ASMDA 2015

Proceedings

16th Applied Stochastic Models and Data Analysis International Conference

With 4th Demographics 2015 Workshop

Editor

Christos H Skiadas

30 June – 4 July 2015

University of Piraeus, Greece

Imprint

Proceedings of the 16th Applied Stochastic Models and Data Analysis International Conference (ASMDA2015) with Demographics 2015 Workshop, 30 June – 4 July 2015, University of Piraeus, Greece
Published by: ISAST: International Society for the Advancement of Science and Technology.
Editor: Christos H Skiadas

e-ISBN: 978-618-5180-05-8

Introduction

XVth Applied Stochastic Models and Data Analysis (ASMDA2015) International Conference
with
4th Demographics 2015 Workshop

30 June – 4 July 2015, University of Piraeus, Greece

Since 1981, ASMDA aims to serve as the interface between Stochastic Modeling and Data Analysis and their real life applications particularly in Business, Finance and Insurance, Management, Production and Reliability, Biology and Medicine.

Our main objective is to welcome papers, both theoretical or practical, presenting new results having potential for solving real-life problems. Another important objective is to present new methods for solving these problems by analyzing the relevant data. Also, the use of recent advances in different fields will be promoted such as for example, new optimization and statistical methods, data warehouse, data mining and knowledge systems, computing-aided decision supports and neural computing.

The role played by ASMDA as interface between theory and practice means that the conferences are of great interest for both the academic and business world. The high standard of the meetings are guaranteed by strong international scientific committees.

The preceding international ASMDA Symposia were organized in Brussels (1981, 1983, 1985), Belgium, in Nancy (1988), France, in Granada (1991), Spain, in Chania (1993), Greece, in Dublin (1995), Ireland, in Anacapri (1997), Italy, in Lisboa (1999), Portugal, in Compiègne (2001), France, in Brest (2005), France, in Chania Crete (2007), Greece, in Vilnius (2009), Lithuania, Rome (2011), Italy and in Mataró (Barcelona 2013), Spain.

In general, the proceeding volumes were published by international scientific publishers or local publishers.

The ASMDA 2015 Conference focuses on new trends in theory and applications of Applied Stochastic Models and Data Analysis.
A special Demographics2015 Workshop is organized under the umbrella ASMDA International. The aim is to gather people interested in improving demography and the related fields of analysis and research including life and physical sciences as well as medical and technical information. We strongly support interdisciplinary studies and the improvement of the analytic tools and research methods
We thank all the contributors to the success of this conference, the ISAST Committee, the Secretary Mary Karadima, and especially the authors of this Proceedings Book.
Special thanks to Yiannis Dimotikalis, ISAST Vice-Chair, for publishing this Volume and Sotiris Bersimis for his valuable work as local ASMDA chair these two years of hard work and dedication to the success of the conference.

October, 2015

Raimondo Manca, Rome, Italy
Sally McClean, Coleraine, UK
Christos H. Skiadas, Chania, Crete, Greece
Mariano Valderrama, Granada, Spain

ASMDA2015 Co-Chairs

ASMDA-2015 Committees

Chairman

Jacques Janssen, Honorary Professor of Université Libre de Bruxelles

Members

Alejandro Aguirre, El Colegio de México, México
Alexander Andronov, Transport and Telecom. Institute, Riga, Latvia
Vladmir Anisimov, Research Statistics Unit. GlaxoSmithKline
Soren Asmussen, University of Aarhus, Denmark
Dimitrios Antzoulakos, University of Piraeus, Greece
Robert G. Aykroyd, University of Leeds, UK
Narayanaswamy Balakrishnan, McMaster University, Canada
Helena Bacelar-Nicolau, University of Lisbon, Portugal
Paolo Baldi, University of Rome "Tor Vergata", Italy
Vlad Stefan Barbu, University of Rouen, France
Henry W. Block, Department of Statistics, University of Pittsburgh, USA
James R. Bozeman, Math. & Comp. Sci. Lyndon State College, Lyndonville, VT, USA
Mark Brown, The City College, CUNY, USA
Ekaterina Bulinskaya, Moscow State University, Russia
Jorge Caiado, Centre Appl. Math., Econ., Techn. Univ. of Lisbon, Portugal
Enrico Canuto, Dipart. di Automatica e Informatica, Politec. di Torino, Italy
Manolis Christodoulou, Technical University of Crete, Chania, Greece
Erhan Çinlar, Princeton University, USA
Maria Mercè Claramunt, Barcelona University, Spain
Marco Dall'Aglio, LUISS Rome, Italy
Guglielmo D'Amico, University of Chieti and Pescara, Italy
Pierre Devolder, Université Catholique de Louvaine, Belgium
Giuseppe Di Biase, University of Chieti and Pescara, Italy
Yiannis Dimotikalis, Technological Educational Institute of Crete, Greece
Dimitris Emiris, University of Piraeus, Greece
N. Farmakis, Aristotle University of Thessaloniki, Greece
Ilia Frenkel, Sami Shamoon College of Engineering, Israel
Leonid Gavrilov, Center on Aging, NORC at the University of Chicago, USA
Natalia Gavrilova, Center on Aging, NORC at the University of Chicago, USA
A. Giovanis, Technological Educational Institute of Athens, Greece
Valerie Girardin, Université de Caen Basse Normandie, France
Joseph Glaz, University of Connecticut, USA
Maria Ivette Gomes, Lisbon University and CEAUL, Lisboa, Portugal
Gerard Govaert, Universite de Technologie de Compiegne, France
Alain Guenoche, University of Marseille, France
Y. Guermeur, LORIA-CNRS, France
Montserrat Guillen University of Barcelona, Spain
Steven Haberman, Cass Business School, City University, London, UK
Diem Ho, IBM Company
Emilia Di Lorenzo, University of Naples, Italy
Aglaia Kalamatianou, Panteion Univ. of Political Sciences, Athens, Greece
Udo Kamps, Inst. fur Stat. und Wirtschaftsmath., RWTH Aachen, Germany

Organizing Committee

Chairmen

Local Committee

Plenary Speakers

Subha Chakraborti
Department of Information Systems, Statistics and Management Science
University of Alabama in Tuscaloosa, USA

Ioannis Kontoyiannis
Department of Informatics
Athens University of Economics and Business, Greece

Markos Koutras
Department of Statistics and Insurance Science
University of Piraeus, Greece

Nikolaos Limnios
Université de Technologie de Compiègne, France
Laboratoire de Mathématiques Appliquées

Jean-Marie Robine
Université Montpellier 2
Place Eugène Bataillon, Montpellier, France

Dmitrii Silvestrov
Professor of Mathematical Statistics
Department of Mathematics
Stockholm University

Fugee Tsung
Department of Industrial Engineering and Logistics Management
Hong Kong University of Science and Technology, China

Contents Page

Infant Mortality reduction in Mexico: Is there a Matthew Effect?

Alejandro Aguirre[1], Fortino Vela-Peón[2]

[1] El Colegio de Mexico , Mexico
(E-mail: aguirre@colmex.mx)
[2] Universidad Autonoma Metropolitana Xochimilco
(E-mail: fvela@correo.xoc.uam.mx)

Abstract. Applying Brass Method we estimated infant mortality rates (IMR) for all the states of Mexico using information on children ever born and children surviving (CEB/CS) from two censuses. With these results we analyzed regional inequalities in the reductions of infant mortality, to test the hypothesis of the "Matthew Effect". Originally labeled by Merton the "Matthew Effect" refers to the biblical verse: "Unto every one that hath shall be given, and he shall have abundance; but from him that hath not shall be taken away even that which he hath" (Matthew 25:29) and refers to an increase in inequalities. We tested the hypothesis that reductions in relative differences in the IMR are greater in the states with a lower baseline. Spearman and Kendall rank correlation coefficients were used to test the hypothesis. Unlike the international experience, results for Mexico do not support the Matthew Effect.

Keywords: Infant Mortality, Brass Technique, The Matthew Effect, Spearman and Kendall Correlation Tests.

1 Introduction

Health and living conditions in the various regions of Mexico have improved considerably since the middle of last century. In this regard, some infectious diseases have been controlled; populations have increased their chances of development and different public health services are now available to a wider number of inhabitants. As a consequence of this, life expectancy has increased and infant mortality has decreased significantly. In particular, the analysis of the reduction in infant mortality in Mexico can be seen from different perspectives according to the level of aggregation of the data used and the characteristics of their distribution over different geographical areas.

In this paper the cross section data on infant mortality are analyzed, i.e. considering the 32 states of the country. Our goal is to test the hypothesis whether reductions in this indicator have been higher in states where child mortality was already low. In other words, we want to know whether

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

inequalities in child mortality have been increasing or decreasing. In the literature on the subject to the hypothesis proposed has been referred to, as the "Matthew Effect" (Dzakpasu, Kramer and Allen, 2000; Bishai and Poon, 2001).

The conceptualization of Matthew Effect is applied to analyze changes in various aspects in the areas of economics, scholar performance and health, in order to establish whether these changes deepen or reduce inequalities. The term was coined by Robert K. Merton (1968), referring to the biblical phrase as follows: "For he that hath, to him shall be given, and he will have abundance; and whoever does not have, even what he has will be taken away "(Gideons International, 2000, chapter 25, verse 29, p. 50) and implies widening inequalities.

Empirically, the existence or not of the Matthew Effect described in the literature has considered at least two methodologies applied to data both internationally and for different regions of a country.

In its application to infant mortality, the basic idea of the Matthew Effect is to verify whether the improvement in the indicator is greater in places where rates are already low, increasing the gap. In this vein, Dzakpasu et al (2000) made a studio by comparing the experience of the twelve provinces of Canada with 133 countries, in relation the behavior in the IMR for the years 1960 and 1995.

To conduct their study they calculated Spearman correlation coefficient on absolute and relative percentage changes in both infant mortality rates and relative difference of the event (infant mortality). Their findings show that while in the Canadian provinces the Matthew Effect was not observed, internationally it does.

2 Methodology

a) The estimation of infant mortality with the method of children ever born /children surviving (CEB/CS) of Brass

William Brass (Brass et al., 1968) developed the indirect method of children ever born /children surviving (CEB/CS) to estimate infant and child mortality, that has allowed to ascertain more accurately the actual levels of infant mortality of about three quarters of the world's population. The CEB/CS method uses information (usually from censuses, although it may come from surveys) of the total number of children ever born (CEB), and the children surviving (CS) that women have had throughout their life, until the moment in which they are interviewed. The information is classified by age of the mother. It is expected (on average) that the older the women, the higher the risk of death for their children, because they have been exposed to the risk of death during a longer period, and thus the proportion of dead children increases with the age of the woman. The method is based on the similarity of the values of the proportions

of deceased children, according to the age of the mother, in five-year age groups within the reproductive period (i. e. 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49) and the probability of dying between birth and the ages of 1 year, 2, 3, 5, 10, 15 and 20 years, respectively.

According to the original formulation of Brass, the method relies on several assumptions. Although none of them is fulfilled completely in any population, its violation in one way or another, in the majority of cases, does not produce very important biases. In general the methodology produces -- in the populations with data incomplete-- estimates closer to reality, compared to the underestimations that produce vital statistics. The assumptions are:

-There is a constant fertility. Although the five-year cohorts at the time of the census have different levels of fertility, ultimately the method is based on ratios of deceased children (DC) in relation to the CEB, DC/CEB, which are not affected by the level of fertility of each cohort in particular.

-There is a constant mortality. This was an assumption that was fulfilled in African populations in the middle of last century, where Brass (1968) worked originally. Given that there was a constant mortality, the method estimated the mortality of the population under study. The application of the method was spreading to other regions of the developing world. Indeed, the method allowed having reasonable estimates of infant mortality in the last quarter of the XX century, for countries where live around three-quarters of the world's population. However, in some of these regions child mortality was in decline. Some variants of the original method including Sullivan (1972), Trussell (1975), and Feeney (1980) were developed. The latter introduced -- considering changes in the mortality -- the location in the time of the estimates of each five-year age group of women who give the information of CEB and CS. In this way, what in the original proposal may have been considered a limitation, with Feeney´s adaptations became a virtue of the method, since it allows having estimates at several points in time and it is possible to know the trends of the indicator.

-There is independence between the mortality of children and the age of the mother. The mortality of children depends, among other variables, on the age of the mother. The children of very young women are more exposed to the risk of dying. This is due to both physical and socio-cultural reasons. On one hand some of these women still do not complete their own development when they are already exerting motherhood, which compromises the survival of their children (and their own survival). On the other hand they may not be prepared optimally for the care of the infant. This is more notorious for the children of women in the 15-19 age group.

However, for the other age groups the assumption is not far from reality.

-There is independence between child mortality and the mortality of their mothers. An orphan child is more likely to die due to the lack of attention and care that his mother could provide. The extreme case occurs when the mother dies when giving birth and the infant is not of breastfed. Fortunately, maternal mortality is a relatively rare event, and even for women in childbearing ages mortality is low. Therefore, there are not so many orphans, and consequently the non-compliance of this assumption does not have a high impact on the estimates.

- The population is closed to migration. The information used for the Brass method reflects the experience of mortality of the children of the informants at any time in the past. Thus, a woman may be reporting in the place where she currently lives, the mortality (or survival) of their children in a place where she lived before. This bias will be important to the extent that there is a considerable differential between the town of origin and that of destination.

For this paper we used the package Q-FIVE: Microcomputer Program for Child Mortality Estimation developed by the Population Division of the United Nations. This program is based on the Trussell´s variant (1975) of the CEB/CS method, and produces results consistent with the four models of the regional life tables of Coale and Demeny (1966), as well as the five mortality patterns of United Nations life tables for developing countries (1982).

From the assumptions of the original formulation of the method by Brass, one of which is more clearly broken, for many countries, and certainly in the case of Mexico, is the one of constant mortality. The Q-FIVE package produces estimates which have a point of reference in the recent past: the estimates derived from data from the women in the seven five-year age groups (between 15 and 49 years of age) are located in the fifteen years prior to the census. Thus, it is possible to observe recent trends in infant mortality.

One of the objectives of this paper was to obtain state-level child mortality estimates for 1990 and 2010 in order to evaluate the declining trends. Therefore, we used information from the XI and the XIII General Population Censuses carried out in 1990 and 2010, respectively. From each of these two sources of information we obtained an estimate of the IMR for every state. We then proceeded in the following manner:

(i) we applied program Q-FIVE to CEB and CS information by five-year age groups (15-19 to 45-49) of women for each state;
ii) Q-FIVE produced seven estimates of IMR with its location in time (approximately from 15 up to 1.5 years before each census).
(iii) with these estimates we projected the trend of IMR until the time of the census, to have estimates in 1990 and 2010.

As it was pointed out above, there is a differential mortality by age of the mother. The children of women 15-19 years, experience an excess mortality. Due to this fact, the estimate for this age group was eliminated from the projection since it breaks with the trend observed in the estimates for the other six five-year age groups (from 20-24 to 45-49 years).

Given that mortality decreases at a slower rate, and to ensure that the projection produced no negative numbers, we adjusted in each case a negative exponential curve for the projection. In some cases in particular perhaps there could have been a better fit, but we preferred to use a standard protocol in the sixty four projections of the IMR by state for 1990 and 2010, that appear in table 1.

b) The Spearman and Kendall coefficients of rank correlation

From the results obtained, in a second stages, we tested the hypothesis under study by calculating the Spearman and Kendall coefficients of rank correlation to determine whether the IMR inequalities among states have declined or not. The Spearman rank correlation coefficient is known in the literature as rho, ρ. The use of ranks indicates the relative magnitude of observations. ρ can take values between + 1 and - 1. A positive coefficient suggests that the two variables are directly related; a negative value indicates that they are inversely related. If the two are independent variables, the correlation coefficient would be close to zero.

To test whether the correlation coefficient is statistically significant population the null hypothesis H0: $\rho \leq 0$ is used, which is contrasted with the alternative Ha: $\rho > 0$, considering that follows a t-student distribution. The rejection of H0 in favor of Ha for the event is interpreted as the absence of Matthew Effect.

An alternative method for determining a correlation coefficient ranges is using the Kendall correlation coefficient, which is denoted by the greek letter τ (tau). Although Kendall τ is suitable for determining the rank correlation with the same type of data used in the Spearman ρ, the two methods use different techniques to determine this correlation, so their values rarely coincide. The formulas used were:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i}{n(n^2 - 1)}$$

where di = xi - yi, is the difference between ranks, and

$$\tau = \frac{(number\ of\ concordant\ pairs) - (number\ of\ discordant\ pairs)}{\frac{1}{2}n(n+1)}$$

3 Results

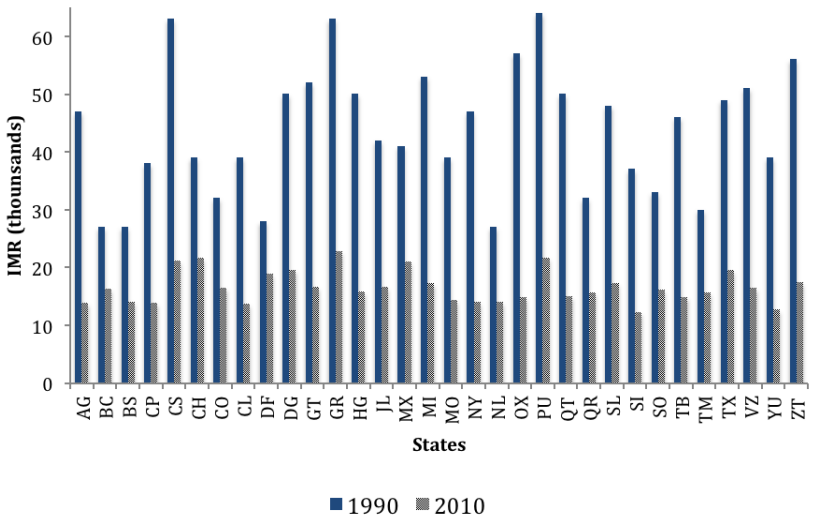
According to the methodology considered in this paper, it is estimated that for 1990 and 2010 the infant mortality rate (IMR), nationally stood at around 38.2 and 17.5 deaths per 1000 live births, respectively. Table 1 shows the indirect estimates of the TMI for each of the 32 states for the years 1990 and 2010. Considering these estimates, the highest value of IMR in 1990 was located in 64 deaths while for 2010 this figure reached the value of 23 deaths. For its part, the minimum values were in the order of 27 to 12 deaths per thousand live births for each of the years shown respectively. Figure 1 shows that this downward trend in the level of the IMR -during the period under study-is present in all the states of our country. Table 2 shows the ranges of both the absolute values of the IMR and the relative differences.

By taking the rank of IMR in 1990 and calculating their correlation in relation to the range of the relative difference for the period 1990-2010, according to the methodology of the Spearman correlation rank coefficient already mentioned, one can feel that small values in differences indicate high chances of a positive correlation coefficient, which can be interpreted as the absence of Matthew Effect (in other words, the relative differences tend to be greater in the states were in 1990 the IMR was high, meaning that inequality is reduced). To envisage this possibility, Figure 2 presents the scatterplot between these two variables (where x = range of IMR in 1990; y = range of the relative difference for the period 1990-2010), appreciating a cloud of scatter points with a positive trend.

If the Matthew Effect existed there would be a negative correlation, i. e. states whith a low IMR in 1990 (and therefore with lower ranks) should be associated with high relative reductions, which do not seem to occur with the data. Instead there is a positive correlation. Calculating the Spearman correlation coefficient confirms this idea producing a positive correlation with a value of 0.7152, being highly significant ($p < 0.000$). In order to corroborate the results obtained and avoid any possibility of miscalculation (caused by the ties in the ranks), we proceeded to estimate the Kendall correlation coefficient, τ, (0.5161 and p-value=0.000). The results obtained confirmed the absence of the Matthew Effect.

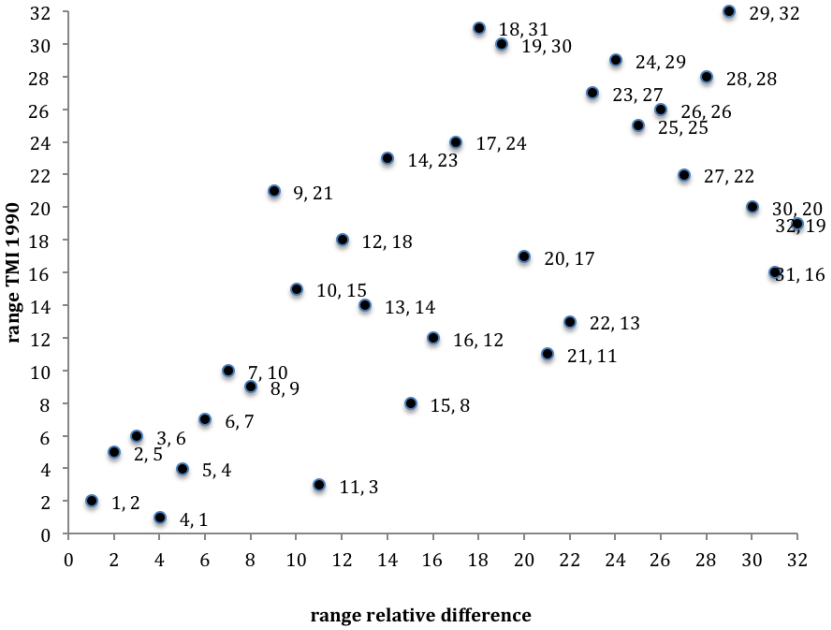
4 Discussion

This paper investigated the possibility of the existence of the Matthew Effect in the IMR in Mexico. To meet this goal we proceeded in two stages. In the first, indirect mortality estimates were made. Here, the results are consistent with the downward trend in the IMR in Mexico. In a second stages two non-parametric statistical tests allowed to know the possible increase in inequality in infant mortality for the period 1990-2010 in Mexico; these tests consisted of calculating the Spearman and Kendall correlation coefficients, which offered clear evidence against the Matthew Effect.



Source: Based on data from Table 1.

Fig, 1. IMR by State in 1990 and 2010



Source: Based on data from Table 2

Fig. 2. Scatterplot between the ranges of IMR 1990 and the relative difference from1990 to 2010

Table 1. Estimates of IMR by state obtained by Brass' children ever born/ children surviving method, 1990 and 2010
(rate per thousand live births)

	IMR		Difference	
State	1990	2010	absolut	relative (%)
Aguascalientes	47.0	13.9	33.1	70.4
Baja California	27.0	16.3	10.7	39.6
Baja California Sur	27.0	14.1	12.9	47.8
Campeche	38.0	13.9	24.1	63.4
Chiapas	63.0	21.1	41.9	66.5
Chihuahua	39.0	21.6	17.4	44.6
Coahuila	32.0	16.5	15.5	48.4
Colima	39.0	13.7	25.3	64.9
Distrito Federal	28.0	18.9	9.1	32.5
Durango	50.0	19.6	30.4	60.8
Guanajuato	52.0	16.6	35.4	68.1
Guerrero	63.0	22.8	40.2	63.8
Hidalgo	50.0	15.8	34.2	68.4
Jalisco	42.0	16.6	25.4	60.5
México	41.0	21.0	20.0	48.8
Michoacán	53.0	17.3	35.7	67.4
Morelos	39.0	14.3	24.7	63.3
Nayarit	47.0	14.0	33.0	70.2
Nuevo León	27.0	14.0	13.0	48.1
Oaxaca	57.0	14.9	42.1	73.9
Puebla	64.0	21.7	42.3	66.1
Querétaro	50.0	15.0	35.0	70.0
Quintana	32.0	15.7	16.3	50.9
San Luis Potosí	48.0	17.3	30.7	64.0
Sinaloa	37.0	12.3	24.7	66.8
Sonora	33.0	16.2	16.8	50.9
Tabasco	46.0	14.9	31.1	67.6
Tamaulipas	30.0	15.7	14.3	47.7
Tlaxcala	49.0	19.5	29.5	60.2
Veracruz	51.0	16.4	34.6	67.8
Yucatán	39.0	12.7	26.3	67.4
Zacatecas	56.0	17.4	38.6	68.9

Source: Own elaboration with data Aguirre and Vela (2012).

Table 2. Ranges of IMR 1990 and ranges of relative difference from 1990 to 2010, by State

State	Ranges of IMR		Ranges of relative difference
	1990	2010	from 1990 to 2010 (%)
Aguascalientes	18	4	31
Baja California	1	17	2
Baja California Sur	2	8	5
Campeche	10	5	15
Chiapas	30	29	20
Chihuahua	11	30	3
Coahuila	6	19	7
Colima	12	3	18
Distrito Federal	4	25	1
Durango	22	27	13
Guanajuato	26	20	26
Guerrero	31	32	16
Hidalgo	23	15	27
Jalisco	16	21	12
México	15	28	8
Michoacán	27	22	22
Morelos	13	9	14
Nayarit	19	6	30
Nuevo León	3	7	6
Oaxaca	29	10	32
Puebla	32	31	19
Querétaro	24	12	29
Quintana	7	13	10
San Luis Potosí	20	23	17
Sinaloa	9	1	21
Sonora	8	16	9
Tabasco	17	11	24
Tamaulipas	5	14	4
Tlaxcala	21	26	11
Veracruz	25	18	25
Yucatán	14	2	23
Zacatecas	28	24	28

Source: Own elaboration with data Table 1.

