation rate, ARCH.

1 Introduction

Prognosis, a complex process, encapsulates models obtained through advanced econometric or data mining techniques and representative data from the dataset being considered. Regarding the degree of representation of the results about to be obtained, prognosis can be achieved by utilizing methods from the spectrum of neuronal calculus and econometric modeling.

By focusing on concise knowledge of results from the future time frame being analyzed, the necessity of using abstract methods and models with a high capability of concentrating future elements becomes apparent. Thus, more and more accent is put on the accuracy of results, in order to ensure minimal, controlled redundancy of the information on the basis of which unadulterated estimates are achieved.

In the last few years, researchers have generated diverse types of neuronal networks while predicting various macroeconomic indicators, wishing to obtain feasible results with major implications that can be compared with ones obtained through classic econometric modeling, using differences of various orders.

Starting from this context, it is imperative that we create neuronal networks with various structures, referring especially to the number of hidden layers, as the latter are those with a major role in handling and disseminating the information in the prediction process.

Researcher interest in the field of macroeconomic prognosis with regards to neuronal networks has reached stronger and stronger points, as neuronal calculus allows the solving of many complex problems efficiently, by applying easily implemented algorithms in automated, complex systems, which allow great accuracy in extracting and manipulating information from the entry data set.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 190–199, 2013.

Generally, neuronal networks are more precise than linear prognosis models, applied to analyzed samples and on financial variables. However, neuronal networks have small disadvantages, and must thus be considered complementary instruments in econometric modeling and prognosis.

Neuronal networks have been successfully applied to macroeconomic variables. Applications in the macroeconomic field are in full development, situated on the border of empiric economic models.

Starting from the idea that information can be brought closer and closer to future reality, in the condition that it depends exponentially from the number of hidden layers in the neuronal network used for prognosis, a study was developed regarding inflation prediction in Romania, presented in section 4.2 of the current paper.

2 Current Stage of Knowledge in Financial Prognosis, Utilizing Neuronal Calculus Techniques

The stage of knowledge in financial prediction by utilizing neuronal networks is still in its incipient phase, as the problem of finding a neuronal network with an architecture that can be utilized in a generalized process of prediction is still present.

The first scientific papers regarding the usage of neuronal networks in predictions date from the late 1980s. In this sense, authors' Hornik K., Stinchombe., and White H. paper, titled "Multilayer feed-forward networks are universal approximators" (1989) highlights the universal approximator quality of neuronal networks.

Another key paper on this topic is entitled „A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks"(1997), written by Swanson N. And White H., in which the estimation of a model on the level of macroeconomic indicators utilizing hidden layers neuronal networks is approached.

In Knodo T.'s paper entitled „GMDH neuronal network algorithm using the heuristic self-organization method and its application to the pattern identification problem" (1999) the GMDH method is approached.

Such a representation is reccommended in modeling connections, or connection points, between entries and exits.

Lamy concluded a prognosis analysis starting from quarterly data regarding real GDP of Canada, the analyzed time frame being 1978 – 1998. Thus, he estimated a linear regression model, statistically significant for future prognosis regarding growth of Canada's real GDP. His model explains in an 88% proportion Canada's real GDP between 1978 and 1998.

Thus, the model contains the following six independent or explanatory variables:

R_{t-1} - the quarterly growth rate of the Canadian Financing Index in rapport with the main economic activity indicators; F_t- the growth rate of work force at time t, F_{t-1} - the growth rate of the work force at time t-1; B_t - the index of consumer trust at time t;

d_{y-9}- the first difference of real long-term interest rates at time t-9 (nine months before time t); d_{t-3}- the first difference of the government's federal funds as part of GDP at time t-3 (three months before time t). In the model, four control dummy variables were introduced at the level of four periods, considered aberrant (1980 - 3rd trimester, 1981 1st trimester, 1986 4th trimester, 1991 3rd trimester). The linear regression model was estimated based on 62 observations (1978 trimester 1 – 1993 trimester 3) and is given by the relation:

$$GDP_t = -1.695+0.075R_{t-1}+0.304F_t+0.25F_{t-1}+0.019B_t-0.175d_{t-9}-0.32d_{t-3}-1.15d^1+1.168d^2 + 0.906d^3-0.814d^4+\varepsilon_t$$

Also, the model's coefficients are statistically significant and, thus, the model can be used for prognosis analysis of different time frames. The model presents the advantage of taking into account multiple variables. By using the prognosis-oriented neuronal network presented (fig. 1), BASED ON THE 6 variables described above, we obtain a real GDP growth of 0.30% for the 2nd trimester of 1998, which is considerably less precise than the prediction achieved with the linear model.

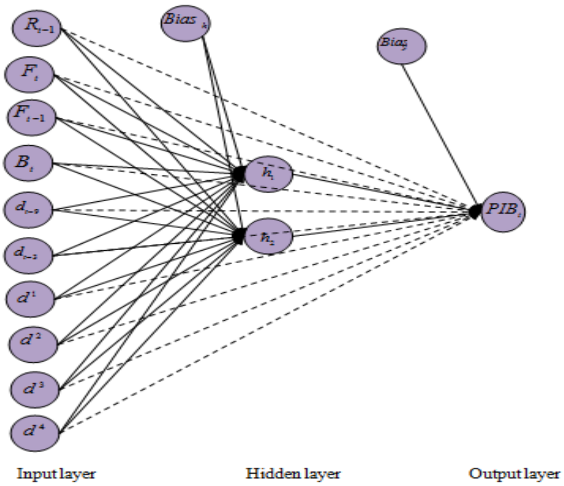


Fig. 1. Predictive neuronal network

3 Some Theoretical Information Regarding Neuronal Networks

3.1 Building a Neuronal Network Utilized in Prognosis

In the process of building a linear regression model, the econometrics specialist must identify the variables being considered and determine the number of necessary lags for a representative prediction, so that the differences between real future values and predicted values are as small as possible. These processes imply a large consumption of resources (time, technical, etc.), which is why a way to make them more efficient has been researched. In this sense, neuronal networks are a complex and cheap estimating model. In building a neuronal network for financial predictions, one must take into account the input set (financial data, macroeconomic data, etc.), but especially, to determine its architecture. It is pointless to mention that the changing of the architecture can lead to alterations in the results obtained after the prediction process. The search for the best architecture implies attempts to minimize errors. The steps of this process are:

Step 1. The explicative variables are given (independent).

Step 2. The network architecture is selected, using one or more hidden layers.

Step 3. The model is estimated using a large number of representative values (generally 100 values).

Step 4. The best network for each architecture is identified, having in its composition from 1 to *N* hidden layers.

Step 5. The neuronal network with the best architecture is identified.

Step 6. The decision of whether or not a variable is added or removed is taken, then return to Step 1.

3.2 Mathematic Description of Utilizing Neuronal Networks in the Prediction Process

Usually, networks utilized in prognosis work with two data sets. Thus, the learning set is used to estimate proportions, and the other set, known as the testing set, is used to evaluate representation or accuracy of the prediction done by the neuronal network. As the testing set does not enter the process of estimating proportions, the predictions are done by starting from the values of the testing set from an ex-post sample of the prognosis process. The main purpose of the networks, then, is to minimize prognosis errors appeared in the training set, taking into account the criterion of medium quadratic error value. In general, dynamic systems contain both an autonomous part (the proprietary component) as well as a part governed by external forces (guidance component). We must mention that relevant external forces can be hard to identify, or there can be a great degree of noise in the data. Thus, an accurate and exact description of the dynamics of the system can be practically impossible. Thus, for an as accurate description as possible of the model, the error model can be used.

A recurrent system (network) can be described at time t by the following functions:

$$q_t=z(q_{t-1}.u_t) \qquad (1)$$

which is to say the function that describes the transitions of the states

$$y_t=k(q_t) \qquad (2)$$

or the equation of the output

where: z and k - are unspecified functions; y_t- is calculated output; q_t- defines the state of the system at time t.

Observation: If the term u_t were to be removed from relations (1) and (2), these would compose an autonomous system.

Parameter optimization implies: $\min_{r,w1,w2,w4}((1/T)\sum^{T}_{t=1}(y_t-o_t)^2)$
where o_t is the observations at time t.

Observation: Only the subtraction y_t-o_t significantly influences o_t.
The neuronal network described is pictured in figure 2:

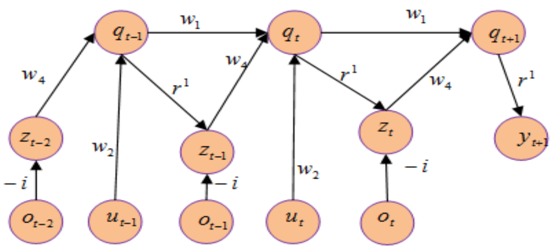


Fig. 2. Mathematical description of the representation of the neuronal network utilized in prognosis

Observation: w_i is the matrix connected to the proportions, and *I* is the fixed identity matrix, also known as the unit matrix.

3.3 Stopping Conditions in the Prediction Process

In order to understand the mechanism used in the prediction process it is imperative that we know the stopping conditions, as this process implies a finite number of iterations established beforehand[1]. Throughout time it has been shown that neuronal networks have the tendency to generate the phenomenon of overfitting[2] whereas the entry data is concerned, thus generating weak prognosis.

In order to diminish this tendency, numerous criterion have been developed, amongst them the criterion of early stopping. This criterion implies splitting the data set in three subsets:

– forming set: used when estimating network proportions;
– testing set: used in evaluating the prognosis sample;
– validating set: has the role of a precision indicator regarding the prediction sample, although it is not used throughout the prediction process.

After every iteration of the estimating process, a sample is generated in which relevant information is stored and the quadratic medium of errors is calculated (QME). Figure 3 presents the evolution of QME in the formation sets and in the validation sets.

---

[1] The number of iterations is known from the start

[2] The overfitting phenomenon takes place when a statistic model is excessively complex, which is to say it has too many parameters compared to the number of observations.

A model which is overfitting would then have a small performance in predictions, as it can exaggerate the minor fluctuations which take place in the data set.

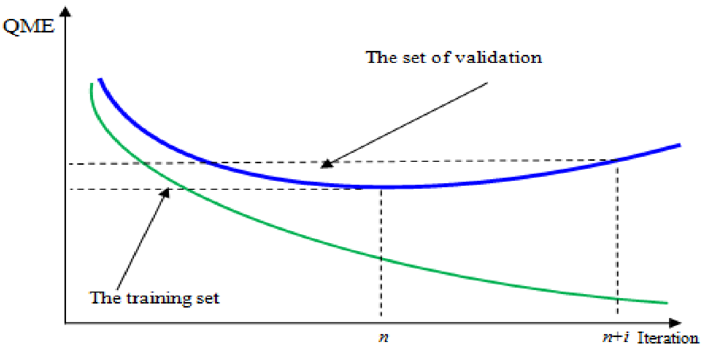


Fig. 3. QME evolution in the forming sets and the validation sets in the estimation process

Based on figure 2 we can deduce that while the number of iterations grows, generally, the sum of quadratic errors in the forming and validating sets decreases. However, after a certain number of iterations, the sum of quadratic errors in the validating set increases as the neuronal network, trained only for specific observations, loses its capacity to generalize estimates for other data.

4 Empiric Study

A study was created regarding the efficiency of applying neuronal networks, with different architectures, in the process of predicting the inflation rate in Romania. Also, estimate results were compared by using the neuronal network with results obtained by applying classic econometric models.

4.1 Methods Used and Methodology

In order to achieve the study, monthly inflation rates **from December 1997 to November 2012** (180 observations[3])were taken into account[4]. For the comparative analysis, the software GMDH Shell was used for predicting with neuronal networks, and Eviews 7 for estimating with econometric models. The inflation rate prognosis had at its basis applying the GMDH[5] algorithm and diverse neuronal network architectures, whereas the number of neurons in hidden layers is concerned. Thus, we could observe how varying the number of hidden neurons influenced the prediction process. The study implied, at the same time, evaluating the efficiency and accuracy of the obtained results. The econometric prediction was achieved using the classic method, ARCH.

---

[3] To differentiate the representation of the econometric model the data was stationarized through 1st degree differentiation.

[4] The source of data: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/

[5] Group Method of Data Handling.

4.2 Empiric Results

Inflation prognosis was done for the following 10 months. Thus, the following data was returned:

Table 1. Estimate inflation rate results using neuronal networks

	Inflation rate (%)					
Number of hidden layers / Month	1	2	5	10	20	100
+1	3.48	3.19	3.22	3.25	3.21	3.22
+2	3.80	3.08	3.12	3.14	3.13	3.14
+3	5.05	4.87	2.71	2.70	2.70	2.66
+4	5.12	4.94	2.79	2.79	2.78	2.75
+5	5.20	5.02	2.88	2.87	2.87	2.83
+6	7.61	7.59	5.02	5.00	5.13	5.91
+7	7.61	7.59	5.02	5.00	5.13	5.91
+8	7.71	7.69	5.13	5.11	5.24	6.01
+9	7.76	7.73	5.19	5.17	5.29	6.06
+10	7.80	7.78	5.25	5.22	5.35	6.11

Efficiency indicators are given in table 2:

Table 2. Efficiency of the inflation rate prediction with the help of neuronal networks

Number of hidden layers	1	2	5	10	20	100
Efficiency	66,9	66%	66%	69%	68.6%	69.2%

The dependency of residuals on the number of hidden layers is revealed by figure 4:

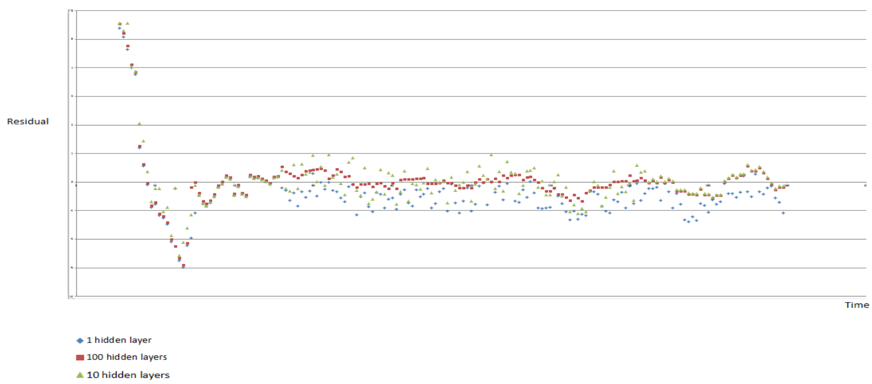


Fig. 4. Evolution of residuals in the prediction process for single hidden layer, 10 hidden layers and 100 hidden layers

Using the ARCH model to express dependency of the inflation rate on first and second degree differences of the time series, the prediction was achieved and obtained, with a degree of efficiency of 49.67%[6], the following results for the next 10 months:

Table 3. Estimate results of inflation rate, obtained via econometric modeling

Month	Estimate results of inflation rate (%)
+1	7.087
+2	6.829
+3	6.958
+4	6.690
+5	6.127
+6	6.404
+7	7.0508
+8	6.3122
+9	6.3122
+10	5.9430

The ARCH model the prognosis is based on is:

$$GARH_t=0.4967+1.0932\cdot\varepsilon^2_{t-1}+0.00550\cdot GARCH_{t-1}$$

Based on table 2 we can deduce that as the number of hidden layers increases, the degree of efficiency increases with it. This is explained by the fact that hidden layers have a major role in refining information and thus increases the level of extracted information and the estimation of viable results.

The neuronal network used for prediction contained 10 neurons in its hidden layer, having been trained with the dataset that contained stationarized inflation data from december 1997 to november 2012. The proportions were automatically generated during the training process. Connections between input and output are represented by the complex form of functional Volterra series:

$$y=b_0+\sum^{M}_{j=1}(b_l u_j)+\sum^{M}_{j=1}\sum^{M}_{i=1}(b_l u_j u_i)+\sum^{M}_{j=1}\sum^{M}_{i=1}\sum^{M}_{p=1}(b_l u_j u_i u_p)+\ldots \quad (3)$$

where: Coefficient b_j, from relation (3) is calculated using the regression method and, thus, the difference between the real output value, y, and the esimated value, ŷ, at the level of each pair of inputs u_j, u_i is minimized.

From figure 3, we may say that the more hidden layers the prognosis network has in its composition, the more stable the residues become, so that differences in the residual series are infinitesimal.

[6] Given by the degree of determination of the model.

5 Conclusions

According to the completed study, the prediction of a macroeconomic indicator with the help of a neuronal network is far more efficient that predictions achieved with econometric modeling (for example in the current case the smallest degree of efficiency on the level of the neuronal prediction is 66% greater than the degree of the econometric prediction, this being of 49.67%).

Although differences in the level of econometrically predicted values are comparably smaller than the level of those predicted with neuronal networks, the latter are the more relevant ones as the prediction process utilizing calculus implies more operations: synthesizing and separating information from the entry data set, filtering the information, generating exit information through a specific procedure (propagation), continual network training until obtaining viable information through minimization of efforts. In other words, for prediction, the neuronal network must train long enough for the adjustment errors to not differ significantly from zero. This implies active training for the neuronal network with different data sets.

The prognosis process utilizing neuronal networks is more realistic as econometric modeling implies prediction rigidities, which is to say that if a model is not adjustable[7], the data set needs to be adjusted through specific mathematical operations: differentiating, logarithms, etc. However predicting with the help of neuronal networks does not necessarily imply transforming operations on a data set level, but on the information extracted from the data set.

Combining neuronal calculus and data mining tends to develop more and more, accentuating bringing real values closer to estimate values through data redundancy.

References

1. Gonzalez, S.: Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models, Working Paper 2000-07 (2000)
2. Hornik, K., Stinchombe, M., White, H.: Multilayer feedforward networks are universal approximators. 2(5), 359–366 (1989)
3. Kondo, C., Kondo, T.: Revised GMDH-type neural network algorithm self-selecting optimum neural network architecture. Artificial Life and Robotics 14(4), 519–523 (2009)
4. Kondo, T.: Nonlinear system identification by feedback GMDH-type neural network with architecture self-selecting function. In: 2010 IEEE International Symposium on Intelligent Control (ISIC), pp. 1521–1526 (2010)
5. Kondo, T., Ueno, J., Takao, S.: Feedback GMDH-type neural network algorithm using prediction error criterion defined as AIC. In: Watada, J., Watanabe, T., Phillips-Wren, G., Howlett, R.J., Jain, L.C. (eds.) Intelligent Decision Technologies, Vol. 2. Smart Innovation, Systems and Technologies, vol. 16, pp. 313–322. Springer, Heidelberg (2012)
6. Stancu, S.: Calcul neuronal. Teorie şi aplicații economice. ASE Publishing House, Bucharest (2012)
7. Stancu, S., Constantin, A.M.: Retele Neuronale Artificiale. ASE Publishing House, Bucharest (2013)

[7] Whereas representation, stationarity and cointegration is concerned.

8. Stancu, S., Constantin, A.M.: Macroeconomie Cantitativa Avansata. ASE Publishing House, Bucharest (2013)
9. Swanson, N.R., White, H.: A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. The Review of Economics and Statistic 79, 540–550 (1997); President and Fellows of Harvard College and the Massachusetts Institute of Technology
10. Stancu, S., Constantin, A.M.: Artificial Neural Networks, Theory and applications. ASE Publishing House, Bucharest (2013)
11. Lam, H.K., Nguyen, H.T., Ling, S.S.H.: Computational Intelligence and Its Applications: Evolutionary Computation, Fuzzy Logic, Neural Network and Support Vector Machine Techniques. World Scientific Pub. Co. Inc. (2011)

PEV: A Computer Program for Fuzzy Sales Forecasting

Salvador Linares-Mustarós[1], José M. Merigó[2], and Joan Carles Ferrer-Comalat[1]

[1] Department of Business Administration, University of Girona,
Campus de Montilivi s/n, 71017 Girona, Spain
{salvador.linares,joancarles.ferrer}@udg.edu
[2] Department of Business Administration, University of Barcelona,
Av. Diagonal 690, 08034 Barcelona, Spain
jmerigo@ub.edu

Abstract. The aim of this work is the development of computer software for sales forecasting based on the experts' opinion but incorporating also the entrepreneur's particular knowledge of the business. The software, called PEV (PROCESSING EXTREME VALUES), provides fuzzy forecasts represented by triangular fuzzy numbers based on the experts' sales forecasting. In a second phase, the program allows removal or mitigation of extreme values based on the business owners' personal opinions enabling them to enrich the knowledge of the company not provided by the experts.

Keywords: OWA operator, triangular fuzzy number, sales forecast.

1 Introduction

A basic problem facing entrepreneurs who want to create a new company and business owners who want to expand the market for their products is determining the sales forecast for the first few years. Correct estimates are essential in identifying, among other things, the initial assets or the initial cash forecast.

In this paper we show a program for incorporating the particular knowledge of an entrepreneur or business owner to calculate a sales forecast by experts. The program offers the option of removing or softening the extreme projections of a data set. To do this, the business owner or entrepreneur must answer the following questions:

Do we consider all the results to be valid? Are there upper and lower limits for the results? Do we want to incorporate extreme values?

The computer program code offers the possibility of obtaining a triangular fuzzy number to represent the sales forecast, taking into consideration the response of each question according to Table 1. The program code was submitted in the International Conference, MS2012 New Rochelle, NY, USA in June 2012 [1].

2 A Program to Remove or Mitigate Extreme Values Based on the Business Owners' Personal Opinions

The program code offers sales forecasts as triangular fuzzy numbers ([2], [3], [4] and [5]) combining OWA aggregation functions ([6], [7], [8], [9], [10], [11], [12], [13] and [14]) and options of approaching or deleting extreme data according to Table 1.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 200–209, 2013.