References

1. Aguirre, Alejandro and Fortino Vela-Peón (2012). "La mortalidad infantil en México, 2010", Papeles de Población, Vol.18, No. 73, México.
2. Bishai, D. y Poon A. (2001). "Does the law of diminishing returns apply to infant mortality decline?". HPC Working paper 03, Hopkins Population Center, University Baltimore.
3. Brass, W. y Coale A. J. (1968). "Methods of Analysis and Estimation" En: Brass W. et al The Demography of Tropical Africa, Princeton University Press.
4. Coale, A. J. y Demeny P. (1966). Regional Model Life Tables and Stable Populations, Princeton, Nueva Jersey.
5. Dzakpasu S.; Joseph K. S., Kramer M. S. y Allen A. C. (2000). "The Matthew effect: infant mortality in Canada and Internationally", Pediatrics; Vol. 106, Núm. 1. http:// www.pediatrics.org/cgi/content/full/106/1/e5 (consulta 23 de septiembre de 2007).
6. Gideons International (2000). El nuevo testamento de nuestro Señor Jesucristo con salmos y proverbios, Estados Unidos.
7. Feeney, G. (1980). "Estimating Infant Mortality Trends from Child Survivorship Data", Populations Studies, Vol. XXXIV, Núm. 1: 109-128.
8. Harnett, D. y Murphy J. L. (1987). Introducción al análisis estadístico, Addison-Wesley, 579-582.
9. Hartmann, M. (1991). "A Parametric Method for Census Based Estimation of Child Mortality", Journal of Official Statistics, Vol. 7, Núm. 1, pp. 45-55. (http://www.jos.nu/ Articles/article.asp consulta 10 de diciembre 2011)
10. INEGI (1990). Estadísticas Históricas de México, tomo I, México.
11. INEGI (1994). Encuesta Nacional de la Dinámica Demográfica, 1992. Principales resultados. México:
12. INEGI (2004). La mortalidad infantil en México 2000. Estimación por entidad federativa y municipio, México.
13. Joseph, K. S. (1989). "The Matthew effect in health development", BMJ, Vol. 298, Núm. 6686, 1497-1518.
14. Mcmahon, M. J. (2004). "The Matthew Effect and Federal Taxation", Boston College Law Review, Vol. 45, 993- 1128.
15. Medoff, M. H. (2006). "Evidence of a Harvard and Chicago Matthew Effect", Journal of Economic Methodology, Vol.13, Núm. 4, 485-506.
16. Merton, R. K. (1968). "The Matthew effect in science", Science, Vol. 159, Núm. 810, 56-63.
17. Naciones Unidas (1982). Tablas de Vida Modelo para Países en Desarrollo, NuevaYork: Autor.
18. QFIVE: Microcomputer Program for Child Mortality Estimation [CD-ROM] (1990). United Nations.
19. Trussell, J. T. (1975). "A Re-estimation of the Multiplying Factors for the Brass Technique for Determining Childhood Survivorship Rates", Population Studies, Vol. XXIX, Núm. 1, 97-108.
20. Wackerly, D. D., Mendenhall, W. y Scheaffer R. L. (2002). Estadística matemática con aplicaciones, 6a. ed.,Thomson, 750-755

Bayesian object identification from image data using the level set method

Robert G Aykroyd and Nada Alfaer

Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK
(e-mail: r.g.aykroyd@leeds.ac.uk)

Abstract. Many image analysis problems can be characterised as finding objects within an image. The problem might involve an unknown number of objects with unknown locations within the image, and with unknown and varying sizes and shapes. In dynamic imaging these might also change with time, in particular with objects merging and splitting, as well as moving and changing size and shape. This paper will study the level-set method for object identification which embeds the object outlines within a higher dimensional surface. The method can easily cope with complex geometric changes whilst maintaining closed and non-overlapping outlines making it particularly useful for dynamic modelling. The Chan-Vese model based on the level-set method will be defined and given a Bayesian re-interpretation. Then, the method will be applied to a sequence of simulated image data, and to 2D image reconstructions obtained from medical SPECT data to identify the outline of the subject.

Keywords: Bayesian methods, Chan-Vese model, Image analysis, Segmentation.

1 Introduction

Consider a generic object identification problem, where the aim is to find the shape of objects against a general background. Further suppose that there is some physical characteristic which can be used to distinguish between objects and background, for example colour, intensity or texture. However, suppose that measurements of this property are subject to noise and so some data processing steps are needed before the objects can be reliably identified.

Traditional approaches to object identification and image segmentation might consider pixel-wise classification using some multivariate analysis technique, but would require further post-processing to guarantee compact regions. Alternative methods could be based on edge-detection but there is no certainty that region boundaries will be closed, or region growing algorithms which are usually slow and accuracy depends heavily on the choice of similarity criterion.

In this paper the level set approach [8] will be considered as a method for locating object outlines; see also [6] and [10]. The method was proposed for modelling temporal changes of flame shapes, but the same approach is relevant when searching for best-fit outlines through some iterative algorithm. The main

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

advantages are that it can handle an unknown number of objects and complicated topological changes in a simple and natural way. For example, when changes are between situations with simply and multiply connected regions.

The paper is organised as follows. Section 2 gives a mathematical description of the object-based model and defines notation to be used later. A short description of the level-set approach is given in Section 3 with a Bayesian reinterpretation in Section 4. The methods are applied to simulated and real data in Section 5. Finally, conclusions and further work are discussed in Section 6.

2 An object-based image model

Let the unknown physical characteristic, which will be called intensity, be defined $x(\mathbf{s}), \mathbf{s} \in \mathcal{S}$, within some domain, $\mathcal{S} \subset \mathbb{R}^2$. Now, the background and the objects partition the domain, and so denote these non-overlapping regions, $\boldsymbol{R} = \{R_k : k = 0 \ldots, m-1\}$ with corresponding characteristic intensities $\boldsymbol{\mu} = \{\mu_k : k = 0, \ldots, m-1\}$. Hence, $R_k \subset \mathcal{S}$, $R_k \cap R_l = \emptyset$ for $k \neq l$ and $R_0 \cup \cdots \cup R_{m-1} = \mathcal{S}$. Further R_0 will be used to denote the background, R_1 the first object etc. Also, let $R_\star = R_1 \cup \cdots \cup R_{m-1}$ represent the overall subset of the domain containing objects, that is $R_\star = R_0^c$.

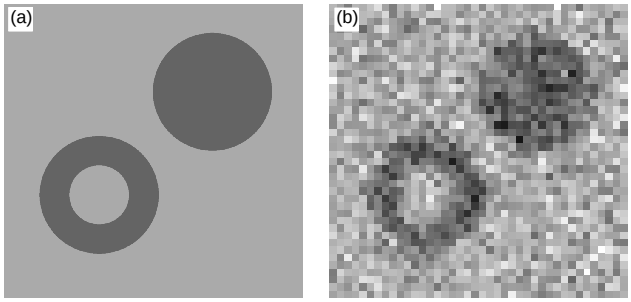


Fig. 1. Grey-level images showing: (a) domain partition and (b) an example dataset.

Suppose that there is a finite set of measurements $\boldsymbol{y} = \{y_i : i = 1, \ldots, n\}$ collected at locations $\boldsymbol{s} = \{\mathbf{s}_i : i = 1, \ldots, n\}$ forming a regular square grid. The intensity function can be discretised to give unknowns $\boldsymbol{x} = \{x_i : i = 1, \ldots, n\}$ with $x_i = x(\mathbf{s}_i)$ for $i = 1, \ldots, n$. See Fig. 1 for an example of a domain partition composed of two objects and an example dataset both shown as grey-level images. For a review of the statistical approach to image analysis see, for example, [1] and references therein. The measurements can be defined in terms of a deterministic component, which only depends on the region, and a stochastic component, thus

$$y_i = x_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where

$$x_i = \begin{cases} \mu_0 & \text{if } \mathbf{s}_i \in R_0, \\ \mu_\star & \text{if } \mathbf{s}_i \in R_\star; \end{cases} \tag{2}$$

and it will be assumed that the errors are independent and normally distributed with constant variance, that is $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, n$. The object identification problem is then one of estimating the domain partition, that is the number of regions and the coverage of each region, along with region intensities and noise variance. This gives a parameter set $\boldsymbol{\Theta} = \{m, R_0, \ldots, R_{m-1}, \mu_0, \mu_\star, \sigma\}$ of length $m + 4$ — which is not fixed since m itself is unknown.

3 The level-set method

The overall aim of a generic segmentation problem is to partition a domain of interest into homogeneous parts, which might be labelled background, R_0, and objects, $R_\star$. Then, let γ be the outline of the objects, that is the interface between R_0 and $R_\star$ — see Figure 2(a).

Suppose that we will use some iterative scheme, starting from an initial solution, which gradually moves towards a final solution. These steps can usefully be thought of as temporal changes, and this links object identification from single images to the original application of tracking a moving flame [7]. Any estimation method based on an iterative scheme must allow smooth changes in the position of the curve. An approach which models the curve explicitly will, however, need additional constraints to guarantee the curve stays well defined.

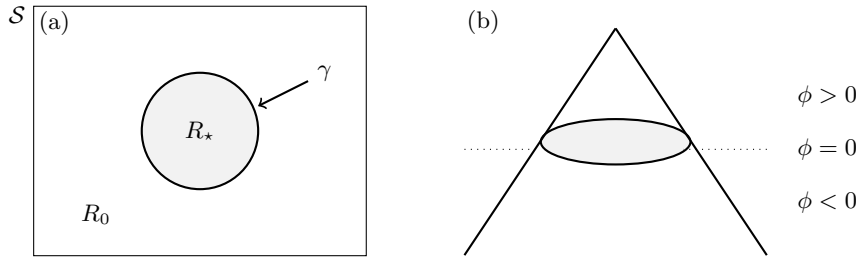


Fig. 2. Diagram showing the relationship between the object outline and the corresponding level-set function.

In the level-set method [7] the curve is embedded in a higher dimensional surface, ϕ, such that $\gamma(t) = \{\mathbf{s} : \phi(\mathbf{s}, t) = 0\}$, see Figure 2(b), with

$$\begin{aligned} \phi(\mathbf{s}, t) &< 0 \quad \text{for} \quad \mathbf{s} \in R_0, \\ \phi(\mathbf{s}, t) &> 0 \quad \text{for} \quad \mathbf{s} \in R_\star. \end{aligned}$$

The function ϕ is called the *level-set function* and γ is the *zero level set*.

One approach to the estimation of the level-set and the corresponding zero level-set [3] defines an energy function which is made-up of two parts: one which measures the model mismatch, and the other, a regularising term which aims to ensure stability of the iterative process. Firstly, the model mismatch is defined as

$$\chi(\boldsymbol{y}, \boldsymbol{x}) = \int_{\mathcal{S}} (y(\mathbf{s}) - x(\mathbf{s}))^2 d\mathbf{s}. \tag{3}$$

Now, the Heaviside function, $H(z)$, takes the value 1 when $z \geq 0$ and zero otherwise, and the Dirac function, $\delta(z)$, is its derivative $\delta(z) = dH(z)/dz$. These give rise to definitions of the area of the objects and the length of the object boundary

$$A_\star = A(\phi(\mathbf{s}) > 0) = \int_{\mathcal{S}} H(\phi(\mathbf{s}))d\mathbf{s} \tag{4}$$

and

$$L_\star = L(\phi(\mathbf{s}) = 0) = \int_{\mathcal{S}} |\nabla H(\phi(\mathbf{s}))|d\mathbf{s} = \int_{\mathcal{S}} \delta(\phi(\mathbf{s}))|\nabla\phi(\mathbf{s})|d\mathbf{s}. \tag{5}$$

The Heaviside function can also be used to re-write the model mismatch as

$$\int_{\mathcal{S}} (y(\mathbf{s}) - x(\mathbf{s}))^2(1 - H(\phi(\mathbf{s})))d\mathbf{s} + \int_{\mathcal{S}} (y(\mathbf{s}) - x(\mathbf{s}))^2 H(\phi(\mathbf{s}))d\mathbf{s}. \tag{6}$$

Putting these together gives an energy function, in terms of the ϕ,

$$\begin{aligned} E(\phi) = &\int_{\mathcal{S}} (y(\mathbf{s}) - \mu_0)^2(1 - H(\phi(\mathbf{s})))d\mathbf{s} + \int_{\mathcal{S}} (y(\mathbf{s}) - \mu_\star)^2 H(\phi(\mathbf{s}))d\mathbf{s} \\ &+ \alpha \int_{\mathcal{S}} H(\phi(\mathbf{s}))d\mathbf{s} + \lambda \int_{\mathcal{S}} \delta(\phi(\mathbf{s}))|\nabla\phi(\mathbf{s})|d\mathbf{s}. \end{aligned} \tag{7}$$

The best-fit curve is then defined as the solution which minimises the energy,

$$\widehat{\phi} = \arg\min_{\phi} E(\phi). \tag{8}$$

It can then be shown [5] that

$$\frac{\partial E}{\partial \phi} = \delta(\phi(\mathbf{s})) \left[-(y(\mathbf{s}) - \mu_0)^2 + (y(\mathbf{s}) - \mu_\star)^2 + \alpha - \lambda \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|} \right) \right]. \tag{9}$$

Further, since $\partial\phi/\partial t = -\partial E/\partial\phi$ the corresponding discrete-time updating equation is

$$\phi^{k+1} = \phi^k + \Delta t \cdot \left. \frac{\partial \phi}{\partial t} \right|_{\phi^k} \quad \text{for } k = 1, \ldots \tag{10}$$

where $\phi^k = \phi(\mathbf{s}, t_k)$ and Δt is a time step. For this method to work well a *reinitialization* step may be required [11] to ensure $|\nabla\phi| = 1$, and spatial smoothing of ϕ to produce smooth object outlines.

4 Bayesian interpretation

The key ingredients in any Bayesian approach are the likelihood function and the prior distribution; which combine to give the posterior distribution. Following on from the previous section, the model parameters are now $\Theta = \{\phi, \mu_0, \mu_\star, \sigma\}$. The likelihood is the conditional distribution of the data given the unknown parameters, denoted as $f(\boldsymbol{y}|\boldsymbol{\Theta})$. The data, $\boldsymbol{y}$, depend on Θ

through a mapping to the corresponding intensity function $\boldsymbol{x}(\Theta)$ and stochastic noise assumed to be well modelled by a normal distribution leading to the likelihood

$$f(\mathbf{y}|\Theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - x_i(\Theta)\right)^2\right\} \qquad \sigma > 0, \tag{11}$$

where $x_i(\Theta) = \mu_k$ if $\mathbf{s}_i \in R_k$. This can be re-written as

$$f(\mathbf{y}|\Theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i:\mathbf{s}_i\in R_0}\left(y_i-\mu_0\right)^2 + \sum_{i:\mathbf{s}_i\in R_\star}\left(y_i-\mu_\star\right)^2\right]\right\}. \tag{12}$$

The prior distribution, denoted $p(\Theta)$, quantifies detailed expert knowledge or, as here, general beliefs about the unknown function. The choice of the exact form of this distribution is much more subjective than is the choice of likelihood. Here consider prior distributions which say that smaller and compact regions are more likely than large or irregular regions. This might suggest modelling the area of the objects or the length of their boundaries. If we choose to describe these using independent exponential random variables, then the corresponding joint prior distribution is

$$p(\Theta) = p(A_\star) \times p(L_\star) = (\alpha\lambda)\exp\{-\alpha A_\star - \beta L_\star\}, \quad A_\star \geq 0, L_\star \geq 0 \tag{13}$$

where $A_\star$ is the area of the objects calculated as in equation (4) and $L_\star$ is the total length of the object boundaries calculated as in equation (5).

For estimation, evidence from the data and from prior beliefs are brought together by combining the likelihood and prior distribution, using Bayes theorem, to form the posterior distribution, defined as

$$p(\Theta|\boldsymbol{y}) = f(\boldsymbol{y}|\Theta)p(\Theta)/f(\boldsymbol{y}). \tag{14}$$

using (12) and (13) gives

$$\log p(\Theta|\boldsymbol{y}) = -\frac{1}{2\sigma^2}\left(\sum_{i:s_i\in R_0}\left(y_i-\mu_0\right)^2 + \sum_{i:s_i\in R_\star}\left(y_i-\mu_\star\right)^2\right) - \alpha A_\star - \lambda L_\star, \tag{15}$$

which is the discrete equivalent of the (negative) of the energy function in (7).

In the Bayesian setting a point estimate can be found as the value which corresponds to the maximum of the posterior distribution, or equivalently the log-posterior, this is called the maximum a posteriori (MAP) estimate

$$\widehat{\Theta}_{MAP} = \arg\max_{\Theta} p(\Theta|\boldsymbol{y}) = \arg\max_{\Theta} \log p(\Theta|\boldsymbol{y}). \tag{16}$$

Hence it is clear that the maximization of the posterior is exactly equivalent to the minimisation of the level-set energy function (7) which means that the same iterative estimation algorithm based on (10) and using (9) can be used. The updates continue until any further changes are deemed insignificant. Here, all calculations are performed in the R language [9] and, of course, in practice all these terms must be approximated by their discrete equivalents—details will not be given here.

5 Numerical results

To illustrate the iterative nature of the level-set method consider Figure 3 which shows the initial zero level-set in (a) and the zero level-set after the first five iterations. After one iteration the zero level-set begins to encircle the two objects. Gradually the outlines are more closely matched and the central region in the annulus appears and grows. By the fourth iteration, (e), the object outlines are well established with little change afterwards.

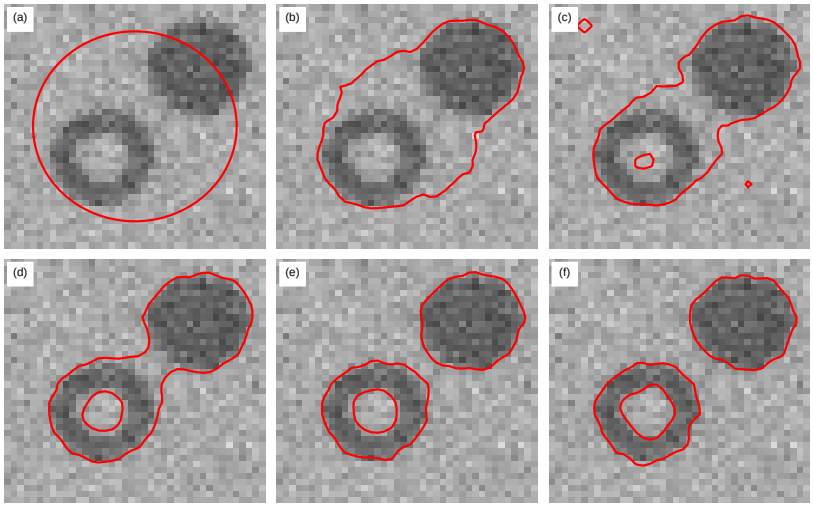


Fig. 3. Results showing the zero-level set for the first few iterative.

Next a dynamic simulation is considered with result in Figure 4 which clearly demonstrate that the method can track the boundaries.

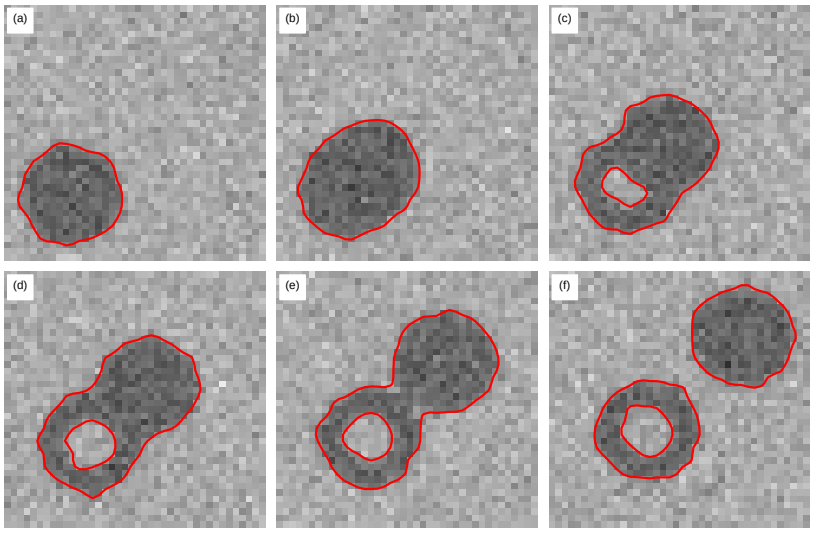


Fig. 4. Results for an image sequence showing data image and final zero-level set.

As an illustration of a real-world application consider the analysis of image reconstructions in SPECT. The medical investigation involved the imaging of the human head where a radioactively tagged chemical was injected into the bloodstream highlighting areas of high blood flow. The dataset used here was first analysed in [4], but the reconstructed images are from [2]; the reader is directed to these papers for further details.

Initially the images had been analysed using a model which wrongly assumed spatial homogeneity, see [2], but this lead to reconstructions which were grossly over smoothed. So a fixed diameter circle was manually chosen as a *region of interest* and the data re-analysed. The level-set approach presented here is a clear alternative and would allow automatic and case specific regions to be determined. Figure 5 shows six such cases equally spaced from the neck upwards to the top of the head. The proposed level-set approach successfully identifies the central region and could be used as a basis for further analysis. Note that the grey band around the zero level-set is due to the over smoothing during image reconstruction which should be largely eliminated if the newly identified boundary is incorporated into a re-analysis.

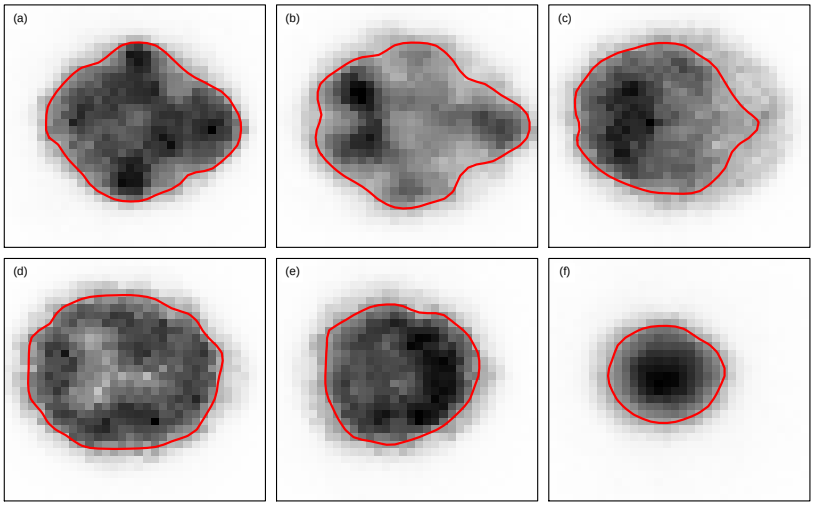


Fig. 5. Results for real data showing reconstructions as data and final zero level-set.

6 Conclusions

Object identification is a common problem in image analysis which can be challenging when the number of objects is unknown and the location, shape and size of the objects also has to be determined. There are further difficulties when the object geometries change with time. The level-set approach provides an ideal option for such situations in that it is fast and can cope with dynamic changes in a natural way. The proposed method has been successfully applied

to both simulated and real data. This technique is not widely used in statistics, but has a simple and direct interpretation in statistical language. Once this has been done the probabilistic framework can easily be seen and understood, and hence adaptations made for tackling other problems.

The work presented here is the first part of an ongoing research project and there are many practical and methodological issues still to be addressed. There are many choices of prior descriptions, both in terms of which model parameters or other summary measures to include in the model, but then also the choice of prior distribution. In the original method the specified regularising function corresponds to an exponential distribution. This, however, has the unusual property that the most likely values are close to zero, which is not a meaningful statement. Instead it might be more appropriate to use a distribution which allows a non-zero mode, such as a gamma distribution. There is then the choice of what prior parameter values should be used. Perhaps these could be estimated in some way which would then allow the approach to be fully automated. It is also important to perform a sensitivity analysis to see how robust the approach is to changes in model aspects which are somewhat arbitrary. This discussion emphasises the great range of generalisations possible which make it clear that this approach has potential use in many important practical applications.

References

[1] Aykroyd, R.G. Statistical image reconstruction in *Industrial tomography: Systems and applications*, pp. 401–427. Woodhead Publishing, 2015.

[2] Aykroyd, R.G. and Zimeras, S. Inhomogeneous prior models for image reconstruction. *J. Am. Stat. Soc.*, 94:934–946, 1999.

[3] Chan, T.F. and Vese, L.A. Active contours without edges. *IEEE Trans. on Image Processing*, 10:266–277, 2001.

[4] Green, P.J. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. on Medical Imaging*, 9:84–93, 1990.

[5] Mitiche, A. and Ben Ayed, I. *Variational and level set methods in image segmentation*, Vol. 5. Springer Science & Business Media, 2010.

[6] Osher, S. and Fedkiw, R. *Level set methods and dynamic implicit surfaces*, Vol. 153. Springer Science & Business Media, 2003.

[7] Osher, S. and Paragios, N. *Geometric level set methods in imaging, vision, and graphics*. Springer Science & Business Media, 2003.

[8] Osher, S. and Sethian, J.A. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comp. Phys.*, 79:1249, 1988.

[9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[10] Sethian, J.A. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, Vol. 3. Cambridge University Press, 1999.

[11] Sussman, M., Smereka, P. and Osher, S. A level set approach for computing solutions to incompressible two-phase flow. *J. Comp. Phys.*, 114:146–159, 1994.

Performance Comparisons of Some Bivariate Dispersion Control Charts

Georgios Bartzis and Sotirios Bersimis

Department of Statistics and Insurance Science, University of Piraeus, Karaoli and Dimitriou, 80, 18534 Piraeus, Greece
(E-mail: bartzis.georgios@gmail.com, sbersim@unipi.gr)

Abstract. Statistical Process Monitoring is one of the most well-known sectors of Statistics. The most advanced procedure of Statistical Process Monitoring is control charting. Control charts are applied to monitor the levels of a variable related to the quality of a product/process. When the quality of a product is related to more than one characteristics, then a Multivariate Statistical Process Monitoring technique is applied. Moreover, when the multivariate dispersion has to be monitored we use a multivariate control chart for the dispersion. The aim of this paper is to compare some well known multivariate control charts for the dispersion of Normally distributed processes.
Keywords: Multivariate Statistical Process Monitoring, Control Charts Comparison, Bivariate Dispersion, Multivariate Dispersion.

1 Introduction

Statistical process monitoring (*SPM*) is a well known method for controlling and improving a product's quality in an industrial or manufacturing process. By using statistical methodology, the researcher can either establish conforming standards or contribute to the maintenance of a desirable product's quality or both. In practice, the quality of a product is not related to one but to more than one characteristics. In other words, it is necessary to monitor more than one characteristics simultaneously to assure product's quality. Jackson [14], has stated that 4 points are necessary to be answered by a multivariate statistical process monitoring (*MSPM*) procedure:

- is the process in-control or out-of-control?
- what is the overall probability for the event "procedure diagnoses an out-of-control state erroneously"?
- is the relation between the variables/attributes taken into account?
- if the process is out-of-control, what is the problem?

Historically speaking, Hotteling [9] applied first the idea of *MSPM* in a dataset regarding bomb sights in World War II. A huge amount of studies followed Hotteling's idea thereafter including Jackson [11][12][13], Montgomery and Wadsworth [21], Alt [1], Crosier [6], Hawkins [7][8], Pignatiello and Runger [22], Tracy *et*

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

al. [25], Lowry and Montgomery [16], Maravelakis *et al.* [19], Koutras *et al.* [15], and Maravelakis and Bersimis [20].
In the area of *SPM* and *MSPM*, the main interest lies in monitoring the mean of the process. Nevertheless, controlling the dispersion of the process is as equally important as controlling the mean. By monitoring the mean in an industrial or manufacturing process, the researcher investigates if the characteristics of the product comply to the predetermined specifications. By assuming that an appropriate statistic is plotted in a point in time between the Upper Control Limit (UCL) and the Lower Control Limit (LCL) when monitoring the mean of the process, then the process is considered to be in-control on the condition that the dispersion of the process has not changed through time. It should be noted that when constructing a control chart (CC) for the mean level of the process, the dispersion of the process is also taken into account indirectly by computing the control limits. In other words, the control limits of a process always depend upon the dispersion and therefore, the dispersion of the process should always be monitored. In the case that the dispersion is out-of-control, the level of the process fluctuates more, leading the process out-of-control.
The purpose of this paper is to present and compare multivariate CCs for the dispersion of the process by assuming normally distributed data. It is known that the CCs are compared by using the Average Run Length (ARL) which is the expected waiting time until the first occurrence of an event creating an out-of-control signal. In the literature there are two distinct cases for the ARL. The in-control ARL and the out-of-control ARL. The in-control ARL is the average number of plotted samples until an out-of-control signal, when the process is truly in-control. On the other hand, the out-of-control ARL is the average number of plotted samples until an out-of-control signal when the process is truly out-of-control. Thus, the aim of the paper is to identify, which is the optimal way for monitoring a dispersion process under certain conditions (these conditions, in our study, correspond to different simulation scenarios). In the case that we have a best suited methodology for each and every possible monitoring scenario (large sample size, small sample size, high dimensional case, case with missing values, bivariate processes, multivariate processes, etc.), we are in place of knowing if the process is in- or out-of-control faster (1^{st} point of Jackson).

2 Methods

In this section, eight well known bivariate and multivariate CCs for the dispersion are presented. These CCs have been constructed for monitoring the dispersion of two or more characteristics simultaneously. The Phase II versions of these CCs are presented since in the comparison study the in-control parameters are assumed known. By following Bersimis *et al.* [5] definition, "*in Phase II, CCs are used for testing whether the process remains in control when future subgroups are drawn. In this phase, the charts are used as aids to the practitioner in monitoring the process for any change from an in-control state. At each sampling stage, the practitioner asks the question -Has the state of the process changed?- ".*

2.1 CC1: A *CC* based on the Moments of the Generalized Variance

Alt [1] proposed the development of a $|\boldsymbol{S}|$-CC by using the first two moments of $|\boldsymbol{S}|$ (Generalized Variance). The statistic that is plotted on this CC is the $|\boldsymbol{S}|$. The 3σ control limits are calculated using:

$$\begin{aligned} UCL &= |\boldsymbol{\Sigma}_0|\left(b_1 + 3\sqrt{b_2}\right), \\ CL &= |\boldsymbol{\Sigma}_0| b_1, \\ LCL &= |\boldsymbol{\Sigma}_0|\left(b_1 - 3\sqrt{b_2}\right), \end{aligned}$$

where $\boldsymbol{\Sigma}_0$ is the known variance-covariance matrix,

$$b_1 = \frac{\prod_{i=1}^{p}(n-i)}{(n-1)^p},$$

$$b_2 = \frac{\prod_{i=1}^{p}(n-i)}{(n-1)^{2p}} \times \left[\prod_{j=1}^{p}(n-j+2) - \prod_{j=1}^{p}(n-j)\right],$$

and n and p are the sample size and the number of variables, respectively. If the LCL is less than zero, then it is replaced by zero.

2.2 W: A *CC* based on the Likelihood Ratio Test (*LRT*)

Another method for monitoring the dispersion of a multivariate process proposed by Alt [1] uses the statistic:

$$W = -pn + pn\ln n - n\ln\left(\frac{|\boldsymbol{A}|}{|\boldsymbol{\Sigma}_0|}\right) + trace\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}\right).$$

Alt, made use of the asymptotic LRT result from Anderson [3] for computing the above statistic. The W statistic is used to test repetitively: $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ vs $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$. In the LRT, the likelihood between the null and the alternative hypothesis are compared. Large difference between the likelihoods suggests that the null hypothesis poorly describes the current situation relatively to the alternative hypothesis.
The α probability control limits for the CC, where α denotes the probability of wrongly reject the H_0, are:

$$\begin{aligned} UCL &= \chi^2_{p(p+1)/2;1-\alpha}, \\ LCL &= 0. \end{aligned}$$

It is noted that $\boldsymbol{A}$ is the sum of squares and cross products matrix, i.e. $\boldsymbol{A} = (n-1)\boldsymbol{S}$. If W plots over UCL then the process is considered out-of-control. $\chi^2_{p(p+1)/2}$ refers to a Chi-Square distribution with $p(p+1)/2$ degrees of freedom and α refers to the upper α percentile of the Chi-Square distribution with $p(p+1)/2$ degrees of freedom.

2.3 CC2: A *CC* based on the Distribution of the Generalized Variance

Alt [1], proposed that the variability of the process can be monitored using a CC that is based on the distribution of $|\boldsymbol{S}|$. The plotted statistic is again the $|\boldsymbol{S}|$. The $CC2$ has the following α probability limits:

$$UCL = \frac{|\boldsymbol{\Sigma}_0|\left(\chi^2_{2n-4;1-\alpha/2}\right)^2}{\left(2\left(n-1\right)\right)^2},$$
$$LCL = \frac{|\boldsymbol{\Sigma}_0|\left(\chi^2_{2n-4;\alpha/2}\right)^2}{\left(2\left(n-1\right)\right)^2},$$

and center line $CL = |\boldsymbol{\Sigma}_0|$ where

$$\frac{2(n-1)|\boldsymbol{S}|^{1/2}}{|\boldsymbol{\Sigma}_0|^{1/2}} \sim \chi^2_{2n-4}.$$

2.4 CC3: A Modified *CC* based on the Distribution of the Generalized Variance

Alt and Smith [2] proposed a modified CC for monitoring simultaneously p quality characteristics, in which $|\boldsymbol{S}|^{1/2}$ is used as the plotted statistic. The α probability control limits are the following:

$$UCL = \frac{|\boldsymbol{\Sigma}_0|^{1/2}\chi^2_{2n-4,1-\alpha/2}}{2\left(n-1\right)},$$
$$LCL = \frac{|\boldsymbol{\Sigma}_0|^{1/2}\chi^2_{2n-4,\alpha/2}}{2\left(n-1\right)},$$

with center line $CL = |\boldsymbol{\Sigma}_0|^{1/2}$.

2.5 VMAX: A *MEWMA* Scheme Based on the *VMAX* Statistic

Machado and Costa [18] proposed a Bivariate Exponentially Weighted Moving Average CC scheme based on $VMAX$ for detecting changes in the covariance matrix $\boldsymbol{\Sigma}_0$ of a bivariate process.
The $EWMA$ CCs introduced by Roberts [24] are used as a tool for detecting small shifts in a manufacturing process. $EWMA$ CCs have no memory, since they take into account information obtained from all previous samples, where each sample is weighted proportionally to the distance from the current sample. The $MEWMA$ scheme based on $VMAX$ uses the following statistics:

$$Z_i = \lambda Y_i + (1-\lambda)\, Z_{i-1},\ i = 1, 2, \ldots$$

A signal is given when $Z_i > UCL$ where:

$$UCL = \frac{\left(\chi^2_{2n-4,\alpha}\right)^2 |\boldsymbol{\Sigma}_0|}{4\left(n-1\right)^2}.$$

It is noted that $Y_i = \max\left\{S^2_{x_i}, S^2_{y_i}\right\}$. $S^2_{x_i}$ and $S^2_{y_i}$ are the sample variances of X and Y, respectively. The starting value Z_0 is often taken to be the expected in-control value of Z as defined by Lucas and Saccucci [17].

2.6 T1 and T2: Two *CC*s for Testing Covariance Changes Without Large Data

Hung and Chen [10] proposed two statistics - $T1$ and $T2$- for monitoring the variance-covariance matrix. Their proposal is based on the fact that $(n-1)\boldsymbol{S}$ follows the Wishart distribution as defined by Wishart [26] with parameters $(n-1)$ and $\boldsymbol{\Sigma}_0$ ($\boldsymbol{S} \sim W_p\left(n-1, \boldsymbol{\Sigma}_0\right)$). Since $\boldsymbol{\Sigma}_0$ is positive definite, there is a matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A}\boldsymbol{\Sigma}_0\boldsymbol{A}' = \boldsymbol{I}_p$ so that $(n-1)\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}' \sim W_p\left(n-1, \boldsymbol{I}_p\right)$. Using the Cholesky's decomposition theorem developed by Cholesky [4], $\boldsymbol{\Sigma}_0$ can be decomposed into $\boldsymbol{M}\boldsymbol{M}'$ with $\boldsymbol{M}$ being a lower triangular matrix with positive diagonal elements. As mentioned by the authors, the aforementioned $\boldsymbol{A}$ can be chosen to be $\boldsymbol{M}^{-1}$. By applying the Cholesky's decomposition theorem once more to the $(n-1)\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}'$ another lower triangular matrix $\boldsymbol{T}$ can be obtained. The elements of the matrix $\boldsymbol{T}$ are mutually independent and are distributed as follows:

$$t^2_{ii} \sim \chi^2_{n-i}, \text{ for } 1 \leq i \leq p,$$

and

$$t_{ij} \sim N\left(0,1\right), \text{ for } 1 \leq j < i \leq p,$$

which is a result from Bartlett's decomposition theorem. The following hypothesis test is considered: $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ versus $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$. Any departure from the null hypothesis will make certain t_{ij} behave abnormally. By using the lower triangular matrix $\boldsymbol{T}$ two statistics can be derived (T_{diag} and $T_{\text{off-diag}}$) which are defined as:

$$T_{\text{diag}} = \sum_{1 \leq i \leq p} t^2_{ii} \sim \chi^2_{(p/2)(2n-p-1)},$$

and

$$T_{\text{off-diag}} = \sum_{1 \leq j < i \leq p} t^2_{ij} \sim \chi^2_{(p/2)(p-1)}.$$

By making use of the elements in the lower diagonal matrix $\boldsymbol{T}$ and the 2 defined statistics (T_{diag} and $T_{\text{off-diag}}$) the following two sets of test statistics can be defined:

$$T1 = \{t^2_{ii} \text{ for } 1 \leq i \leq p, T_{\text{off-diag}}\},$$

and

$$T2 = \{T_{\text{diag}}, T_{\text{off-diag}}\}.$$

The set $T1$ with all t_{ii}^2's and $T_{\text{off-diag}}$ as plotted statistics (*p+1* statistics) signals when one of the following occurs:

$$\begin{gathered}
\{t_{11}^2 < \chi^2_{n-1,(a_1/2)} \quad \text{or } t_{11}^2 > \chi^2_{n-1,1-(a_1/2)}\}, \\
\text{or} \\
\vdots \\
\text{or} \\
\{t_{pp}^2 < \chi^2_{n-1,(a_p/2)} \quad \text{or } t_{pp}^2 > \chi^2_{n-1,1-(a_p/2)}\}, \\
\text{or} \\
\{T_{\text{off-diag}} > \chi^2_{(p/2)(p-1),1-\alpha_{\text{off-diag}}}\}.
\end{gathered}$$

On the other hand, the set $T2$ with two plotted statistics (T_{diag} and $T_{\text{off-diag}}$) signals when one of the following occurs:

$$\begin{gathered}
\{T_{\text{diag}} < \chi^2_{(p/2)(2n-p-1),\left(\alpha_{\text{diag}}/2\right)}\}, \\
\text{or} \\
\{T_{\text{diag}} > \chi^2_{(p/2)(2n-p-1),1-(\alpha_{\text{diag}}/2)}\}, \\
\text{or} \\
\{T_{\text{off-diag}} > \chi^2_{(p/2)(p-1),1-\alpha_{\text{off-diag}}}\}.
\end{gathered}$$

By setting the overall type-I error for each set of statistics equal to α an $\alpha_{base} = 1-(1-\alpha)^{(1/p(n-1))}$ is defined. The α_i, $\alpha_{\text{off-diag}}$ and α_{diag} are defined as:

$$\begin{aligned}
a_i &= 1-(1-\alpha_{\text{base}})^{n-i}, \\
a_{\text{off-diag}} &= 1-(1-\alpha_{\text{base}})^{p(p-1)/2}, \\
a_{\text{diag}} &= 1-(1-\alpha_{\text{base}})^{p(2n-p-1)/2}.
\end{aligned}$$

2.7 VMIX: A *CC* Based on the *VMIX* Statistic

Quinino *et al.* [23] proposed a statistic for controlling the covariance matrix of a bivariate process with known means and variances. The *CC* is known as *VMIX CC* and the monitoring statistic *VMIX* is:

$$VMIX = \frac{\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} Y_i^2}{2n}.$$

The *CC* is defined by considering X^* and Y^* as two quality characteristics of interest with means μ_{X^*} and μ_{Y^*}, respectively. The variances are defined as

$\sigma^2_{X^*}$ and $\sigma^2_{Y^*}$ and the covariance is defined as $\sigma_{X^*Y^*}$. If all the parameters are known, the new variables can be defined as:

$$X_i = \frac{(X_i^* - \mu_{X^*})}{\sigma_{x^*}}$$

and

$$Y_i = \frac{(Z_t^* - \rho X_i)}{\sqrt{1-\rho^2}},$$

where $Z_i^* = \frac{(Y_i^* - \mu_{Y^*})}{\sigma_{Y^*}}$. When the process is in-control, X_t and Y_t follow the standardized normal distribution and become free of the correlation parameter ρ. The CC signals when $VMIX > UCL$. For determining the UCL, the following has been used:

$$P(VMIX \leq UCL) = P(T_X + T_Y \leq 2nUCL) = 1 - \alpha,$$

where $T_X = \sum_{i=1}^n X_i^2$ and $T_Y = \sum_{i=1}^n Y_i^2$. The statistic $T = T_X + T_Y$ follows a χ^2 distribution with $2n$ degrees of freedom so:

$$P(T \leq 2nUCL) = 1 - \alpha.$$

3 Simulation Study and Results

3.1 Simulation Study

Every case that can be considered in real life is special and differs with any other. A practitioner may deal with various problems that can occur during production. In the case that the process is out-of-control, the practitioner should be capable to recognize an out-of-control signal as soon as possible. In other words, he should monitor the process with the CC that best suits the problem. This section, deals with this problem. Various scenarios based on the available sample and possible shift have been considered.
Regarding the comparison of the various CCs, the control limits of the CCs were computed for achieving an in-control ARL equal to 200. The number of variables was set equal to $p = 2$. Furthermore, scenarios for different sample sizes have been considered with $n = 5$, 10, 20. In addition, the scenarios were made for simulating a process with mean vector $\boldsymbol{\mu} = (0, 0)$ and variance-covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \sigma_{11}\rho\sigma_{22} \\ \sigma_{11}\rho\sigma_{22} & 1 \end{bmatrix}$ with $\rho = -0.75, -0.3, 0$ and 0.5. Finally, the out-of-control ARL are compared for a shift in one or two variances and the shifts had the form $k\sigma^2$ with $k = 1, 1.1, 1.2, \ldots, 2$. In the plots, the volume of the shift and the $\ln(ARL)$ are presented. The number of simulations were set to be 10^5.

3.2 Results

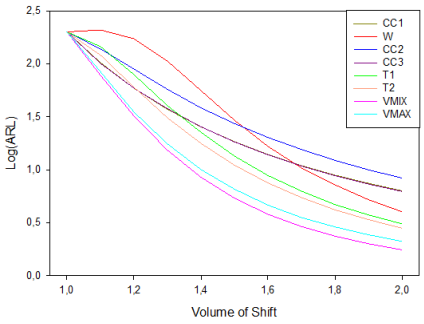
Scenario with $\rho = -0.75$ The first scenario assumes that the correlation between the variables is -0.75 meaning that the variables have a strong negative correlation. Figures 2(a), 2(c), and 2(e) show the shift in variability of **only one** variable for sample sizes equal to **5**, **10** and **20**, respectively. For a sample sizes n= **5**, **10**, **20** and shift in the dispersion of **both** variables, we have Figures 2(b), 2(d), and 2(f).

From Figure 1 it appears that the *VMIX CC* performs better regardless the sample size and the shift. A high negative correlation between the two variables indicates that the *VMIX CC* should be applied for monitoring the dispersion of the process. Figures 2(a), 2(c), 2(e) also show that the *VMAX CC* performs better for a shift in one variable regardless the volume of the shift. More specifically as the sample size increases the *VMAX CC* approximates the performance of the *VMIX CC*. *VMAX CC* can also be selected for monitoring the process if the sample size is small (n=5) when the shift occurs in both variables (see Figure 2(b)). On the contrary, *W CC* is the worst chart for monitoring the process especially if a shift in both variables takes place (Figures 2(b), 2(d), and 2(f)). In the case of a large shift ($1.7\sigma^2$) in one variance, *W CC* has similar performance to *VMIX* and *VMAX* for large sample sizes (n=10, 20) (Figures 2(c), 2(e)). $T2$ *CC* should be considered for shifts in both variables regardless the sample size (Figures 2(b), 2(d), and 2(f)) but preferably for shifts over $1.4\sigma^2$. The $T2$ *CC* can also be selected for shift in one variable with a small sample size (n=5) because it has the third best performance recorded (Figure 2(a)). $T1$ *CC* is able to detect a shift in one variable when the sample size is more than 10 (Figures 2(c), 2(d)). Finally, the $CC2$ chart has a moderate performance since it does not perform best with any specific sample size (Figures 2(a) - 2(f)).

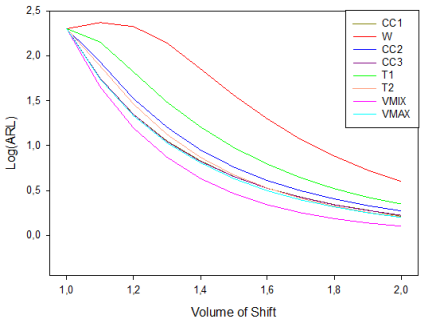
Scenario with $\rho = -0.30$ The second scenario assumes that the correlation between the variables is -0.30 meaning that the variables have a moderate negative correlation. Figures 3(a), 3(c), and 3(e) show the performance of the *CC*s for the case of a shift in the variance of **one** variable and for sample sizes 5, 10 and 20 respectively. Figures 3(b), 3(d), and 3(f) show for sample sizes 5, 10 and 20 the performance of the *CC*s when there is shift in the variances of **both** variables.

From Figure 2 it seems that the *VMAX CC* is the best for detecting a shift in only one variable regardless the sample size and the volume of the shift (Figures 3(a), 3(c), and 3(e)) while *VMIX CC* outperforms all *CC*s for shift in both variables regardless the sample size and the volume of the shift (Figures 3(b), 3(d), and 3(f)). In this scenario, the *W CC* performs really good only in big sample sizes ($n = 20$) and great shifts ($> 1.8\sigma^2$) in only one variable (Figure 3(e)). $T1$ *CC* in this case seems to have one of the best performances when it comes to a shift in one variable regardless the sample size (Figures 3(a), 3(c), and 3(e)). For sample sizes equal to 5, $T1$ should be preferred for shifts over

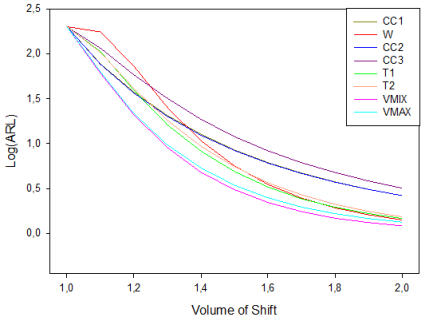
Fig. 1. CC Performance for $\rho = -0.75$



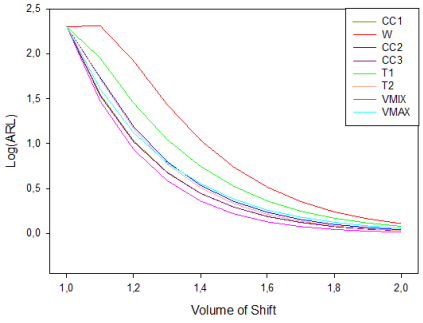
(a) $n = 5$ and shift in one variable



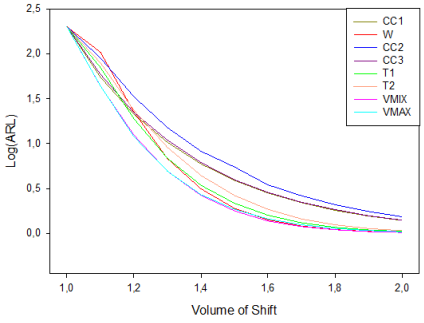
(b) $n = 5$ and shift in two variables



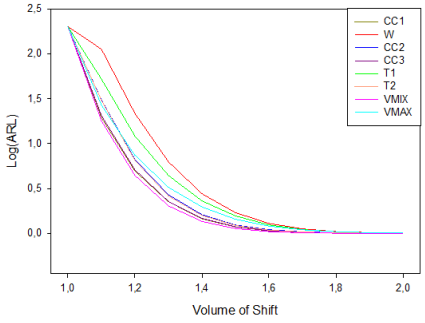
(c) $n = 10$ and shift in one variable



(d) $n = 10$ and shift in two variables

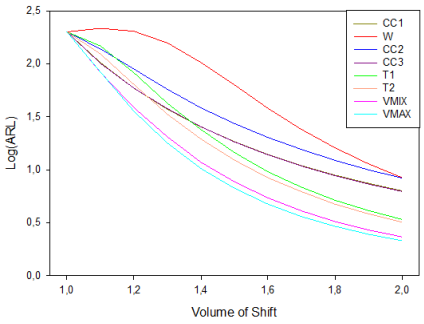


(e) $n = 20$ and shift in one variable

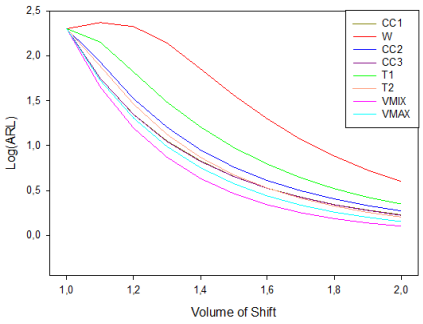


(f) $n = 20$ and shift in two variables

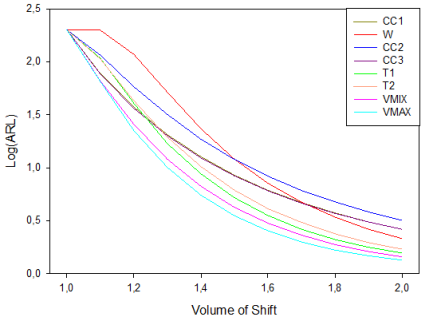
Fig. 2. CC Performance for $\rho = -0.30$



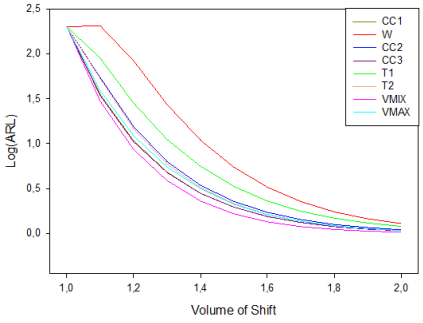
(a) $n = 5$ and shift in one variable



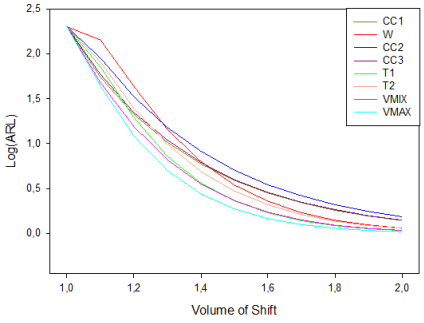
(b) $n = 5$ and shift in two variables



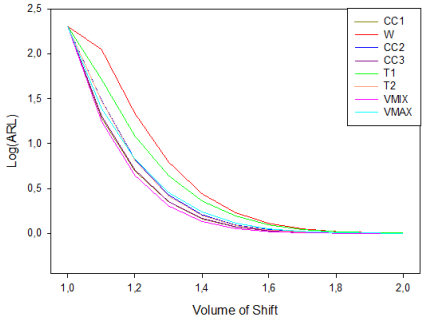
(c) $n = 10$ and shift in one variable



(d) $n = 10$ and shift in two variables



(e) $n = 20$ and shift in one variable



(f) $n = 20$ and shift in two variables

$1.4\sigma^2$ (Figure 3(a)), for sample size equal to 10, $T1$ should be preferred for shifts over $1.3\sigma^2$ (Figure 3(c)) and for big sample sizes ($n = 20$), $T1$ should be preferred for shifts over $1.2\sigma^2$ (Figure 3(e)). $T2$ CC should be considered for a shift in one variable regardless the sample size but for shifts over $1.2\sigma^2$ (Figures 3(a), 3(c), and 3(e)). For a shift in both variables, $T2$ should be preferred if the sample size is 5 and the shift is over $1.6\sigma^2$ (Figure 3(b)), for sample size 10 when the shift is over $1.5\sigma^2$ (Figure 3(d)) or for sample size of 20 if the shift is over $1.2\sigma^2$ (Figure 3(f)). $CC1$ and $CC3$ should be considered only for shift in both variables when sample size is 5 and the volume of the shift is less than $1.6\sigma^2$ (Figure 3(b)) or for sample sizes 10 and 20 regardless the shift (Figures 3(d), 3(d)). Again, $CC2$ is not exceptional in any case so it should not be considered (Figures 3(a) - 3(f)).