Table 1. Method to soften or delete extreme values

<table>
<tr><td colspan="3">Do we consider all the results as valid?</td><td></td><td>Proposition of triangular fuzzy number</td></tr>
<tr><td>Yes</td><td></td><td></td><td></td><td>$\tilde{P}$</td></tr>
<tr><td rowspan="2"></td><td colspan="3">Are there upper and lower limits for the results?</td><td rowspan="2"></td></tr>
<tr><td></td><td colspan="2">Do we want to incorporate extreme values?</td></tr>
<tr><td rowspan="4">No</td><td rowspan="2">No</td><td>Yes</td><td></td><td>$\tilde{P}_{IQR\updownarrow}$</td></tr>
<tr><td>No</td><td></td><td>$\tilde{P}_{IQRx}$</td></tr>
<tr><td rowspan="2">Yes</td><td>Yes</td><td></td><td>$\tilde{P}_{glbLUB\updownarrow}$</td></tr>
<tr><td>No</td><td></td><td>$\tilde{P}_{glbLUBx}$</td></tr>
</table>

Where if $a_1, a_2,..., a_n$ is a series of sales forecasts obtained by a group of experts

$\tilde{P} = (OWA_*(a_1, a_2,..., a_n), OWA_{ave}(a_1, a_2,..., a_n), OWA^*(a_1, a_2,..., a_n))$
whose membership function can be written as follows:

$$\mu_{\tilde{P}}(x) = \begin{cases} 0 & \text{if } x < b_1 \\ \dfrac{x - b_1}{\frac{1}{n}\sum_{j=1}^{n} b_i - b_1} & \text{if } b_1 \le x \le \frac{1}{n}\sum_{j=1}^{n} b_i \\ \dfrac{b_n - x}{b_n - \frac{1}{n}\sum_{j=1}^{n} b_i} & \text{if } \frac{1}{n}\sum_{j=1}^{n} b_i \le x \le b_n \\ 0 & \text{if } x > b_n \end{cases}$$

$\tilde{P}_{IQRx} = (OWA_*(b_{s+1}, b_{s+2},..., b_{n-t-1},b_{n-t}), OWA_{ave}(b_{s+1}, b_{s+2},..., b_{n-t-1},b_{n-t}), ,OWA^*(b_{s+1}, b_{s+2},..., b_{n-t-1},b_{n-t}))$
whose memebership function can be written in the following manner:

$$\mu_{\tilde{P}_{IQRx}}(x) = \begin{cases} 0 & \text{if } x < b_{s+1} \\ \dfrac{x - b_{s+1}}{\dfrac{1}{n-s-t}\displaystyle\sum_{i=s+1}^{n-t} b_i - b_{s+1}} & \text{if } b_{s+1} \le x \le \dfrac{1}{n-s-t}\displaystyle\sum_{i=s+1}^{n-t} b_i \\ \dfrac{b_{n-t} - x}{b_{n-t} - \dfrac{1}{n-s-t}\displaystyle\sum_{i=s+1}^{n-t} b_i} & \text{if } \dfrac{1}{n-s-t}\displaystyle\sum_{i=s+1}^{n-t} b_i \le x \le b_{n-t} \\ 0 & \text{if } x > b_{n-t} \end{cases}$$

$\tilde{P}_{IQR^{\downarrow}_{\uparrow}} = (OWA_{n*}, OWA_{n\ ave}, OWA_n^{\ *})$

whose membership function depends on the number of extreme values modified. Thus, we can distinguish between four cases:

i) If we have not modified any values, the membership function has the form:

$$\mu_{\tilde{P}_{IQR^{\downarrow}_{\uparrow}}}(x) = \begin{cases} 0 & \text{if } x < b_1 \\ \dfrac{x - b_1}{\left(\dfrac{1}{n}\displaystyle\sum_{j=1}^{n} b_j\right) - b_1} & \text{if } b_1 \le x \le \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} b_j \\ \dfrac{b_n - x}{b_n - \left(\dfrac{1}{n}\displaystyle\sum_{j=1}^{n} b_j\right)} & \text{if } \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} b_j \le x \le b_n \\ 0 & \text{if } x > b_n \end{cases}$$

ii) Suppose we have only increased the s > 0 earlier predictions. Then the membership function can be calculated using the following expression:

$$\mu_{\tilde{P}_{IQR^{\downarrow}_{\uparrow}}}(x) = \begin{cases} 0 & \text{if } x < q_1 - 1{,}5\cdot RI \\ \dfrac{x - \left(q_1 - 1{,}5\cdot RI\right)}{\dfrac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + \displaystyle\sum_{j=s+1}^{n} b_j\right) - \left(q_1 - 1{,}5\cdot RI\right)} & \text{if } q_1 - 1{,}5\cdot RI \le x \le \dfrac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + \displaystyle\sum_{j=s+1}^{n} b_j\right) \\ \dfrac{b_n - x}{b_n - \dfrac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + \displaystyle\sum_{j=s+1}^{n} b_j\right)} & \text{if } \dfrac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + \displaystyle\sum_{j=s+1}^{n} b_j\right) \le x \le b_n \\ 0 & \text{if } x > b_n \end{cases}$$

iii) Suppose we have only decreased the t > 0 last predictions. Thus, the membership function can be calculated using the following expression:

$$\mu_{\tilde{P}_{IQR^{\downarrow}_{\uparrow}}}(x) = \begin{cases} 0 & \text{if } x < b_1 \\ \dfrac{x - b_1}{\dfrac{1}{n}\left(t\cdot(q_3 + 1{,}5\cdot RI) + \displaystyle\sum_{j=1}^{n-t} b_j\right) - b_1} & \text{if } b_1 \le x \le \dfrac{1}{n}\left(t\cdot(q_3 + 1{,}5\cdot RI) + \displaystyle\sum_{j=1}^{n-t} b_j\right) \\ \dfrac{\left(q_3 + 1{,}5\cdot RI\right) - x}{\left(q_3 + 1{,}5\cdot RI\right) - \dfrac{1}{n}\left(t\cdot(q_3 + 1{,}5\cdot RI) + \displaystyle\sum_{j=1}^{n-t} b_j\right)} & \text{if } \dfrac{1}{n}\left(t\cdot(q_3 + 1{,}5\cdot RI) + \displaystyle\sum_{j=1}^{n-t} b_j\right) \le x \le q_3 + 1{,}5\cdot RI \\ 0 & \text{if } x > q_3 + 1{,}5\cdot RI \end{cases}$$

iv) Suppose that we have increased the s> 0 earlier predictions and decreased the t > 0 last predictions. Then the membership function can be calculated using the following expression:

$$\mu_{\tilde{P}_{IQR\updownarrow}}(x) = \begin{cases} 0 & \text{if } x < q_1 - 1{,}5\cdot RI \\ \dfrac{x - (q_1 - 1{,}5\cdot RI)}{\frac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + t\cdot(q_3 + 1{,}5\cdot RI) + \sum_{j=s+1}^{n-t} b_j\right) - (q_1 - 1{,}5\cdot RI)} \quad i & \text{if } q_1 - 1{,}5\cdot RI \le x \le \frac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + t\cdot(q_3 + 1{,}5\cdot RI) + \sum_{j=s+1}^{n-t} b_j\right) \\ \dfrac{(q_3 + 1{,}5\cdot RI) - x}{(q_3 + 1{,}5\cdot RI) - \frac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + t\cdot(q_3 + 1{,}5\cdot RI) + \sum_{j=s+1}^{n-t} b_j\right)} & \text{if } \frac{1}{n}\left(s\cdot(q_1 - 1{,}5\cdot RI) + t\cdot(q_3 + 1{,}5\cdot RI) + \sum_{j=s+1}^{n-t} b_j\right) \le x \le p_{n-t} \\ 0 & \text{if } x > q_3 + 1{,}5\cdot RI \end{cases}$$

$$\tilde{P}_{minMAXx} = (OWA_*(p_{s+1}, p_{s+2}, \ldots, p_{n-t-1}, p_{n-t}),\ OWA_{ave}(p_{s+1}, p_{s+2}, \ldots, p_{n-t-1}, p_{n-t}),\ OWA^*(p_{s+1}, p_{s+2}, \ldots, p_{n-t-1}, p_{n-t}))$$

The membership function can be obtained from the following expression:

$$\mu_{\tilde{P}_{glbLUBx}}(x) = \begin{cases} 0 & \text{if } x < b_{s+1} \\ \dfrac{x - b_{s+1}}{\frac{1}{n-s-t}\sum_{j=s+1}^{n-t} b_j - b_{s+1}} & \text{if } b_{s+1} \le x \le \frac{1}{n-s-t}\sum_{j=s+1}^{n-t} b_j \\ \dfrac{b_{n-t} - x}{b_{n-t} - \frac{1}{n-s-t}\sum_{j=s+1}^{n-t} b_j} & \text{if } \frac{1}{n-s-t}\sum_{j=s+1}^{n-t} b_j \le x \le b_{n-t} \\ 0 & \text{if } x > b_{n-t} \end{cases}$$

$$\tilde{P}_{glbLUB\updownarrow} = (\ OWA_*, OWA_{ave}, OWA^*\ )$$

whose membership function, just as in the previous case, also depends on the number of extreme values modified. So, we will distinguish four cases:

i) If we have not already modified any values, the membership function has the form:

$$\mu_{\tilde{P}_{glbLUB\updownarrow}}(x) = \begin{cases} 0 & \text{if } x < b_1 \\ \dfrac{x - b_1}{\left(\frac{1}{n}\sum_{j=1}^{n} b_j\right) - b_1} & \text{if } b_1 \le x \le \frac{1}{n}\sum_{j=1}^{n} b_j \\ \dfrac{b_n - x}{b_n - \left(\frac{1}{n}\sum_{j=1}^{n} b_j\right)} & \text{if } \frac{1}{n}\sum_{j=1}^{n} b_j \le x \le b_n \\ 0 & \text{if } x > b_n \end{cases}$$

ii) Suppose we have only increased the s > 0 earlier predictions. Then the membership function can be calculated using the following expression:

$$\mu_{\tilde{P}_{glbLUB\downarrow\uparrow}}(x) = \begin{cases} 0 & \text{if } x < glb \\ \dfrac{x - glb}{\frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n} b_j\right) - glb} & \text{if } glb \le x \le \frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n} b_j\right) \\ \dfrac{b_n - x}{b_n - \frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n} b_j\right)} & \text{if } \frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n} b_j\right) \le x \le b_n \\ 0 & \text{if } x > b_n \end{cases}$$

iii) Suppose we have only decreased the t > 0 last predictions. Then the membership function can be calculated using the following expression:

$$\mu_{\tilde{P}_{glbLUB\downarrow\uparrow}}(x) = \begin{cases} 0 & \text{if } x < b_1 \\ \dfrac{x - b_1}{\frac{1}{n}\left(\sum_{j=1}^{n-t} b_j + t \cdot LUB\right) - b_1} & \text{if } b_1 \le x \le \frac{1}{n}\left(\sum_{j=1}^{n-t} b_j + t \cdot LUB\right) \\ \dfrac{LUB - x}{LUB - \frac{1}{n}\left(\sum_{j=1}^{n-t} b_j + t \cdot LUB\right)} & \text{if } \frac{1}{n}\left(\sum_{j=1}^{n-t} b_j + t \cdot LUB\right) \le x \le LUB \\ 0 & \text{if } x > LUB \end{cases}$$

iv) Finally, suppose that we have increased the s> 0 earlier predictions and decreased the t > 0 last predictions. Then the membership function can be calculated using the following expression:

$$\mu_{\tilde{P}_{glbLUB\downarrow\uparrow}}(x) = \begin{cases} 0 & \text{if } x < glb \\ \dfrac{x - glb}{\frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n-t} b_j + t \cdot LUB\right) - glb} & \text{if } glb \le x \le \frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n-t} b_j + t \cdot LUB\right) \\ \dfrac{LUB - x}{LUB - \frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n-t} b_j + t \cdot LUB\right)} & \text{if } \frac{1}{n}\left(s \cdot glb + \sum_{j=s+1}^{n-t} b_j + t \cdot LUB\right) \le x \le LUB \\ 0 & \text{if } x > LUB \end{cases}$$

3 PEV: A Computer Program

In order to test the method, we have developed a computer program that provides the different fuzzy numbers from the aggregation problem. Moreover, it also gives the possibility of obtaining any forecasted value for any of the fuzzy numbers identified. The program can be downloaded from the following address:

http://web2.udg.edu/grmfcee/PEV.exe

It is possible to download the open access version, if the reader wants to explore in this software in greater depth. The program has been prepared in Adobe Flash Professional CS5. This can be downloaded here:

http://web2.udg.edu/grmfcee/PEV.fla
http://web2.udg.edu/grmfcee/functions.as

In order to understand its usability, let us give a simple numerical example. We use the forecasts in units for a specific product of a company. The results given by the experts are as follows:

Expert 1: 300	Expert 2: 250	Expert 3: 300	Expert 4: 240	Expert 5: 100
Expert 6: 320	Expert 7: 290	Expert 8: 340	Expert 9: 260	Expert 10: 330

When the program is run, a box appears for users to enter the number of ex-perts forecasts, followed by a continue button (Fig. 1). To facilitate ease of use, all secondary buttons, images and text inputs have been hidden.

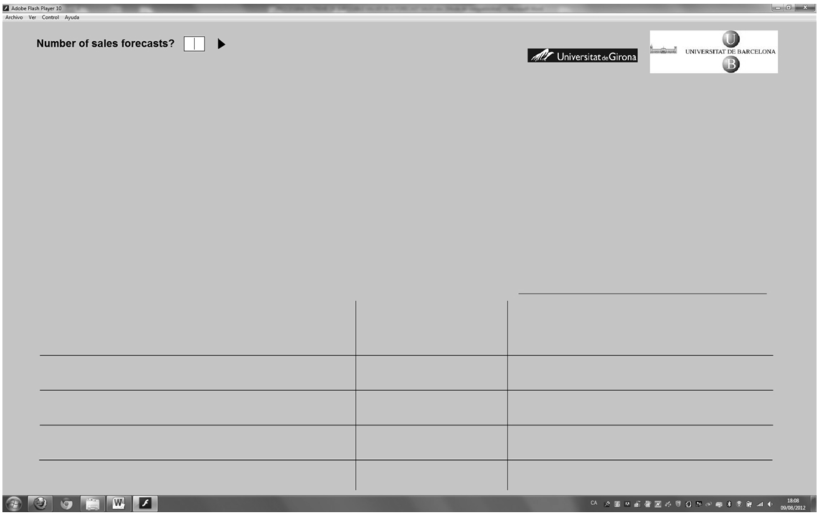


Fig. 1. Initial interface

In our case, we enter the value 10 and click the continue button. This number of boxes appear for entering the sales forecast values and the continue button appears in a new position (Fig. 2).

We fill the boxes with the forecasts and click the continue button.

We observe that the program reorders the forecasts from minimum to maxi-mum values, offering the quartile values and the possibility of introducing supper and lower limits. The continue button appears to the right of these values (Fig.3). By default, limits are always the maximum and minimum data. These values can be modified using the keyboard after selecting the textbox.

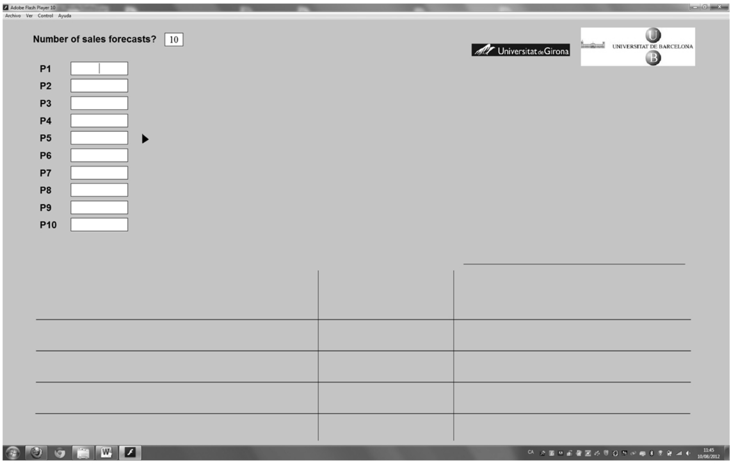


Fig. 2. Interface with new boxtexts

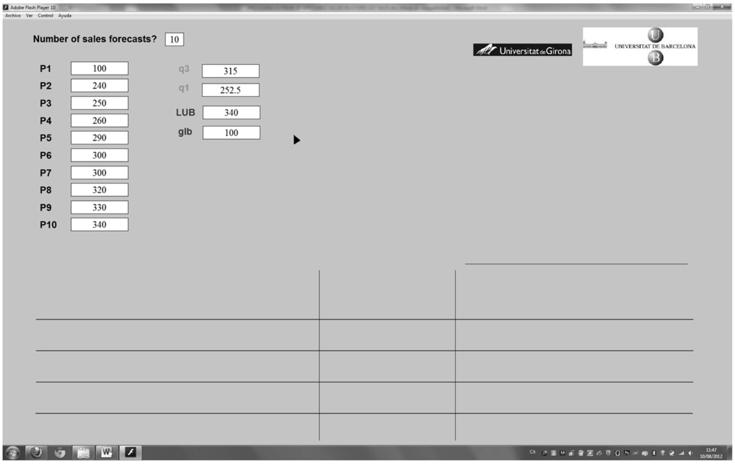


Fig. 3. Interface with the first and third quartiles and the supper and lower limits

In our case, we assume that having analyzed all the sales representatives' arguments, the entrepreneur or decision maker considers it impossible to produce under 200 and over 300 units per day. Thus, we modify the LUB and glb values to 300 and 200, respectively.

As we can see, once we have clicked the continue button, the program sisplas all the forecasts on a scale, together with the five triangular numbers obtained from the method presented in this paper (Fig. 4). Clicking each button with these numbers displays, the discarded or modified forecasts on the graph (Fig. 5).

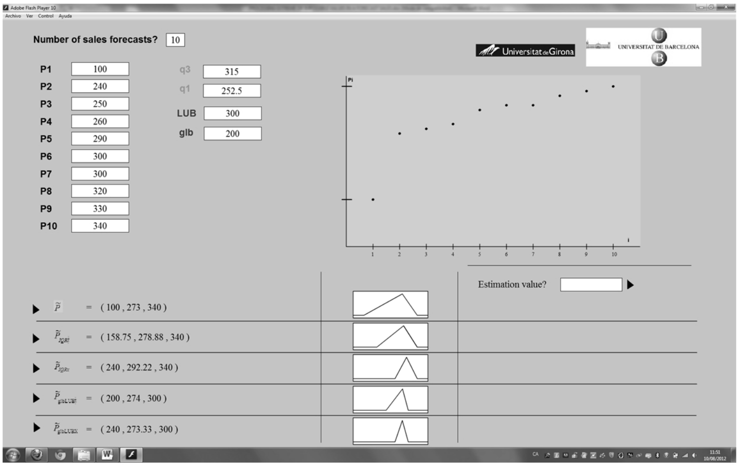


Fig. 4. Interface with the $\tilde{P}$ selected

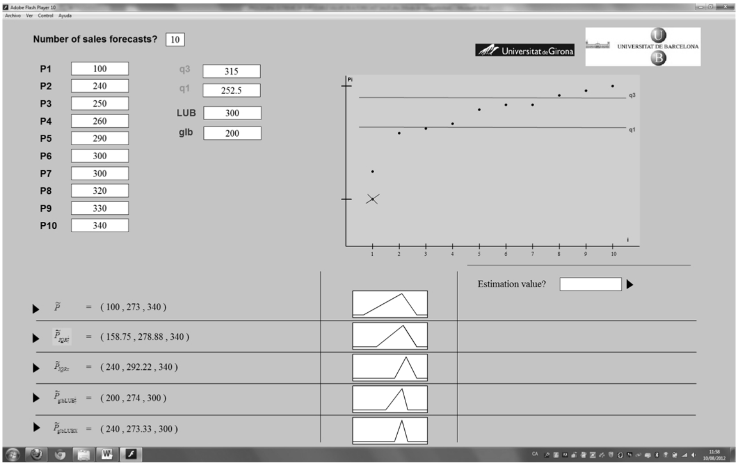


Fig. 5. Interface with the $\tilde{P}_{IQR_{\uparrow}^{\downarrow}}$ selected

In our case we can see that the program modifies the first value in the predicted values image by q1-1.5(q3-q1) due to its distance from the quartiles.

Finally, we would like to demonstrate the utility of the program's estimation value. By entering a value in the box "Estimation value" and clicked its associat-ed button, the program will show the presumption level of this value in the different fuzzy numbers. We can see, for example, that a production of 250 units has an associated possibility of 68 in 100 in the case of taking into account all views by locating the extreme limits cases (Fig. 6). We also observe that a production of 300 units has a slightly high associated possibility (0.6, 0.65, 0.84). If the entrepreneur or business owner believes there is limited production with an upper limit of 300, will control sales because it is easily possible to reach sales for this production (Fig. 7).

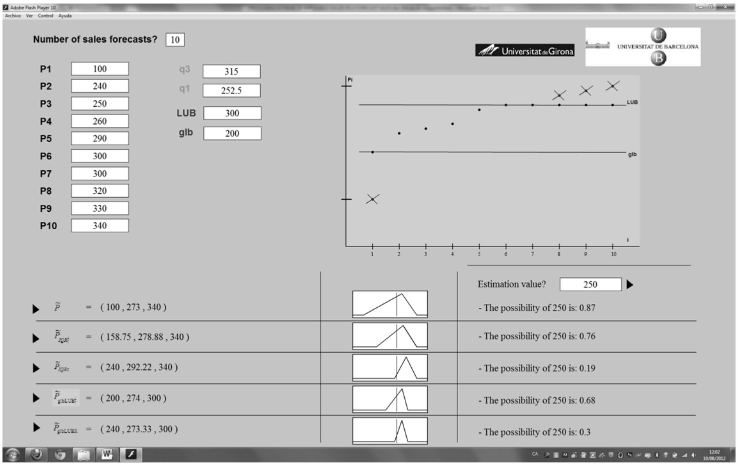


Fig. 6. Possibility levels with a production of 250

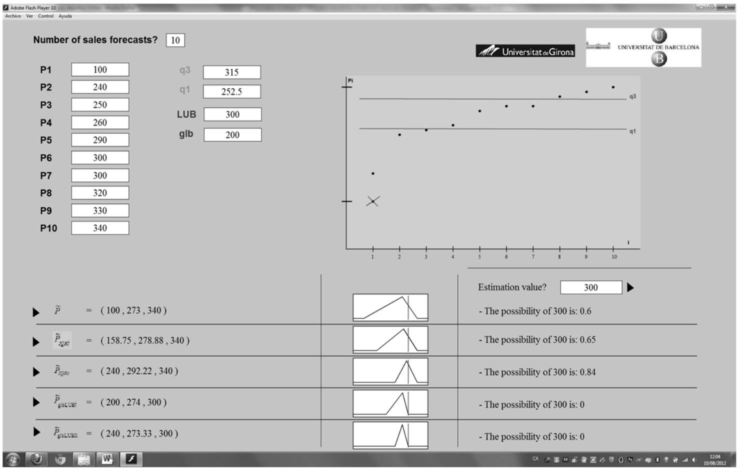


Fig. 7. Possibility levels with a production of 300

4 Conclusions

This paper provides a tool to assess, on the basis of forecasts obtained from experts, the presumption level to which the sales forecasting of a company is a specific value. It suggests different types of triangular fuzzy numbers whose vertices are obtained by aggregation functions that act on experts' sales forecasts. The method also offers the option for entrepreneurs or business owners to remove or mitigate extreme values based on their personal opinion, thus enabling them to provide knowledge of the company unknown to the experts. By allowing entrepreneurs or business owners to decide whether to incorporate extreme values or not, it paves the way for them to decide the final function of the sales forecast, making it a completely personal estimate.

Acknowledgements. We would like to thank the anonymous referees for their valuable comments and suggestions, which led to an improved version of the paper.

References

1. Linares-Mustarós, S., Merigó, J.M., Ferrer-Comalat, J.C.: A method for uncertain sales forecast by using triangular fuzzy numbers. In: Engemann, K.J., Gil-Lafuente, A.M., Merigó, J.M. (eds.) MS 2012. LNBIP, vol. 115, pp. 98–113. Springer, Heidelberg (2012)
2. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
3. Kaufmann, A., Gil-Aluja, J.: Introducción de la Teoría de los Subconjuntos Borrosos a la Gestión de las Empresas. Milladoiro, Santiago de Compostela (1986) (in Spanish)
4. Kaufmann, A., Gupta, M.M.: Introduction to Fuzzy Arithmetic. Theory and Applications. Van Nostrand Reinhold, New York (1991)
5. Kaufmann, A., Gil-Aluja, J., Terceño, A.: Matemáticas para la Economía y la Gestión de Empresas, vol.1: Aritmética de la incertidumbre. Foro Científico, Barcelona (1994) (in Spanish)
6. Yager, R.R.: On Ordered Weighted Averaging Aggregation Operators in Multi-Criteria Decision Making. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190 (1988)
7. Yager, R.R.: Families of OWA Operators. Fuzzy Sets and Systems 59, 125–148 (1993)
8. Yager, R.R.: Generalized OWA Aggregation Operators. Fuzzy Optimization and Decision Making 3, 93–107 (2004)
9. Yager, R.R., Kacprzyk, J.: The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer Academic Publishers, Norwell (1997)
10. Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Springer, Berlin (2007)
11. Keighobadi, J., Yazdanpanah, M., Kabganian, M.: An Enhanced Fuzzy H8 Estimator Applied to Low-Cost Attitude-Heading Reference System. Kybernetes 40(4), 300–326 (2011)
12. Merigó, J.M.: New Extensions to the OWA Operator and its Application in Decision Making. PhD Thesis (2008) (in Spanish)
13. Merigó, J.M.: Fuzzy Multi-Person Decision Making With Fuzzy Probabilistic Aggregation Operators. International Journal of Fuzzy Systems 13(3), 163–174 (2011)
14. Merigó, J.M., Gil-Lafuente, A.M.: Fuzzy Induced Generalized Aggregation Operators and its Application in Multi-Person Decision Making. Expert Systems with Applications 38(8), 9761–9772 (2011)

Decision Making with Fuzzy Moving Averages and OWA Operators

José M. Merigó, Jian-Bo Yang, and Dong-Ling Xu

Manchester Business School, University of Manchester,
Booth Street West, M15 6PB Manchester, UK
jose.merigolindahl@mbs.ac.uk

Abstract. A new framework for multi-period decision making is presented. It deals with moving averages and the ordered weighted average (OWA) in an imprecise environment that can be assessed with fuzzy numbers. Several new aggregation operators are presented including the fuzzy ordered weighted moving average (FOWMA) and the fuzzy ordered weighted averaging weighted moving average (FOWAWMA). Their main advantage is that they can provide a parameterized family of aggregation operators between the minimum and the maximum in a dynamic environment assessed with the moving average and fuzzy numbers. An application of the new approach is presented in national multi-period decision making problems concerning the fixation of the main interest rate of a country.

Keywords: Fuzzy numbers; OWA operator; Moving average; Multi-period decision making.

1 Introduction

A wide range of aggregation operators have been studied in the literature in order to aggregate the information [1]. A very well-known aggregation operator is the ordered weighted average (OWA) [2]. It is an aggregation operator that provides a parameterized family of aggregation operators from the minimum to the maximum. Since it appearance, it has been studied by a lot of authors [3-4].

Sometimes, the available information is imprecise and cannot be assessed with precise numbers. The reason for this is that when dealing with the real world, there are a lot of uncertainties and it is not easy to quantify the information. For example, if we want to analyze the inflation rate of a country the next year, it is very difficult to quantify a precise number because a lot of variations can happen during one year. Thus, it is better to provide a more flexible tool such as an interval number [5] or a fuzzy number [6]. For example, we could predict that the inflation of a country for the next year is [1, 4]%. That is, in any case it will not be below 1% or above 4%. By using these techniques, we can assess the information in a more complete way because we are not losing information in the analysis. By using fuzzy numbers, we can represent the minimum and the maximum and the possibility that the internal results

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 210–221, 2013.

will occur. Thus, the study is more in accordance with the reality because it considers all the different scenarios that can occur and those scenarios that are expected to happen.

The use of fuzzy information in the OWA operator produces the fuzzy OWA (FOWA) operator. It is very similar to the OWA operator with the difference that can deal with imprecise information assessed with fuzzy numbers. Several authors have studied different extensions and generalizations of this approach. For example, Chen and Chen [7] extended this approach by using induced aggregation operators. Merigó and Casanovas [8] developed a generalization by using generalized and quasi-arithmetic means. Merigó and Gil-Lafuente [9] presented a further extension by using the induced OWA (IOWA) operator [10]. Other authors studied this framework by using intuitionistic fuzzy sets [11] and a wide range of different researchers have studied other extensions including the use of harmonic means [12] and distance measures [13].

Another interesting technique for aggregating the information is the moving average. It is widely known for its usefulness in time series forecasting [14]. Yager [15] suggested the use of the OWA operator in the moving average obtaining the ordered weighted moving average. Its main advantage is that it can provide a parameterized family of aggregation operators from the minimum to the maximum in a subset of arguments that move towards a bigger set of arguments. Recently, Merigó and Yager have suggested further extensions and generalizations of this approach by using distance measures and induced aggregation operators [16] and a unified framework between the OWA operator and the weighted average [17].

The aim of this paper is to present a framework that is able to deal with OWA operators, moving averages and in an imprecise environment assessed with fuzzy numbers. For doing so, the fuzzy ordered weighted moving average (FOWMA) is introduced. It is an aggregation operator that provides a family of aggregation operators between the fuzzy minimum and the fuzzy maximum in a dynamic process assessed with the moving average. Some of its main properties are studied and a wide range of particular cases are presented including the fuzzy moving average (FMA), the fuzzy moving maximum and the fuzzy moving minimum.

A further generalization is suggested for dealing with environments that contain weighted averages and OWA operators in the same problem. In this paper, this problem is assessed with the ordered weighted averaging weighted average (OWAWA) [18] that unifies the OWA operator and the weighted average in the same formulation and considering the degree of importance that each concept has in the analysis. However, in the literature, there are other methods for unifying the OWA operator and the weighted average such as the weighted OWA (WOWA) operator [19], the hybrid average [20] and the importance OWA operator [21]. By using the OWAWA operator, it is formed the fuzzy ordered weighted averaging weighted moving average (FOWAWMA). Its main advantage is that it can assess a dynamic process with the moving average and in a fuzzy environment and using weighted averages and OWA operators in the same formulation and taking into account the degree of importance that each concept has in the analysis. That is, we can assess subjective information and the attitudinal character of the decision maker in the same aggregation process and considering the relevance that each of them has in the problem. Some of the main particular cases are also studied and further extensions by using induced aggregation operators are introduced.

An application of the new approach in a decision making problem is presented. The focus is on the assessment of information that is provided in different periods of time and produce forecasts by using the FOWMA and the FOWAWMA operators. An illustrative example in national decision making is studied by analyzing the interest rate of a country. By using this new approach, it is possible to consider a more complete picture of the different scenarios that may occur and select the scenario that seems to be in closest accordance with the interests of the decision maker.

This paper is organized as follows. Section 2 describes some basic concepts including fuzzy numbers, the moving average, the OWA operator and the OWAWA operator. Section 3 introduces the FOWMA and FOWAWMA operator. Section 4 presents an application in multi-period decision making and Section 5 summarizes the main conclusions of the paper.

2 Preliminaries

2.1 Fuzzy Numbers

The FN was introduced by Chang and Zadeh [22] and Zadeh [23] and it has been studied and applied by a lot of authors [6]. Its main advantage is its flexibility because it can consider the maximum and the minimum and the possibility that the internal values may occur.

Definition 1. Let R be $(-\infty, \infty)$, the set of all real numbers. A FN is a fuzzy subset [24] of R with membership function m: $R \rightarrow [0, 1]$ satisfying the following conditions:

- Normality: There exists at least one number $a_0 \in R$ such that $m(a_0) = 1$.
- Convexity: $m(t)$ is nondecreasing on $(-\infty, a_0]$ and nonincreasing on $[a_0, \infty)$.

Note that the FN may be considered as a generalization of the interval number although it is not strictly the same because the interval numbers may have different meanings. In the literature, we find a wide range of FNs [9,11] such as triangular FNs, trapezoidal FNs, interval-valued FNs, intuitionistic FNs and type-2 FNs.

For example, a trapezoidal FN (TpFN) A of a universe of discourse R can be characterized by a trapezoidal membership function $A = (\underline{a}, \overline{a})$ such that

$$\begin{aligned} \underline{a}(\alpha) &= a_1 + \alpha(a_2 - a_1), \\ \overline{a}(\alpha) &= a_4 - \alpha(a_4 - a_3). \end{aligned} \tag{1}$$

where $\alpha \in [0, 1]$ and parameterized by (a_1, a_2, a_3, a_4) where $a_1 \leq a_2 \leq a_3 \leq a_4$, are real values. Note that if $a_1 = a_2 = a_3 = a_4$, then, the FN is a crisp value and if $a_2 = a_3$, the FN is represented by a triangular FN (TFN). Note that the TFN can be parameterized by (a_1, a_2, a_4).

In the following, we are going to review the FNs arithmetic operations as follows. Let A and B be two TFNs, where $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$. Thus:

1. $A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$,
2. $A - B = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$,
3. $A \times k = (k \times a_1, k \times a_2, k \times a_3)$; for $k > 0$.

Among the wide range of methods existing in the literature for ranking FNs, we recommend the use of the methods commented by Merigó [9] such as the use of the value found in the highest membership level ($\alpha = 1$) and if it is an interval, the average of the interval.

2.2 The OWA Operator

The OWA operator [2] is an aggregation operator that provides a parameterized family of aggregation operators that include the arithmetic mean, the maximum and the minimum. It can be defined as follows.

Definition 2. An OWA operator of dimension n is a mapping OWA: $R^n \rightarrow R$ that has an associated weighting vector W of dimension n with the sum of the weights equal to 1 and $w_j \in [0, 1]$, such that:

$$OWA\ (a_1, a_2,\ldots, a_n) = \sum_{j=1}^{n} w_j b_j \, , \tag{2}$$

where b_j is the jth largest of the a_i.

From a generalized perspective of the reordering step, we can distinguish between the descending OWA (DOWA) operator and the ascending OWA (AOWA) operator [2]. The OWA operator is commutative, monotonic, bounded and idempotent. For further information concerning the OWA and its applications, see for example [3-4].

2.3 The Moving Average

A moving average is a usual average that aggregates a subset of arguments in a bigger set of information and moves towards the whole set. More generally, the moving average can be seen as a moving aggregation operator. Its main advantage is that it permits to consider the results of some part of the sample and make comparisons when modifying the partial sample selected. It is widely known for solving time-series forecasting problems [14]. A weighted moving average (WMA) can be formulated as follows.

Definition 3. A WMA of dimension m is a mapping WMA: $R^m \rightarrow R$ that has an associated weighting vector W of dimension m with $W = \sum_{i=1+t}^{m+t} w_i = 1$ and $w_i \in [0, 1]$, such that:

$$WMA\ (a_{1+t}, a_{2+t},\ldots, a_{m+t}) = \sum_{i=1+t}^{m+t} w_i a_i \, , \tag{3}$$

where a_i is the ith argument, m is the total number of arguments considered from the whole sample and t indicates the movement done in the average from the initial analysis.

If $w_i = 1/m$ for all i, the WMA becomes the MA aggregation. Yager [15] introduced the use of the OWA operator as a moving average. Thus, we are able to provide a parameterized family of moving aggregation operators between the minimum and the maximum. The ordered weighted moving average (OWMA) can be defined as follows.

Definition 4. An ordered weighted moving average (OWMA) of dimension m is a mapping $OWMA$: $R^m \to R$ that has an associated weighting vector W of dimension m with $W = \sum_{j=1+t}^{m+t} w_j = 1$ and $w_j \in [0, 1]$, such that:

$$OWMA\ (a_{1+t}, a_{2+t}, \ldots, a_{m+t}) = \sum_{j=1+t}^{m+t} w_j b_j\ , \tag{4}$$

where b_j is the jth largest argument of the a_i, m is the total number of arguments considered from the whole sample and t indicates the movement done in the average from the initial analysis.

2.4 The OWAWA Operator

The ordered weighted averaging – weighted average (OWAWA) is a model that unifies the OWA operator and the weighted average in the same formulation considering the degree of importance that each concept has in the analysis. Therefore, both concepts can be seen as a particular case of a more general one. It can be defined as follows.

Definition 5. An OWAWA operator of dimension n is a mapping $OWAWA$: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$OWAWA\ (a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j\ , \tag{5}$$

where b_j is the jth largest of the a_i, each argument a_i has an associated weight (WA) v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = \beta w_j + (1-\beta) v_j$ with $\beta \in [0, 1]$ and v_j is the weight (WA) v_i ordered according to b_j, that is, according to the jth largest of the a_i.

Note that if $\beta = 1$, we get the OWA operator and if $\beta = 0$, the WA. The OWAWA operator accomplishes similar properties than the usual aggregation operators including the distinction between descending and ascending orders, the use of mixture operators and so on [18].

3 Fuzzy Moving Averages and OWA Operators

3.1 The Fuzzy Ordered Weighted Moving Average

The fuzzy ordered weighted moving average (FOWMA) operator is an aggregation operator that provides a parameterized family of operators between the fuzzy moving minimum and the fuzzy moving maximum. Thus, it can consider the attitudinal character (or degree of orness) of a decision maker in a dynamic process where additional information is continuously added to the aggregation. It can be defined as follows.

Definition 6. Let Ψ be the set of FNs. A fuzzy ordered weighted moving average (FOWMA) of dimension m is a mapping $FOWMA$: $\Psi^n \rightarrow \Psi$ that has an associated weighting vector W of dimension m with $W = \sum_{j=1+t}^{m+t} w_j = 1$ and $w_j \in [0, 1]$, such that:

$$FOWMA\ (\tilde{a}_{1+t}, \tilde{a}_{2+t}, \dots, \tilde{a}_{m+t}) = \sum_{j=1+t}^{m+t} w_j b_j\ , \tag{6}$$

where b_j is the jth largest argument of the $\tilde{a}_i$, m is the total number of arguments considered from the whole sample, t indicates the movement done in the average from the initial analysis and $\tilde{a}$ are the argument variables represented in the form of FNs.

Note that different types of FNs can be used in the aggregation including TFNs, TpFNs and interval-valued FNs. If the weighting vector is not normalized because it can also be fuzzy, i.e., $\tilde{W} = \sum_{j=1}^{n} \tilde{w}_j \neq 1$, then, the FOWMA operator can be expressed as:

$$FOWMA\ (\tilde{a}_{1+t}, \tilde{a}_{2+t}, \dots, \tilde{a}_{m+t}) = \frac{1}{\tilde{W}} \sum_{j=1+t}^{m+t} w_j b_j\ . \tag{7}$$

The FOWMA operator is commutative, monotonic, idempotent and bounded. It is commutative because $FOWMA\ (a_{1+t}, a_{2+t}, \dots, a_{m+t}) = FOWMA\ (e_{1+t}, e_{2+t}, \dots, e_{m+t})$ for any permutation of the arguments. It is monotonic because if $a_i \geq e_i$, for all $i \in \{1 + t, 2 + t, \dots, m + t\}$, then: $FOWMA\ (a_{1+t}, a_{2+t}, \dots, a_{m+t}) \geq f\ (e_{1+t}, e_{2+t}, \dots, e_{m+t})$. It is idempotent because if $a_i = a$, for all $i \in \{1 + t, 2 + t, \dots, m + t\}$, then: $FOWMA\ (a_{1+t}, a_{2+t}, \dots, a_{m+t}) = a$. It is bounded by the minimum and the maximum because $\text{Min}\{a_i\} \leq FOWMA\ (a_{1+t}, a_{2+t}, \dots, a_{m+t}) \leq \text{Max}\{a_i\}$.

A further interesting issue are different measures for characterrising the weighting vector such as the degree of orness, the entropy of dispersion, the divergence of W and the balance operator [2,9,18]. The degree of orness is defined as follows.

$$\alpha(W) = \sum_{j=1+t}^{m+t} w_j \left(\frac{m-j}{m-1} \right). \tag{8}$$

It can be shown that $\alpha \in [0, 1]$. Note that for the maximum criteria $\alpha(W) = 1$ and for the minimum criteria $\alpha(W) = 0$.

Finally, note that a lot of other properties could be studied following the available literature concerning the OWA operator and the aggregation operators [2-4,9,18].

3.2 Families of FOWMA Operators

The main advantage of the FOWMA operator is that it can consider a wide range of aggregation operators ranging from the fuzzy moving minimum to the fuzzy moving maximum. Thus, it can provide a more complete representation of the information because it permits to see all the different situations that may occur and focus on those cases that are in accordance to the interests of the decision maker. For example, we can obtain the fuzzy moving maximum, the fuzzy moving minimum and the step-FOWMA operator.

Remark 1. The fuzzy moving maximum is found when $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$. The fuzzy moving minimum is formed when $w_n = 1$ and $w_j = 0$ for all $j \neq n$. More generally, the step-FOWMA is formed when $w_k = 1$ and $w_j = 0$ for all $j \neq k$. Note that if $k = 1$, the step-FOWMA becomes the fuzzy moving maximum and if $k = n$, into the fuzzy moving minimum.

Remark 2. Using the orness measure and the entropy of dispersion explained in the previous section, we can develop another group of methods for obtaining the FOWMA weights including the maximal entropy FOWMA (MEFOWMA). The aim of this method is to maximize the entropy, subject to an established degree of or-ness. It can be solved by using the following mathematical programming problem:

$$\text{maximize} \quad H(\hat{V}) = -\left(\beta \sum_{j=1+t}^{m+t} w_j \ln(w_j) + (1-\beta) \sum_{i=1+t}^{m+t} v_i \ln(v_i) \right)$$

subject to

$$\beta \sum_{j=1+t}^{m+t} w_j \left(\frac{m+t-j}{m-1} \right) + (1-\beta) \sum_{j=1+t}^{m+t} v_j \left(\frac{m+t-j}{m-1} \right) = \alpha(\hat{V})$$

$$0 \leq \alpha(\hat{V}) \leq 1, \tag{9}$$

$$\sum_{j=1+t}^{m+t} w_j = 1, \quad \sum_{j=1+t}^{m+t} v_j = 1, \quad 0 \leq w_j, v_j \leq 1, \quad j = 1, 2, \ldots, n.$$

Remark 3. Further aggregations could be developed following the recent literature for obtaining OWA weights [2-4,9,18]. Note that in this paper we consider dynamic aggregations but the methodology is the same. For example, we can consider the following particular cases.

- Median-FOWMA: If n is odd we assign $w_{(n+1)/2} = 1$ and $w_{j*} = 0$ for all others. If n is even we assign for example, $w_{n/2} = w_{(n/2)+1} = 0.5$ and $w_j = 0$ for all others.
- Centered-FOWMA: If it is symmetric, strongly decaying and inclusive. That is, if $w_j = w_{j+n-1}$; when $i < j \leq (n + 1)/2$ then $w_i < w_j$ and when $i > j \geq (n + 1)/2$, $w_i < w_j$; and if $w_j > 0$.
- Olympic-FOWMA: When $w_1 = w_n = 0$, and for all others $w_{j*} = 1/(n - 2)$.
- General olympic-FOWMA: When $w_j = 0$ for $j = 1, 2, \ldots, k, n, n - 1, \ldots, n - k + 1$, and for all others $w_{j*} = 1/(n - 2k)$, where $k < n/2$.
- Contrary case of the general olympic-FOWMA: When $w_j = (1/2k)$ for $j = 1, 2, \ldots, k, n, n - 1, \ldots, n - k + 1$, and $w_j = 0$, for all other values, where $k < n/2$.

A further generalization is the FOWAWMA operator. It is an aggregation operator that that uses weighted averages and OWA operators in the same formulation in a moving aggregation process. Moreover, it also uses fuzzy numbers in order to represent imprecise information that cannot be quantified in a simple numerical way. It can be defined as follows.

Definition 7. Let Ψ be the set of FNs. A FOWAWMA operator of dimension m is a mapping $FOWAWMA$: $\Psi^n \rightarrow \Psi$ that has an associated weighting vector W of dimension m with $W = \sum_{j=1+t}^{m+t} w_j = 1$ and $w_j \in [0, 1]$, and a weighting vector V with $\sum_{i=1+t}^{m+t} v_i = 1$ and $v_i \in [0, 1]$, such that:

$$FOWAWMA(\tilde{a}_{1+t},\ldots,\tilde{a}_{m+t}) = \beta \sum_{j=1+t}^{m+t} w_j b_j + (1-\beta) \sum_{i=1+t}^{m+t} v_i \tilde{a}_i , \tag{10}$$

where b_j is the jth largest argument of the $\tilde{a}_i$, $\beta \in [0, 1]$, m is the total number of arguments considered from the whole sample, t indicates the movement done in the average from the initial analysis and $\tilde{a}$ are the argument variables represented in the form of FNs.

If $\beta = 0$, we get the FWMA, and if $\beta = 1$, the FOWMA operator. Note that when β decreases, we are giving more importance to the FWMA operator, and when β increases, we give more importance to the FOWMA.

Further measures could be studied as shown in Section III. The FOWAWMA operator can be further extended by using a wide range of aggregation operators.

4 Application in Multi-Period Decision Making

The FOWMA and FOWAWMA operators can be applied in a wide range of problems. The main reason is that they include the classical moving average as a particular case. Therefore, all the previous studies that have been developed with the moving average can be revised and extended with this new approach. For example, we can apply it in a wide range of problems in statistics, economics and engineering. This paper develops an application in multi-period decision making. It considers a national decision making problem regarding the fixation of the optimal interest rate of a country. Particularly, we provide an example in the European Central Bank.

Assume that the European Central Bank (ECB) is planning its general interest rate for the next year and has five alternatives.

- A_1 = Increase the interest rate 0.5%.
- A_2 = Increase the interest rate 0.25%.
- A_3 = Do not make any change.
- A_4 = Decrease the interest rate 0.25%.
- A_5 = Decrease the interest rate 0.5%.

In order to evaluate these alternatives, the group of experts of the ECB considers that the key variable that determines this problem is the economic situation of the European region for the next year. They assume that three states of nature may occur:

- S_1 = Worse economic situation than the previous year.
- S_2 = Same economic situation than the previous year.
- S_3 = Better economic situation than the previous year.