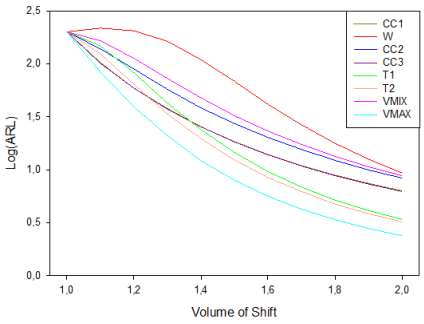
Scenario with $\rho = 0$ The third scenario assumes that the correlation between the variables is 0 meaning that the variables are uncorrelated. The ARL curves for shift in the variance of one and two variables can be seen in Figure 3.

It seems that $VMAX$ CC is the best chart for detecting shifts in any case (Figures 4(a) - 4(f)). $VMIX$ CC performs near optimally for a shift in the dispersion of both variables but for large sample sizes (n=10 or 20) (Figures 4(d), 4(f)). $T1$ can be chosen for detecting shifts over $1.2\sigma^2$ in one variable for sample size over 10 (Figures 4(e), 4(b)) and should not be preferred at all if the shift occurs in both variances simultaneously (Figures 4(b), 4(d), and 4(f)). In the case of shift in one variable, $T2$ can be considered when the shift is over $1.3\sigma^2$ (Figures 4(a), 4(c), and 4(e)). In the case of shift in both variances, $T2$ can be preferred when the shift is larger than $1.3\sigma^2$ and the sample size is up to 10 (Figures 4(b), 4(d)) or if the sample size is large ($n = 20$) regardless the volume of the shift (Figure 4(f)). $CC1$ and $CC3$ can be chosen for shift in both variances regardless the sample size (Figures 4(b), 4(d), and 4(f)).

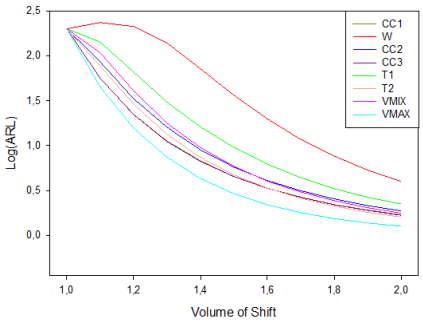
Scenario with $\rho = 0.5$ The fourth scenario assumes that the correlation between the variables is 0.5 meaning that the variables have a moderate positive correlation. Figure 4 shows the ARL curves for shift in the variance of one and two variables under the fourth scenario.

The correlation in the fourth scenario is moderately positive ($\rho = 0.5$). It seems that the $VMAX$ CC should be chosen for a shift in one variable regardless the sample size (Figures 4(a), 4(c), and 4(e)) or for shift in both variables when the sample size is small ($n = 5$) (in this case it has the second better performance) (Figure 4(b)). $VMIX$ CC has the best performance when the shift occurs in both variables regardless the sample size (Figures 4(b), 4(d), and 4(f)) and has the second better performance for a shift in one variable (Figures 4(a), 4(c), and 4(e)). $T1$ CC has a good performance when it comes to a shift in one variable but for sample sizes of 10 or 20 (Figures 4(c), 4(e)). It best performs for a shift over $1.3\sigma^2$ when the sample size is 10 (Figure 4(c)) and for a shift over $1.2\sigma^2$ when the sample is size equal to 20 (Figure 4(e)). $T2$ performs good for shifts

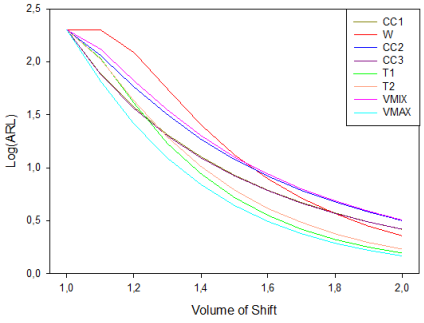
Fig. 3. CC Performance for $\rho = 0$



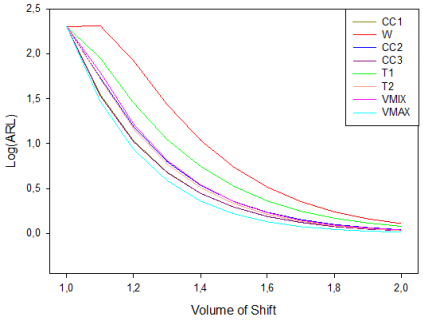
(a) $n = 5$ and shift in one variable



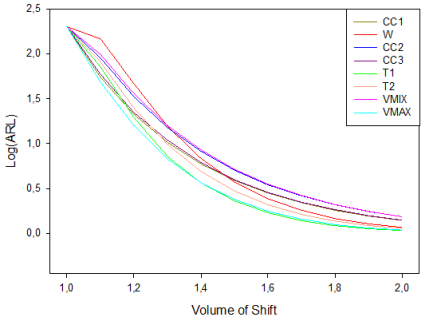
(b) $n = 5$ and shift in two variables



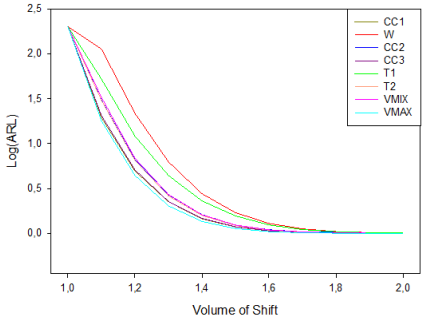
(c) $n = 10$ and shift in one variable



(d) $n = 10$ and shift in two variables

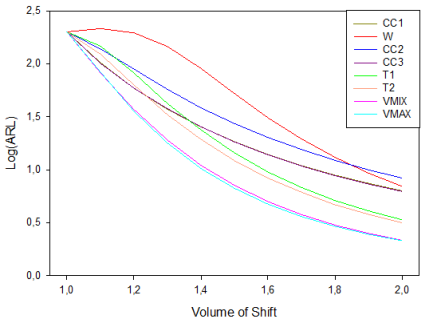


(e) $n = 20$ and shift in one variable

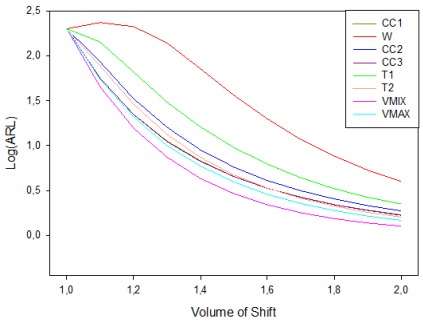


(f) $n = 20$ and shift in two variables

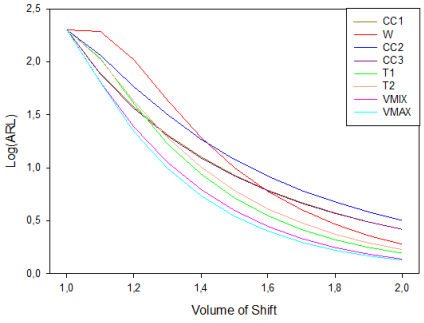
Fig. 4. CC Performance for $\rho = 0.5$



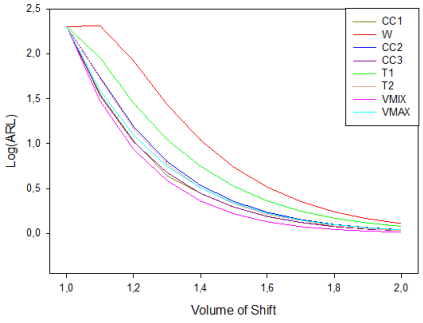
(a) $n = 5$ and shift in one variable



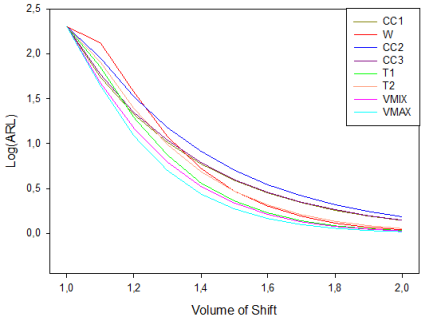
(b) $n = 5$ and shift in two variables



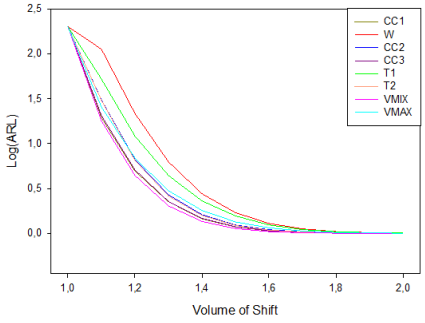
(c) $n = 10$ and shift in one variable



(d) $n = 10$ and shift in two variables



(e) $n = 20$ and shift in one variable



(f) $n = 20$ and shift in two variables

in both variables when the sample size is 5 and the shift is over $1.6\sigma^2$ (Figure 4(b)). Also when the sample size is 10 and the shift is over $1.5\sigma^2$ (Figure 4(d)) and finally when the sample size is 20 and the shift is over $1.2\sigma^2$ (Figure 4(f)). It also performs well for sample size of 5 and shift over $1.2\sigma^2$ in one variable (Figure 4(a)). Again $CC1$ and $CC3$ perform good for shifts in both variables regardless the sample size (Figures 4(b), 4(d), and 4(f)). Also for small shifts ($< 1.2\sigma^2$) in one variable regardless the sample size (Figures 4(a), 4(c), and 4(e)). Finally, there is no case in which $CC2$ should be preferred.

4 Discussion and further work

In this paper some bivariate and multivariate CCs were presented. Moreover, these CCs were compared for determining the most efficient CCs for any given scenario which involves a bivariate process.
It was clear that almost in any considered case, the *VMAX CC* by Machado and Costa [18] and the *VMIX CC* by Quinino *et al.* [23] had the best performance. Both charts are bivariate CCs and should be preferred if the process consists of 2 variables. From the remaining CCs, $T2$ developed by Hung and Chen [10] performs really good in large shifts either in one or both variables especially for big samples. $T1$ also developed by Hung and Chen [10] performs better for a shift in one variable and also for big sample size regardless the correlation between the variables. $CC1$ and $CC3$ perform good for shifts in both variables regardless the correlation between the variables. Finally, *W CC* seems to perform well when big shifts occur in one variable especially for high correlation between the variables.
While Jackson's [14] list is right on what information should be obtained from *MSPM*, it is probably incomplete and should be expanded by one information that should be provided to the researcher. The researcher should be in position to answer to the question "Am I using the optimal way to monitor the process?". The reason is really simple. If the best *CC* is not used for the scenario encountered then it is not sure for the researcher to know in any given time whether the process is in-control (first information from Jackson [14]) leading to ignorance. A large-scale study should be done with main objective to determine the best option for every scenario that can be encountered. This painful study should include scenarios for different number of variables, different sample sizes, different correlations and different shifts in variances. Also scenarios with different in control ARLs should be included.

References

1. F.B. Alt. Multivariate quality control, *Encyclopedia of Statistical Sciences*, 6, 110–122, 1985.
2. F.B. Alt and N.D. Smith. Multivariate process control. *Handbook of Statistics*, 7, 333-351,1988.
3. T.W. Anderson. *An introduction to multivariate statistical analysis*, John Wiley and Sons, New York, 1958.

4. Cdt. Benoît, Note sur une méthode de résolution des équations normales provenant de lápplication de la méthode des moindres carrés á un systéme d'équations linéaires en nombre inférieur á celui des inconnues, (Procédé du Commandant Cholesky), *Bulletin Géodésique*, 2, 67-77, 1924.
5. S. Bersimis, S. Psarakis and J. Panaretos. Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23, **5**, 517–543, 2007.
6. R.B. Crosier. Multivariate generalizations of cumulative sum quality control scheme, *Technometrics*, 30, 291–303, 1988.
7. D.M. Hawkins. Multivariate quality control based on regression adjusted variables, *Technometrics*, 33, 61–75, 1991.
8. D.M. Hawkins. Regression adjustment for variables in multivariate quality control, *JI. Quality Technology*, 25, 170–186, 1993.
9. H. Hotelling. Multivariate quality control. *Techniques of Statistical Analysis*, C.Eisenhart et. al. eds., 111–184, 1947.
10. H. Hung and A. Chen. Test of covariance changes without a large sample and its application to fault detection and classification, *Journal of Process Control*, 22, 1113–1121, 2012.
11. J.E. Jackson. Quality control methods for two related variables, *Industrial Quality Control*, 12, 2–6, 1956.
12. J.E. Jackson. Quality control methods for several related variables, *Technometrics*, 4, 359–377, 1959.
13. J.E. Jackson. Multivariate quality control, *Communication in statistics- Theory and Methods*, 14, 2657–2688, 1985.
14. J.E. Jackson. *A user's guide to principal components*, John Wiley and Sons, New York, 1991.
15. M.V. Koutras, S. Bersimis and D. Antzoulakos. Improving the performance of the chi-square control chart via runs ruls. *Methodology and Computing in Applied Probability*, 8, **3**, 409–426, 2006.
16. C.A. Lowry and D.C. Montgomery. A review of multivariate control charts. *IIE Transactions*, 27, **6**, 800–810, 1995.
17. J.M. Lucas and M.S. Saccucci. Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics*, 32, 1-?16, 1990.
18. M.A. Machado and A.F. Costa. The double sampling and the EWMA charts based on the sample variances, *International Journal of Production Economics*, 114, 134–148, 2008.
19. P.E. Maravelakis, S. Bersimis, S. Panaretos and S. Psarakis. Identifying the out of control variable in a multivariate control chart. *Communication in Statistics-Theory and Methods*, 31, **12**, 2391–2408, 2002.
20. P.E. Maravelakis and S. Bersimis. The use of Andrews curves for detecting the out-of-control variables when a multivariate control chart signals. *Statistical Papers*, 50, **1**, 51–65, 2009.
21. D.C. Montgomery and H.M. Wadsworth. Some techniques for multivariate quality control applications, *ASQC Technical Conference Transactions*, Washington DC, 427–435, 1972.
22. J.J. Pignatiello Jr and G.C. Runger. Comparison of multivariate CUSUM charts, *JI. Quality Technology*, 22, 173–186, 1990.
23. R. Quinino, A. Costa and L. Lee Ho. A single statistic for monitoring the Covariance matrix of bivariate processes, *Quality Engineering*, 24, 423–430, 2012.
24. S.W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1, **3**, 239-250, 1959.

25. N.D. Tracy, J.C. Young and R.L. Mason. Multivariate control charts for individual observations. *JI. Quality Technology*, 24, 88–95, 1992.
26. J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 32–52, 1928.

Second Order Asymptotic Expansion for Pricing European Options in a Model with Two Stochastic Volatilities

Betuel Canhanga[1], Anatoliy Malyarenko[2], Ying Ni[2], and Sergei Silvestrov[2]

[1] Faculty of Sciences, Department of Mathematics and Computer Sciences, Eduardo Mondlane University, Box 257, Maputo, Mozambique (e-mail: betuel.canhanga@mdh.se)

[2] Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, SE-721 23 Västerås, Sweden (e-mail: anatoliy.malyarenko@mdh.se, ying.ni@mdh.se, sergei.silvestrov@mdh.se)

Abstract. Asset price processes with stochastic volatilities have been actively used by researchers in financial mathematics for valuing derivative securities. This type of models allows characterizing the uncertainty of the volatility variable in the asset price process in financial markets. In a recent paper Chiarella and Ziveyi [3] analyzed a model with two stochastic volatility variables of mean reversion type with one variable changing fast and the other changing slowly. They used method of characteristics to solve the obtained partial differential equation and determine the price of an American option. Fouque et al [6], [7] and [8] presented also a similar model in which the volatility of the underlying asset is governed by two diffusion processes which are not of mean reversion type. They developed a first-order asymptotic expansion for the European option price via a perturbation method.

In this paper we consider the model given in Chiarella and Ziveyi [3]. Instead of pricing American options we price European options by generalizing the techniques presented in Fouque et al [6], [7] and [8] to a more complex model with mean reverting stochastic volatility factors. We analyse both regular and singular perturbations to obtain an asymptotic expansion up to second order which can serve as an approximation for the price of non-path-dependent European options. Similar work is done in authors earlier work Canhanga et al [1] in which a first-order asymptotic expansion has been developed. Involving the second order terms has the advantage of capturing more accurately the effects of volatility smile and skew on the option pricing. Analytical approximation formula for pricing European Option is presented.

Keywords: Financial market, Mean reversion volatility, asymptotic expansion, Stochastic Volatilities, Regular perturbation, Singular perturbation, European option.

1 Introduction

The idea that in the financial market there is no free lunch suggests that one should not get for free rights without obligations. An option is a financial contract that gives the buyer the right but not the obligation to buy, in case of a call option, or sell, in case of a put option, an underlying asset at the maturity time (for European options) or any time up to the maturity (for American options) at a predetermined *strike* price. Thus, there is an amount of money, namely *premium* or price of the option, that the buyer of a option

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

has to pay in order to acquire this right. It is clear that this price depends on the stochastic process followed by the price of the underlying asset.

Following the fairness principle that financial models must assume, from the beginning of history of financial mathematics up to now, there has been a lot of efforts to develop models that capture all characteristics of asset price behaviors. The well-known Black-Scholes-Merton model values options based on the geometric Brownian motion for the asset price process under constant volatility. With a closed-form and easy-to-use option pricing formula this model has the weakness of not considering the variability through time for the volatility variable. Later Cox and Ross [15] introduced *constant elasticity of variance model* which did not consider a proper stochastic process for the volatility but instead, introduced a constant to characterize the changes of the volatility variable according to the increasing or decreasing of the underlying asset price. Following ideas from Vasicek [5] and Cox and Ross [15], Heston [2] brought a solution to the Black-Scholes option pricing model extended to the setting of non-constant and non-deterministic volatility. Heston [2] considered in his model a single factor volatility of mean reversion type in order to confine volatility distribution to realistic level. The mean reversion property of the the volatility process is justified by the long term equilibrium observed in real life movement of volatility variable of an underlying asset. He presented a semi-analytical formula for pricing European options.

Afterwards numerous researchers have increasingly realized that asset prices are influenced at least by two volatilities of mean reversion type with different speeds (see for example Fouque et al [8] and LeBaron [11]), thus the lightness from the single-factor Heston model was dropped and new research took place trying to capture all effects on the volatilities behavior to the option pricing.

One of this research was done by Christoffersen et al [12]. They improved the single factor Heston model by considering the variance of the underlying asset price as the sum of two uncorrelated stochastic variances of mean-reverting type, which relaxed the assumption of a single volatility factor in the Heston model. The volatility factors, more accurately the variance rate processes, have different reverting rates with one factor reverting fast and the other one reverting slowly. Using the method of characteristics and Fourier inversion they determined prices for European call options. These results were extended by Chiarella and Ziveyi [3] for the American option pricing problem.

In the present paper we consider the aforementioned model with two stochastic volatility factors. Instead of American options we consider pricing European options. The fact that it has a fast and a slow changing volatilities makes the determination of the option price a very tedious process. We replace the method of characteristics by constructing an approximating solution using asymptotic expansion method. Such method was described in Fouque et al [6], [7] and [8] for a similar but simpler model of asset price process. In the authors earlier work (Canhanga et al [1]), singular and regular perturbation analysis was performed to develop an asymptotic expansion up to the first order which can serve as an approximation to European option prices. The perturbation parameters represent the rate of reverting for the two variance rate processes. The present paper follows this line of research and perform a second-order singular and

regular perturbation analysis to capture more accurately the skew and smile of volatilities. The obtained results on second order asymptotic expansion provide more accurate approximation to the options prices.

2 The Problem

Let $dW_i,\ i = \{1, 2, 3, 4\}$ be independent Wiener processes, we consider an asset process $(S_t)_{0\leq t\leq T}$ driven by the following stochastic differential equation

$$dS_t = \mu S_t dt + \sqrt{V_{1,t}} S_t\ dW_1 + \sqrt{V_{2,t}} S_t\ dW_2, \tag{1}$$

where μ is the expected return of the asset. Under the assumption $0 < \varepsilon << 1$, $0 < \delta << 1, |\rho_{13}| < 1, |\rho_{24}| < 1$, the finite variance rates $V_{1,t}$ and $V_{2,t}$ are stochastically governed by

$$\begin{aligned} dV_{1,t} &= \frac{1}{\varepsilon}(\theta_1 - V_{1,t})dt + \rho_{13}\sqrt{\frac{1}{\varepsilon}V_{1,t}}\ dW_1 + \sqrt{\frac{1}{\varepsilon}(1-\rho_{13}^2)V_{1,t}}\ dW_3, \\ dV_{2,t} &= \delta(\theta_2 - V_{2,t})dt + \rho_{24}\sqrt{\delta V_{2,t}}\ dW_2 + \sqrt{\delta(1-\rho_{24}^2)V_{2,t}}\ dW_4, \end{aligned} \tag{2}$$

where the parameters θ_1, θ_2 are reverting means and $\frac{1}{\varepsilon}$ and δ refer to the rates/speeds of mean-reversions. It can be shown that the parameter ρ_{13} (ρ_{24}) is the constant instantaneous correlation coefficient between S_t and $V_{1,t}$ (between S_t and $V_{2,t}$). Note that since $dW_i,\ i = \{1, 2, 3, 4\}$ are independent Wiener processes, the variance rate processes $V_{1,t}, V_{2,t}$ are uncorrelated. Since $0 < \varepsilon << 1$ and $0 < \delta << 1$, the variance rate $V_{1,t}$ will be changing fast on the direction of its long run mean θ_1, while the variance rate $V_{2,t}$ will change slowly on the direction of its long run mean θ_2. The speeds of changing for these two variance rates, $V_{1,t},\ V_{2,t}$, are given by $\frac{1}{\varepsilon}$ and δ respectively and the corresponding volatilities (volatilities of volatilities) are $\sqrt{\frac{1}{\epsilon}}$ and $\sqrt{\delta}$. We will hereafter call $V_{1,t}$ the fast volatility factor and $V_{2,t}$ the slow volatility factor.

Moreover the assumption $|\rho_{13}| < 1$ and $|\rho_{24}| < 1$ will make the system to fulfill the condition proposed in Feller [4], which guarantees that the stochastic differential equation for the underlying asset has an analytic solution under real world probability measure. It also guarantees that there exists an equivalent solution to the underlying asset price equation under risk neutral probability measure. Since our system is under real world probability measure we need to transform it to a risk neutral probability measure. With the same thought as Chiarella and Ziveye [3] we consider the market prices of risk $\lambda_1,\ \lambda_2$ associated to Wiener processes dW_1 and dW_2 respectively to be constants. The other two market prices of risk are given in terms of λ_1 and λ_2 by

$$\lambda_{3,t} = \frac{\lambda_1\sqrt{V_{1,t}}}{\sigma_1\sqrt{1-\rho_{13}^2}} \text{ and } \lambda_{4,t} = \frac{\lambda_2\sqrt{V_{2,t}}}{\sigma_2\sqrt{1-\rho_{24}^2}},$$

where σ_1 and σ_2 are the instantaneous volatilities of $V_{1,t}$ and $V_{2,t}$ per unit time respectively. The use of Girsanov theorem (see for example Kijima [5]) will

transform our system to a new system under risk neutral probability measure

$$\begin{aligned}
\mathrm{d}S_t &= (r-q)S_t\,\mathrm{d}t + \sqrt{V_{1,t}}S_t\,\mathrm{d}W_1^* + \sqrt{V_{2,t}}S_t\,\mathrm{d}W_2^*,\\
\mathrm{d}V_{1,t} &= \left(\frac{1}{\varepsilon}(\theta_1 - V_{1,t}) - \frac{1}{\sqrt{\varepsilon}}\lambda_{3,t}\sqrt{V_{1,t}(1-\rho_{13}^2)}\right)\mathrm{d}t + \frac{1}{\sqrt{\varepsilon}}\sqrt{V_{1,t}}\rho_{13}\,\mathrm{d}W_1^*\\
&\quad + \frac{1}{\sqrt{\varepsilon}}\sqrt{V_{1,t}(1-\rho_{13}^2)}\,\mathrm{d}W_3^*,\\
\mathrm{d}V_{2,t} &= \left(\delta(\theta_2 - V_{2,t}) - \sqrt{\delta}\lambda_{4,t}\sqrt{V_{2,t}(1-\rho_{24}^2)}\right)\mathrm{d}t + \sqrt{\delta}\sqrt{V_{2,t}}\rho_{24}\,\mathrm{d}W_2^*\\
&\quad + \sqrt{\delta}\sqrt{V_{2,t}(1-\rho_{24}^2)}\,\mathrm{d}W_4^*,
\end{aligned} \tag{3}$$

where $dW_i^*,\ i = \{1,2,3,4\}$ are the Wiener processes under risk neutral probability measure. Details of this transformation and computation to the new system under risk neutral probability measure can be seen in Canhanga et al [1]. Consider an European option on an underlying asset governed by model (3) with maturity T, and a given payoff function $h(S_T)$ when S_T is the underlying price at maturity time T. Let $U(t,S,v_1,v_2)$ denotes the options price at time $t<T$ when the underlying has price S and the spot fast volatility factor and slow volatility factor are v_1 and v_2 respectively.

The option price $U(t,S,v_1,v_2)$ is then the expected payoff under risk-neutral probability measure discounted to time t, i.e.

$$U(t,S,v_1,v_2) = e^{-r\tau}E^*[h(S_T)|S,v_1,v_2], \quad \tau := T-t \tag{4}$$

For the system of stochastic differential equation given in this paper, the calculation of the above expected value will require analysis of many parameters which makes it difficult. Andersen and Piterbarg [13] proves that the option price given by (4) is equal to the unique solution of the following parabolic partial differential equation

$$\begin{aligned}
rU - \tfrac{\partial U}{\partial t} &= (r-q)S\tfrac{\partial U}{\partial S} + \left[\tfrac{1}{\varepsilon}(\theta_1 - v_1) - \lambda_1 v_1\right]\tfrac{\partial U}{\partial v_1} + [\delta(\theta_2 - v_2) - \lambda_2 v_2]\tfrac{\partial U}{\partial v_2} +\\
&\quad \tfrac{1}{2}\left[(v_1+v_2)S^2\tfrac{\partial^2 U}{\partial S^2} + \tfrac{1}{\varepsilon}v_1\tfrac{\partial^2 U}{\partial v_1^2} + \delta v_2\tfrac{\partial^2 U}{\partial v_2^2}\right] +\\
&\quad \tfrac{1}{\sqrt{\varepsilon}}\rho_{13}Sv_1\tfrac{\partial^2 U}{\partial S\partial v_1} + \sqrt{\delta}\rho_{24}Sv_2\tfrac{\partial^2 U}{\partial S\partial v_2},
\end{aligned} \tag{5}$$

with boundary conditions $U(T,S,v_1,v_2) = h(S_T)$.

Thus, instead of the difficult computation involved in using formula (4), we will search for the option price via the approximate solution of (5) using asymptotic expansion method, simplifying even more the calculation and using much less parameters than those imposed by calculating expected values. On the asymptotic expansion procedure, the results become asymptotically better when we increase the order of expansion. In our case we have extended the order of asymptotic expansion to the second order, expecting that our solution will be more adjusted to the real values of the option price. We will capture

the skewness and smiles of volatilities during the life time of the option and it will make the option price to express closely the real markets situation.

Let us define the following operators

$$\mathcal{L}_2 = \frac{\partial}{\partial t} + (r-q)S\frac{\partial}{\partial S} + \frac{1}{2}(v_1+v_2)S^2\frac{\partial^2}{\partial S^2} - r - \lambda_1 v_1\frac{\partial}{\partial v_1} - \lambda_2 v_2\frac{\partial}{\partial v_2},$$
$$\mathcal{L}_1 = \rho_{13}Sv_1\frac{\partial^2}{\partial S\partial v_1},$$
$$\mathcal{L}_0 = (\theta_1 - v_1)\frac{\partial}{\partial v_1} + \frac{1}{2}v_1\frac{\partial^2}{\partial v_1^2},$$
$$\mathcal{M}_1 = \rho_{24}Sv_2\frac{\partial^2}{\partial S\partial v_2},$$
$$\mathcal{M}_2 = (\theta_2 - v_2)\frac{\partial}{\partial v_2} + \frac{1}{2}v_2\frac{\partial^2}{\partial v_2^2}.$$

and

$$\mathcal{L}^{\varepsilon} = \frac{1}{\varepsilon}\mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}}\mathcal{L}_1 + \mathcal{L}_2.$$

Equation (5) can be written for short as

$$\left(\mathcal{L}^{\varepsilon} + \sqrt{\delta}\mathcal{M}_1 + \delta\mathcal{M}_2\right)U = 0 \tag{6}$$

where

- $\frac{1}{\varepsilon}\mathcal{L}_0$ is the infinitesimal generator of the process $V_{1,t}$;
- $\mathcal{L}_1$ express the correlation between the asset price S and the volatility $V_{1,t}$;
- $\mathcal{L}_2$ is the multidimensional Black-Scholes operator presented in Conze et al [9] and Sin et al [14] with volatility level given by $f(v_1, v_2) = \sqrt{v_1} + \sqrt{v_2}$ and can be denoted as $\mathcal{L}_{BS}$;
- $\mathcal{M}_1$ express the correlation between the asset price and the volatility process $V_{2,t}$;
- $\delta\mathcal{M}_2$ is the infinitesimal generator of process $V_{2,t}$.

In equation (6) the operators assigned with ε are diverging in the small ε limit and the operators assigned to δ are converging to zero in the small δ. This drives us to a singular and a regular perturbation about the operator $\mathcal{L}_{BS}$ which is neither assigned to ε nor to δ. It is possible to present the single and regular perturbation at the same time, but for better explanation we will present the perturbation analysis stepwise starting from the regular to the singular perturbation. It could also be done in the reverse order without affecting the solution.

3 Regular perturbation

Consider the option price U as function of ε and δ. In the regular perturbation we expand operators and the solution $U^{\varepsilon,\delta}$ in (6) in terms of powers of $\sqrt{\delta}$. Let

us assume that our solution can be expressed in the following form

$$U = U^{\varepsilon,\delta} = \sum_{j\geq 0} \sqrt{\delta}^j U_j^\varepsilon = U_0^\varepsilon + \sqrt{\delta} U_1^\varepsilon + \delta U_2^\varepsilon + \delta\sqrt{\delta} U_3^\varepsilon + \cdots \tag{7}$$

where the omitted terms are of order $o(\delta\sqrt{\delta})$. Using this expansion in (6), collecting all terms with the same power of $\sqrt{\delta}$ and considering the fact that $0 < \delta << 1$, gives that equation (6) holds only if all coefficients of $\sqrt{\delta}^j, \quad j = \{0, 1, 2, \cdots\}$ equals zero with its final conditions. The first three lowest order components of the regular expansion will be

$$\begin{aligned} \mathcal{L}^\varepsilon U_0^\varepsilon = 0, \qquad & U_0^\varepsilon(T, S, v_1, v_2) \quad = h(S); \\ \mathcal{L}^\varepsilon U_1^\varepsilon + \mathcal{M}_1 U_0^\varepsilon = 0, \qquad & U_1^\varepsilon(T, S, v_1, v_2) \qquad = 0; \\ \mathcal{L}^\varepsilon U_2^\varepsilon + \mathcal{M}_1 U_1^\varepsilon + \mathcal{M}_2 U_0^\varepsilon = 0, \qquad & U_2^\varepsilon(T, S, v_1, v_2) \qquad = 0. \end{aligned} \tag{8}$$

We could produce infinitely many equations of this type. However only the first tree equations are needed since we are interested in performing the second-order asymptotic expansion. Indeed from the fourth equation that we would obtain, all contained terms would be of $o(\delta\sqrt{\delta})$) and hence not of our interest. In the regular perturbation we are looking for $U_0^\varepsilon,\ U_1^\varepsilon$ and U_2^ε. Then the singular perturbation analysis will be performed on each of these three solution-components to produce our approximation.

4 Singular Perturbation

Operators in (8) and $U_0^\varepsilon,\ U_1^\varepsilon,\ U_2^\varepsilon$ have to be submitted to a singular perturbation. We do it by expanding $U_j^\varepsilon,\ j = \{0,\ 1,\ 2\}$ in terms of powers of $\sqrt{\varepsilon}$ according to the formula

$$U_j^\varepsilon = \sum_{i\geq 0} \sqrt{\varepsilon}^i U_{j,i}. \tag{9}$$

We need to obtain all $U_{i,j}$ for $i + j \leq 2$ to determine the expansion for option price. Applying the above formula to (8) and solving them gives us the desired $U_{i,j}$. Putting the resulting expansions for $U_0^\varepsilon,\ U_1^\varepsilon,\ U_2^\varepsilon$ in (7), our asymptotic expansion for the option price will take the following form

$$U^{\varepsilon,\delta} = U_{0,0} + \sqrt{\epsilon} U_{1,0} + \epsilon U_{2,0} + \sqrt{\delta} U_{0,1} + \delta U_{0,2} + \sqrt{\delta\varepsilon} U_{1,1} + \cdots \tag{10}$$

4.1 Singular perturbation on $\mathcal{O}(1)$

Here we use (9) to do the singular perturbation on the first equation of (8)

$$U_0^\varepsilon = U_{0,0} + \sqrt{\epsilon} U_{1,0} + \epsilon U_{2,0} + \epsilon\sqrt{\epsilon} U_{3,0} + \cdots . \tag{11}$$

Let us for simplicity from now on denote $U_{0,0} = U_0$. Applying the expansion (11) in the first equation of (8), collecting terms with the same power of $\sqrt{\varepsilon}$ and using the fact that $0 < \varepsilon << 1$, the resulting equation will be attended only if

$$
\begin{aligned}
\mathcal{L}_0 U_0 = 0, \quad & U_0(T,S,v_1,v_2) = h(S_T); \\
\mathcal{L}_0 U_{1,0} + \mathcal{L}_1 U_0 = 0, \quad & U_{1,0}(T,S,v_1,v_2) = 0; \\
\mathcal{L}_0 U_{2,0} + \mathcal{L}_1 U_{1,0} + \mathcal{L}_2 U_0 = 0, \quad & U_{2,0}(T,S,v_1,v_2) = 0; \\
\mathcal{L}_0 U_{3,0} + \mathcal{L}_1 U_{2,0} + \mathcal{L}_2 U_{1,0} = 0, \quad & U_{3,0}(T,S,v_1,v_2) = 0; \\
\mathcal{L}_0 U_{4,0} + \mathcal{L}_1 U_{3,0} + \mathcal{L}_2 U_{2,0} = 0, \quad & U_{4,0}(T,S,v_1,v_2) = 0.
\end{aligned} \tag{12}
$$

- Operator $\mathcal{L}_0$ contains only derivatives on v_1 and is a second order linear ordinary differential operator. It can be solved and will have solutions with exponential form that will cause our system to diverge. It also contradicts the mean reversion approach that characterizes the variances in the system. For our system to converge we have to consider the trivial solution of this type of equation, i.e, $U_0 = U_0(t,S,v_2)$.
- Operator $\mathcal{L}_1$ contain derivative with respect to v_1, therefore, applying any $U_{i,j}(t,S,v_2)$ to $\mathcal{L}_1$ gives $\mathcal{L}_1 U_{i,j}(t,S,v_2) = 0; \qquad \forall i,j \in \mathbf{N} : \ 0 \leq i+j \leq 2;$
- If we have Poisson equation, the solvability condition imposes that the non-homogeneous part of the equation must be centered with respect to the invariant distribution of the process generated by $\mathcal{L}_0$. This is equivalent to saying that the average of the non-homogeneous source with respect to invariant distribution of the process V_1 must equal zero. More clearly, letting $\langle \cdot \rangle$ be the average with respect to invariant distribution $N(m, v^2)$ of the fast volatility factor process V_1. If we have $\mathcal{L}_0 \chi + z = 0$, then, the solvability condition says that

$$
\langle z \rangle = \frac{1}{v\sqrt{2\pi}} \int_{\mathcal{R}} z(v_1) e^{\frac{-(m-v_1)^2}{2v^2}} dv_1 = 0;
$$

- Operator $\mathcal{L}_2$ is the two-dimensional Black-Scholes operator meaning that if we apply $U_{i,j}$ to $\mathcal{L}_2$ and we realize that it equals zero, then $U_{i,j}$ is the Black-Scholes option price. We denote this price by $U_{BS(\overline{\sigma}(v_2))}$ with volatility given by $\overline{\sigma}(v_2)$ defined below.
- The averaged effective volatility $\overline{\sigma}(v_2)$ is defined as

$$
\overline{\sigma}^2(v_2) = \langle v_1 + v_2 \rangle = \int (v_1 + v_2) \Pi(dv_1)
$$

  where Π is the invariant distribution of the process $V_{1,t}$.

With this in mind, the first three equations in the system (12) gives

$$
U_0 = U_{BS}, \tag{13}
$$

where U_{BS} is the solution to the corresponding two-dimensional Black-Scholes model, when the volatility is given by $\overline{\sigma}(v_2)$. Before we compute the other components of the approximation, let us first introduce the notation

$$
D_k = S^k \frac{\partial^k}{\partial S^k}, \quad k = 1, 2, 3, \cdots
$$

and define an operator that will help on our calculation.

Suppose that the function $\phi(v_1, v_2)$ is the solution of

$$\mathcal{L}_0\phi(v_1, v_2) = f^2(v_1, v_2) - \sigma^2(v_2)$$

then

$$\mathcal{L}_2 - \langle\mathcal{L}_2\rangle = \frac{1}{2}\left[f(v_1, v_2) - \sigma^2(v_2)\right] D_2 \tag{14}$$

can be expressed as

$$\mathcal{L}_2 - \langle\mathcal{L}_2\rangle = \frac{1}{2}\mathcal{L}_0\phi(v_1, v_2)D_2$$

which implies that

$$\mathcal{L}_0^{-1}\left(\mathcal{L}_2 - \langle\mathcal{L}_2\rangle\right) = \frac{1}{2}\phi(v_1, v_2)D_2.$$

If we apply $\mathcal{L}_1$ operator on both side of the above equation we obtain

$$\mathcal{L}_1\mathcal{L}_0^{-1}\left(\mathcal{L}_2 - \langle\mathcal{L}_2\rangle\right) = \mathcal{L}_1\frac{1}{2}\phi(v_1, v_2)D_2$$

or

$$\langle\mathcal{L}_1\mathcal{L}_0^{-1}\left(\mathcal{L}_2 - \langle\mathcal{L}_2\rangle\right)\rangle = \left\langle\frac{1}{2}\rho_{13}v_1\frac{\partial\phi(v_1, v_2)}{\partial v_1}D_1D_2\right\rangle.$$

Then we define the operator

$$\mathcal{B}^\varepsilon = -\Upsilon_2^\varepsilon(v_2)D_1D_2 \tag{15}$$

where

$$\Upsilon_2^\varepsilon(v_2) = -\frac{\sqrt{\varepsilon}\rho_{13}}{2}\left\langle v_1\frac{\partial\phi(v_1, v_2)}{\partial v_1}\right\rangle. \tag{16}$$

Now, considering that the fourth equation on the System (12) is a Poisson equation on $U_{3,0}$ and the solvability condition impose that

$$\langle\mathcal{L}_1U_{2,0}\rangle + \langle\mathcal{L}_2\rangle U_{1,0} = 0; \tag{17}$$

together with the fact that $\mathcal{L}_1U_{1,0} = 0$ the third equation in the system (12) will be transformed into

$$\mathcal{L}_0U_{2,0} + \mathcal{L}_2U_0 = 0$$

which is again Poisson equation; its solvability condition will impose that $\langle\mathcal{L}_2U_0\rangle = 0$. On the other hand

$$\mathcal{L}_0U_{2,0} = -\mathcal{L}_2U_0$$

implies that

$$U_{2,0} = -\mathcal{L}_0^{-1}\left[\mathcal{L}_2 - \langle\mathcal{L}_2\rangle\right]U_0. \tag{18}$$

Using the above definition of $U_{2,0}$ in equation (17) we obtain

$$\langle\mathcal{L}_2\rangle U_{1,0} = \mathcal{B}^\varepsilon U_0 \tag{19}$$

for the operator $\mathcal{B}^\varepsilon$ defined in equation (15). In case we consider European options, the equation (19) can be expressed as

$$\mathcal{L}_{BS}(\sigma(v_2))U_{1,0}(t,s,v_2)^\varepsilon = \mathcal{B}^\varepsilon U_{BS}; \quad U^\varepsilon_{1,0}(T,s,v_2) = 0, \tag{20}$$

therefore

$$U^\varepsilon_{1,0} = -(T-t)\mathcal{B}^\varepsilon U_{BS}. \tag{21}$$

Using definitions of the operators presented in equation (18) one can compute $U_{2,0}$ and obtain

$$U_{2,0} = -\frac{1}{2}\phi D_2 U_0 + c_{2,0}(t,S,v_2), \tag{22}$$

where $c_{2,0}$ is constant with respect to v_2 which is obtained from the integration of $U_{2,0}$.

Computation of $c_{2,0}(t,S,v_1)$

To determine the constant $c_{2,0}$, we use last equation in (12) which is a Poisson equation, therefore

$$\langle \mathcal{L}_1 U_{3,0} + \mathcal{L}_2 U_{2,0}\rangle = 0 \Rightarrow \langle \mathcal{L}_1 U_{3,0}\rangle + \langle \mathcal{L}_2 U_{2,0}\rangle = 0.$$

We need to compute the two terms on the left side of the last equation,

$$\langle \mathcal{L}_2 U_{2,0}\rangle = \langle \mathcal{L}_2 \ (-\frac{1}{2}\phi D_2 U_0 + c_{2,0}(t,S,v_1))\rangle.$$

The fact that the average is with respect to invariant distribution of $V_{1,t}$ allows us to rewrite

$$\langle \mathcal{L}_2 U_{2,0}\rangle = -\frac{1}{2}\langle \phi\mathcal{L}_2\rangle D_2 U_0 + \langle \mathcal{L}_2\rangle c_{2,0}(t,S,v_1).$$

Considering that U_0 does not depend on the fast volatility factor v_1

$$\langle \mathcal{L}_2 U_{2,0}\rangle = -\frac{1}{2}\langle\phi\rangle\left(\frac{\partial}{\partial t} + (r-q)S\frac{\partial}{\partial S} - \lambda_2 v_2 \frac{\partial}{\partial v_2} - r\right) D_2 U_0$$

$$-\frac{1}{4}\langle\phi(v_1+v_2)\rangle D_2^2 U_0 + \langle\mathcal{L}_2\rangle c_{2,0}(t,S,v_2)$$

which is

$$\langle \mathcal{L}_2 U_{2,0}\rangle = -\tfrac{1}{2}\langle\phi\rangle\left(\tfrac{\partial}{\partial t} + (r-q)S\tfrac{\partial}{\partial S} - \lambda_2 v_2 \tfrac{\partial}{\partial v_2} + \tfrac{1}{2}\langle(v_1+v_2)\rangle D_2 - r\right) D_2 U_0 -$$

$$\tfrac{1}{4}\left[\langle\phi(v_1+v_2)\rangle D_2 - \langle\phi\rangle\langle(v_1+v_2)\rangle D_2\right] D_2 U_0 + \langle\mathcal{L}_2\rangle c_{2,0}(t,S,v_2);$$

thus

$$\langle \mathcal{L}_2 U_{2,0}\rangle = \mathcal{A}_2(v_2) D_2 U_0 + \langle\mathcal{L}_2\rangle c_{2,0}(t,S,v_2) \tag{23}$$

where $\mathcal{A}_2(v_2)$ is defined as

$$\mathcal{A}_2(v_2) = -\frac{1}{4}\left(\langle\phi(v_1+v_2)\rangle D_2 - \langle\phi\rangle\langle(v_1+v_2)\rangle D_2\right).$$

The third equation in (12) implies that

$$\begin{aligned}
\mathcal{L}_0 U_{3,0} &= -\mathcal{L}_1 U_{2,0} + \langle\mathcal{L}_1 U_{2,0}\rangle - \mathcal{L}_2 U_{1,0} - \langle\mathcal{L}_1 U_{2,0}\rangle \\
&= -\mathcal{L}_1 U_{2,0} + \langle\mathcal{L}_1 U_{2,0}\rangle - (\mathcal{L}_2 - \langle\mathcal{L}_2\rangle) U_{1,0} \\
&= -\mathcal{L}_1\left(-\frac{1}{2}\phi D_2 U_0 + c_{2,0}(t,S,v_2)\right) + \left\langle \mathcal{L}_1\left(-\frac{1}{2}\phi D_2 U_0 + c_{2,0}(t,S,v_2)\right)\right\rangle - \frac{1}{2}\mathcal{L}_0\phi D_2 U_{1,0} \\
&= -\frac{1}{2}\rho_{13}\left(v_1\phi\frac{\partial}{\partial v_1} - \left\langle v_1\phi\frac{\partial}{\partial v_1}\right\rangle\right) D_1 D_2 U_0 - \frac{1}{2}\mathcal{L}_0\phi D_2 U_{1,0} \\
&= \frac{1}{2}\rho_{13}\mathcal{L}_0\psi D_1 D_2 U_0 - \frac{1}{2}\mathcal{L}_0\phi D_2 U_{1,0}
\end{aligned}$$

where ϕ and ψ are functions that satisfies

$$\frac{1}{2}\mathcal{L}_0\phi(v_1,v_2)D_2 = \mathcal{L}_2 - \langle\mathcal{L}_2\rangle;$$

$$\mathcal{L}_0\psi = v_1\frac{\partial}{\partial v_1}\phi - \langle v_1\frac{\partial}{\partial v_1}\phi\rangle.$$

$$U_{3,0} = \frac{1}{2}\rho_{13}\psi D_1 D_2 U_0 - \frac{1}{2}\phi D_2 U_{1,0} + c_{3,0}$$

where $c_{3,0}$ is a constant on v_1.