They consider the information available from the last four years in order to provide an appropriate forecast that permits to reach the optimal decision. They evaluate each alternative analyzing the benefits they could give each year. Since the information is very imprecise, they use triangular FNs. Note that the information is calculated assuming the occurrence of different states of nature and then formulate the expected benefit. The information is shown in Tables 1, 2, 3 and 4. Note that the information are general evaluations given by the experts between 0 and 100, where 0 is the worst situation and 100 the highest possible benefit.

Table 1. Payoff matrix – Year 1

	S_1	S_2	S_3
A_1	[22,26,29]	[45,49,52]	[66,68,73]
A_2	[24,27,32]	[48,51,56]	[67,69,72]
A_3	[21,25,28]	[58,60,64]	[64,70,73]
A_4	[32,36,38]	[49,53,56]	[46,49,54]
A_5	[30,34,37]	[44,47,49]	[40,45,50]

Table 2. Payoff matrix – Year 2

	S_1	S_2	S_3
A_1	[37,40,46]	[52,54,57]	[63,65,69]
A_2	[32,34,39]	[56,57,59]	[60,64,68]
A_3	[36,38,43]	[52,55,57]	[56,58,59]
A_4	[40,44,46]	[48,51,56]	[59,60,63]
A_5	[39,44,50]	[50,53,57]	[66,69,75]

Table 3. Payoff matrix – Year 3

	S_1	S_2	S_3
A_1	[40,46,50]	[55,59,62]	[60,64,67]
A_2	[52,58,63]	[60,63,67]	[66,69,70]
A_3	[48,49,50]	[54,57,59]	[62,66,68]
A_4	[53,54,56]	[60,62,66]	[70,74,75]
A_5	[50,52,58]	[55,57,60]	[69,70,73]

Table 4. Payoff matrix – Year 4

	S_1	S_2	S_3
A_1	[36,38,41]	[52,55,57]	[66,67,69]
A_2	[33,38,39]	[55,57,60]	[62,64,67]
A_3	[28,32,37]	[54,58,60]	[66,67,69]
A_4	[30,32,36]	[58,62,65]	[70,72,74]
A_5	[33,40,44]	[60,61,62]	[66,69,71]

This information can be aggregated in a wide range of ways. In this example, it is considered the aggregation with the FOWMA and FOWAWMA operators. First, it is considered the FOWMA operator with the weighting vector $W = (0.3, 0.3, 0.2, 0.2)$. The results are shown in Table 5.

Table 5. Payoff matrix – Forecast Year 5 with FOWMA

	S_1	S_2	S_3
A_1	[34.7,38.6,42.8]	[51.5,54.8,57.5]	[64.2,66.3,69.8]
A_2	[36.7,41,44.8]	[55.4,57.6,61.1]	[64.3,67,69.6]
A_3	[35.2,37.5,40.9]	[54.8,57.8,60.4]	[62.6,65.9,68]
A_4	[40.3,43,45.4]	[54.8,58,61.7]	[63,66,68.1]
A_5	[39.3,43.6,48.6]	[53.3,55.4,57.8]	[61.7,64.5,68.6]

Once it is formed a single matrix, it is possible to aggregate the information again in order to form results that permit to analyze those alternatives closest to our interests. In this case, it is used the weighting vector $W = (0.4, 0.3, 0.3)$. It is considered the minimum, the maximum and a FOWMA operator as shown in Table 6.

Table 6. Aggregated results with FOWMA

	Min	*FOWMA*	*Max*
A_1	[34.7,38.6,42.8]	[51.54,54.54,58.01]	[64.2,66.3,69.8]
A_2	[36.7,41,44.8]	[53.35,56.38,59.61]	[64.3,67,69.6]
A_3	[35.2,37.5,40.9]	[52.04,54.95,57.59]	[62.6,65.9,68]
A_4	[40.3,43,45.4]	[53.73,56.7,59.37]	[63,66,68.1]
A_5	[39.3,43.6,48.6]	[52.46,55.5,59.36]	[61.7,64.5,68.6]

The optimal choice with the FOWMA operator depends on the type of aggregation operator used. With the minimum and the FOWMA, A_4 is the best alternative and with the maximum it is A_2. However, depending on the particular type of aggregation operator used, the results may lead to different rankings between the alternatives.

A similar analysis is also developed with the FOWAWMA operator as shown in Table 7 and 8. It is assumed the following weighting vector for the weighted average: $V = (0.1, 0.2, 0.3, 0.4)$ and the parameter $\beta = 0.4$. First, the analysis is focused on making a forecast for the fifth year and based on the previous four years.

Table 7. Payoff matrix – Forecast Year 5 with FOWAWMA

	S_1	S_2	S_3
A_1	[35.48,39.2,43.22]	[51.92,55.16,57.8]	[63.84,66,69.2]
A_2	[37.24,41.66,45.22]	[55.76,57.96,61.34]	[63.7,66.4,69]
A_3	[35.02,37.56,41.08]	[54.32,57.5,59.86]	[62.6,65.48,67.46]
A_4	[39.58,42.04,44.68]	[55.34,58.54,62.24]	[64.44,67.14,69.3]
A_5	[39.12,43.72,48.66]	[54.26,56.24,58.58]	[63.25,65.94,69.62]

Once the forecasts for the fifth year are available, it is possible to aggregate the information as shown in Table 8. For the weighted average, the weighting vector used is: $V = (0.3, 0.4, 0.3)$. Note that it is used the minimum, the maximum and the FOWAWMA operator.

Table 8. Aggregated results with FOWAWMA

	Min	*FOWAWMA*	*Max*
A_1	[35.48,39.2,43.22]	[51.04,54.05,57.30]	[63.84,66,69.2]
A_2	[37.24,41.66,45.22]	[52.90,55.93,59.10]	[63.7,66.4,69]
A_3	[35.02,37.56,41.08]	[51.34,54.23,56.81]	[62.6,65.48,67.46]
A_4	[39.58,42.04,44.68]	[53.70,56.51,59.37]	[64.44,67.14,69.3]
A_5	[39.12,43.72,48.66]	[52.77,55.78,59.35]	[63.25,65.94,69.62]

With the FOWAWMA operator, the optimal choice is A. Note that the main advantage of using FOWMA and FOWAWMA operators is that they provide a more complete picture of the problem since it is possible to consider the most optimistic and pessimistic scenarios and inter medium situations that are in closest accordance with our interests.

Note that in this example, only one year forecast has been considered. However, the use of FOWMA and FOWAWMA operators can also be used in a more dynamic process that considers 1 year, 2 years, 3 years and more. Furthermore, note that in this example, instead of using the moving average for one variable, it is used for the whole matrix. Thus, this paper represents complex structures in a dynamic process.

As we can see, each aggregation operator may provide different results. However, in this example it seems clear that in most of the cases expert 3 (E_3) is the one who makes the highest benefits in this financial problem since it has the lowest aggregated errors.

5 Conclusions

A new framework for fuzzy decision making in multi-period environments has been suggested. For doing so, it has been introduced the FOWMA and FOWAWMA operators. Their main advantage is that they can represent the information in a more complete and flexible way considering dynamic processes that account for different periods of time and in an imprecise environment that can be assessed with FNs. Several particular cases and some of their main properties have been studied. A further generalization by using induced aggregation operators has also been presented. Its main advantage is that it can deal with complex information in a more efficient way.

An application of the new approach in multi-period decision making has been developed. By using moving averages, it is possible to deal with information from different periods of time and form the most appropriate decisions. The paper has focused in national decision making problems where a country has to fix the general interest rate. Particularly, we have studied the general interest rate of the European Central Bank (ECB) in the European Union (EU). We have seen that this approach permits to consider historical information and form a complete forecast that considers optimistic and pessimistic scenarios and those that are in closest accordance to the interests of the decision maker.

In future research, it is expected that further developments to this approach will be developed by using further extensions and generalizations such as the use of probabilities and generalized aggregation operators. Moreover, other applications will be considered including juridical management and times series forecasting.

Acknowledgments. Support from the European Commission through the project PIEF-GA-2011-300062 is gratefully acknowledged.

References

1. Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: A guide for practitioners. Springer, Berlin (2007)
2. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. Syst. Man Cybern. B 18, 183–190 (1988)
3. Yager, R.R., Kacprzyk, J.: The ordered weighted averaging operators: Theory and applications. Kluwer Academic Publishers, Norwell (1997)
4. Yager, R.R., Kacprzyk, J., Beliakov, G.: Recent developments on the ordered weighted averaging operators: Theory and practice. Springer, Heidelberg (2011)
5. Moore, R.E.: Interval Analysis. Prentice Hall, Englewood Cliffs (1966)
6. Dubois, D., Prade, H.: Fuzzy sets and systems: Theory and applications. Academic Press, New York (1980)
7. Chen, S.J., Chen, S.M.: A new method for handling multi-criteria fuzzy decision making problems using FN-IOWA operators. Cybern. & Syst. 34, 109–137 (2003)
8. Merigó, J.M., Casanovas, M.: The fuzzy generalized OWA operator and its application in strategic decision making. Cybernetics & Syst. 41, 359–370 (2010)
9. Merigó, J.M., Gil-Lafuente, A.M.: Fuzzy induced generalized aggregation operators and its application in multi-person decision making. Expert Syst. Applic. 38, 9761–9772 (2011)
10. Yager, R.R., Filev, D.P.: Induced ordered weighted averaging operators. IEEE Trans. Syst. Man Cybern. B 29, 141–150 (1999)
11. Zhao, H., Xu, Z.S., Ni, M., Liu, S.: Generalized aggregation operators for intuitionistic fuzzy sets. Int. J. Intell. Syst. 25, 1–30 (2010)
12. Xu, Z.S.: Fuzzy harmonic mean operators. Int. J. Intell. Syst. 24, 152–172 (2009)
13. Zeng, S.Z., Su, W.: Intuitionistic fuzzy ordered weighted distance operator. Knowledge-Based Syst. 24, 1224–1232 (2011)
14. Evans, M.K.: Practical business forecasting. Blackwell Publishers, Hong Kong (2002)
15. Yager, R.R.: Time series smoothing and OWA aggregation. IEEE Trans. Fuzzy Syst. 16, 994–1007 (2008)
16. Merigó, J.M., Yager, R.R.: Generalized moving averages, distance measures and OWA operators, GIEGI Working Papers, GIEGI-10-11, University of Barcelona, Spain (2010)
17. Merigó, J.M., Yager, R.R.: Moving averages in a unified model between the OWA operator and the weighted average. In: World Congress of Soft Computing, Berkeley, San Francisco (2012)
18. Merigó, J.M.: A unified model between the weighted average and the induced OWA operator. Expert Syst. Applic. 38, 11560–11572 (2011)
19. Torra, V.: The weighted OWA operator. Int. J. Intell. Syst. 12, 153–166 (1997)
20. Xu, Z.S., Da, Q.L.: An overview of operators for aggregating information. Int. J. Intelligent Syst. 18, 953–969 (2003)
21. Yager, R.R.: Including importances in OWA aggregation using fuzzy systems modelling. IEEE Trans. Fuzzy Syst. 6, 286–294 (1998)
22. Chang, S.S.L., Zadeh, L.A.: On fuzzy mapping and control. IEEE Trans. Syst. Man Cybern. 2, 30–34 (1972)
23. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Part 1, Inform. Sci. 8, 199–249 (1975); Part 2, Inform. Sci. 8, 301–357 (1975); Part 3, Inform. Sci. 9, 43–80 (1975)
24. Zadeh, L.A.: Fuzzy sets, Inform. Control 8, 338–353 (1965)

A Generalization of the Variance by Using the Ordered Weighted Average

José M. Merigó[1,2], Montserrat Guillén[1], and José M. Sarabia[3]

[1] Department of Econometrics, Riskcenter – IREA, University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain
[2] Manchester Business School, University of Manchester, Booth Street West, M15 6PB Manchester, UK
[3] Department of Economics, University of Cantabria, Santander, Spain
jose.merigolindahl@mbs.ac.uk, mguillen@ub.edu, sarabiaj@unican.es

Abstract. The ordered weighted average is an aggregation operator that provides a parameterized family of aggregation operators between the minimum and the maximum. This paper analyzes the use of the ordered weighted average in the variance. It presents several extensions by using a unified framework between the weighted average and the ordered weighted average. Furthermore, it also develops other generalizations with induced aggregation operators and by using quasi-arithmetic means.

Keywords: Variance, Ordered weighted average, Induced aggregation operators, Quasi-arithmetic mean.

1 Introduction

The variance is a basic concept in statistics for measuring the dispersion of data. It is fundamentally related to the notion of distance and, as such, it has been at the center of virtually all data analysis applications [1]. Usually, the variance is defined as an averaging process of the individual dispersions. The most common tools for doing so are the simple average and the weighted average.

However, it is possible to introduce some additional information on the definition of weights use more sophisticated tools. For example, the ordered weighted average (OWA) [2,3] can be used for that purpose. The OWA provides a parameterized family of aggregation operators between the minimum and the maximum. Since its introduction, it has been studied and generalized by many authors. For example, Yager and Filev [4] introduced the induced OWA (IOWA) operator by using complex ordering processes assessed with order inducing variables. Fodor et al. [5] presented the quasi-arithmetic OWA (Quasi-OWA) operator that generalized a wide range of aggregation operators including the OWA and the quadratic OWA operator. Merigó and Gil-Lafuente [6] suggested a unified framework of the previous approaches named induced generalized OWA (IGOWA) operator. Some other authors such as Merigó and Casanovas [7] and Zeng and Su [8] have studied several extensions under imprecise information that can be assessed with interval and fuzzy numbers.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 222–231, 2013.

The aim of this paper is to analyze the use of the OWA operator in the variance. The main advantage of doing so is that the variance can be studied considering a wide range of scenarios from the minimum to the maximum. That is, from the most optimistic to the most pessimistic scenario. Currently, the main studies in this direction have been developed by Yager [3,9] with the use of the OWA and the IOWA operator in the variance and by Merigó [10] that focused on the use of the OWAWA and the IOWAWA operator. This work reviews these approaches and suggests some additional ones by using the weighted OWA (WOWA) [11], the hybrid average [12] and the immediate weights [13]. Furthermore, it is also introduced the use of generalized aggregation operators by using quasi-arithmetic means. Therefore, it is possible to consider a wide range of particular cases including quadratic and cubic aggregations.

This paper is organized as follows. Section 2 reviews the OWA operator, some of its extensions and its implementation in the variance. Section 3 studies the use of the OWAWA operator and some related extensions in the variance and Section 4 summarizes the main results of the paper.

2 Preliminaries

This Section presents a brief description of some basic aggregation operators, the variance and the use of the OWA operator in the variance.

2.1 Aggregation Operators

The Weighted Average

The weighted average (WA) is one of the most well-known aggregation operators. It has been used in a wide range of applications including statistics, economics and engineering. It can be defined as follows.

Definition 1. A WA operator of dimension n is a mapping WA: $R^n \rightarrow R$ that has an associated weighting vector V, with $v_i \in [0, 1]$ and $\sum_{i=1}^{n} v_i = 1$, such that:

$$WA\ (a_1, \ldots, a_n) = \sum_{i=1}^{n} v_i a_i \ , \tag{1}$$

where a_i represents the argument variable.

The WA operator satisfies the common properties of aggregation operators. For further reading on different extensions and generalizations, see for example [14-15].

An important issue when integrating the weighted average with the OWA operator is that one of them has to adapt his initial ordering to the other one. Therefore, it is useful to see how the weighted average would be formulated if it has to adapt his ordering to the OWA operator [10]. Note that this is a key feature for unifying the OWA with the WA and will be explained in Section 2.1.3.

The OWA Operator

The OWA operator [2] is an aggregation operator that provides a parameterized family of aggregation operators between the minimum and the maximum. It can be defined as follows.

Definition 2. An OWA operator of dimension n is a mapping OWA: $R^n \rightarrow R$ that has an associated weighting vector W of dimension n with $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, such that:

$$OWA\,(a_1, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j \,, \tag{2}$$

where b_j is the jth smallest of the a_i.

Note that different properties could be studied including the distinction between descending and ascending orders, different measures for characterizing the weighting vector and different families of OWA operators. For further reading, refer to [16-17].

In most of the OWA literature, the arguments are reordered from lowest to highest according to a weighting vector. However, it is also possible to reorder the weighting vector according to the initial positions of the arguments a_i [18].

The OWAWA Operator

The ordered weighted averaging – weighted average (OWAWA) operator is an aggregation operator that integrates the OWA operator and the weighted average in the same formulation and considering the degree of importance that each concept has in the analysis. It can be defined as follows.

Definition 3. An OWAWA operator of dimension n is a mapping $OWAWA$: $R^n \rightarrow R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$OWAWA\,(a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j \,, \tag{3}$$

where b_j is the jth smallest of the a_i, each argument a_i has an associated weight v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = \beta w_j + (1-\beta) v_j$ with $\beta \in [0, 1]$ and v_j is the weight v_i ordered according to b_j, that is, according to the jth smallest of the a_i.

As we can see, if $\beta = 1$, we get the OWA operator and if $\beta = 0$, the weighted average. The OWAWA operator accomplishes similar properties than the usual aggregation operators. Note that we can distinguish between descending and ascending orders, extend it by using mixture operators, and so on.

Note that Eq. (5) has been presented adapting the ordering of the weighted average to the OWA operator. However, it is also possible to formulate the OWAWA operator integrating the ordering of the OWA operator to the weighted average as:

$$OWAWA\ (a_1, \ldots, a_n) = \sum_{i=1}^{n} \hat{v}_i a_i\ , \tag{4}$$

where each argument a_i has an associated weight v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_i = \beta w_i + (1-\beta) v_i$ with $\beta \in [0, 1]$ and w_i is the weight w_j ordered according to the ordering of the arguments a_i.

The IOWAWA Operator

The induced ordered weighted averaging – weighted average (IOWAWA) operator [10] is a model that unifies the IOWA operator and the WA in the same formulation and considering a complex reordering process based on order-inducing variables. Therefore, both concepts can be seen as a particular case of a more general one. It can also be seen as a unification between decision-making problems under uncertainty (with IOWA operators) and under risk (with probabilities). Note that the motivation for using this approach instead of the OWAWA operator is especially useful when dealing with complex interpretations of the information. It can be defined as follows.

Definition 4. An IOWAWA operator of dimension n is a mapping $IOWAWA$: $R^n \times R^n \rightarrow R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$IOWAWA\ (\langle u_1,a_1\rangle, \langle u_2,a_2\rangle, \ldots, \langle u_n,a_n\rangle) = \sum_{j=1}^{n} \hat{v}_j b_j\ , \tag{5}$$

where b_j is the a_i value of the IOWAWA pair $\langle u_i,a_i\rangle$ having the jth smallest u_i, u_i is the order inducing variable and a_i is the argument variable, each argument a_i has an associated weight (WA) v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = \beta w_j + (1-\beta) v_j$ with $\beta \in [0, 1]$ and v_j is the weight (WA) v_i ordered according to b_j, that is, according to the jth smallest u_i.

If $\beta = 1$, it becomes the IOWA operator [4] and if $\beta = 0$, the weighted average. For further reading, see Merigó [10].

2.2 The Variance

Introduction

The variance and the covariance are two statistical measures of variability. The variance measures how far the numbers lie from the mean. It has been applied in a wide range of problems and it can be defined as:

$$Var\ (a_1, \ldots, a_n) = \sum_{i=1}^{n} v_i (a_i - \mu)^2\ , \tag{6}$$

where a_i is the argument variable, μ is the average and each argument $(a_i - \mu)^2$ has an associated weight (WA) v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$.

Note that in this formulation it is implicitly assumed that the variance uses a weighted average. However, it is also possible to consider it with arithmetic means, that is, when $v_i = 1/n$, for all i. Furthermore, the variance is often transformed into the standard deviation. Note that further measures could be studied as it is commonly used in statistics [1].

Variance with the OWA Operator

The OWA operator can also be used in the variance forming a measure that provides a parameterized family of aggregation operators that range from the minimum dispersion to the maximum one. The use of the OWA operator in the variance was suggested by Yager [3] and it can be defined as follows:

$$Var\text{-}OWA\ (a_1, \ldots, a_n) = \sum_{j=1}^{n} w_j D_j , \tag{7}$$

where D_j is the jth smallest of the $(a_i - \mu)^2$, a_i is the argument variable, μ is the average (in this case, the OWA operator), $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$.

The Var-OWA can be extended by using induced aggregation operators [9] in order to deal with complex reordering processes as follows:

$$Var\text{-}IOWA\ (\langle u_1,a_1\rangle, \langle u_2,a_2\rangle, \ldots, \langle u_n,a_n\rangle) = \sum_{j=1}^{n} w_j D_j , \tag{8}$$

where D_j is the $(a_i - \mu)^2$ value of the IOWA pair $\langle u_i,x_i\rangle$ having the jth smallest u_i, u_i is the order inducing variable, a_i is the argument variable, μ is the average (in this case, the OWA operator), $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$.

In the case of ties between order inducing variables the aggregation needs an adjustment. Yager and Filev [4] suggested replacing the arguments of the tied order inducing variables by their average. Thus, in the case of ties in the Var-IOWA, the arguments $(a_i - \mu)^2$ with tied inducing variables can be replaced by the average, that is, $[(a_i - \mu)^2 + (a_k - \mu)^2]/2$. The Var-OWA and the Var-IOWA operator accomplish similar properties than other OWA and IOWA operators [4,19] such as commutativity, the boundary condition, idempotency, monotonicity and reflexivity. Note that the Var-IOWA includes the classical variance with the weighted average when the reordering of the order inducing variables is equal to the ordering of the arguments a_i. Moreover, it includes the classical variance as a particular case when $w_j = 1/n$ for all i.

3 OWAWA Operators in the Variance

Recently, Merigó [10] has suggested a new aggregation approach that integrates the OWA operator with the weighted average considering the degree of importance that each concept has in the specific problem considered. He called it the OWAWA operator. He also briefly showed that it is possible to use the OWAWA in the variance as follows:

$$Var\text{-}OWAWA\ (a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j D_j \ , \tag{9}$$

where D_j is the jth smallest of the $(a_i - \mu)^2$, a_i is the argument variable, μ is the average (in this case, the OWAWA operator), $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, each argument $(a_i - \mu)^2$ has an associated weight (WA) v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_i = \beta w_i + (1-\beta) v_i$ with $\beta \in [0, 1]$ and v_j is the weight (WA) v_i ordered according to D_j, that is, according to the jth smallest of the $(a_i - \mu)^2$.