Now since we know $U_{3,0}$, we are ready to compute the average with respect to the invariant distribution of $V_{1,t}$ for $\mathcal{L}_1 U_{3,0}$, i.e.

$$\begin{aligned}
\langle\mathcal{L}_1 U_{3,0}\rangle &= \left\langle \rho_{13} D_1 v_1 \frac{\partial}{\partial v_1}\left(\frac{1}{2}\rho_{13}\psi D_1 D_2 U_0 - \frac{1}{2}\phi D_2 U_{1,0} + c_{3,0}\right)\right\rangle \\
&= \left\langle \rho_{13} D_1 v_1 \frac{\partial}{\partial v_1}\left(\frac{1}{2}\rho_{13}\psi D_1 D_2 U_0 - \frac{1}{2}\phi D_2 U_{1,0}\right)\right\rangle \\
&= \frac{1}{2}\rho_{13}^2\left\langle v_1\frac{\partial}{\partial v_1}\psi\right\rangle D_1^2 D_2 U_0 - \frac{1}{2}\rho_{13}\left\langle v_1\frac{\partial}{\partial v_1}\phi\right\rangle D_1 D_2 U_{1,0}.
\end{aligned}$$

If we make

$$\mathcal{A}_1(v_2) = \frac{1}{2}\rho_{13}^2\left\langle v_1\frac{\partial}{\partial v_1}\psi\right\rangle; \tag{24}$$

the average of $\mathcal{L}_1 U_{3,0}$ will be

$$\langle\mathcal{L}_1 U_{3,0}\rangle = \mathcal{A}_1(v_2) D_1^2 D_2 U_0 - \mathcal{B}^\varepsilon U_{1,0}. \tag{25}$$

Combining (25) with (16) gives

$$\langle\mathcal{L}_2\rangle c_{2,0}(t,S,v_2) = -\mathcal{A}_1(v_2) D_1^2 D_2 U_0 + \mathcal{B}^\varepsilon U_{1,0} - \mathcal{A}_2(v_2) D_2 U_0.$$

Lemma 2.1 from Fouque et al [8] together with the fact that

$$\langle U_{2,0}(T,S,v_2)\rangle = 0 \quad \text{and} \quad \langle \phi(\cdot,v_2)\rangle = 0$$

implies that

$$\langle \mathcal{L}_2\rangle c_{2,0}(t,S,v_2) = \mathcal{A}_3(v_2)\langle \mathcal{L}_2\rangle \tau U_{BS} + \left(\mathcal{B}^{\varepsilon}\right)^2 \langle \mathcal{L}_2\rangle \frac{\tau^2}{2} U_{BS}$$

for

$$\mathcal{A}_3(v_2) = \mathcal{A}_1(v_2)D_1^2 D_2 + \mathcal{A}_2(v_2)D_2$$

and

$$c_{2,0}(t,S,v_2) = \left(\tau \mathcal{A}_3(v_2) + \frac{\tau^2}{2}\left(\mathcal{B}^{\varepsilon}\right)^2\right) U_{BS}. \tag{26}$$

4.2 Singular perturbation on $\mathcal{O}(\sqrt{\delta})$

So far we have determined the first three of the six terms in expansion (10). We are seeking to find the next three terms with coefficients $U_{0,1}, U_{0,2}$ and $U_{1,1}$. Let us compute first the coefficients $U_{0,1}$ and $U_{1,1}$. This is possible by using the expansion of U_1 according to formula (9), i.e

$$U_1^{\varepsilon} = U_{0,1} + \sqrt{\varepsilon}U_{1,1} + \varepsilon U_{2,1} + \cdots . \tag{27}$$

Applying (27) in the second equation of the system (8) together with formula (9) gives

$$\mathcal{L}^{\varepsilon}\left(U_{0,1} + \sqrt{\varepsilon}U_{1,1} + \varepsilon U_{2,1} + \cdots\right) + \mathcal{M}_1\left(U_{0,1} + \sqrt{\varepsilon}U_{1,1} + \varepsilon U_{2,1} + \cdots\right) = 0.$$

Collecting all terms with the same power of $\sqrt{\varepsilon}$ and again, considering the fact that our fast reverting speed is never zero will imply that, in the obtained equation, all coefficients of $\sqrt{\varepsilon}^i$ with its final conditions, must equal zero; i.e

$$\begin{aligned}
\mathcal{L}_0 U_{0,1} &= 0, & U_{0,1}(T,S,v_1,v_2) &= 0;\\
\mathcal{L}_1 U_{0,1} + \mathcal{L}_0 U_{1,1} &= 0, & U_{1,1}(T,S,v_1,v_2) &= 0;\\
\mathcal{L}_2 U_{0,1} + \mathcal{L}_1 U_{1,1} + \mathcal{L}_0 U_{2,1} + \mathcal{M}_1 U_0 &= 0, & U_{2,1}(T,S,v_1,v_2) &= 0;\\
\mathcal{L}_2 U_{1,1} + \mathcal{L}_1 U_{2,1} + \mathcal{L}_0 U_{3,1} + \mathcal{M}_1 U_{1,0} &= 0, & U_{3,1}(T,S,v_1,v_2) &= 0.
\end{aligned} \tag{28}$$

We constructed the above system of partial differential equation only with the four equations, because the other equations that we could obtain, do not contain any of the six terms of our expansion (10), i.e, terms with coefficients $U_{i,j}, \quad 0 \le i+j \le 2$. The values of U_0 and $U_{1,0}$ are already computed and given in (13) and (21). From the first equation in (28), and the same reasons presented in the system (12) for U_0, the component $U_{0,1}$ will depend on the time t, the underlying asset price and the slow changing volatility factor v_2, i.e. $U_{0,1} = U_{0,1}(t,S,v_2)$. The fact that $\mathcal{L}_1$ contains derivatives with respect to the fast changing volatility factor v_1 implies that $\mathcal{L}_0 U_{1,1} = 0$ which also for

the convergence reasons imply that $U_{1,1} = U_{1,1}(t, S, v_2)$. These conditions will transform (28) into

$$\begin{aligned} \mathcal{L}_0 U_{2,1} + \mathcal{L}_2 U_{0,1} + \mathcal{M}_1 U_0 = 0, \qquad & U_{2,1}(T, S, v_1, v_2) = 0; \\ \mathcal{L}_0 U_{3,1} + \mathcal{L}_2 U_{1,1} + \mathcal{L}_1 U_{2,1} + \mathcal{M}_1 U_{1,0} = 0, \qquad & U_{3,1}(T, S, v_1, v_2) = 0. \end{aligned} \tag{29}$$

We have two Poisson differential equations, respectively for $U_{2,1}$ and $U_{3,1}$ with respect to the invariant distribution of the process $V_{1,t}$. The solvability conditions of Poisson equation impose that the non-homogeneous part of it must be in the null complement of operator $\mathcal{L}_0$, therefore

$$\langle \mathcal{L}_2 U_{0,1} + \mathcal{M}_1 U_0 \rangle = 0, \tag{30}$$

and

$$\langle \mathcal{L}_2 U_{1,1} + \mathcal{L}_1 U_{2,1} + \mathcal{M}_1 U_{1,0} \rangle = 0. \tag{31}$$

Since both $U_{0,1}$ and U_0 do not depends on v_1 and the averaging procedure $\langle \cdot \rangle$ consists of integration only with respect to v_1, equation (30) can be written as

$$\langle \mathcal{L}_2 \rangle U_{0,1}^{\delta} = -\sqrt{\delta} \langle \mathcal{M}_1 \rangle U_0,$$

where the right part is

$$\begin{aligned} -\langle \mathcal{M}_1 \rangle U_0 &= -\sqrt{\delta} \left\langle \rho_{24} S v_2 \frac{\partial^2}{\partial S \partial v_2} \right\rangle U_0 \\ &= -\sqrt{\delta} \rho_{24} \langle v_2 \rangle D_1 \frac{\partial}{\partial v_2} U_0 \\ &= -\sqrt{\delta} \rho_{24} \langle v_2 \rangle D_1 \frac{\partial}{\partial \sigma(v_2)} \frac{\partial \sigma(v_2)}{\partial v_2} U_0 \\ &= -\sqrt{\delta} \rho_{24} \langle v_2 \rangle D_1 \sigma'(v_2) \frac{\partial}{\partial \sigma(v_2)} U_0 \\ &= -2 \mathcal{A}^{\delta} U_{BS} \end{aligned}$$

where

$$\mathcal{A}^{\delta} = \Theta_1^{\delta}(v_2) D_1 \frac{\partial}{\partial \sigma(v_2)}, \quad \Theta_1^{\delta}(v_2) = \frac{1}{2} \sqrt{\delta} \rho_{24} \langle v_2 \rangle \frac{\partial \sigma(v_2)}{\partial v_2}. \tag{32}$$

It follows then

$$\mathcal{L}_{BS} U_{0,1}^{\delta} = -2 \mathcal{A}^{\delta} U_{BS}; \quad U_{0,1}^{\delta}(T, s, v_2) = 0$$

which means that

$$U_{0,1}^{\delta} = (T - t) \mathcal{A}^{\delta} U_{BS} \tag{33}$$

Now we want to compute the factor $U_{1,1}$. From (31) the independence of $U_{1,0}$ and $U_{1,1}$ from v_1 allow to re-write it as

$$\langle \mathcal{L}_2 \rangle U_{1,1} = -\langle \mathcal{L}_1 U_{2,1} \rangle - \langle \mathcal{M}_1 \rangle U_{1,0}. \tag{34}$$

Since we do not have interest in $U_{2,1}$, we will have to express it in terms of components that are part of our approximation. Going back to the first equation in system (29) and using (30) yields

$$\begin{aligned}\mathcal{L}_0 U_{2,1} &= -\mathcal{L}_2 U_{0,1} - \mathcal{M}_1 U_0 \\ &= -\mathcal{L}_2 U_{0,1} - \mathcal{M}_1 U_0 + \langle \mathcal{L}_2 U_{0,1} + \mathcal{M}_1 U_0 \rangle \\ &= -(\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle) U_{0,1} - (\mathcal{M}_1 - \langle \mathcal{M}_1 \rangle) U_0.\end{aligned}$$

As in (14), we determine the difference between $\mathcal{M}_1$ and its average as follows

$$\begin{aligned}\mathcal{M}_1 - \langle \mathcal{M}_1 \rangle &= \rho_{24} S v_2 \frac{\partial^2}{\partial S \partial v_2} - \left\langle \rho_{24} S v_2 \frac{\partial^2}{\partial S \partial v_2} \right\rangle \\ &= \rho_{24}(v_2 - \langle v_2 \rangle) D_1 \frac{\partial}{\partial v_2}. \\ &= 0.\end{aligned}$$

On the other hand, the difference between the second-order Black-Scholes operator and its average was defined in (14), therefore

$$\begin{aligned}\mathcal{L}_0 U_{2,1} &= -\frac{1}{2}\left((v_1 + v_2) - \langle v_1 + v_2 \rangle\right) D_2 U_{0,1} \\ &= -\frac{1}{2}\mathcal{L}_0 \phi D_2 U_{0,1}\end{aligned}$$

or

$$U_{2,1} = -\frac{1}{2}\phi D_2 U_{0,1} + c_{2,1}(t, S, v_2).$$

Now we calculate

$$\begin{aligned}\langle \mathcal{L}_1 U_{2,1} \rangle &= \left\langle \mathcal{L}_1 \left(-\frac{1}{2}\phi D_2 U_{0,1} + c_{2,1}(t, S, v_2) \right) \right\rangle \\ &= -\left\langle \rho_{13} S v_1 \frac{\partial^2}{\partial S \partial v_1} \left(\frac{1}{2}\phi D_2 U_{0,1} + c_{2,1}(t, S, v_2) \right) \right\rangle \\ &= -\mathcal{B}^\varepsilon U_{0,1}\end{aligned}$$

that can be applied in (34)

$$\begin{aligned}\langle \mathcal{L}_2 \rangle U_{1,1} &= \frac{1}{2}\rho_{13} \left\langle v_1 \frac{\partial}{\partial v_1} \right\rangle D_1 D_2 U_{0,1} \\ &= \mathcal{B}^\varepsilon U_{0,1}.\end{aligned}$$

Since $U_{0,1}$ is defined in (33) we have

$$\begin{aligned}\langle \mathcal{L}_2 \rangle U_{1,1} &= \mathcal{B}^\varepsilon U_{0,1} \\ &= \mathcal{B}^\varepsilon \tau \mathcal{A}^\delta U_{BS} \\ &= -\frac{\tau^2}{3}\mathcal{B}^\varepsilon \Theta_1^\delta(v_2) \langle \mathcal{L}_2 \rangle \frac{\partial}{\partial \sigma} U_{BS}\end{aligned}$$

or

$$U_{1,1} = -\frac{\tau^2}{3}\mathcal{B}^\varepsilon \Theta_1^\delta(v_2) \frac{\partial}{\partial \sigma} U_{BS}. \tag{35}$$

4.3 Singular perturbation on $\mathcal{O}(\delta)$

To complete our approximation we need to do the singular perturbation for the δ component of the regular perturbation. We already have five coefficients for the expansion of our solution (10). It remains to calculate the $U_{0,2}$ component. We consider the last equation on (8). Expanding U_2^ε according to (9) and using expansion of U_1^ε and U_0^ε presented in (11) and (27) respectively gives

$$\begin{aligned}
\mathcal{L}^\varepsilon(U_{0,2}+\sqrt{\varepsilon}U_{1,2}+\varepsilon U_{2,2}+\varepsilon\sqrt{\varepsilon}U_{3,2}+\cdots)+\\
\mathcal{M}_1(U_{0,1}+\sqrt{\varepsilon}U_{1,1}+\varepsilon U_{2,1}+\varepsilon\sqrt{\varepsilon}U_{3,1}+\cdots)+\\
\mathcal{M}_2(U_0+\sqrt{\varepsilon}U_{1,0}+\varepsilon U_{2,0}+\varepsilon\sqrt{\varepsilon}U_{3,0}+\cdots)+\\
=0;\\
U_{i,2}(T,S,v_1,v_2)=0;\ \forall i=\{0,1,2,\cdots\}.
\end{aligned}$$

Collecting terms with the same power of $\sqrt{\varepsilon}$ our system can be written as

$$\begin{aligned}
\frac{1}{\varepsilon}\mathcal{L}_0U_{0,2}+\\
\frac{1}{\sqrt{\varepsilon}}\left(\mathcal{L}_1U_{0,2}+\mathcal{L}_0U_{1,2}\right)+\\
\left(\mathcal{L}_2U_{0,2}+\mathcal{L}_1U_{1,2}+\mathcal{L}_0U_{2,2}+\mathcal{M}_1U_{0,1}+\mathcal{M}_2U_0\right)+\\
\sqrt{\varepsilon}\left(\mathcal{L}_1U_{2,2}+\mathcal{L}_2U_{0,2}+\mathcal{L}_0U_{3,2}+\mathcal{M}_1U_{1,1}+\mathcal{M}_2U_{1,0}\right)+\\
\varepsilon\left(\mathcal{L}_0U_{4,2}+\mathcal{L}_1U_{3,2}+\mathcal{L}_2U_{2,2}+\mathcal{M}_1U_{2,1}+\mathcal{M}_2U_{2,0}\right)+\\
\cdots=0,
\end{aligned}$$

from where we generate the system of partial differential equations

$$\begin{aligned}
\mathcal{L}_0U_{0,2}=0, \qquad & U_{0,2}(T,S,v_1,v_2)=0;\\
\mathcal{L}_1U_{0,2}+\mathcal{L}_0U_{1,2}, \qquad & U_{1,2}(T,F,v_1,v_2)=0;\\
\mathcal{L}_2U_{0,2}+\mathcal{L}_1U_{1,2}+\mathcal{L}_0U_{2,2}+\mathcal{M}_1U_{0,1}+\mathcal{M}_2U_0, \qquad & U_{2,2}(T,S,v_1,v_2)=0.
\end{aligned} \tag{36}$$

From the first equation of the above system, choosing $U_{0,2}$ to be independent of the fast changing volatility will guarantee the convergence of the equation. This will imply that also $U_{1,2}$ is independent of the fast changing volatility factor and the last equation of our system will be transformed into

$$\mathcal{L}_2U_{0,2}+\mathcal{L}_0U_{2,2}+\mathcal{M}_1U_{0,1}+\mathcal{M}_2U_0=0$$

which is Poisson equation on $U_{2,2}$ with respect to variable v_1 . The solvability condition impose that

$$\langle\mathcal{L}_2U_{0,2}+\mathcal{M}_1U_{0,1}+\mathcal{M}_2U_0\rangle=0$$

where $\mathcal{M}_1,\ \mathcal{M}_1,\ U_{0,1},\ U_{0,2}$ and U_0 do not depends on v_1. Therefore

$$\begin{aligned}\langle\mathcal{L}_2\rangle U_{0,2} &= -\mathcal{M}_1 U_{0,1} - \mathcal{M}_2 U_0 \\ &= -\rho_{24} D_1 v_2 \frac{\partial}{\partial v_2} U_{0,1} - \left((\theta_2 - v_2)\frac{\partial}{\partial v_2} + \frac{1}{2} v_2 \frac{\partial^2}{\partial v_2^2}\right) U_0 \\ &= -2\Theta_1^{\delta}(v_2) D_1 \frac{\partial U_{0,1}}{\partial \sigma} - \left(\frac{(\theta_2 - v_2)\sigma'\partial}{\partial \sigma} + \frac{1}{2} v_2 \frac{\partial}{\partial v_2}\left(\frac{\sigma'\partial}{\partial \sigma}\right)\right) U_0 \\ &= \left[-2\Theta_1^{\delta}(v_2) D_1 \frac{\partial}{\partial \sigma}\left(\tau\mathcal{A}^{\delta}\right) - \mathcal{N}(v_2)\frac{\partial}{\partial \sigma} - \frac{1}{2} v_2 (\sigma')^2 \frac{\partial^2}{\partial \sigma^2}\right] U_{BS}\end{aligned}$$

for

$$\mathcal{N}(v_2) = (\theta_2 - v_2)\sigma' + \frac{1}{2} v_2 \sigma''.$$

The use of Lemma 2.1 from Fouque et al [8] gives

$$\begin{aligned}U_{0,2} = &\left[\tfrac{2}{3}\tau^2 \mathcal{A}^{\delta}\mathcal{A}_3'(v_2)\tfrac{\partial}{\partial\sigma} + \tfrac{\tau^2}{2}\mathcal{A}_3^2(v_2)\left(\tfrac{\partial^2}{\partial\sigma^2} + \tfrac{1}{3\sigma}\tfrac{\partial}{\partial\sigma}\right)\right] U_{BS} \\ &+ \left[\tfrac{\tau}{2}\mathcal{N}(v_2)\tfrac{\partial}{\partial\sigma} + \tfrac{\tau}{6} v_2 (\sigma')^2\left(\tfrac{\partial^2}{\partial\sigma^2} + \tfrac{1}{2\sigma}\tfrac{\partial}{\partial\sigma}\right)\right] U_{BS}.\end{aligned} \tag{37}$$

5 Pricing Formula For European Options

We now summarize the asymptotical results obtained in the previous sections into the following main theorem.

Theorem 1. *Consider an asset whose price evolves according to* (1) *and* (2) *where $V_{1,t}$ and $V_{2,t}$ are stochastic variance processes of mean reversion type. Consider also that the rates of reversion of the two variance processes are given by $1/\varepsilon$ and δ respectively where $0 < \varepsilon << 1$ and $0 < \delta << 1$. If $h(S_T)$ is the payoff of an European option on this asset with maturity time T, then the price of this option can be approximated by $U^{\varepsilon,\delta}$ given below.*

$$\begin{aligned}U^{\varepsilon,\delta} = &\, U_{BS} - \sqrt{\varepsilon}\tau\mathcal{B}^{\varepsilon} U_{BS} + \\ &\varepsilon\left[-\tfrac{1}{2}\phi D_2 + \left(\tau\mathcal{A}_3(v_2) + \tfrac{\tau^2}{2}\left(\mathcal{B}^{\varepsilon}\right)\right)\right] U_{BS} + \\ &\sqrt{\delta}\tau\mathcal{A}^{\delta} U_{BS} - \sqrt{\varepsilon\delta}\tfrac{\tau^2}{3}\mathcal{B}^{\varepsilon}\tfrac{\partial}{\partial\sigma} U_{BS} + \\ &\delta\left[\tfrac{2}{3}\tau^2\mathcal{A}^{\delta}\mathcal{A}_3'(v_2)\tfrac{\partial}{\partial\sigma} + \tfrac{\tau^2}{2}\mathcal{A}_3^2(v_2)\left(\tfrac{\partial^2}{\partial\sigma^2} + \tfrac{1}{3\sigma}\tfrac{\partial}{\partial\sigma}\right)\right] U_{BS} + \\ &\delta\left[\tfrac{\tau}{2}\mathcal{N}(v_2)\tfrac{\partial}{\partial\sigma} + \tfrac{\tau}{6} v_2 (\sigma')^2\left(\tfrac{\partial^2}{\partial\sigma^2} + \tfrac{1}{2\sigma}\tfrac{\partial}{\partial\sigma}\right)\right] U_{BS}\end{aligned}$$

which is the same as $U^{\varepsilon,\delta} \cong U_0 + \sqrt{\varepsilon}U_{1,0} + \varepsilon U_{2,0} + \sqrt{\delta}U_{0,1} + \sqrt{\varepsilon\delta}U_{1,1} + \delta U_{0,2}$ for $U_{i,j}$ given in (13), (21), (22), (33), (35) *and* (37). *The approximated option price converges to the Black-Scholes price when $\varepsilon \to 0$ and $\delta \to 0$.*

6 Acknowledgements

This work was partially supported by Swedish SIDA Foundation International Science Program. Betuel Canhanga thanks Division of Applied Mathematics, School of Education, Culture and Communication, Mlardalen University for creating excellent research and educational environment.

References

1. Canhanga B., Malyarenko, A., Ni, Y. and Silvestrov S. "Perturbation methods for pricing European options in a model with two stochastic volatilities". *3rd SMTDA Conference Proceedings.* 11-14 June 2014,Lisbon Porturgal, C. H. Skiadas (Ed) 489-500 (2014).
2. Heston, S.L. "A closed-form solution for options with stochastic volatility with applications to bonds and currency options". *Rev. Finan. Stud.* 6:327–343 (1993).
3. Chiarella, C, and Ziveyi, J. "American option pricing under two stochastic volatility processes". *J. Appl. Math. Comput.* 224:283–310 (2013).
4. Feller, W. "Two singular diffusion problems". *Ann. Math.* 54:173–182 (1951).
5. Kijima, M. "Stochastic Processes with Applications to Finance", Second Edition, Capman and Hall/CRC, (2013).
6. Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K. "Singular Perturbation in Option Pricing" (2011a).
7. Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K. "Multiscale stochastic volatility for equity, interest rate and credit derivatives", Cambridge University Press, Cambridge (2011b).
8. Fouque, J.-P., Lorig, M., Sircar, R. "Second order multiscale stochastic volatility asymptotics: stochastic terminal layer analysis and calibration", University of California (2014).
9. Conze, A., Lantos, N., and Pironneau, O. "The forward Kolmogorov equation for two dimenssional options"; University Pierre et Marie Curie; (2010).
10. John C. Hull "Options, Futures, and Other Derivatives (5th Edition)"; Pearson; (2009).
11. LeBaron, B. "Stochastic Volatility as a simple generator of apparent financial power laws and long memory." Quantitative Finance, 1(6):621-631, (2001).
12. Christoffersen, P.; Heston, S.; Jacobs, K. "The shape and term structure of the index option smirk: why multifactor stochastic volatility models work so well" Manage.Sci.55(2) 1914-1932; (2009).
13. L. Andersen and V. Piterbarg. "Interest rate modeling Volume 1 - foundations and vanilla models"; Atlantic Financial Press, (2010).
14. Sin, K., Sulaiman, J. and Mail, R. "Numerical Solution for 2D European Option Pricing Using Quarter-Sweep Modified Gauss-Seidel Method"; Journal of Mathematics and Statistics 8(1) 129-135 (2012).
15. Cox, J., Ross, S. "The valuation of Options for alternative stochastic processes"; Journal of financial economics 3:146-166, (2014).

Numerical Studies on Asymptotics of European Option under Multiscale Stochastic Volatility

Betuel Canhanga[1], Anatoliy Malyarenko[2], Jean-Paul Murara[3], Ying Ni[2], and Sergei Silvestrov[2]

[1] Faculty of Sciences, Department of Mathematics and Computer Sciences, Eduardo Mondlane University, Box 257, Maputo, Mozambique
(e-mail: betuel.canhanga@mdh.se)

[2] Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, SE-721 23 Västerås, Sweden
(e-mail: anatoliy.malyarenko@mdh.se, ying.ni@mdh.se, sergei.silvestrov@mdh.se)

[3] Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, SE-721 23 Västerås, Sweden, and
College of Science and Technology,School of Sciences, Department of Applied Mathematics, University of Rwanda, Kigali-Rwanda
(e-mail: jean-paul.murara@mdh.se)

Abstract. Multiscale stochastic volatilities models relax the constant volatility assumption from Black-Scholes option pricing model. Such model can capture the smile and skew of volatilities and therefore describe more accurately the movements of the trading prices. Christoffersen et al. [3] presented a model where the underlying price is governed by two volatility components, one changing fast and another changing slowly. Chiarella and Ziveyi [2] transformed Christoffersen's model and computed an approximate formula for pricing American options. They used Duhamel's principle to derive an integral form solution of the boundary value problem associated to the option price. Using method of characteristics, Fourier and Laplace transforms, they obtained with good accuracy the American options prices. In a previous research of the authors (Canhanga et al. [1]), a particular case of Chiarella and Ziveyi [2] model is used for pricing of European options. The novelty of this earlier work is to present an asymptotic expansion for the option price. The present paper provides experimental and numerical studies on investigating the accuracy of the approximation formulae given by this asymptotic expansion. We present also a procedure for calibrating the parameters produced by our first-order asymptotic approximation formulae. Our approximated option prices will be compared to the approximation obtained by Chiarella and Ziveyi [2].

Keywords: Financial market, Mean reversion volatility, asymptotic expansion, Stochastic Volatilities, Regular perturbation, Singular perturbation, European option.

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

1 Introduction

Consider an asset process given by S_t whose dynamic when $0 \leq t \leq T$ is described by the following system of stochastic differential equations

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_{1,t}} S_t \; dW_1 + \sqrt{V_{2,t}} S_t \; dW_2, \\ dV_{1,t} &= \frac{1}{\varepsilon}(\theta_1 - V_{1,t})dt + \rho_{13}\sqrt{\frac{1}{\varepsilon}V_{1,t}} \; dW_1 + \sqrt{\frac{1}{\varepsilon}(1-\rho_{13}^2)V_{1,t}} \; dW_3, \\ dV_{2,t} &= \delta(\theta_2 - V_{2,t})dt + \rho_{24}\sqrt{\delta V_{2,t}} \; dW_2 + \sqrt{\delta(1-\rho_{24}^2)V_{2,t}} \; dW_4, \end{aligned} \tag{1}$$

where $W_i, \; i = \{1,2,3,4\}$ are independent Wiener processes; μ is the expected return of the asset. θ_1, θ_2 are reverting means and $\frac{1}{\varepsilon}$, δ refer to the rates or speeds of mean-reversions. ρ_{13} and ρ_{24} are the constant instantaneous correlation coefficients between S_t and the finite uncorrelated variances rates $V_{1,t}$ and $V_{2,t}$ respectively. Assume that $0 < \varepsilon << 1$, $0 < \delta << 1, |\rho_{13}| < 1, |\rho_{24}| < 1$, and that the variance rates $V_{1,t}$ changes fast around its long run mean θ_1, with speed $\frac{1}{\varepsilon}$ while the variance rate $V_{2,t}$ changes slowly around its long run mean θ_2 with speed δ. We hereafter call $V_{1,t}$ the fast volatility factor and $V_{2,t}$ the slow volatility factor.