Note that the Var-OWAWA operator can also be formulated separating the part that affects the OWA operator and the part concerning the weighted average as:

$$Var\text{-}OWAWA\ (a_1, \ldots, a_n) = \beta \sum_{j=1}^{n} w_j D_j + (1-\beta) \sum_{i=1}^{n} v_i (a_i - \mu)^2 \ , \tag{10}$$

Moreover, it is also possible to adapt the reordering of the OWA operator to the weighted average as:

$$Var\text{-}OWAWA\ (a_1, \ldots, a_n) = \sum_{i=1}^{n} \hat{v}_i (a_i - \mu)^2 \ , \tag{11}$$

where a_i is the argument variable, μ is the average (in this case, the OWAWA operator), $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, each argument $(a_i - \mu)^2$ has an associated weight v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_i = \beta w_i + (1-\beta) v_i$ with $\beta \in [0, 1]$ and w_i is the weight w_j ordered according to the ordering of the arguments a_i and j is the ordering of the jth smallest $(a_i - \mu)^2$.

Obviously, once we have the variance, it is straightforward to obtain the standard deviation (S.D.) with the OWAWA operator by using:

$$\text{S.D.} = \sqrt{\sum_{j=1}^{n} \hat{v}_j D_j} \ , \tag{12}$$

The Var-OWAWA operator accomplishes similar properties as the OWAWA operator including the boundary and semi boundary condition, monotonicity and idempotency. It is worth noting that if $\beta = 1$, it becomes the Var-OWA operator and if $\beta = 0$, the classical variance with the weighted average. The more of β located to the top, the more importance we give to the OWA aggregation and vice versa.

An important issue when analyzing the Var-OWAWA operator is the use of measures for characterizing the weighting vector. In the OWA literature [2,10], there are several measures for doing so including the or-ness measure, the entropy of dispersion, the balance operator and the divergence. The or-ness measure for an OWAWA operator can be formulated as follows:

$$\alpha(\hat{V}) = \beta \sum_{j=1}^{n} w_j \left(\frac{j-1}{n-1} \right) + (1-\beta) \sum_{j=1}^{n} v_j \left(\frac{j-1}{n-1} \right). \tag{13}$$

It is straightforward to calculate the and-ness measure by using the dual: $Andness(\hat{V}) = 1 - \alpha(\hat{V})$. Note that $\alpha \in [0, 1]$.

There are different methods for defining the entropy of dispersion [2,10,20]. In this study, the combined entropy is defined as follows:

$$R(\hat{V}) = -\left(\beta \sum_{j=1}^{n} w_j \ln(w_j) + (1-\beta) \sum_{i=1}^{n} v_j \ln(v_j) \right). \tag{14}$$

Note that if $\beta = 1$, this measure becomes the Yager [2] entropy of dispersion and if $\beta = 0$, the Shannon [20] entropy.

Another interesting feature of the Var-OWAWA is that it includes a wide range of particular cases by selecting a different manifestation in the weighting vector. Among others, the following cases shown in Table 1 are included as particular cases.

Table 1. Families of Var-OWAWA operators

Weights	Particular case
$w_j = 1/n$ and $v_i = 1/n$, for all i, j	Simple variance
$w_1 = 1$ and $w_j = 0$ for all $j \neq 1$	Minimum weighted variance
$w_n = 1$ and $w_j = 0$ for all $j \neq n$	Maximum weighted variance
$w_j = 1/n$, for all j	Arithmetic weighted variance
$v_i = 1/n$, for all i	Arithmetic Var-OWA
$w_k = 1$ and $w_j = 0$ for all $j \neq k$	Step-Var-OWA weighted variance
$w_1 = 1-\alpha$, $w_n = \alpha$ and $w_j = 0$ for all $j \neq 1, n$	Hurwicz weighted variance
If n is odd we assign $w_{(n+1)/2} = 1$ and $w_j = 0$ for all others	Median-Var-OWA weighted variance
If n is even we assign for example, $w_{n/2} = w_{(n/2)+1} = 0.5$ and $w_j = 0$ for all others	Median-Var-OWA weighted variance
$w_1 = w_n = 0$, and for all others $w_j = 1/(n-2)$	Olympic-Var-OWA weighted variance

Moreover, note that in the literature there are other methods for integrating the OWA operator and the weighted average in the same formulation. The main approaches are the weighted OWA (WOWA) [11], the hybrid average [12] and the immediate weights [13]. Thus, these approaches could also be considered when constructing new types of variance measures. By using the WOWA operator, the variance can be expressed as follows.

Let P and W be two weighting vectors of dimension n [$P = (p_1, p_2, \ldots, p_n)$], [$W = (w_1, w_2, \ldots, w_n)$], such that $p_i \in [0, 1]$ and $\sum_{i=1}^{n} p_i = 1$, and $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$. In this case, a mapping *Var-WOWA*: $R^n \rightarrow R$ is a Var-WOWA operator of dimension n if:

$$Var\text{-}WOWA\ (a_1, \ldots, a_n) = \sum_{i=1}^{n} \omega_i D_{\sigma(i)}\ , \tag{15}$$

where $\{\sigma(1), \ldots, \sigma(n)\}$ is a permutation of $\{1, \ldots, n\}$ such that $D_{\sigma(i-1)} \geq D_{\sigma(i)}$ for all $i = 2, \ldots, n$. (i.e. $D_{\sigma(i)}$ is the ith largest in the collection $D_1, \ldots, D_n$ and $D_i = (a_i - \mu)^2$), and the weight ω_i is defined as:

$$\omega_i = w^*\left(\sum_{j \leq i} p_{\sigma(j)}\right) - w^*\left(\sum_{j < i} p_{\sigma(j)}\right), \tag{16}$$

with w^* a monotonically increasing function that interpolates the points $(i/n, \sum_{j \leq i} w_j)$ together with the point (0, 0). w^* is required to be a straight line when the points can be interpolated in this way.

By using the hybrid average it becomes the hybrid averaging variance (Var-HA) and it can be formulated as follows.

A Var-HA operator of dimension n is a mapping *Var-HA:* $R^n \rightarrow R$ that has an associated weighting vector W of dimension n with $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, such that:

$$Var\text{-}HA\ (a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j D_j\ , \tag{17}$$

where D_j is the jth largest of the $\hat{a}_i$ ($\hat{a}_i = n\omega_i(a_i - \mu)^2$, $i = 1,2,\ldots,n$), $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ is the weighting vector of the a_i, with $\omega_i \in [0, 1]$ and the sum of the weights is 1.

Finally, the immediate weighted average [13] may form the immediate variance (Var-IWA) in the following way.

A Var-IWA operator of dimension n is a mapping *Var-IWA*: $R^n \rightarrow R$ that has an associated vector W of dimension n with $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, such that:

$$Var\text{-}IWA\ (a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j D_j\ , \tag{18}$$

where D_j is the jth largest of the $(a_i - \mu)^2$, each a_i has associated a weight v_i, v_j is the weight v_i ordered according to b_j, and $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$.

It is worth noting that there are other methods for dealing with OWA operators and weighted averages in the same formulation that could be considered [17].

The variance of a population (discrete case) using the IOWAWA operator can be expressed with the following formulation:

$$Var\text{-}IOWAWA\ (\langle u_1,a_1\rangle, \langle u_2,a_2\rangle, \ldots, \langle u_n,a_n\rangle) = \sum_{j=1}^{n} \hat{v}_j D_j\ , \tag{19}$$

where D_j is the $(a_i - \mu)^2$ value of the IOWAWA pair $\langle u_i, a_i\rangle$ having the jth smallest u_i, u_i is the order inducing variable, a_i is the ith argument variable of the set X, μ is the average, $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, each argument $(a_i - \mu)^2$ has an associated weight (WA) v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_i = \beta w_i + (1-\beta) v_i$ with $\beta \in [0, 1]$ and v_j is the weight (WA) v_i ordered according to D_j, that is, according to the jth smallest u_i.

Obviously, it is straightforward to obtain the standard deviation (S.D.) with the IOWAWA operator as it has been explained in Eq. (12). In the case of ties between order inducing variables in the reordering process of the arguments $(a_i - \mu)^2$, it is recommended the use of the common policy for solving these situations that consists in replacing the tied arguments by their average [4].

Furthermore, it is possible to use generalized aggregation operators in the analysis. Thus, it is possible to form the generalized Var-OWAWA by using generalized means [6] and the quasi-arithmetic Var-OWAWA (Var-QOWAWA).

$$Var\text{-}QOWAWA\ (a_1, \ldots, a_n) = \beta g^{-1}\left(\sum_{j=1}^{n} w_j g(D_j)\right) + (1-\beta) h^{-1}\left(\sum_{i=1}^{n} v_i h((a_i - \mu)^2)\right), \tag{20}$$

where D_j is the jth smallest of the $(a_i - \mu)^2$, a_i is the argument variable, μ is the average, $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, each argument $(a_i - \mu)^2$ has an associated weight v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\beta \in [0, 1]$, v_j is the weight v_i ordered according to D_j, that is, according to the jth smallest of the $(a_i - \mu)^2$ and g and h are strictly continuous monotonic functions.

4 Conclusions

This paper has presented an overview regarding the use of the OWA operator in the variance and some fundamental extensions. The main advantage of this approach is that it provides a parameterized family of aggregation operators between the minimum and the maximum. Thus, in uncertain environments it is possible to reconsider the traditional variance under a wide range of scenarios that may occur from the most pessimistic to the most optimistic one. Moreover, it has been demonstrated that the classical variance is included as a particular case. Several extensions and generalizations have been introduced by using the OWAWA, the WOWA, the hybrid average and immediate weights. Furthermore, some extensions with induced aggregation operators and quasi-arithmetic means have been presented.

In future research, several generalizations will be studied in the analysis by using additional statistical tools and aggregation operators. Moreover, applications will be studied in a wide range of fields including economics and engineering.

Acknowledgments. Support from the European Commission through the project PIEF-GA-2011-300062, ECO2010-21787-C03-01 from the Spanish Government and ICREA Academia, is gratefully acknowledged.

References

1. McClave, J.T., Sincich, T.: Statistics, 9th edn. Prentice Hall, Upper Saddle River (2003)
2. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. Syst. Man Cybern. B 18, 183–190 (1988)
3. Yager, R.R.: On the inclusion of variance in decision making under uncertainty. Int. J. Uncert. Fuzz. Knowledge-Based Syst. 4, 401–419 (1996)
4. Yager, R.R., Filev, D.P.: Induced ordered weighted averaging operators. IEEE Trans. Syst. Man Cybern. B 29, 141–150 (1999)
5. Fodor, J., Marichal, J.L., Roubens, M.: Characterization of the ordered weighted averaging operators. IEEE Trans. Fuzzy Syst. 3, 236–240 (1995)
6. Merigó, J.M., Gil-Lafuente, A.M.: The induced generalized OWA operator. Inform. Sci. 179, 729–741 (2009)
7. Merigó, J.M., Casanovas, M.: The uncertain induced quasi-arithmetic OWA operator. Int. J. Intelligent Systems 26, 1–24 (2011)
8. Zeng, S.Z., Su, W., Le, A.: Fuzzy generalized ordered weighted averaging distance operator and its application to decision making. Int. J. Fuzzy Systems 14, 402–412 (2012)
9. Yager, R.R.: Generalizing variance to allow the inclusion of decision attitude in decision making under uncertainty. Int. J. Approximate Reasoning 42, 137–158 (2006)
10. Merigó, J.M.: A unified model between the weighted average and the induced OWA operator. Expert Syst. Applic. 38, 11560–11572 (2011)
11. Torra, V.: The weighted OWA operator. Int. J. Intelligent Syst. 12, 153–166 (1997)
12. Xu, Z.S., Da, Q.L.: An overview of operators for aggregating information. Int. J. Intelligent Syst. 18, 953–969 (2003)
13. Merigó, J.M., Gil-Lafuente, A.M.: Decision making techniques in business and economics based on the OWA operator. SORT – Stat. Oper. Res. Trans. 36, 81–101 (2012)
14. Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: A guide for practitioners. Springer, Berlin (2007)
15. Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E.: Aggregation functions: Means. Inform. Sci. 181, 1–22 (2011)
16. Yager, R.R.: Families of OWA operators. Fuzzy Sets Syst. 59, 125–148 (1993)
17. Yager, R.R., Kacprzyk, J., Beliakov, G.: Recent developments on the ordered weighted averaging operators: Theory and practice. Springer, Heidelberg (2011)
18. Yager, R.R.: New modes of OWA information fusion. Int. J. Intelligent Syst. 13, 661–681 (1998)
19. Merigó, J.M., Casanovas, M.: Decision making with distance measures and induced aggregation operators. Computers & Indust. Engin. 60, 66–76 (2011)
20. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Technical J. 27, 379–423 (1948)
21. Zhou, L.G., Chen, H.Y., Liu, J.P.: Generalized power aggregation operators and their applications in group decision making. Computers & Indust. Engin. 62, 989–999 (2012)

Application of Finite Spectrum Assignment Method on Distillation Column

Djordje N. Dihovicni

Visoka tehnicka skola, Bulevar Zorana Djindjica 152-a, 11000 Belgrade, Serbia
ddihovicni@gmail.com

Abstract. For multivariable multi-delay systems which is represented by distillation column, is applied a frequency- domain approach to an arbitrary finite spectrum assignment for time-delay systems. By this approach the unstable open-loop systems may be stabilized and therefore controlled by means of some well- established methods for time delay open-loop stable systems. It is proved that the control law achieves arbitrary finite spectrum assignment and is physically realizable.

Keywords: time-delay, pole assignment, finite spectrum.

1 Introduction

The effective control of many industrial processes is very difficult because of the presence of considerable time delays. The main idea is to convert the infinite spectrum design problem to a finite one, through the removal of all delays from the characteristic equation. However, time-delay compensators (for example Smith predictor for scalar systems) are unable to stabilize open-loop unstable systems. Therefore, new methods for stabilization of such systems were introduced, mainly in state- space domain. In recent years there has been an increasing interest in frequency analysis and synthesis for linear multivariable systems and there has been an intention to apply frequency domain techniques used in un-delayed systems for delay systems, [2].

In time-delay systems, delay might occur in state, control and as well in state and in control, when arises the principal difficulty in the control loop such as the increased phase lag which leads to unstable control system behavior at relatively low controller gains. These problems are getting even more complex in case of distributed parameter systems with time delay or hereditary systems, such as distillation, chemical reactors and paper making. Distillation column control is good example of this type of problem.

2 Model of Distillation Column

Distillation column models are shown with partial different equations, and solution of these systems depends on time and space coordinate, so besides distribution part, it is

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 232–238, 2013.

presented also time delay component. Controller design procedures for distributed parameter with time delay usually involve using a prediction device in the control loop to compensate for time delays. In that case it is used standard ODE multivariable controller design procedure. Time delay occurs in state [9], output or in control and methods for their compensation involve simple procedures such as Smith predictor, or statistical state-estimation procedures.

The control strategy for the multi-side stream distillation column includes very high-order time-domain model of the column involving concentrations and temperature on every tray. The experimentally determined transfer function includes pure time delays which are often very small.

The binary distillation column used for methanol-water separation, studied by Wood and Berry [10], Shah and Fisher [8], an Meyer et al [1].

The model of transfer function is given by:

$$\bar{y} = G(s)\bar{u}(s) + G_d(s)\bar{d}(s). \tag{1}$$

The transfer function is determined from data in reference [1].

$$G(s) = \begin{bmatrix} \dfrac{12.8e^{-s}}{16.7s+1} & \dfrac{18.9e^{-3s}}{21.0s+1} \\ \dfrac{6.6e^{-7s}}{10.9s+1} & \dfrac{19.4e^{-3s}}{14.4s+1} \end{bmatrix}. \tag{2}$$

If the multidelay compensator is applied, the characteristic equation contains no time delays. In case of applying finite spectrum assignment method, we are keeping real model and get desired dynamic characteristics of the system, by choosing poles of the system on pre-dominant places.

The schematic diagram is shown in figure 1.

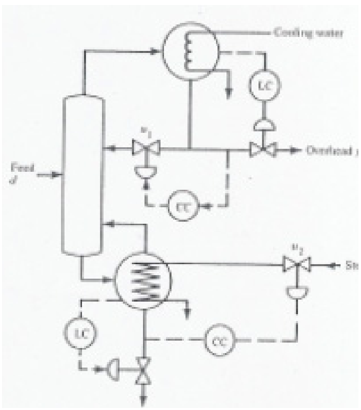


Fig. 1. Schematic diagram of the methanol distillation column with conventional two point column control system

3 Finite Spectrum Assignment Method

In frequency domain we are not interested in time delay nature (transport, technology or information delay) as shown in the reference [7], so the frequency domain technique is very powerful for these kinds of problems.

Consider multivariable system with multiple delays described by:

$$Y(s) = G(s)U(s) . \tag{3}$$

In the above equation Y and U are p- dimensional and m- dimensional output and input vectors, τ is time delay, and the transfer function can be presented as:

$$G = [G_r], \; r = ij, \; i = 1,2,..., p, \quad j = 1,2,...,m . \tag{4}$$

$$G_{ij} = \sum_{k=1}^{v_{ijk}} G_{ijk} \exp(-\tau_{ijk} s). \tag{5}$$

It is assumed that for any i, j and k, Gijk are strictly proper rational functions with real coefficients which has no multiple poles.The transfer function with time delay Gijk, have the following partial fraction form:

$$G_{ijk} = \sum_{k=1}^{\mu_{ijk}} \frac{1}{s - \lambda_{ijk}} \alpha_{ijkl} . \tag{6}$$

From a given G, we define a matrix G^0 with the same dimension as G:

$$\begin{aligned} G_{ij}{}^0 &= \sum_{k=1}^{\mu_{ij}} G_{uk}{}^0 \\ G_{ijk} &= \sum_{l=1}^{v_{ijk}} \frac{1}{s - \lambda_{ijk}} \beta_{ijkl} \\ \beta_{ijkl} &= \alpha_{ijkl} \exp(-\lambda \bullet \tau_{ijk} \lambda_{ijk}) \end{aligned} \quad . \tag{7}$$

It follows that G^0 can be factorized as a right co-prime polynomial matrix function:

$$G^0(s) = R(s)P^{\ 1}(s) . \tag{8}$$

It is possible as shown in the reference [2], to determine a polynomial matrix $P_c(s)$ which has column degree equal to those of P(s), such that $[\det P_c(s)]^{-1}$ has a pre-assigned set of closed- loop poles to be achieved by an appropriate control law, and matrix F has column degree strictly less than those of P, and is defined by:

$$F = P - P_c . \tag{9}$$

For the purpose of finite spectrum assignment [5], the following control law is introduced, where D is the differential operator with respect to time t, and V is a reference vector:

$$Q(D)\overline{U}(t) = K(D) \bullet \overline{U}(t) + H(D)\{\overline{Y}(t) + T[\overline{U}(t)]\} + Q(D) \bullet V(t). \tag{10}$$

The polynomial matrices Q, K and H satisfy following equation:

$$K(s)P(s) + H(s)R(s) = Q(s)F(s). \tag{11}$$

With [det Q(s)]$^{-1}$ being stable, and $Q^{-1}H$ proper, which is given in reference [3], it yields:

$$\begin{aligned} t_i &= \sum_{j=1}^{m} t_{ij} \\ t_{ij} &= \sum_{k=1}^{\mu_{ij}} \int_0^{\tau_{ijk}} \sum_{l=1}^{\nu_{ijk}} \beta_{ijk} \bullet \exp(\chi_{ijk} \bullet \sigma) \bullet u_j(t-\sigma) \bullet d\sigma \end{aligned}. \tag{12}$$

4 Control of Distillation Column

For controlling the distillation column by using finite spectrum assignment method, we started from the transfer function given by equation 2, and by appropriate transformations we finally obtained desired form of the denominator.

By definition it is obtained the transfer function with no delays:

$$G^0(s) = \begin{bmatrix} \dfrac{12.8}{s+0.05} & \dfrac{18.9}{s+0.04} \\ \dfrac{6.6}{s+0.05} & \dfrac{19.4}{s+0.04} \end{bmatrix}. \tag{13}$$

Transfer function G^0 is factorized as matrix function:

$$G^0(s) = R(s)P^{\ 1}(s) = \begin{bmatrix} 12.8 & 18.9 \\ 6.6 & 19.4 \end{bmatrix} \cdot \begin{bmatrix} s+0.05 & 0 \\ 0 & s+0.04 \end{bmatrix}. \tag{14}$$

Then we chose polynomial matrix:

$$P_c(s) = \begin{bmatrix} s+3 & 0 \\ 0 & s+0.04 \end{bmatrix}, Q = \begin{bmatrix} s+9 & 0 \\ 0 & s+0.04 \end{bmatrix}. \tag{15}$$

Matrix F is given as:

$$F = P - P_c - = \begin{bmatrix} -1.62 & 0 \\ 0 & -0.006 \end{bmatrix}. \tag{16}$$

Polynomial matrix K, H, P and Q are given as:

$$P = P_1 s + P_0,\ Q = Q_1 s + Q_0,\ R = R_0, F = F_0 . \tag{17}$$

To solve the equation KP + HR = QF, with $P_1=Q_1=I_2$, it yields:

$$\begin{aligned} P_0 &= \begin{bmatrix} 0.05 & 0 \\ 0 & 0.04 \end{bmatrix}, Q_0 = \begin{bmatrix} 9 & 0 \\ 0 & 0.04 \end{bmatrix}. \\ H &= I_2 s + H_0,\ K = K_0 \end{aligned} \tag{18}$$

If we put equation (18) and (17) into equation KP + HR = QF, [2] then we obtains:

$$\begin{aligned} & K_0 + R_0 = F_0 \\ & K_0 P_0 + H_0 R_0 = Q_0 F_0 \\ & K = K_0 = \begin{bmatrix} 1.92 & 0.76 \\ 0.11 & 2.4 \end{bmatrix} . \\ & H_0 = \begin{bmatrix} 0.48 & 0.26 \\ 0.72 & 2.4 \end{bmatrix} \\ & H = I_2 s + H_0 = \begin{bmatrix} s+0.48 & 0.26 \\ 0.72 & s+2.4 \end{bmatrix} \end{aligned} \tag{19}$$

This is compensation method where $U=PP_c^{-1}V$ and Y=GU, then the closed-loop transfer matrix is:

$$G_{yv} = GPP_c^{-1} = \begin{bmatrix} \dfrac{12.8}{s+3} & \dfrac{18.9}{s+005} \\ \dfrac{6.6}{s+3} & \dfrac{19.4}{s+0.05} \end{bmatrix}. \tag{20}$$

Simulation in Matlab programming language is conducted, and from engineering view of point, as shown in reference [4], it has shown in figures 2 and 3, the efficiency of the results.

The step responses for initial transfer function with time delay are presented in figure 2.

In figure 3, are shown the step responses for closed-loop transfer matrix. The time delay component is eliminated from the transfer function, and the poles are assigned at desired places, hence the dynamic characteristics might be easily controlled.

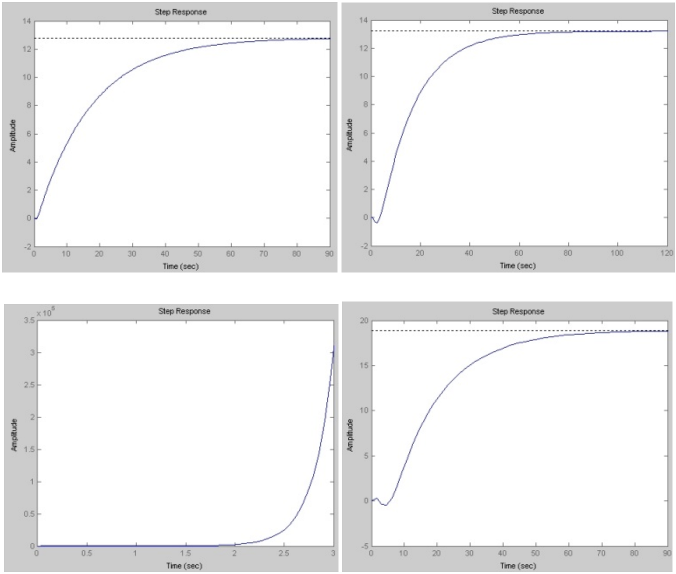


Fig. 2. The step responses for initial transfer function with time delay

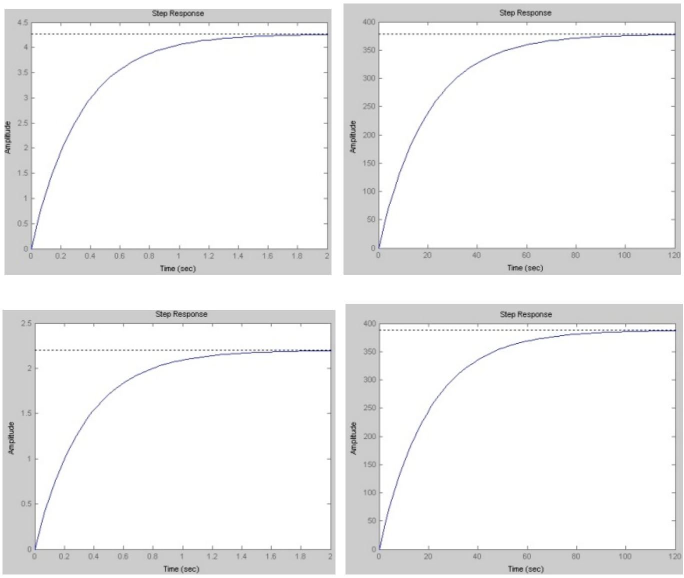


Fig. 3. The step responses for closed-loop transfer matrix

5 Conclusion

The binary distillation column used for methanol-water separation is controlled by using a frequency domain approach to an arbitrary finite spectrum assignment for multivariable time-delay systems. The process model is degenerative that makes the effective control very difficult and its dynamic performances not acceptable. Improving these process characteristics can be successfully obtained by using finite spectrum assignment method for multivariable delay systems in frequency domain. Spectrum can be arbitrary assigned by suitable choice of matrices Q and P_c. Program support for these kind of problems is developed by using symbolic package Maple, and it is shown in reference [2]. After synthesis, transfer function of process model becomes non-degenerative, system delay remains the same and remarkable features of the system could be seen by analyzing its responses.

6 Discussion

In this scientific paper, a frequency domain approach to an arbitrary finite spectrum assignment for multivariable time delay system is used, and it is given the graphic presentation. Taking into account, that the most modern chemical systems should be described by using partial different equations of higher order, it is very important focusing on development of more complex and faster program supports, for solving them. Combining time delay component and as well distributed parameter component, would open the field for discussion of applying adequate approaches for solving these kind of problems, [6].

References

1. Meyer, C.D., Seborg, D.E., Wood, R.K.: AlChE. J. 25, 24 (1979)
2. Dihovicni, D., Debeljkovic, D., Milinkovic, S., Ilic, V.: Program Support for Multivariable Time-Delay Systems in the Frequency Domain. In: Proc. ITHURS 1996, Leon- Spain, pp. 399–403 (1996)
3. Dihovicni, D., Nedic, N.: Program Support for Distributed Control Systems on Finite Space Interval. International Journal of Intelligent Technology 2(1), 49–54 (2007) ISSN 1305-6417
4. Dihovicni, D., Medenica, M.: Mathematical Modelling and Simulation of Pneumatic Systems. In: Advances in Computer Science and Engineering, ch. 9, India, pp. 161–186 (2011) ISBN 978-953-307-173-2
5. Dihovicni, D., Nedic, N.: Simulation, Animation and Program Support for a High Performance Pneumatic Force Actuator System. Mathematical and Computer Modelling 48, 761–768 (2008)
6. Dihovicni, D.: Constructing knowledge database in decision making. In: Communications, Zelina University, Slovakia, vol. 3, pp. 74–77 (2011)
7. Dihovicni, D.: Constructing knowledge database in decision making. In: Communications, Zelina University, Slovakia, vol. 3, pp. 74–77 (2011)
8. Shah, S.I., Fisher, D.G.: Proc. Joint Automatic Control Conference, Phila (1978)
9. Wang, Q.G., Sun, Y.X., Zhou, C.H.: Finite Spectrum Assignment for Multivariable Systems in the Frequency Domain. Int. J. Control 47(3), 729–734 (1988)
10. Wood, R.K., Berry, M.W.: Chem. Eng. Sci. 28, 1707 (1973)

Kinetic Modelling of Sphalerite in Sulfuric Acid Solution

Abdelhakim Begar[1], M.A. Djeghlal[2], and M. Ould Hamou[3]

[1] Department of Mechanical Engineering, University of Biskra, Biskra BP145, Algeria
begarabdelhakim@yahoo.fr
[2] LSGM Laboratory, Department of Metallurgy, National Polytechnic School,
10 Avenue Pasteur B.P182, El-Harrach Algiers, Algeria
[3] Department of Mining and Mineral Engneering, National Polytechnic School,
10 Avenue Pasteur B.P182, El-Harrach Algiers, Algeria
ouldhamoumalek@yahoo.fr

Abstract. The dissolution of sphalerite in aqueous sulfuric acid is studied in the presence of oxygen in an autoclave at 200 C. In the presence of oxygen, and an oxidized hydrogen sulfide there's formation of elemental sulphur and sulfuric acid. The kinetic of the reaction is studied as a function of the mass of zinc sulfide, the concentration of sulfuric acid, the oxygen partial pressure and temperature. A kinetic model taking into account those factors is hence proposed. On the other hand, the role of hydrogen sulfide on the progress of dissolution is highlighted.

Keywords: sphalerite, sulfuric acid, activity coefficients, leaching, kinetic model.

1 Introduction

Zinc is the third largest non-ferrous metal in the world and is conventionally produced from sphalerite. Strict environmental restrictions are imposed on sulfide smelters and the necessity of using complex deposits have stimulated the development of alternative methods in particular hydrometallurgical ones, as to avoid the production of SO2, a pollutant. In the course of this four decades, attention has been granted to the leaching of zinc sulfide ores concentrates with sulfuric acid (Forward and Veltman, 1959; Parker, 1961; Demopoulos and Baldwin, 1999) [1,2, 3], nitric acid (Bjorling, 1954) [4], hydrochloric acid (Mizoguchi and Habashi, 1981; Majima and al. 1981[5,6], and acidified solutions containing ferric ions (Bobeck and Su, 1985; Palencia Perez and Dutrizac, 1991) [7,8].

The hydrometallurgical pretreatment of sulfides can be performed using different methods (Havlik and Kammel, 2000, Havlik and al, 2001a, b) [9, 10, 11].

Weisener and al. (2003, 2004) [12, 13] suggested the rate of dissolution of sphalerite in the acid, the solution being controlled by oxygen diffusion through a porous nS° or polysulfide surface layer. As a result of their batch experiences at pH 2.5 in the presence of O2, Abraitis and al (2004) [14] have found that the main dissolution reaction for sphalerite in these conditions was a none-oxidization. Finally, Malmström

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 239–249, 2013.

and Collin (2004) [15] have studied for a long period the dissolution ofsphalerite, using air purged and dissolution experiments by batch at pH values of 1 to 4 and showed that the dissolution rate of sphalerite is dependent on the pH.

(Balaz and Ebert, 1991) [16] Studied the correlation between the changes on the surface and the properties of sphalerite because of the mechanic of activation with a rate of leaching and oxidation of minerals in presence. Hydrogen peroxide was selected as a strong oxidizing leaching model for leaching. This selective leaching agent for copper dissolves all the iron which is often present in the mineral in a large quantity as a precipitate (Anthony and al., 1990) [17] rushes.

This study is about the application of the non-selective leaching agent (H 2 SO 4), which dissolves the zinc and iron from the sphalerite (15.53% Fe), and to evaluate the possible influence of sphalerite; activation by milling on the leaching selectivity defined as Zn / Fe ratio mass.

The Direct leaching of sphalerite concentrate with ferric iron under atmospheric pressure has been described elswhere (Kammel and al. 1987; Crundwell, 1987; Suni and al. 1989; Palencia Perez and Dutrizac, 1991, Cheng and al. 1994; Pedlik and Lochmann, 1995) [18, 19, 20, 21, 22, 23]. As the zinc is leached, the ferric iron is reduced to ferrous iron by the sulfur in the zinc sulfide according to the following overall reaction stoichiometry:

$$Fe2\ (SO4)3(aq) + ZnS(s) \rightarrow 2FeSO4\ (aq) + ZnSO4\ (aq) + S(s)$$

The solid–liquid reaction takes place in acidic environment and elemental sulfur is the solid product formed. In currently applied large-scale zinc dissolution processes, the produced ferrous iron is re-oxidized in order to continue the leaching. However, if the aim is to reduce the ferric iron, the zinc sulfide concentrate can also be used as a reduction agent and is then leached simultaneously.

2 Experimental

Leaching under pressure of sphalerite of Algeria in an aqueous sulfuric acid was studied in an autoclave at temperatures up to 200 ° C with the presence of oxygen. Many parameters were involved in this reaction: influence of time, of the initial mass of zinc sulfide, of the oxygen pressure, of the temperature and of the concentration of sulfuric acid. A kinetic model taking into account these factors is proposed. In the other hand, the role of hydrogen sulfide on the progress of dissolution is highlighted. The sample used as a 98.495 in sphalerite. Analysis by atomic absorption of the sample employees gave the following composition s= 32.15% and Zn = 66,345%.

3 Results and Comments

3.1 Influence of Time

The experiment consists in attacking 5g of concentrate ZnS with 400ml of initial concentration of sphalerite (0.125 M and 0.5 M) at a pressure of 106Pa for 2 to 120 minutes at a temperature ranging from 50 °C to 200 °C.

In the sulfuric acid medium of 0.125M, the concentration of total dissolved zinc is plotted versus time in Figure (1). This graph shows that the part of the curves corresponding to the oxidation reactions is linear between 15 and 120 min and at temperatures at or below 180 ° C.

In case where the temperature is 130°C, and time of less than 15 minutes, the curve represents the dissolution of zinc sulfide with sulfuric acid

$$ZnS + H_2SO_4 \rightleftharpoons ZnSO_4 + H_2S \qquad \textbf{R-1}$$

This reaction is too fast, and the measures are not precise as to obtain a significant initial velocity.

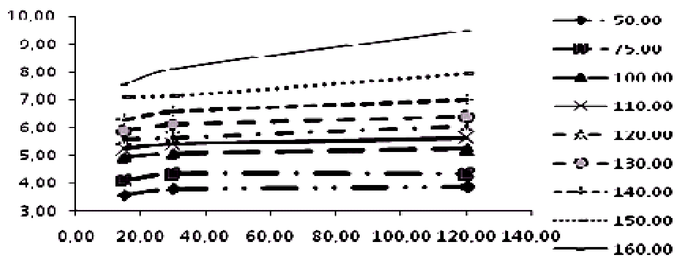


Fig. 1. Dissociation of Sphalerite as function of time between 50 and 160°C

From 15 minutes time, the slope of this line is used to define an average rate of reaction which corresponds to the linear part of the dissolution curves

This equation is of the form [Zn +2] t = Kt + [Zn +2] eq E-1

Thus the reaction is very slow, which correspond to the oxidation of hydrogen sulfide and sulfur.

$$H_2S + \frac{1}{2}O_2 \rightarrow H_2O + S° \qquad R2$$

$$S° + \frac{3}{2}O_2 + H_2O \rightarrow H_2SO_4 \qquad R3$$

$$H_2S + H_2SO_4 \rightarrow H_2O + H_2SO_3^- + S° \qquad R4$$

At 190 and 200 ° C, the curves in Figure (2) have a slight concavity turned toward the time axis. The equilibrium corresponding to the first reaction can be considered as reached.

In case of the temperature at 190 ° C, the concentration of the total dissolved zinc is plotted versus the time in Figure (2). This graph shows that the slope of this line is used to define the average rate of reaction.

The relationship [Zn 2 +] t = [Zn 2 +] eq + kt is not well verified because of the strong contribution of the oxidation reactions of hydrogen sulfide and sulfur. The reactions of sulfur oxidation and decomposition of the hydrogen sulfide are added to the oxidation reaction of hydrogen sulfide.

In a sulfuric medium of 0, 5M, the concentration of total dissolved zinc, is plotted versus time in Figure (3).

This graph shows that the curves of the [Zn 2 +] = f (t), corresponding to the oxidation reactions is linear between 15 and 120 min and at temperatures at or below 150°C. In addition, the curves are more concave, but the average rate of reaction will always be taken as

$$\Delta [Zn\ 2\ +] = ([Zn2\ +]\ 120mn - [Zn\ 2\ +]\ eq). \qquad E2$$

The curves observed here are in good agreement with the decrease in total moles of gas system over time. After a period (20 minutes) of establishment of equilibrium, there is an almost linear decrease in the number of moles versus time for temperatures below 170-180 ° C.

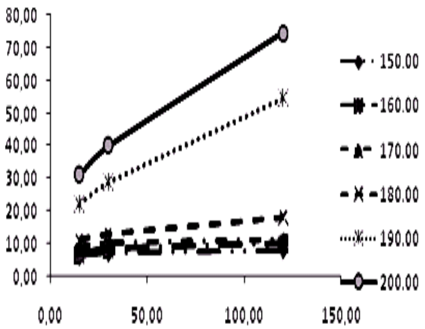


Fig. 2. Dissociation of Sphalerite as function of time between 150 and 200°C

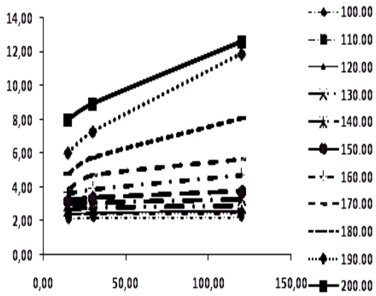


Fig. 3. Dissociation of sphalerite as function of time between 100 and 200°C

3.2 Influence of the Initial Mass in Zinc Sulfide

The experiment consists in attacking 5 to 50g of concentrate ZnS with 400ml of initial concentration of sphalerite (0.125 M) at a pressure of 106Pa for 120 minutes at a temperature ranging from 130 °C to 200 °C.

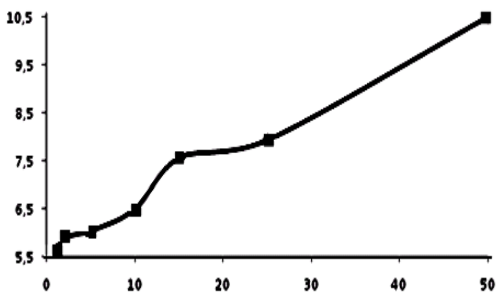


Fig. 4. Dissociation of Sphalerite as function of the initial mass of ZnS at 130°C

The figure (4) shows that the reaction of dissolution of zinc sulfide presents, within the temperature range studied, a single regime as function of the initial mass of solid:

$$[Zn\ 2+]\ t = [Zn\ 2+]\ eq + kmt \qquad E3$$

[Zn 2 +] eq is independent of the initial mass of zinc sulfide provided that it is sufficient.

After rapid dissolution (time t0), the reaction is strongly inhibited by the presence of hydrogen sulfide and the reaction rate is of order 1 in relation with the initial mass of zinc sulfide.

3.3 Influence of Oxygen Partial Pressure

The experiment consists in attacking 5 g of concentrate ZnS with 400ml of initial concentration of sphalerite (0.125 M et 0.5M) at a pressure of 105Pa to 4 106Pa for 120 minutes at a temperature ranging from 130 °C - 150°C and to 180 °C.

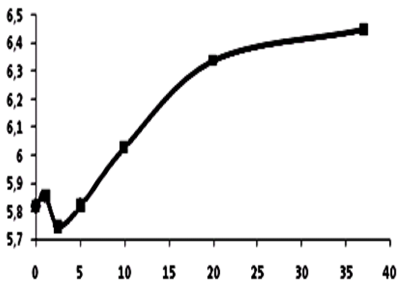


Fig. 5. Dissolution of sphalerite as function of partial pressure of oxygen at 130°c in an aqueous sulfuric medium at 0.125 M

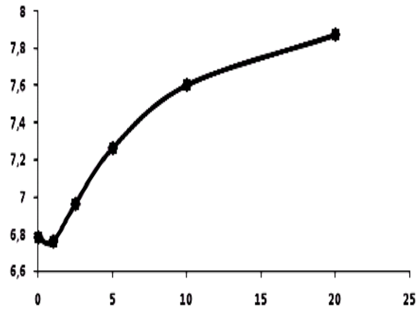


Fig. 6. Dissolution of sphalerite as function of partial pressure of oxygen at 150°c in an aqueous sulfuric medium at 0.125M

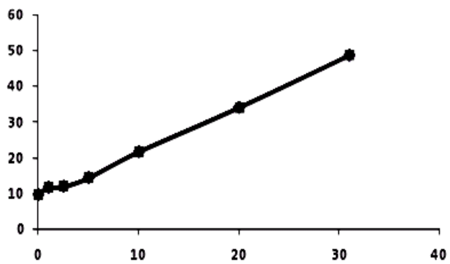


Fig. 7. Dissolution of sphalerite as function of partial pressure of oxygen at 180°c in an aqueous sulfuric medium at 0.125 M

Figures (5), (6) and (7) show that regardless to the temperature, the rate of dissolution of zinc sulfide is directly proportional to the oxygen pressure in the pressure range of (0 to 4.10^6 Pa) and the maximum temperature considered (180°C). At low concentrations and at temperatures below 150-160°C, dissolution increases slightly with oxygen pressure ($1.36\ 10^{-6}$ mol dissolved by additional 10^5 Pa). At 180°C, the oxidation reactions are strongly activated, dissolution increases significantly with the oxygen pressure (1.01 mole dissolved by additional 10^5 Pa).

In all cases: $[Zn^{2+}]_t - [Zn^{2+}]_{eq} = kP_{tO_2}$ E4

Time, temperature, concentration and the initial mass are fixed.

The dissolution rate obtained corresponds to the following reactions:

$$H_2S + 1/2\, O_2 \rightarrow H_2O + S^\circ \quad \text{R5}$$

$$H_2O_2 + O_2 \rightarrow H_2SO_4 \quad \text{R6}$$

$$S^\circ + H_2O + 3/2\, O_2 \rightarrow H_2SO_4 \quad \text{R7}$$

The curves obtained do not pass through the origin.

3.4 Influenced of the Initial Concentration of Sulfuric Acid

The experiment consists in attacking 5g of concentrate ZnS with 400ml of initial concentration of sphalerite (0 M et 1M) at a pressure of 106Pa for 120 minutes at a temperature ranging from 130°C to 180 ° C.

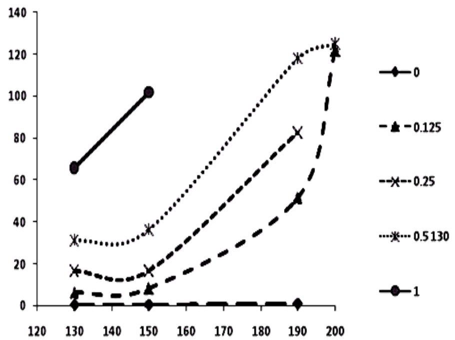


Fig. 8. Dissolution of sphalerite as function of initial concentrations for in sulfuric acid

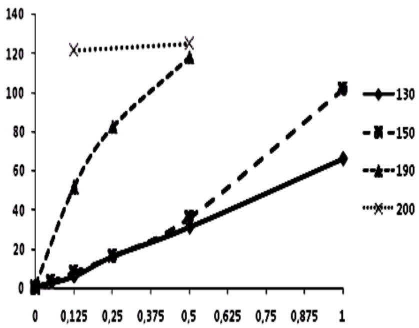


Fig. 9. Dissolution of sphalerite as function of temperature for different initial concentrations in sulfuric acid

The figures (8 and 9) show that for the temperature 130 ° C, the concentration of dissolved zinc is quad when the acidity passes from 0.125 to 0.5 M.

The analyses show that the solutions at 180 ° C are always accompanied by the production of sulfur which increases with acidity. At 130 ° C, the production of hydrogen sulfide increases as a function of acidity. At low temperatures, the dissolution is held following an electrolytic process, the rate of the reaction depends, for low acidities, on the acidity of the solution. The increase of H + ions in the solution helps the dissolution.

At high temperatures (> 160 C), the overall reaction is strongly activated; hydrogen sulphide is almost completely oxidized as sulphur for sulfuric acid concentrations above 0.5 M.

The rate of reaction, whatever the temperature below 200°C, and taking into account the dissolution without oxygen, can be written

$$[Zn^{2+}] = [Zn^{2+}]_{eq} + k [H_2SO_4] t \quad \text{E5}$$

$[Zn^{2+}]_{eq}$ is proportional to the initial concentration of sulfuric acid.

3.5 Influence of Hydrogen Sulfide in the Solution

The test is to attack an amount of ZnS (5g) by a sulfuric acid solution (0.125 M) at temperatures between 130,150 and 180 ° C in an autoclave at a partial pressure of oxygen of 106 Pa for a time 120 min.

Figures (10 and 11) show that whatever is the temperature, the dissolution of zinc sulfide increase greatly.

Activated carbon acts as an adsorbent for hydrogen sulfide and allows a significant dissolution, even at relatively low temperatures (90-130°C), shifting the equilibrium reaction in the direction of dissolution. On contrary, the presence of hydrogen sulfide at high pressure delays considerably the dissolution and almost stops it when the pressure is of 10 6 Pa. At 180°C, hydrogen sulfide reacts with sulfuric acid to give mainly sulfur. Sulfur, which coats the grains of sulfur, creates an additional resistance to diffusion which delays the dissolution.