The assumption $|\rho_{13}| < 1$ and $|\rho_{24}| < 1$ makes the system to fulfill the conditions proposed in Feller [6], which guarantees that the stochastic differential equation for the underlying asset has a solution under the real-world probability measure and also guarantees that there exists an equivalent solution to the underlying asset price equation under risk neutral probability measure. In the authors earlier research, Canhanga et al. [1] considered an European option on an underlying asset governed by system (1) with maturity T, and a given payoff function $h(S_T)$ when S_T is the underlying price at maturity. Let us denote $U(t, S_t, v_1, v_2)$ as the options price at time $t < T$ when the underlying has spot price S_t, the spot fast and slow volatility factors are v_1 and v_2 respectively, Canhanga et al. [1] used Girsanov and Feynman-Kac theorems (presented for example in Kijima [9]) to express the European option price as the solution of the following partial differential equation

$$\left(\frac{1}{\varepsilon}\mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}}\mathcal{L}_1 + \mathcal{L}_2 + \sqrt{\delta}\mathcal{M}_1 + \delta\mathcal{M}_2\right)U = 0 \tag{2}$$

for

$$\mathcal{L}_0 = (\theta_1 - V_{1,t})\frac{\partial}{\partial V_{1,t}} + \frac{1}{2}V_{1,t}\frac{\partial^2}{\partial V_{1,t}^2}, \tag{3}$$

$$\mathcal{L}_1 = \rho_{13}S_t V_{1,t}\frac{\partial^2}{\partial S_t \partial V_{1,t}},$$

$$\mathcal{L}_2 = \frac{\partial}{\partial t} + (r-q)S_t\frac{\partial}{\partial S_t} + \frac{1}{2}(V_{1,t}+V_{2,t})S_t^2\frac{\partial^2}{\partial S_t^2} - r - \lambda_1 V_{1,t}\frac{\partial}{\partial V_{1,t}} - \lambda_2 V_{2,t}\frac{\partial}{\partial V_{2,t}},$$

$$\mathcal{M}_1 = \rho_{24}S_t V_{2,t}\frac{\partial^2}{\partial S_t \partial V_{2,t}},$$

$$\mathcal{M}_2 = (\theta_2 - V_{2,t})\frac{\partial}{\partial V_{2,t}} + \frac{1}{2}V_{2,t}\frac{\partial^2}{\partial V_{2,t}^2}.$$

Here r and q refer to the constant risk free interest rate and the continuous compounded dividend yield respectively. The market prices of risk λ_1 and λ_2 are given by

$$\lambda_1 = \frac{\mu - (r-q)}{2\sqrt{V_{1,t}}}; \quad \lambda_2 = \frac{\mu - (r-q)}{2\sqrt{V_{2,t}}}.$$

Assumption that the solution $U = U^{\varepsilon,\delta}$ can be expressed in the following form

$$U^{\varepsilon,\delta} = U_{0,0} + \sqrt{\delta}U_{0,1} + \sqrt{\varepsilon}U_{1,0} + \delta U_{0,2} + \varepsilon U_{2,0} + \cdots \tag{4}$$

leads to the asymptotic expansion method. Such method was described in Fouque et al [7] for a similar but simpler model of asset price process. After putting the expansion (4) into the partial differential equation (2), we obtain systems of differential equations solutions to which yield the unknown coefficients $U_{i,j}$ for expansion (4).

The first order approximation presented with details by Canhanga et al. [1] gives the European option price by

$$\begin{aligned}\widetilde{U}^{\varepsilon,\delta} &= U_{0,0} + U_{1,0}^{\varepsilon} + U_{0,1}^{\delta} \\ &= U_{BS} + (T-t)\left(\Theta^{\delta}(v_2)D_1\tfrac{\partial}{\partial\overline{\sigma}(v_2)} + \Upsilon^{\varepsilon}(v_2)D_1D_2\right)U_{BS}\end{aligned} \tag{5}$$

where $U_{0,0} = U_{BS}$. The notation U_{BS} stands for the approximate solution presented by Conze et al. [4] of the corresponding two dimensional Black-Scholes model and

$$\begin{aligned}U_{1,0} &= \sqrt{\varepsilon}(T-t)\Upsilon^{\varepsilon}(v_2)D_1D_2U_{BS}, \\ U_{0,1} &= \sqrt{\delta}(T-t)\Theta^{\delta}(v_2)D_1\tfrac{\partial\overline{\sigma}(v_2)}{\partial v_2}U_{BS},\end{aligned} \tag{6}$$

for

$$\Theta^{\delta}(v_2) = \frac{1}{2}\rho_{24}\langle v_2\rangle\frac{\partial\overline{\sigma}(v_2)}{\partial v_2}; \qquad \Upsilon^{\varepsilon}(v_2) = \frac{1}{2}\rho_{13}\langle v_1\frac{\partial\phi}{\partial v_1}\rangle; \quad D_i = x^i\frac{\partial^i}{\partial x^i}.$$

Here the brackets $\langle\cdot\rangle$ denotes the averaging over the invariant distribution Π of the process $V_{1,t}$, i.e.

$$\langle f(\cdot, v_2)\rangle = \int f(\cdot, v_2)\Pi(dv_1).$$

Using the above definition, the average effective volatility is represented by

$$\overline{\sigma}^2(v_2) = \int (v_1 + v_2)\Pi(dv_1).$$

The function ϕ which depend on v_1 and v_2 is a smooth function that solves the following equation

$$\mathcal{L}_0\phi(v_1, v_2) = f^2(v_1, v_2) - \overline{\sigma}^2(v_2).$$

To obtain U_{BS} we need to have $\overline{\sigma}(v_2)$ which is estimated from the spot market data. To obtain $U_{0,1}^{\delta}$ and $U_{1,0}^{\varepsilon}$ we need to have parameters $\Theta^{\delta}(v_2)$ and $\Upsilon^{\varepsilon}(v_2)$ which depends on the slow volatility v_2. These parameters must be calibrated to the observed implied volatility surface. The term U_{BS} contribute to the approximation independently of the two volatilities components. The influence of the fast time scale is expressed by $\Upsilon^{\varepsilon}(v_2)$. On the other hand, the slow time scale comes on the parameter $\Theta^{\delta}(v_2)$. The parameter $\Upsilon^{\varepsilon}(v_2)$ also express the influence of correlation between Wiener process W_1 and W_3 while $\Theta^{\delta}(v_2)$ does the same for Wiener processes W_2 and W_4.

In this paper we study the accuracy of the approximation formula (5). We start with calibrating the parameters needed for our first order approximation formula (5) using a procedure well explained by Fouque et al. [7]. Then we compute European option prices using our model and compare our results to the approximated European options prices obtained by Chiarella and Ziveyi [2].

2 Accuracy of first order approximation

Procedure on studying the accuracy

1. Construct the residual;
2. Represent probabilistically the residual using Feynman-Kac formula and therefore obtain boundaries;
3. Study the influence on the asset price S_t, of the fast volatility factor $V_{1,t}$ and the slow volatility factor $V_{2,t}$ for $0 < \delta << 1$ and $0 < \varepsilon << 1$ and fixed (tS_t, v_1, v_2).

Lemma 1. *Assume that the process $V_{1,t}$ with infinitesimal generator given by equation* (3) *admits moments of any order uniformly in t; i.e*

$$\sup_t E^*\{|V_{1,t}|^k\} \leq C(k).$$

Assume also that the process $V_{2,t}$ admits moments of any order uniformly in $t \leq T$; i.e

$$\sup_{t \leq T} E^*\{|V_{2,t}|^k\} \leq C(T,k)$$

then, if $J(v_1, v_2)$ is polynomially growing for any v_1 and v_2 and $t \leq T$ there exists a constant C such that for $\varepsilon \leq 1$ and $\delta \leq 1$

$$E^*_{S_t,\ v_1,\ v_2}\,|J(V_{1,t},\ V_{2,t})| \leq C.$$

The proof of this lemma can be found in Fouque et al. [7].

Let us impose the following assumptions:

1. Processes $S_t,\ V_{1,t},\ V_{2,t}$ exist and are unique for fixed $(\varepsilon,\ \delta)$ and are solutions of the system under risk neutral probability measure which is equivalent to solution of the system described in equations (1).
2. The market prices of risk are bounded and the conditions of Lemma 1 hold.

3. The process $V_{1,t}$ is mean reverting and has a unique invariant distribution.
4. Function $f(v_1, v_2)$ is smooth such that ϕ (the solution of Poisson equation) is polynomially growing.
5. The payoff function $h(x)$ and its derivatives are smooth and bounded.
6. The family J must include functions

$$\phi,\ \sqrt{V_{1,t}},\ \theta_1 - V_{1,t},\ \sqrt{V_{2,t}},\ \theta_2 - V_{2,t}.$$

Theorem 1. *Under the above assumptions, for small enough ε and δ there is a constant C such that*

$$\left|U^{\varepsilon,\delta} - \widetilde{U}^{\varepsilon,\delta}\right| \leq C(\varepsilon + \delta)$$

for fixed (t, S_t, v_1, v_2) where $U^{\varepsilon,\delta}$ and $\widetilde{U}^{\varepsilon,\delta}$ are respectively the exact and the approximate solution of problem (1).

Proof Before we prove the theorem, let us consider a higher-order approximation for the solution of problem (1)

$$\begin{aligned}\hat{U}^{\varepsilon,\delta} &= \widetilde{U}^{\varepsilon,\delta} + \varepsilon U_{2,0} + \varepsilon\sqrt{\varepsilon}U_{3,0} + \sqrt{\delta}(\sqrt{\varepsilon}U_{1,1} + \varepsilon U_{2,1}) \\ &= U_{0,0} + \sqrt{\varepsilon}U_{1,0} + \varepsilon U_{2,0} + \varepsilon\sqrt{\varepsilon}U_{3,0} + \sqrt{\delta}(U_{0,1} + \sqrt{\varepsilon}U_{1,1} + \varepsilon U_{2,1})\end{aligned}$$

where $U_{0,0} = U_{BS}$, $U_{1,0}$ and $U_{0,1}$ are defined in equation (6). The other $U_{i,j}$ coefficients of the above approximation are defined by

$$U_{2,0} = -\tfrac{1}{2}\phi D_2 U_{BS} + C_{2,0}(t, S_t, v_2)$$

$$U_{1,1} = \tfrac{(T-t)}{3\sqrt{\varepsilon\delta}} D_1 D_2 \Upsilon^{\varepsilon}(v_2)\Theta^{\delta}(v_2)\tfrac{\partial U_{BS}}{\partial\bar{\sigma}(v_2)}$$

$$U_{2,1} = -\tfrac{1}{2}\phi D_2 U_{0,1} + C_{2,1}(t, S_t, v_2)$$

$$U_{3,0} = \tfrac{1}{2}\rho_{13}\psi D_1 D_2 U_{BS} - \tfrac{1}{2}\phi D_2 U_{1,0} + C_{3,0}(t, S_t, v_2).$$

Here the function ψ is the solution of the following partial differential equation

$$\mathcal{L}_2\psi = v_1\frac{\partial}{\partial v_1}\psi - \langle v_1 - \frac{\partial\psi}{\partial v_1}\rangle$$

and $C_{2,0},\ C_{2,1},\ C_{3,0}$ are integration constants obtained from the integration of $U_{2,0},\ U_{2,1}$ and $U_{3,0}$ respectively.

Now we define the residual

$$R^{\varepsilon,\delta} = U^{\varepsilon,\delta} - \hat{U}^{\varepsilon,\delta}$$

and we apply the operator

$$\mathcal{L}^{\varepsilon,\delta} = \frac{1}{\varepsilon}\mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}}\mathcal{L}_1 + \mathcal{L}_2 + \sqrt{\delta}\mathcal{M}_1 + \delta\mathcal{M}_2$$

on both sides of the residual, and considering that from the Feynman- Kac theorem

$$\mathcal{L}^{\varepsilon,\delta}U^{\varepsilon,\delta} = 0.$$

We then obtain

$$\mathcal{L}^{\varepsilon,\delta}R^{\varepsilon,\delta} + \mathcal{L}^{\varepsilon,\delta}\widehat{U}^{\varepsilon,\delta} = 0.$$

The above can be written as

$$\mathcal{L}^{\varepsilon,\delta}R^{\varepsilon,\delta} + \varepsilon R_1^{\varepsilon} + \sqrt{\varepsilon\delta}R_2^{\varepsilon} + \delta(R_3^{\varepsilon} + R_4^{\delta}) = 0$$

where

$$\begin{aligned}
R_1^{\varepsilon} &= \left(\mathcal{L}_1 U_{3,0} + \mathcal{L}_2 U_{2,0} + \sqrt{\varepsilon}\mathcal{L}_2 U_{3,0}\right), \\
R_2^{\varepsilon} &= \left(\mathcal{M}_1 U_{1,0} + \mathcal{L}_1 U_{2,1} + \mathcal{L}_2 U_{1,1} + \varepsilon\mathcal{M}_1 U_{3,0} + \sqrt{\varepsilon}\mathcal{L}_2 U_{2,1}\right), \\
R_3^{\varepsilon} &= \left[\mathcal{M}_1 U_{0,1} + \mathcal{M}_2 U_{0,0} + \sqrt{\varepsilon}\left(\mathcal{M}_1 U_{1,1} + \mathcal{M}_2 U_{1,0}\right)\right], \\
R_4^{\delta} &= \delta\sqrt{\varepsilon}\left(\varepsilon\mathcal{M}_2 U_{3,0} + \sqrt{\delta}\mathcal{M}_2 U_{1,1}\right) + \delta\sqrt{\delta}\left(\mathcal{M}_2 U_{0,1} + \varepsilon\mathcal{M}_2 U_{2,1}\right).
\end{aligned}$$

Using terminal conditions for $U_{0,0},\ U_{0,1},\ U_{1,0}$ at maturity T the residual will be

$$R^{\varepsilon,\delta}(T, S_t, v_1, v_2) = -\varepsilon(U_{2,0}+\sqrt{\varepsilon}U_{3,0})(T, S_t, v_1, v_2) - \sqrt{\varepsilon\delta}(U_{1,1}+\sqrt{\varepsilon}U_{2,1})(T, S_t, v_1, v_2)$$

that can be expressed as

$$R^{\varepsilon,\delta}(T, S_t, v_1, v_2) = \varepsilon G_1^{\varepsilon}(S_t, v_1, v_2) + \sqrt{\varepsilon\delta}G_2^{\varepsilon}(S_t, v_1, v_2)$$

for

$$G_1^{\varepsilon}(S_t, v_1, v_2) = -(U_{2,0} + \sqrt{\varepsilon}U_{3,0})(T, S_t, v_1, v_2)$$

and

$$G_2^{\varepsilon}(S_t, v_1, v_2) = -(U_{1,1} + \sqrt{\varepsilon}U_{2,1})(T, S_t, v_1, v_2).$$

If we use Feynman-Kac probability representation formula we can express the residual in terms of the expected values of discounting time t of $G_1^{\varepsilon}(S_t, v_1, v_2)$ and $G_2^{\varepsilon}(S_t, v_1, v_2)$ functions given $S_t,\ V_{1,t},\ V_{2,t}$ under risk neutral probability measure, i.e.

$$\begin{aligned}
R^{\varepsilon,\delta} = {} & \varepsilon E^* e^{-r(T-t)} G_1^{\varepsilon}(S_T, V_{1,T}, V_{2,T}) + \\
& \varepsilon E^* \int_t^T e^{-r(x-t)} R_1^{\varepsilon}(x, S_x, V_{1,x}, V_{2,x}) dx | S_t, V_{1,t}, V_{2,t} + \\
& \sqrt{\varepsilon\delta} E^* e^{-r(T-t)} G_2^{\varepsilon}(S_T, V_{1,T}, V_{2,T}) + \\
& \sqrt{\varepsilon\delta} E^* \int_t^T e^{-r(x-t)} R_2^{\varepsilon}(x, S_x, V_{1,x}, V_{2,x}) dx | S_t, V_{1,t}, V_{2,t} + \\
& \delta\left(\int_t^T e^{-r(x-t)} R_3^{\varepsilon}(x, S_x, V_{1,x}, V_{2,x}) dx | S_t, V_{1,t}, V_{2,t}\right) + \\
& \int_t^T e^{-r(x-t)} R_4^{\delta}(x, S_x, V_{1,x}, V_{2,x}) dx | S_t, V_{1,t}, V_{2,t}.
\end{aligned}$$

Since R_i are sums of $U_{i,j}$ they are smooth functions of time, asset price and the two volatility components. For $0 < \delta << 1$ and $0 < \varepsilon << 1$ the functions R_i are bounded by smooth functions of time, asset price and the two volatility components. Such functions are independent of ε and δ; uniformly bounded in $t,\ S_t,\ v_2$ and at most linearly growing in v_1.

From the assumptions in our model, the residuals R_i and the functions $G_1^{\varepsilon}(S_t, v_1, v_2),\ G_2^{\varepsilon}(S_t, v_1, v_2)$ are bounded in S_t and at most polynomially growing in $v_1,\ v_2$. Therefore they satisfy conditions from Lemma 1 and hence

$$\begin{aligned}
\left|U^{\varepsilon,\delta} - \widetilde{U}^{\varepsilon,\delta}\right| &= \left|U^{\varepsilon,\delta} - \hat{U}^{\varepsilon,\delta} + \hat{U}^{\varepsilon,\delta} - \widetilde{U}^{\varepsilon,\delta}\right| \\
&\leq \left|R^{\varepsilon,\delta}\right| + \left|\hat{U}^{\varepsilon,\delta} - \widetilde{U}^{\varepsilon,\delta}\right| \\
&\leq \varepsilon \left|R_1^{\epsilon}\right| + \sqrt{\varepsilon\delta}\left|R_2^{\epsilon}\right| + \delta\left|R_3^{\epsilon}\right| + \left|R_4^{\delta}\right| \\
&\leq C_1\varepsilon + C_2\sqrt{\varepsilon\delta} + C_3\delta + C_4 \\
&\leq (\varepsilon + \delta)C_5.
\end{aligned}$$

3 Adjustment to volatilities and calibration

In this section we calibrate our model parameters in our first order approximation formula (5) to real market data obtained from Nasdaq Nordic Exchange website on ABB stock call options. For illustrative proposes in Table 1 we present part of the used data. The notation $E_{ij}\ j = 1, \cdots, m$ stands for the jth exercise price for an option with time-to-maturity $\tau_i = T_i - t,\ i = 1 \cdots, n$, where m and n are positive integers. We consider only call options with time to maturity less than two years and exercise price varying from 130 to 150 SEK (Swedish Kronor). The initial ABB stock price is $S_0 = 148.3$ and the risk-free interest rates is constant and we assume to be equal 5%.

Since we relaxed the Black-Scholes assumptions of constant volatility by introducing stochastic volatility, we will adjust the observed option price to the volatility by converting the option price into its *implied volatility* (the value of volatility for an underlying asset that in an option pricing model returns the current market price of the option).

3.1 The calibration procedure

From Table 1 at a certain time to maturity τ_i we have exercise prices E_{ij}, $j = 1 \cdots, m$. For a fixed τ_i, the implied volatility for exercise price $E_{i,j}$ is denoted by $I(\tau_i, E_{i,j})$.

Holding the time to maturity constant, different exercise prices result into different market option prices. Same applies if we hold exercise price constant

Option Nr.	i	j	τ_i	$E_{i,j}$	Market Price C_{ij}
1	1	1	0.13	132.5	16.75
2	1	2	0.13	135	14.5
3	1	3	0.13	137.5	12.0
4	1	4	0.13	140	9.25
5	2	1	0.27	135	12.5
6	2	2	0.27	137.5	10.25
7	2	3	0.27	140	8.25
8	2	4	0.27	142.5	6.25
9	3	1	0.43	135	12.75
10	3	2	0.43	137.5	11.5
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
104	26	4	0.89	147.5	7.5

Table 1. A part of the used data on ABB call options prices, $S_0 = 148.3$

but let the time to maturity vary. Therefore we compute the matrix of log-moneyness to maturity ratio $LMMR_{ij}$ for option with time to maturity τ_i and exercise price E_{ij}

$$LMMR_{ij} = \frac{1}{\tau_i} \ln\left(\frac{E_{ij}}{S_t}\right) \tag{7}$$

for $1 \leq i \leq n$, $1 \leq j \leq m$ and usually $n < m$ since the changes are less influenced by time comparing with respect to price. For the set of points

$$\{I(\tau_1, E_{1,j}),\ I(\tau_2, E_{2,j}), \cdots, I(\tau_n, E_{n,j})\}$$

we can apply the log-moneyness to maturity ratio matrix given by (7) and obtain an equation for the implied volatility and solve it using the least squares method.

3.2 Calibration of the first order approximation

Let us define the implied volatility as the solution for I in the following equation

$$U_{BS}(I) = C_{i,j}, \tag{8}$$

where $U_{BS}(I)$ is the option price given by Black-Scholes two-dimensional formula at volatility level I and $C_{i,j}$ is the market call option price with time to maturity τ_i and exercise price $E_{i,j}$. Under the assumption that the system of stochastic differential equations (1) is a realistic model of asset price dynamic, the market price C_{ij} should be equal to the price given by system (1). To adjust the model results to the traded values we need to study the difference $I - \overline{\sigma}$ between the implied volatility and the volatility used to compute the two dimensional Black-Scholes prices,

$$I - \overline{\sigma} = \sqrt{\varepsilon} I_{1,0} + \sqrt{\delta} I_{0,1} + \cdots. \tag{9}$$

For simplicity hereafter we write τ, $LMMR$, $\overline{\sigma}$, E, as shortened form for τ_i, $LMMR_{ij}$, $\overline{\sigma}(v_2)$, E_{ij}, respectively. Therefore

$$U_{BS}(I) = U_{BS}(\overline{\sigma}) + (\sqrt{\varepsilon}I_{1,0} + \sqrt{\delta}I_{0,1})\frac{\partial U_{BS}(\overline{\sigma})}{\partial \overline{\sigma}} + \cdots .$$

Considering that the prices given by the above equation must be the same obtained by equation (5) we will have

$$U_{BS}(I) = U_{BS} + \left(\tau\Theta^{\delta}D_1 + \frac{1}{\overline{\sigma}}\Upsilon^{\varepsilon}D_1\right)\frac{\partial U_{BS}}{\partial \overline{\sigma}} + \cdots .$$

The previous equation is true only if

$$\sqrt{\varepsilon}I_{1,0}\frac{\partial U_{BS}(\overline{\sigma})}{\partial \overline{\sigma}} = \frac{1}{\overline{\sigma}}\Upsilon^{\varepsilon}D_1\frac{\partial U_{BS}}{\partial \overline{\sigma}} \tag{10}$$

and

$$\sqrt{\delta}I_{0,1}\frac{\partial U_{BS}(\overline{\sigma})}{\partial \overline{\sigma}} = \tau\Theta^{\delta}D_1\frac{\partial U_{BS}}{\partial \overline{\sigma}}. \tag{11}$$

The two equations above can also be expressed with the uses of Vega - Gamma relations; i.e.

$$\frac{\partial U_{BS}(\overline{\sigma})}{\partial \overline{\sigma}} = \tau\overline{\sigma}S_t^2\frac{\partial^2 U_{BS}(\overline{\sigma})}{\partial S_t^2}$$

as

$$\sqrt{\varepsilon}I_{1,0} = \frac{\Upsilon^{\varepsilon}}{2\overline{\sigma}}\left(1 - \frac{2r}{(\overline{\sigma})^2}\right) + \frac{\Upsilon^{\varepsilon}}{(\overline{\sigma})^3}\frac{\ln\left(\frac{E}{S_t}\right)}{\tau} \tag{12}$$

and

$$\sqrt{\delta}I_{0,1} = \tau\frac{\Theta^{\delta}}{2}\left(1 - \frac{2r}{(\overline{\sigma})^2}\right) + \tau\frac{\Theta^{\delta}}{(\overline{\sigma})^2}\frac{\ln\left(\frac{E}{s}\right)}{\tau} \tag{13}$$

hold. The difference between the implied volatility and the volatility on the approximation given in equation (9) and the representations of $I_{0,1}$ and $I_{1,0}$ presented in equations (12) and (13) allow as to express the implied volatility as

$$I = a_1^{\varepsilon} + \tau a_0^{\delta} + \left(\tau a_2^{\delta} + a_3^{\varepsilon}\right) LMMR \tag{14}$$

where

$$a_1^{\varepsilon} = \overline{\sigma} + \frac{\Upsilon^{\varepsilon}}{2\overline{\sigma}}\left(1 - \frac{2r}{(\overline{\sigma})^2}\right); \quad a_0^{\delta} = \frac{\Theta^{\delta}}{2}\left(1 - \frac{2r}{(\overline{\sigma})^2}\right); \quad a_2^{\delta} = \frac{\Theta^{\delta}}{(\overline{\sigma})^2}; \quad a_3^{\varepsilon} = \frac{\Upsilon^{\varepsilon}}{(\overline{\sigma})^3}.$$

Since there is a correlation between the underlying asset price and the volatility factors, a_1^{ε} will correct $\overline{\sigma}$ in order to incorporate the effects of the fast volatility factor while a_0^{δ} does the same for the effects of the slow volatility factor.