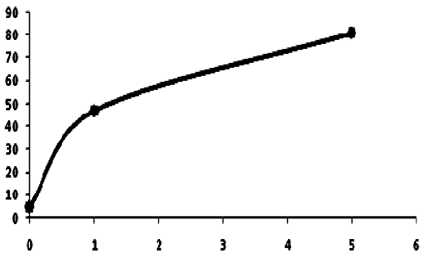


Fig. 10. Dissolution of sphalerite as function of mass activated carbon

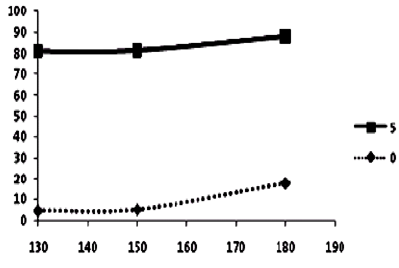


Fig. 11. Dissolution of only sphalerite in presence of activated carbon as function of temperature

Figures (12 and 13) show that whatever is the temperature, the reaction of dissolution is strongly delayed when the partial pressure of hydrogen sulphide increases. A pressure of 106 Pa of hydrogen sulphide completely stops completely the reaction for the considered temperatures.

For a temperature equal to 130°C, the final concentration of acid is considerably more important than the initial concentration.

For a temperature equal to 180°C, the final concentration of acid is also greater than the initial concentration excepted for the partial pressures of 106 Pa of hydrogen sulfide.

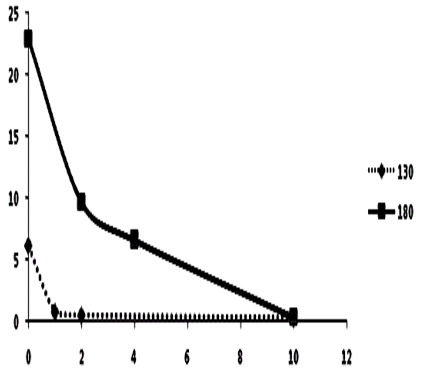


Fig. 12. Dissolution of sphalerite as function of partial pressure of hydrogen sulfide at 130°C and 180°C

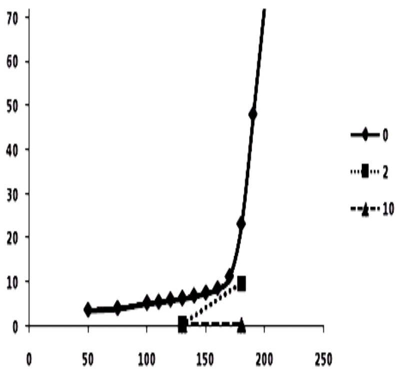


Fig. 13. Dissolution of sphalerite as function of temperature for different partial pressures of hydrogen sulfide

3.6 Effect of Temperature

The test is to attack an amount of ZnS (5g) by a 400ml sulfuric acid solution (0.125 M and 0.5 M) at temperatures of 50 to 180 ° C in an autoclave at a partial pressure of oxygen of 106 Pa for a time of 120 min.

The resulting graph (Figure14) is divided into two areas as for the previous concentrations of sulfuric acid.

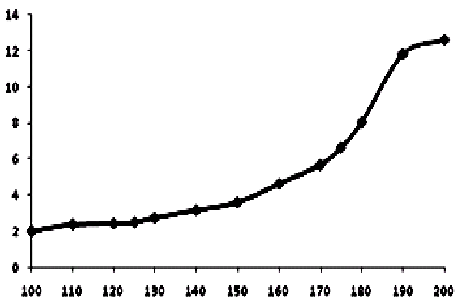


Fig. 14. Dissolution of sphalerite as function of temperature for different partial pressures of hydrogen sulfide

3.6.1 In the Area 120-170 ° C

The concentration of dissolved zinc by the reaction of dissolved oxygen under pressure can be expressed as a function of concentration [Zn 2 +] eq corresponding to the balanced relationship in the absence of oxygen:

$$[Zn^{+2}] = [Zn^{+2}]_{eq+}3.5.m.t.P_{O_2}[H_2SO_4].e^{\frac{-11.21}{T}} \quad \text{E6}$$

With P O2 (Pa), m (g) t (min) [H 2 SO 4] mole of sulfuric acid per liter;

The concentration [Zn2+] eq can be calculated from the relation giving the equilibrium constant of the reaction without oxygen. This concentration depends only on the temperature and the initial concentration of sulfuric acid.

This model corresponds to the slow reaction of oxidation of hydrogen sulfide. The dissolution rate is of order 1 with regard to the oxygen pressure in the initial mass of zinc sulfide and sulfuric acid concentration. The reaction takes place according to an electrochemical oxidation of hydrogen sulfide, but it is also limited by diffusion through a layer of liquid sulfur on the surface of the grains.

3.6.2 Above 160 ° C to 200 ° C

The kinetic model corresponding to the second step is of the same form as before:

$$[Zn^{+2}] = [Zn^{+2}]_{eq+}m.t.P_{O_2}[H_2SO_4].e^{\frac{-20.58}{T}+22.765} \quad \text{E7}$$

With P O2 (Pa), m (g) t (min) [H 2 SO 4] mole of sulfuric acid per liter;

The reaction is faster than temperatures at 160 ° C. Hydrogen sulfide and elemental sulfur are rapidly oxidized as sulfuric acid. The oxidation of hydrogen sulfide and the complex reactions of hydrogen sulfide in sulfuric acid constitute the limiting steps of the dissolution reaction.

4 Conclusion

This present study is related to a dissolution process of sphalerite in aqueous solution of sulfuric acid. Leaching experiments of this study are covering the temperature range of 25 ° C to 200 ° C,

In the presence of oxygen, two areas above 120 ° C can be distinguished.

Between 120 and 160 ° C the reaction takes place according to an electrochemical process. It is divided into two stages. The first corresponds to the dissolution of zinc sulfide which quickly reaches its equilibrium. And hydrogen sulfide is oxidized very slowly. The kinetic model thus proposed in this second step is

$$[Zn^{+2}] = [Zn^{+2}]_{eq+}3.5.m.t.P_{O_2}[H_2SO_4].e^{\frac{-11.21}{T}}$$

[Zn 2+] eq is the concentration of dissolved zinc in equilibrium, corresponding to the equilibrium reaction without oxygen. This concentration depends only on the initial concentration of sulfuric acid and temperature. The overall reaction is limited mainly by the reaction of oxidation of hydrogen sulfide and diffusion through a layer of liquid sulfur on the surface of the grains.

Between 160 and 200 ° C the reactions of oxidation are faster, and the kinetic model is

$$[Zn^{+2}] = [Zn^{+2}]_{eq+} m.t.P_{O_2}[H_2SO_4].e^{\frac{-20.58}{T}+22.765}$$

The overall reaction is limited by parallel chemical reactions oxidation of hydrogen sulfide, the reaction of hydrogen sulfide with sulfuric acid superimposed on the direct oxidation reactions. Elemental sulfur resulting from these reactions is itself oxidized as sulfuric acid.

References

[1] Forward, F.A., Veltman, H.: Direct leaching of zinc sulphide concentrate by Sherrit Gordon. J. Met. 11, 836–840 (1959)
[2] Parker, E.G.: Oxidative pressure leaching of zinc concentrate. CIM Bull. 74(5), 145–150 (1961)
[3] Demopoulos, G.P., Baldwin, S.: Stoichiometric and kinetic aspects on the pressure leaching of zinc concentrates. In: Mishra, B. (ed.) TMS Annual Meeting, San Diego, pp. 567–583 (1999)
[4] Bjorling, G.: Lixiviation of sulphidic minerals under oxygen pressure. Metallurgie 8, 781–784 (1954)
[5] Mizoguchi, T., Habashi, F.: The aqueous oxidation of complex sulfide concentrates in hydrochloric acid. Int. J. Miner. Process. 8, 177–193 (1981)
[6] Majima, H., Awakaura, Y., Misaki, N.: A kinetic study on nonoxidative dissolution of sphalerite in aqueous hydrochloric acid solution. Metall. Trans. 12B, 645–649 (1981)
[7] Bobeck, G.E., Su, H.: The kinetics of dissolution of sphalerite in ferric chloride solutions. Metall. Trans. 16B, 413–424 (1985)
[8] Palencia Perez, I., Dutrizac, J.E.: The effect of the iron content of sphalerite on its rate of dissolution in ferric sulphate and ferric chloride media. Hydrometallurgy 26, 211–232 (1991)
[9] Havlik, T., Kammel, R.: Procedure for selective copperrecovery from tetraedrite. Metall. 54, 26–29 (2000)
[10] Havlik, T., Popovičová, M., Ukašík, M.: Use of microwave energy for chalcopyrite leaching. Metall. 55, 332–335 (2001)
[11] Havlik, T., Miškufová, A., Tatarka, P.: Modern methods of oxidative chalcopyrite leaching. Acta Metall. Slovaca 4, 62–68 (2001)
[12] Weisener, C., Smart, R., Gerson, A.: Kinetics and mechanisms of the leaching of low Fe sphalerite. Geochim. Cosmochim. Acta 67, 823–830 (2003)
[13] Weisener, C., Smart, R., Gerson, A.: A comparison of the kinetics and mechanism of acid leaching of sphalerite containing low and high concentrations of iron. Int. J. Miner. Process. 74, 239–249 (2004)
[14] Abraitis, P., Pattrick, R., Kelsall, G., Vaughan, D.: Acid leaching and dissolution of major sulphide ore minerals: processes and galvanic effects in complex systems. Miner. Mag. 68, 343–351 (2004)

[15] Malmström, M.E., Collin, C.: Sphalerite weathering kinetics: effect of ph and particle size. In: Wanty, R.B., Seal II, R.R. (eds.) Proc. 11th Symp. Water–Rock Interaction, vol. 1, pp. 849–852 (2004)

[16] Balaz, P., Ebert, I.: Oxidative leaching of mechanically activated sphalerite. Hydrometallurgy 27(2), 141–150 (1991)

[17] Anthony, J.W., Bideaux, R.A., Bladh, K.W., Nichols, M.C.: Handbook of Mineralogy, Elements, Sulfides, Sulfosalts, vol. I. Mineral Data Publishing, Tuscon (1990)

[18] Kammel, R., Pawlek, F., Simon, M., Xi-Ming, L.: Oxidizing leaching of sphalerite under atmospheric pressure. Metall. 41, 158–161 (1987)

[19] Crundwell, F.K., Verbaan, B.: Kinetics and mechanisms of the non-oxidative dissolution of sphalerite (zinc sulphide). Hydrometallurgy 17, 369–384 (1987)

[20] Suni, J., Henein, H., Warren, G.W., Reddy, D.: Modelling the leaching kinetics of a sphalerite concentrate size distribution in ferric chloride solution. Hydrometallurgy 22, 25–38 (1989)

[21] Palencia Perez, I., Dutrizac, J.E.: The effect of the iron content of sphalerite on its rate of dissolution in ferric sulphate and ferric chloride media. Hydrometallurgy 26, 211–232 (1991)

[22] Cheng, C.Y., Clarkson, C.J., Manlapig, E.V.: The leaching of zinc sulphide concentrates in sulphate–chloride solutions with ferric ions. The AusIMM Proceedings 2, 57–62 (1994)

[23] Lochmann, J., Pedlik, M.: Kinetic anomalies of dissolution of sphalerite in ferric sulfate solution. Hydrometallurgy 37, 89–96 (1995)

Distributed and Time-Delay Analyse of Pneumatic Signals in Long Pipelines

Djordje N. Dihovicni

Visoka Tehnicka Skola, Bulevar Zorana Djindjica 152-a,
11000 Belgrade, Serbia
ddihovicni@gmail.com

Abstract. Transient in long pneumatic lines is considered from distributed parameter and as well time delay view of point, and are presented time responses. By using Simulink module of program Matlab, efficiency is simulated pneumatic system with time delay. It is proved that the control law achieves arbitrary finite spectrum assignment and is physically realizable.

Keywords: long pipelines, pneumatic systems, time-delay.

1 Introduction

The pneumatic systems have a lot of advantages if we compare them with the same hydraulic types; they are suitable for clean environments, and much safer. In accordance with project and space conditions, valves are positioned at relatively large distance from pneumatic cylinder. Considering real pneumatic systems, it is crucial to describe them with time delay, non-linearity's, with attempt of not creating only academic model. Despite of these problems, development of fast algorithms and using the numerical methods for solving partial different equations, as well as enhanced simulation and animation techniques become the necessity. Pneumatic cylinder systems significantly depend of behavior of pneumatic pipes [12], thus it is very important to analyse the characteristics of the pipes connected to a cylinder. Mathematical model of this system is described by partial different equations, and it is well known fact that it is distributed parameter system. These systems appear in various areas of engineering [11], and one of the special types is distributed parameter system with distributed control. Methods of analyse of control systems and simulation methods, which are used for observing dynamic behavior of linear dynamic systems with time delay, and distributed parameter systems, based on linear algebra, operation calculus, functional analyse, integral differential equations and linear matrix non-equations are applied. Signal transient in long pneumatic lines is analysed from time delay and parameter distribution view of point. The pressure or flow changing phenomena in pneumatic control systems is very complex, and has a significant effect on the stability, response and construction issues of the system and its components. Up to now, the published papers have not been shown complete analyse of this phenomena and as well have not presented the adequate control system as shown in the reference [3].

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 250–256, 2013.

2 Mathematical Model of Pneumatic Transmission Line

The Figure 1 shows a schematic diagram of pneumatic transmission line. Detailed procedure of creating this mathematical model is described in reference [4].

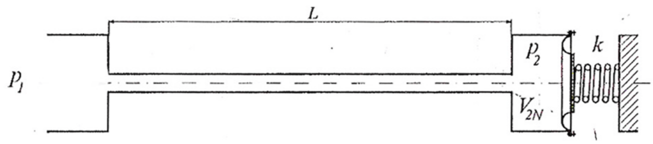


Fig. 1. Schematic diagram of pneumatic system

It is well known fact that pneumatic signal takes a finite time to traverse the length of the line [8]. Second problem arises with the standing waves. They can occur in the pneumatic line, and the phase lag for harmonic inputs exceeds 180 degrees in moment when the frequency passes the natural frequency of the column of gas. Taking into consideration that a pressure pulse at one end of the line and does not cause the pressure at the other end to begin to rise constantly, and that the assumption of a constant pressure throughout the volume leads to a first-order differential equation without delay, the pressure variations along the line must be considered to even approximate the effects of these phenomena according to reference [1].

The figure 1 present's pneumatic line of length L, input pressure is denoted with p_1 and presents the pressure signal, and p_2 is the output pressure. This kind of problem is further investigated in reference [1], Browns arrives at one- dimensional equation by simplifying the more general equations, with included radial and axial heat transfer and radial velocities.

Combining the equations in which figure density, pressure, and velocity, the summation of forces on an annulus of width *dr* to the mass of gas acceleration yields to equation (1), and in case of laminar flow to the equation (2):

$$\sum F_y = -2\pi r dr dP + 2\pi dy d(r\tau) = 2\pi r dr dy \frac{P}{gRT}\frac{\partial u}{\partial t} \quad . \tag{1}$$

$$\frac{\partial P}{\partial y} = -\frac{32 \quad \eta}{D^2} \bullet v - \frac{P}{g \bullet R \bullet T} \bullet \frac{\partial v}{\partial t} \tag{2}$$

The flow equation and fully developed laminar flow lead to transfer functions expressed in hyperbolic sine and cosine of a function of the operator *s*, [10].

3 Analysing the Transmission of Pneumatic Signal

By increasing the length of the transmission pipeline, it is increasing dumping ratio ζ, but decreasing β. In case that length is increasing between 45 and 60 m, has occured critical damping ratio, and then it is used aproximation of the first order. Test results has shown that aproximation of the first order is good and they take into account transport time delay.

These results could be obtained in mathematical metter, by usung adequate aproximations, given with:

$$\cosh\alpha = e^{s/\beta} \bullet \left(1 + \frac{1}{2} \bullet E \bullet \frac{s}{\beta}\right) . \tag{3}$$

$$\alpha \bullet \sinh\alpha = e^{s/\beta} \bullet \left(E_2 \bullet \frac{s}{\beta}\right) . \tag{4}$$

$$\frac{1}{\alpha}\sinh\alpha = e^{s/\beta} \bullet \left(1 + \frac{1}{6} E_2 \bullet \frac{s}{\beta}\right) . \tag{5}$$

The results of above equations show very good matching with results of experiment. Diferential equation of behaving in complex domain is given with:

$$\frac{p_2}{p_1} = \frac{e^{\,s/\beta}}{1 + \frac{1}{2} + \frac{\overline{V}}{AL} \bullet E_2 \bullet \frac{s}{\beta}} . \tag{6}$$

System responses on step input, when we are taking into account distributed parameter system and time-delay view of perspective are shown down below.

The schematic diagram is shown in figure 1.

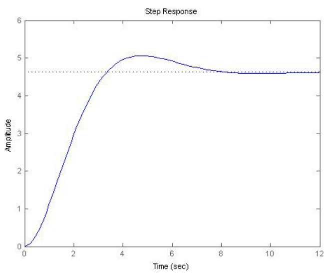


Fig. 2. Schematic diagram of pneumatic system response considered from distributed parameter view of point

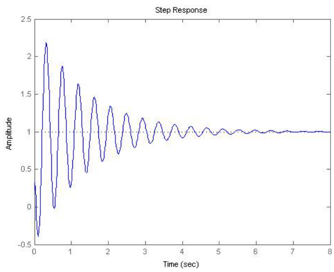


Fig. 3. Schematic diagram of pneumatic system response considered from time-delay view of point

4 Characteristic Indicators of Step Responses

Firstly it is required from technical systems certain quality of dynamic behavior, which manifests through the indicators of the step or frequent responses [5].

The main influence on appearance of transfer process has the poles of system transfer function which are the closest to imaginary axe of a complex plane. These poles are called the dominant poles, and for systems of higher order in case when the real part is six times higher than the imaginary part, the influence on dynamic performance can be neglected in accordance with the reference [7]. Dynamic characteristics of the system shown through certain indicators in time domain are investigated on step responses of transmission signal in long pneumatic pipelines, from time-delay and distributed parameter perspective.

Dynamic pair of poles when the transmission signal in long pneumatic pipelines is observing from distributed parameter system perspective is:

$$s = \quad 6.8 \pm j\,11.6\,. \tag{7}$$

Complex number *s* might be presented as:

$$s = \sigma + jw, \qquad j^2 = 1\,. \tag{8}$$

In above equation σ represents real part, and *w* imaginary part of the complex number s, and in polar coordinates it could be written as:

$$|s| = \sqrt{Re^2(s) + Im^2(s)} = \sqrt{\sigma^2 + w^2} = w_n\,. \tag{9}$$

The un-damped frequency of the system is shown as:

$$w_n = \sqrt{\sigma^2 + w^2} = \sqrt{(6.8)^2 + (11.6)^2} = 13.45\,s^{\ 1}\,. \tag{10}$$

The frequency of the system is given with:

$$\varsigma = \frac{\sigma}{w_n} = \frac{6.8}{13.45} = 0.5\,. \tag{11}$$

A dominant time constant is calculated as T_d:

$$T_d = \frac{1}{|\sigma|} = \frac{1}{|6.8|} = 0.147\,s\ . \tag{12}$$

The settling time is:

$$T_s = \frac{4}{\varsigma \bullet w_n} = \frac{4}{0.5 \bullet 13.45} = 0.59\,s\ . \tag{13}$$

The appearance of probing is given with:

$$t_{\Pi} = \frac{\pi}{w_n\sqrt{1-\varsigma^2}} = \frac{\pi}{13.45 \bullet \sqrt{1-0.5^2}} = 0.311\ s\ . \tag{14}$$

A period of oscillation is shown as:

$$T = 2 \quad t_{\Pi} = 0.622\,s\ . \tag{15}$$

The number of oscillation is calculated:

$$N = \frac{T_s}{\tau} = \frac{2 \bullet \sqrt{1-\varsigma^2}}{\varsigma \bullet \pi} = \frac{2 \bullet \sqrt{1-0.5^2}}{0.5 \bullet \pi} = 0.95\ . \tag{16}$$

When the transmission of the pneumatic signal is concerned from time-delay view of point, the dominant pole is:

$$s = \quad 14.25 \pm j\ 8.24\ . \tag{17}$$

The un-damped frequency of the system taking into account the equation (10) is given with:

$$w_n = \sqrt{\sigma^2 + w^2} = \sqrt{(14.25)^2 + (8.24)^2} = 16.46\,s^{1}. \tag{18}$$

The system damping is shown as:

$$\varsigma = \frac{\sigma}{w_n} = \frac{14.26}{16.46} = 0.866\ . \tag{19}$$

The dominant time constant is calculated as:

$$T_d = \frac{1}{|\sigma|} = \frac{1}{|\quad 14.26|} = 0.07\,s\ . \tag{20}$$

The settling time is:

$$T_s = \frac{4}{\varsigma \bullet w_n} = \frac{4}{0.866 \bullet 16.46} = 0.28\,s\,. \tag{21}$$

The appearance of probing is shown as:

$$t_{\Pi} = \frac{\pi}{w_n\sqrt{1-\varsigma^2}} = \frac{\pi}{1646\bullet\sqrt{1-0.866^2}} = 0.521\,s\,. \tag{22}$$

The period of oscillation is shown as:

$$T = 2 \quad t_{\Pi} = 1.024\,s\,. \tag{23}$$

The number of oscillation is calculated:

$$N = \frac{T_s}{\tau} = \frac{2 \quad \sqrt{1-\varsigma^2}}{\varsigma \bullet \pi} = \frac{2 \quad \sqrt{1-0.5^2}}{0.5 \bullet \pi} = 0.269\,. \tag{24}$$

5 Conclusion

It is obvious that phenomena of transient of the pressure and the flow in pneumatic control systems, especially with long pneumatic lines have character of time delay and parameter distribution, and that further analyse should be implemented. This paper describes the simulation of pneumatic system with long pipelines, observing the problem from time delay and parameter distribution perspective. Mathematical models of these systems are described by partial different equations, but apart from distributed phenomena we can’t neglect system time delay. The step responses from distributed and time delay view of point are compared, and it is determined that speed of response is better with time-delay systems.

References

1. Andersen, B.W.: The Analysis and Design of Pneumatic Systems. John Wiley & Sons, Inc. (1969)
2. Debeljkovic, L.D., Milojkovic, B.R.: Design of Linear Systems Faculty of Mechanical Engineering, Belgrade (1987)
3. Dihovicni, D., Nedic, N.: Program Support for Distributed Control Systems on Finite Space Interval. International Journal of Intelligent Technology 2(1), 49–54 (2007) ISSN 1305-6417
4. Dihovicni, D., Medenica, M.: Mathematical Modelling and Simulation of Pneumatic Systems. In: Advances in Computer Science and Engineering, ch. 9, India, pp. 161–186 (2011) ISBN 978-953-307-173-2

5. Dihovicni, D., Nedic, N.: Simulation, Animation and Program Support for a High Performance Pneumatic Force Actuator System. Mathematical and Computer Modelling 48, 761–768 (2008)
6. Dihovicni, D.: Constructing knowledge database in decision making. In: Communications, Zelina University, Slovakia, vol. 3, pp. 74–77 (2011)
7. Dihovicni, D.: Constructing knowledge database in decision making. In: Communications, Zelina University, Slovakia, vol. 3, pp. 74–77 (2011)
8. Ikegami, F., Kagawa, T.: Pneumatic cylinder with meter-out velocity control. In: 9th Symposium on Fluid Control Matsuayama, Japan (1994)
9. Richer, E., Hurmuzlu, Y.: A High Performance Pneumatic Force Actuator System. ASME Journal of Dynamic Systems Measurement and Control, 416–425 (September 2000)
10. Manual of Hydraulics and Pneumaticsm, The Japan Hydraulics and Pneumatics Society (1989)
11. Thayse, A., Gribont, P.: Approche Logique de l'Inteligence Artificielle l De la Logique. Bordas (1997)
12. Waton, J.: Fluid Power Systems. Prentice Hall (1989)

Comparison between Two Methods for Characterization of a Patch Antenna Array: Experimental and by Simulation

Abdelhakim Elouadih[1], Ahmed Oulad-Said[1], and Moha Mrabet Hassani[2]

[1] Royal Air Academy (ERA), Telecoms and Hyperfrequencies Laboratory, Marrakesh, Morocco
[2] University of Sciences Semlalia (FSSM), Instrumentation and Electronics Laboratory, Marrakesh, Morocco
elouadih@gmail.com,
a_ouladsaid@hotmail.com,
hassani@ucam.ac.ma

Abstract. This paper describes the characterization of a symmetrical probe feed patch antenna array. The characterization consists on finding the most important parameters of the antenna. It is done by two methods. The first one is a simulation by HFSS simulator release 10. The second method is based on experiences using network analyzer and a hyper frequency transmission bench. A comparison of both methods results shows some light differences that can be discussed.

Keywords: patch antenna array, HFSS, characterization.

1 Introduction

Wireless communications have progressed very rapidly in recent years, and many mobile units are becoming smaller and smaller. To meet the miniaturization requirement, the antennas employed in mobile terminals must have their dimensions reduced accordingly. Planar antennas, such as patch (microstrip) and printed antennas have the attractive features of low profile, small size, and conformability to mounting hosts and are very promising candidates for satisfying this design consideration. For this reason, compact, broadband and wideband design technique for planar antennas have been attracted much attention from antenna researchers. Planar antennas are also very attractive for applications in communication devices for wireless local area network, wide mobile telecommunications, aeronautics and embedded systems [1].

This paper processes on an antenna array used in Air Academy telecommunications Laboratory. The problem is that antenna was furnished without any document or workshop sheet. It was then mandatory to characterize it in order to use it in Laboratory student's practical work and projects. It is also an opportunity to test some HFSS

[1] Corresponding author.

M.Á. Fernández-Izquierdo, M.J. Muñoz-Torres, R. León (Eds.): MS 2013, LNBIP 145, pp. 257–266, 2013.

performance and efficiency. By this characterization, the author needs to determine its most important parameters such as resonant frequency, bandwidth, VSWR, S_{11} parameter, gain and radiation pattern. One choice was to use the HFSS simulator that gave some characteristics. Another method was to use a network analyzer to determine the S_{11} parameter and to use an Hyper frequency transmitter receiver bench using other kinds of antennas to determine the diagram pattern. The first step is to determine theoretically the resonant frequency by using a transmission line model. Then, the simulation and experiences results are exposed before making a comparison of the obtained results.

2 Presentation of the Studied Antenna Array

The studied antenna consists of a double-sided printed circuit board (PCB). One side forms the antenna pattern and the other a ground plane. The pattern is consisting of 16 rectangular patches disposed by 4*4 on the x and y axes. Each patch has a width W and length L =W equal to 8.9 mm. The patches are printed on a PCB that has a substrate with a relative permittivity equal to 2.62. The other side of the PCB in made by a perfect conductor presenting a finite ground plane with dimensions of 95 mm and 120 mm. the high of antenna is 1.6 mm and the distance between patches is 10.4 mm. All of patches are symmetrically supplied by line and a probe feed in the centre as shown in figure 1.

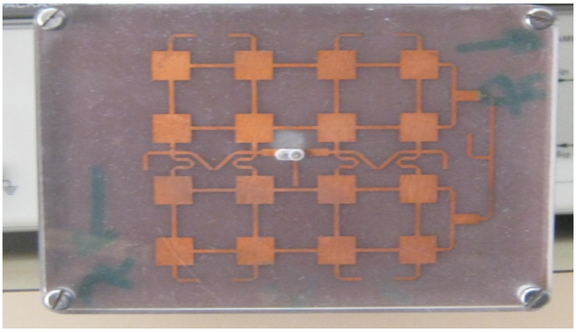


Fig. 1. The front side of the studied patch

3 Estimating of the Theoretical Resonant Frequency

It is very important for simulation by HFSS to estimate the resonant frequency that help the simulator to make a refinement mesh in a band around the resonant frequency and then give more precise values. For one element of the array as shown in figure 2 [1]:

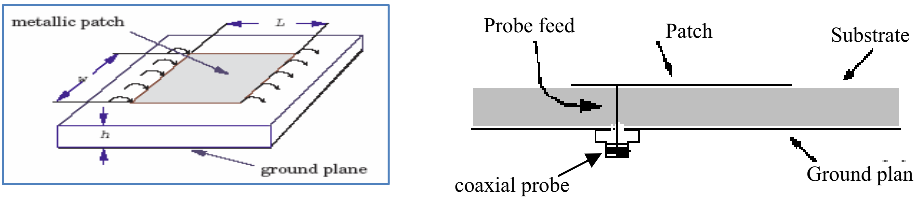


Fig. 2. One of the 16 elements of patch

The element has length L equal to width W=8.96 mm, a height of 1.6 mm and a permittivity of 2.62. We must firstly determine the effective permittivity by the equation (1):

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2} \cdot \frac{1}{\sqrt{1 + \frac{12H}{W}}} \quad (1)$$

The ε_{eff} obtained then is **2.266**.

After, the length extension ΔL is computed by equation 2:

$$\Delta L = 0.412h \frac{(\varepsilon_{eff} + 0.3)\left(\frac{w}{h} + 0.264\right)}{(\varepsilon_{eff} - 0.258)\left(\frac{w}{h} + 0.8\right)} \quad (2)$$

ΔL = 0.77 mm, After an effective length L_{eff} is equal to L+ 2ΔL,

L_{eff} = 10.44 mm. The equation 3 gives the relation between L_{eff} and the resonant frequency:

$$L_{eff} = \frac{c}{2f_r\sqrt{\varepsilon_{eff}}} \quad (3)$$

With a light velocity equal to c= 310^8 m/s, the obtained **f_r is 9.55GHz**.

4 Simulation by HFSS of the Antenna Array

4.1 The Choice of the Simulator HFSS

HFSS (High Frequency Simulator Structure) is a high performance full wave electromagnetic (EM) field simulator for arbitrary 3D volumetric passive device modeling that takes advantage of the familiar Microsoft Windows graphical user interface. It integrates simulation, visualization, solid modeling, and automation in an easy to learn environment where solutions to 3D EM problems are quickly and accurate obtained. Ansoft HFSS employs the Finite Element Method (FEM), adaptive meshing, and brilliant graphics to give unparalleled performance and insight to all of 3D EM problems. Ansoft HFSS can be used to calculate parameters such as S-Parameters, Resonant Frequency and Fields. Typical uses include Package Modeling, PCB Board Modeling , Mobile Communications (Patches, Dipoles, Horns, Conformal Cell Phone Antennas), Specific Absorption Rate (SAR), Infinite Arrays, Radar Section (RCS), Frequency Selective Surface (FSS) and filters sush Cavity Filters, Microstrip, Dielectric. HFSS is an interactive simulation system whose basic mesh element is a tetrahedron. T. In industry, Ansoft HFSS is an the tool of choice for High productivity research, development, and virtual prototyping [2].

The HFSS is then used in the simulation; the results are then exposed in the following.

4.2 Antenna Design

The studied antenna was firstly designed in the HFSS as shown in figure 3.

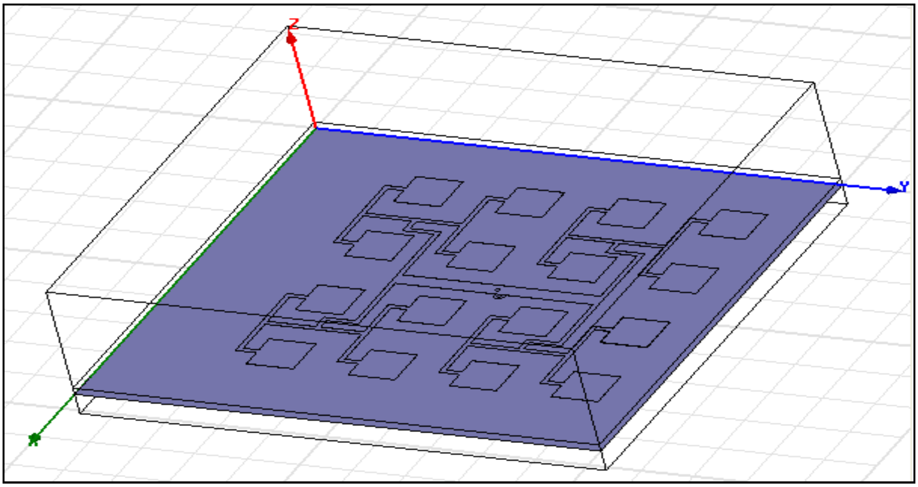


Fig. 3. The design of the studied array antenna in HFSS

The design in HFSS consists firstly on tracing the different components of the antenna (16 patches, substrate, ground plan), and secondly on designing the symmetrical excitation and boundaries. It finishes by configuring the setup solution with different sweeps.

4.3 The Reflection Coefficient

The reflection coefficient is the ratio of (Za-Zo) to (Za+Zo) where Za is the antenna impedance and Zo is the feeding line characteristic impedance. It measures the reflected part of an incident wave. S_{11} is 20 log (reflection coefficient) .The simulation allows the estimation of S_{11} parameter function of the frequency. The obtained result is shown in figure 4.

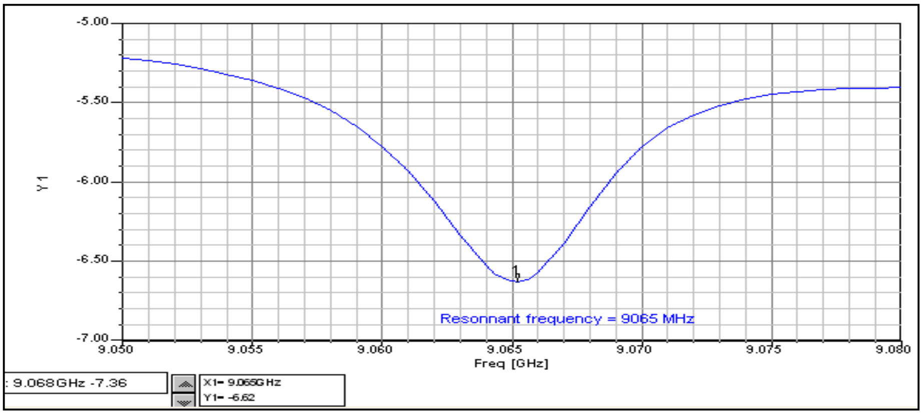


Fig. 4. The parameter S_{11} depending on frequency

A peak is seen around the frequency of 9.065 GHz. We can run the simulation by refining the sweep interval for more precision and as shown in the table of figure 5

	Freq [GHz]	dB(S(WavePort1,WavePort1)) [db] Setup1 : Sweep2
1	9.064000	-6.524604
2	9.064200	-6.554398
3	9.064400	-6.579702
4	9.064600	-6.599882
5	9.064800	-6.614385
6	9.065000	-6.622767
7	9.065200	-6.624727
8	9.065400	-6.620127
9	9.065600	-6.609005
10	9.065800	-6.591575
11	9.066000	-6.568220

Fig. 5. The results table of frequencies with the S_{11} values

The table gives a resonant frequency of 9.0652 GHz with a S_{11} = -6.63 dB.

4.4 The VSWR and Bandwidth

The VSWR (Voltage Standing Wave Ratio) is the ratio of maximum voltage to minimum voltage in a line (VSWR = Vmax/Vmin). The figure 6 indicates a VSWR minimal of 2.75 and a 1:1.2 VSWR bandwidth of 30 MHz.

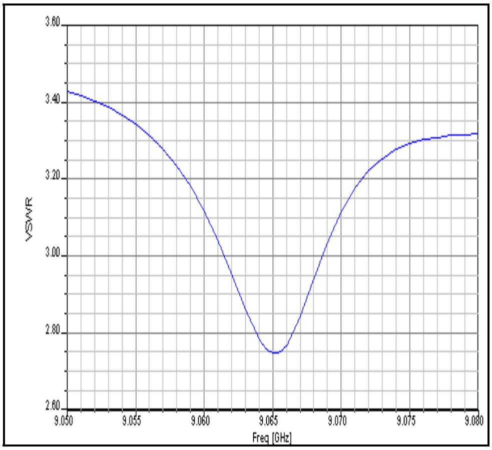


Fig. 6. The VSWR (Frequency)

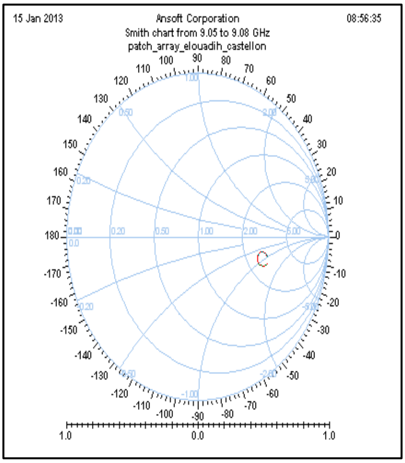


Fig. 7. The Smith Chart

4.5 The Smith Chart

The impedances depending on frequencies are shown on figure 7 via the Smith Chart.

4.6 The Impedance in the Feeding Point of the Array Antenna

Freq	Port Zo
9.057 [GHz]	(54.953, 0)
9.058 [GHz]	(54.953, 0)
9.059 [GHz]	(54.953, 0)
9.06 [GHz]	(54.953, 0)
9.061 [GHz]	(54.953, 0)
9.062 [GHz]	(54.953, 0)
9.063 [GHz]	(54.953, 0)
9.064 [GHz]	(54.953, 0)
9.065 [GHz]	(54.953, 0)
9.066 [GHz]	(54.953, 0)
9.067 [GHz]	(54.953, 0)

Fig. 8. Impedance of the feeding point

Antenna Parameters:

Quantity	Value	Units
Max U	0.40694	W/sr
Peak Directivity	7.0048	
Peak Gain	7.1514	
Peak Realized Gain	5.1139	
Radiated Power	0.73005	W
Accepted Power	0.71509	W
Incident Power	1	W
Radiation Efficiency	1.0209	
Decay Factor	0	

Maximum Field Data:

rE Field	Value	Units	At Phi	At Theta
Total	17.517	V	360deg	0deg
X	2.454	V	220deg	30deg
Y	17.484	V	360deg	0deg
Z	7.9503	V	90deg	60deg
Phi	17.484	V	360deg	0deg
Theta	17.484	V	270deg	0deg
LHCP	12.215	V	130deg	10deg
RHCP	12.94	V	350deg	0deg
Ludwig3/X dominant	3.8872	V	190deg	130deg
Ludwig3/Y dominant	17.484	V	80deg	0deg

Fig. 9. Antenna parameters

Around the resonant frequency, the impedance in the feeding point is equal to 55 Ω. It Is not so far to 50Ω. The antenna is appropriately adapted.

4.7 The Diagram Pattern and the Antenna Parameters

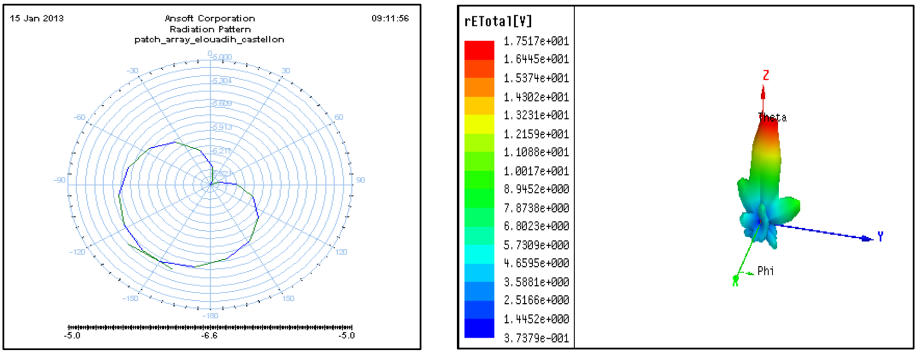


Fig. 10. 2D and 3D Diagram pattern

The simulations results shown in figure 10 give the antenna parameters and the fields maximas. The obtained peak gain G is 7.15 dBi and the radiation efficiency is 1.03. Also, it‘s confirmed that the E-plan of the antenna is the (yz) plan. Consequently, the figure 10 shows the radiation pattern of the antenna.

After exposing all simulation results, the second part presents the experimental measures done in Laboratory.

5 Experimental Methods

5.1 Using a Network Analyzer

The 8719C HP network analyzer was used to measure the S_{11} parameter and the resonant frequency. The network analyzer works in a band from 50 MHz to 13.5 GHz as shown in figure 11. [3].

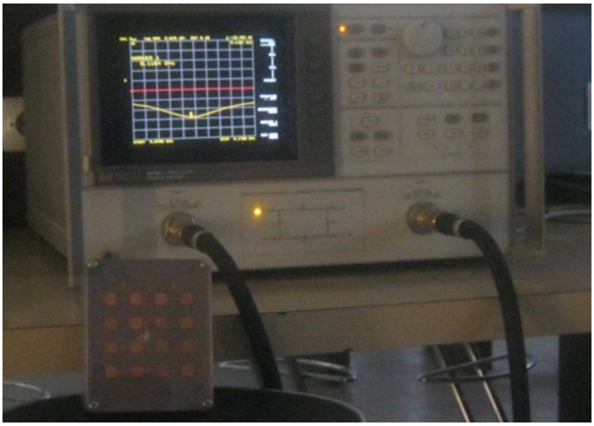
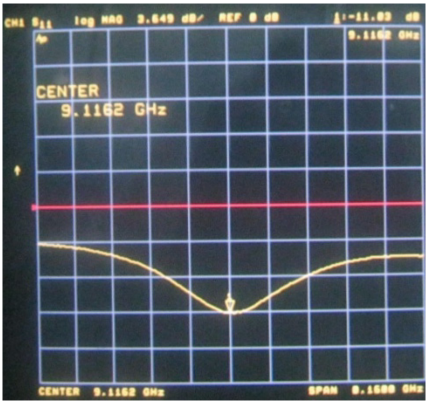


Fig. 11. The measure using the network analyzer

Before doing the measures, the analyzer was calibrated using a short circuit, open circuit and broadband load as shown in figures 12

Fig. 12. The used open circuit , short circuit and load

The network analyzer gives the results as shown in figure 13.



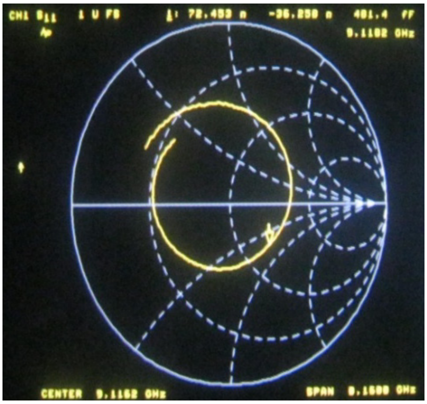


Fig. 13. S11 = f (frequency)

Fig. 14. The Smith Chart

The peak corresponds to the resonant frequency f_r = 9.1162 GHz with a S_{11} = -11 dB.

Also, the measured VSWR minimal was 2 and the 1:1.2 VSWR bandwidth is equal to 60 MHz. The Smith chart for a sweep from 8 GHz to 10 GHz is obtained in figure 14.

Fig. 15. The Experience set

Fig. 16. The Resonant frequency as Indicated by the frequency meter

5.2 Using an Hyper Frequency Transmission Bench

As shown in figure 15, The hyper frequency transmitter [4] is compound of a Gunn oscillator (PM 7015X) that delivers 5 mW in the range 8.5-9.6 GHz supplied by a power supply (PM 7815), a first variable attenuator (PM7110x), a PIN modulator (PM 7026X/01), a frequency meter (PM7072X), a second variable attenuator (HP X 375A), a waveguide (PM 7366X), a rotary joint (PM7888X) and the studied array antenna in the extremity of the transmitter.

The receiver is compound of a horn antenna (PM7320X/01), a detector (PM7195X) linked to SWR meter (PM7833).

The principle is to detect by receiver of a maximum power for a frequency range and in different directions. The maxima is obtained for a frequency indicated by a frequency meter equal to 9105 MHz as shown in figure 16.

6 Comparison of Both Methods

The table 1 summarizes the most important results obtained by simulation method and by measures.

Table 1. Results synthesis

Parameter	Experimental method	Simulation method
Resonant Frequency	**9116 MHz by analyzer** **9105 MHz by experience**	**9065 MHz**
S_{11} Parameter	**-11 dB**	**-7 dB**
1:1.2 Bandwidth	**60 MHz**	**30 MHz**

The comparison shows that results are very close especially for resonant frequency.
The gap can be explained by following causes:

- ✓ The matching between the patch and the feeding point is ideal in HFSS but in reality the matching is not perfect, it adds impedance and a frequency gap.
- ✓ The theoretical frequency computed by using the transmission line model (9.55 GHz) was used in HFSS in the setup solution to make a mesh refinement around it because HFSS uses finite element method. The difference between delivered frequency and theoretical frequency adds some imprecision.
- ✓ Also, a non perfect calibration of the network analyzer generates some errors. In fact, connectors by the repetitive use add a small S_{11} difference.
- ✓ The frequency meter in the transmission string behaves as a resonant cavity (different to transmission line model).It adds some errors.

In conclusion, I can consider the results given by HFSS so important and confident. The HFSS simulator can be considered an interesting way to simulate designed modern antennas.

References

1. Milligan, T.A.: Modern antenna design, ch. 6. J. Wiley and sons, New Jersey (2005)
2. Ansoft corporation: HFSS 10.0 user's guide: Ansoft. rev 1.06, Pittsburg (2005)
3. HP: Service Manual, HP company, Santa Rosa, USA (September 1993)
4. SIVERSIMA: Technical notices, AB Grafiska Gruppen, Stockholm (1989)

Author Index