The definition of a_1^{ε} and the fact that the considered terms have orders up to $\sqrt{\varepsilon}$ and $\sqrt{\delta}$ gives

$$a_1^{\varepsilon} = \overline{\sigma} + O(\sqrt{\varepsilon})$$

or

$$\overline{\sigma} = a_1^{\varepsilon} + a_3^{\varepsilon}\left(\frac{\left(a_1^{\varepsilon}\right)^2}{2} - r\right). \tag{15}$$

On the other hand the fact that

$$\frac{\Upsilon^{\varepsilon}}{\left(\overline{\sigma}\right)^3} = a_3^{\varepsilon}$$

implies that

$$\Upsilon^{\varepsilon}(v_2) = a_3^{\varepsilon}\left(\overline{\sigma}\right)^3 = a_3^{\varepsilon}\left(a_1^{\varepsilon}\right)^3, \tag{16}$$

and

$$\Theta^{\delta}(v_2) = a_2^{\delta}\left(\overline{\sigma}\right)^2. \tag{17}$$

Using the matrix defined by equation (7) into equation (14) it remains to estimate the slope $a_0^{\delta} + a_2^{\delta} LMMR$ and the intercept $a_1^{\varepsilon} + a_3^{\varepsilon} LMMR$.

From equation (14) it is obvious that $I(\tau_i, E_{ij})$ depends on $LMMR_{ij}$. Using the least squares method we can find a line that will fit the set of points $(I(\tau_i, E_{ij}),\ LMMR_{ij})$. We start by searching for α^i and β^i such that

$$Q(\alpha^i, \beta^i) = \sum_{j}^{m}\left(I(\tau_i, E_{ij}) - \alpha^i LMMR_{ij} - \beta^i\right)^2.$$

The $\alpha_{\min}^i$ and $\beta_{\min}^i$ which minimize $Q(\alpha^i, \beta^i)$ are given by

$$\beta_{\min}^i = \overline{I}(\tau_i, E_{ij}) - \alpha_{\min}^i \overline{LMMR}_{ij}$$

and

$$\alpha_{\min}^i = \frac{\sum\limits_{j=1}^{m} LMMR_{ij}\left(I(\tau_i, E_{ij}) - \overline{I}(\tau_i, E_{ij})\right)}{\sum\limits_{j=1}^{m} LMMR_{ij}\left(LMMR_{ij} - \overline{LMMR}_{ij}\right)}.$$

where $\overline{LMMR}_{ij}$ and $\overline{I}(\tau_i, E_{ij})$ are respectively the averages of $LMMR$ and implied volatilities for a given i. For each τ_i there is one $\alpha_{\min}^i$, therefore we can construct with least squares method an linear estimator for such relation; i.e, we can determine the intercept

$$a_{\min}^{\varepsilon} = \overline{\alpha_{\min}^i} - a_2^{\delta}\overline{\tau}_i$$

and the slope

$$a_{\min}^{\delta} = \frac{\sum\limits_{j=1}^{m} \tau_i \alpha_{\min}^i - \overline{\alpha_{\min}^i}}{\sum\limits_{j=1}^{m} \tau_i(\tau_i - \overline{\tau}_i)}.$$

which minimizes

$$R(a^{\delta},\ a^{\varepsilon}) = \sum_{j=1}^{m}\left(\alpha_{\min}^i - (a^{\delta}\tau_i + a^{\varepsilon})\right)^2.$$

The same approach for $\beta^i_{\min}$ gives

$$b^{\varepsilon}_{\min} = \overline{\beta^i_{\min}} - b^{\delta}_{\min}\overline{\tau}_i$$

and

$$b^{\delta}_{\min} = \frac{\sum_{j=1}^{m} \tau_i \beta^i_{\min} - \overline{\beta^i_{\min}}}{\sum_{j=1}^{m} \tau_i(\tau_i - \overline{\tau}_i)}.$$

In equations (16) and (17) parameters $a^{\varepsilon}_1,\ a^{\delta}_2,\ a^{\varepsilon}_3$ are replaced by $b^{\varepsilon}_{\min},\ a^{\delta}_{\min}, a^{\varepsilon}_{\min}$ respectively and are used to express Θ^{δ} and Υ^{ε} that are used to construct our approximate solution.

3.3 Numerical Analysis

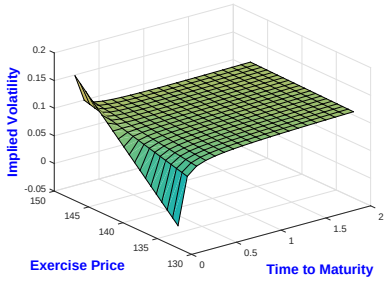
We consider $S_0 = 148.3,\ r = 0.05$ and exercise prices changing from 130 SEK to 150 SEK. In Figure 1(a) we show the dependence of the implied volatility on the time to maturity and the exercise prices. As expected, the volatility changes very smoothly while far from the maturity time, and when approximating the maturity time, it can be seen that the implied volatility changes are more significant. Although we are using different data set and different option pricing model, the behaviour of the implied volatility that we obtain is very similar to the volatility behaviour obtained by Fouque et al. [7].

We check the pattern between the implied volatilities and the option prices. Since the values of implied volatility are too small comparing to those of the options price, to be able to obtain them on the same plot and observe their patterns, in Figure 1(b), instead of options prices and volatilities we consider normalized options prices and volatilities. As expected, these two components are changing inversely, this can be seen in the Figure 1(b).

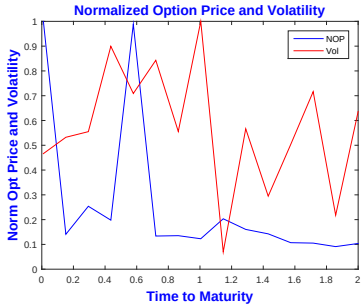
The validation of our model is done by comparing the results obtained from the asymptotic expansion approximation and those obtained by Chiarella and Ziveyi [2]. From the setup of our model, the values of ε and δ must be considered to be small. If we make ε smaller than δ then it becomes more clear that the speed of reversion of $V_{1,t}$ will be much faster than the reverting speed for the process $V_{2,t}$. We observe that when we fix δ but make ε smaller the option price approximation becomes closer to those obtained by Chiarella and Ziveyi. Figure 2(a) depicts such situation. A similar scenario is observed when ε is fixed and δ is decreasing. We also observe that ε has more influence than δ on fitting our results to those presented by Chiarella and Ziveyi. That is, even for small values of δ as shown in Figure 2(b), so long as the parameter ε is not sufficiently small, our model produces results that are not converging to Chiarella and Ziveyi results.

We define the relative difference of our approximation and Chiarella and Ziveyi method by

$$\text{error}(\varepsilon, \delta) = \left| \frac{\widetilde{U}^{\varepsilon,\delta} - U_{CZ}}{max\left(\widetilde{U}^{\varepsilon,\delta}, U_{CZ}\right)} \right|$$

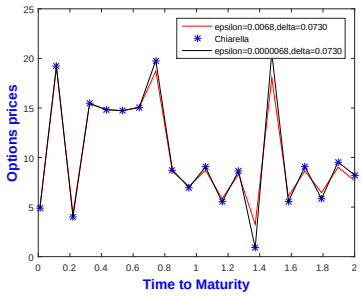


(a) Implied volatility surface estimated according to equation (14) when time to maturity goes from zero to two and the exercise prices goes from 130 to 150. The obtained parameters are $(a_0^\delta, a_1^\varepsilon, a_2^\delta, a_3^\varepsilon) = (-0.3222, 0.5235, 0.5490, 0.4630)$.

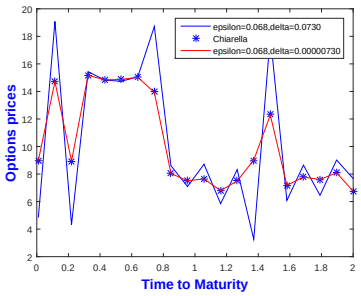


(b) Implied volatility and normalized options prices given by our asymptotic expansion method (with the same input arguments used in the Figure 1(a))

Fig. 1.



(a) Comparison of Chiarella and Ziveyi method with our asymptotic expansion method (varying ε but fixed δ)



(b) Comparison of Chiarella and Ziveyi method with aour asymptotic expansion method (varying δ bux ε fixed)

Fig. 2.

where U_{CZ} gives the option price obtained by using Chiarella and Ziveyi [2] approach. The relative difference in the options prices between both approaches for various pairs of ε and δ are presented in Table 2. Each column lists out the $error(\varepsilon, \delta)$ for options under consideration for a particular pair of (ε, δ).

From Table 2 we can see that for small values of ε and δ the options prices given by asymptotic expansion approximation formula are very close to those presented when we consider Chiarella and Ziveyi model. This fact confirm that our asymptotic expansion approximation gives plausible results.

$\varepsilon = 0.008$ $\delta = 0.2$	$\varepsilon = 0.000008$ $\delta = 0.2$	$\varepsilon = 0.00008$ $\delta = 0.0002$
0.01712	0.0015	0.0009
0.0118	0.0001	0.0011
0.0245	0.0022	0.0000
0.0021	0.0014	0.0007
0.0460	0.0010	0.0017
0.0111	0.0017	0.0004
0.0514	0.0056	0.0001
0.0167	0.0002	0.0002
0.0221	0.0001	0.0081
0.0030	0.0037	0.0013

Table 2. Illustration of the error obtained for different values of ε and δ

4 Conclusions and future work

The calibration procedure for our first-order asymptotic expansion approximation formula has shown to be convenient. The accuracy of the obtained approximation results is also plausible. With small margin of errors our approximation formula price is fitting the European option price calculated by the Chiarella and Ziveyi approach. We recommend that more extensive studies should be done, by considering a wider selection of stocks and options, in order to confirm the efficiency and accuracy of our method. Further numerical studies can be carried out to analyze how the parameters affect the approximation accuracy and also to study how much improvement one can get by using a second-order asymptotic expansion.

Acknowledgements

This work was partially supported by Swedish SIDA Foundation - International Science Program. Betuel Canhanga and Jean-Paul Murara thank Division of Applied Mathematics, School of Education, Culture and Communication, Malardalen University for creating excellent research and educational environment.

References

1. Canhanga B., Malyarenko, A., Ni, Y. and Silvestrov S. "Perturbation methods for pricing European options in a model with two stochastic volatilities". *3rd SMTDA Conference Proceedings.* 11-14 June 2014,Lisbon Porturgal, C. H. Skiadas (Ed) 489-500 (2014).
2. Chiarella, C, and Ziveyi, J. "American option pricing under two stochastic volatility processes". *J. Appl. Math. Comput.* 224:283–310 (2013).
3. Christoffersen, P.; Heston, S.; Jacobs, K. "The shape and term structure of the index option smirk: why multifactor stochastic volatility models work so well" Manage.Sci.55(2) 1914-1932; (2009).

4. Conze, A., Lantos, N., and Pironneau, O. "The forward Kolmogorov equation for two dimenssional options"; University Pierre et Marie Curie; (2010).
5. Cox, J., Ross, S. "The valuation of options for alternative stochastic processes"; Journal of financial economics 3:146-166, (2014).
6. Feller, W. "Two singular diffusion problems". *Ann. Math.* 54:173–182 (1951).
7. Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K. "Multiscale stochastic volatility for equity, interest rate and credit derivatives", Cambridge University Press, Cambridge (2011b).
8. Fouque, J.-P., Lorig, M., Sircar, R. "Second order multiscale stochastic volatility asymptotics: stochastic terminal layer analysis and calibration", University of California (2014).
9. Kijima, M. "Stochastic Processes with Applications to Finance", Second Edition, Capman and Hall/CRC, (2013).
10. Omx Nasdag Nordic Exchange website Retrieved from `http://www.nasdaqomxnordic.com/` cited on October 21; 2015

Estimation of Lévy Processes through Stochastic Programming and Kalman Filtering

Mark Anthony Caruana[1]

[1] Department Statistics and Operations Research, Faculty of Science, University of Malta, Msida, Malta
(E-mail: mark.caruana@um.edu.mt)

Abstract. The estimation of Lévy process has received a lot of attention in recent years. Evidence of this is the extensive amount of literature concerning this problem which can be classified in three categories: the nonparametric approach, the semi-parametric approach and the parametric approach. In this paper, we shall concentrate on the latter, and in particular the parameters will be estimated within a stochastic programming framework. To be more specific, the first derivative of the characteristic function and its empirical counterpart shall be used in objective function. Furthermore, the parameter estimates are recursively estimated by making use of a modified extended Kalman filter (MEKF). Some properties of the parameter estimates are studied. Finally, a number of simulations will be carried out and the results are presented and discussed.

Keywords: Lévy Processes, Kalman filter, Stochastic Programming, Characteristic Function

1 Introduction

It is a known fact that Lévy processes are capable of modeling process which are not only continuous but which may also contain jumps. This feature has made it possible for a great deal of reaserchers to apply Lévy processes successfully in many areas, most prominently in finance.
As a result, the problem concerning the statistical inference of Lévy processes has received considerable attention. The literature pertaining to this area is divided into two major categories: the parametric techniques and the non-parametric methods. The latter have a long history, in fact, some early works were published as of the late 1950's. Furthermore, in recent years several distinguished athors have contributed to this field. As stated in Comte and Genon-Catalot[3], these methods are particularly useful because in many cases the distribution of the independent incremens of the Lévy process is unkown, or has no known formula. In such cases the Lévy process is specified by its Lévy Triple (drift, Gaussian component and Lévy measure). On the other hand the parametric techniques are particularly useful when the distribution (or the corresponding characteristic function) of the increments of a Lévy process is known and has a *closed form expression*. A number of these Lévy processes have been proposed over the years and have been successfully applied to

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

financial data. These include the CGMY, Meixner, NIG and stable processes. In such cases, Figueroa-Lopez[4] argues that parametric methods would amongst other things, give estimates that enjoy a smaller standard error or a greater power.

A good percentage of the parametric methods of estimation make use of the characteristic function (CF) and its empirical counterpart (ECF). As discussed in Sant and Caruana[9], the use of the EFC causes a number of computational issues which are primarily caused by oscillatory integrands. To counteract this issue, the said authors reformulate the parameter estimation problem into the stochastic programming framework, and together with the use of weight functions, the problems caused by the oscillatory integrands are reduced.

In this paper the stochastic programming framework is retained, however the objective function will be totally reformulated and will make use of the first derivative of the characteristic function and its empirical counterpart. This alteration will allow the use of a modified extended kalman filter, in this way the parameters can be estimated recursively as more data points become available.

The rest of this paper is organised as follows: in the next section the stochastic programming framework is discussed and the objective function clearly defined. Moreover, a number of results pertaining to the parameter estimates will be presented. In section 3 the modified extended kalman filter is discussed and applied to the objective function to solve the stochastic programming problem. In section 4, a number of simulation results are presented and discussed. Finally, the last section contains a number of concluding remarks.

2 The Stochastic Programming Framework

We shall commence this section by defining a Lévy process, its increments and the characteristic function together with it first derivative. Afterwards we shall move a step forward by expressing the parameter estimation problem as a stochastic programming problem. Given a Lévy process Z_s, $s \in \mathbb{R}_+$ with independent and identically distributed increments X_j, $j = 1,\ldots,N$, we define the corresponding characteristic function by $\varphi(t,\beta(s)) = \mathbb{E}[\exp(itZ_s)]$ where $\beta(s) \in \mathbb{R}^n$ is the vector of parameters for each time point *s*. The derivative of the characteristic function with respect to *t* is defined by $\varphi'(t,\beta(s)) = \mathbb{E}[iZ_s \exp(itZ_s)]$.

From this point onwards we shall work with $\varphi(t,\beta(\delta s))$, where δs is a constant time increment separating the previously defined X_j's whose characteristic function will now be denoted for simplicity by $\varphi(t,\beta)$ and the corresponding derivative by $\varphi'(t,\beta)$. We will assume throughout this text that φ and φ' are continuous with respect to β on some compact subset $M \subseteq \mathbb{R}^n$.

Next, the parameter estimation problem will be written as a stochastic program that involves the derivative of the characteristic function and its empirical counterpart.
In general a stochastic program may be expressed as follows:

$$\arg\min_{\beta}\left\{\mathbb{E}[F(X,\beta)]\right\} \tag{1.1}$$

were the expectation is done with respect to the random variable *X,* and (1.1) is used to find the true parameter β_0 .
For $\beta \in M$, we define the following random function :

$$F(\beta,\omega)=\left|\int_0^{\infty}\left\{X\exp(itX)-\varphi^x(t,\beta)\right\}w(t)\,dt\right|^2 \tag{1.2}$$

where $\varphi^x(t,\beta)=-i\varphi'(t,\beta)=-i\mathbb{E}[\partial\exp(itX)/\partial t]=\mathbb{E}[X\exp(itX)]$. By setting the weight function $w(t)=\exp(-|t|)$ we expand (1.2) to obtain:

$$\begin{aligned}&F(\beta,\omega)=\\&\left(\frac{X}{1+X^2}\right)^2-2\left(\int_0^{\infty}X\cos(tX)w(t)\,dt\right)\int_0^{\infty}\varphi_r^x(t,\beta)w(t)dt+\left(\int_0^{\infty}\varphi_r^x(t,\beta)w(t)dt\right)^2\\&\qquad+\\&\left(\frac{X^2}{1+X^2}\right)^2-2\left(\int_0^{\infty}X\sin(tX)w(t)\,dt\right)\int_0^{\infty}\varphi_I^x(t,\beta)w(t)dt+\left(\int_0^{\infty}\varphi_I^x(t,\beta)w(t)dt\right)^2\end{aligned} \tag{1.3}$$

where $\varphi_r^x(t,\beta)=\mathrm{Re}[\varphi^x(t,\beta)]$ and $\varphi_I^x(t,\beta)=\mathrm{Im}(\varphi^x(t,\beta))$. Furthermore the specific choice of weight function gives the following results:

$$\int_0^{\infty}X\cos(tX)w(t)\,dt=\frac{X}{1+X^2}\qquad\text{and}\qquad\int_0^{\infty}X\sin(tX)w(t)\,dt=\frac{X^2}{1+X^2}$$

Taking the expectation of (1.3) we obtain:

$$\begin{aligned}&\mathbb{E}[F(\beta,\omega)]=\\&\Psi[\beta_0]-2\left(\int_0^{\infty}\varphi_r^x(t,\beta_0)w(t)\,dt\right)\int_0^{\infty}\varphi_r^x(t,\beta)w(t)dt+\left(\int_0^{\infty}\varphi_r^x(t,\beta)w(t)dt\right)^2\\&\qquad+\\&\Gamma[\beta_0]-2\left(\int_0^{\infty}\varphi_I^x(t,\beta_0)w(t)\,dt\right)\int_0^{\infty}\varphi_I^x(t,\beta)w(t)dt+\left(\int_0^{\infty}\varphi_I^x(t,\beta)w(t)dt\right)^2\end{aligned} \tag{1.4}$$

where $\Psi[\beta_0]=\mathbb{E}[(X/(1+X^2))^2]$ while $\Gamma[\beta_0]=\mathbb{E}[(X^2/(1+X^2))^2]$. It is evident that $\Psi[\beta_0]$ and $\Gamma[\beta_0]$ are finite. The expectation in (1.4) can be approximated by making use of the set of previously defined increments X_j as follows:

$$F_N(\beta,\omega)=\frac{1}{N}\sum_{j=1}^{N}\left|\int_0^{\infty}\left\{X_j\exp(itX_j)-\varphi^x(t,\beta)\right\}w(t)\,dt\right|^2 \tag{1.5}$$

where $w(t)$ is as defined above.

The continuity assumption of φ and φ□ with respect to β implies that the functions F and F_n will also satisfy this assumption. This means that these functions are elements of $C(M)$, the space of continuous functions on M. This set can be considered as a separable Banach space under the supermom norm, which we denote by $||\,.\,||$ such that $||\,F\,|| = \sup_{\beta} |\,F(\beta)\,|$.

We now state and prove a result which determines the asymptotic behavior of (1.5).

Theorem 1.

(i) The sequence $F_N(\beta,\omega)$ converges $\mathbb{P}$ *a.s.* to $\mathbb{E}[F(\beta,\omega)]$.

(ii) If there exist constants C_1, C_2 such that $\forall \beta_1, \beta_2 \in M$:

$$\left|\int_0^\infty \varphi_r^x(t,\beta_2)w(t)dt - \int_0^\infty \varphi_r^x(t,\beta_1)w(t)\,dt\right| \le C_1\left\|\beta_2 - \beta_1\right\|$$

$$\left|\int_0^\infty \varphi_I^x(t,\beta_2)w(t)dt - \int_0^\infty \varphi_I^x(t,\beta_1)w(t)\,dt\right| \le C_2\left\|\beta_2 - \beta_1\right\|$$

the $\sqrt{N}(F_N(\beta,\omega) - \mathbb{E}[F(\beta,\omega)])$ converges in distribution to a Gaussian random element.

Proof.

We start by proving (i):

The following inequality:

$$\left\|\left(\left|\int_0^\infty \left\{X_j \exp(itX_j) - \varphi^x(t,\beta)\right\} w(t)\,dt\right|^2\right)\right\| = \sup_{\beta}\left(\left|\int_0^\infty \left\{X_j \exp(itX_j) - \varphi^x(t,\beta)\right\} w(t)\,dt\right|^2\right) \le 1$$

ensures that $\forall j$, $\mathbb{E}\left[\left\|\left(\left|\int_0^\infty \left\{X_j \exp(itX_j) - \varphi^x(t,\beta)\right\} w(t)\,dt\right|^2\right)\right\|^p\right] < \infty,\ \forall p \ge 0$.

Hence by using the Strong Law of Large numbers for iid random variables in Banach spaces we can say that $F_N(\beta,\omega)$ converges $\mathbb{P}$ *a.s.* to $\mathbb{E}[F(\beta,\omega)]$.

Proof of (ii):

It can be easily shown that the random function $\beta \to F(\beta,\omega)$ is always less than or equal to some constant. Furthermore the said function can be proven to be Lipchitz continuous as shown below:

$$
\begin{aligned}
&| F(\beta_2,\omega) - F(\beta_1,\omega) | \leq \\
&\quad 2\left|\int_0^\infty \varphi_r^x(t,\beta_2)w(t)dt - \int_0^\infty \varphi_r^x(t,\beta_1)w(t)\,dt\right| + \\
&\quad 2\left|\int_0^\infty \varphi_r^x(t,\beta_2)w(t)dt - \int_0^\infty \varphi_r^x(t,\beta_1)w(t)\,dt\right| + \\
&\quad \left|\left(\int_0^\infty \varphi_r^x(t,\beta_2)w(t)dt\right)^2 - \left(\int_0^\infty \varphi_r^x(t,\beta_1)w(t)\,dt\right)^2\right| + \\
&\quad \left|\left(\int_0^\infty \varphi_I^x(t,\beta_2)w(t)dt\right)^2 - \left(\int_0^\infty \varphi_I^x(t,\beta_1)w(t)\,dt\right)^2\right|
\end{aligned}
$$

which means that $| F(\beta_2,\omega) - F(\beta_1,\omega) | \leq 4(C_1 + C_2) \,||\, \beta_2 - \beta_1 \,||$.
The result allows us to use the CLT in Banach spaces as proposed in Jain and Marcus[5], from which it follows that $\sqrt{N}(F_N(\beta,\omega) - \mathbb{E}[F(\beta,\omega)])$ converges in distribution to a Gaussian random element found in $C(M)$.

□

From the above theorem we know that $F_N(\beta,\omega)$ converges $\mathbb{P}$ a.s. to $\mathbb{E}[F(\beta,\omega)]$ which means that it makes sense to solve the following:

$$
\arg\min_{\beta}\left(\frac{1}{N}\sum_{j=1}^{N}\left|\int_0^\infty \left\{X_j \exp(itX_j) - \varphi^x(t,\beta)\right\} w(t)\,dt\right|^2\right) \tag{1.6}
$$

instead of $\arg\min_{\beta}\left(\mathbb{E}[F(\beta,\omega)]\right)$.

Although we have shown that $F_N(\beta,\omega)$ converges $\mathbb{P}$ a.s. to $\mathbb{E}[F(\beta,\omega)]$, we still have to prove that as the number of increments increases $\hat{\beta} = \arg\min_{\beta}\left(F_N(\beta,\omega)\right)$ will approach the true value $\beta_0 = \arg\min_{\beta}\left(\mathbb{E}[F(\beta,\omega)]\right)$.
To prove this result we shall adapt a result found in Sant and Caruana[9].
However, before stating this result it is important to note that the link between $\arg\min_{\beta}\left(\mathbb{E}[F(\beta,\omega)]\right)$ and (1.6) is given by the functional:

$$
\Lambda(g) = \inf_{\beta \in M}[g(\beta)]. \tag{1.7}
$$

Although not linear, (1.7) still shares some properties with the norm. In fact (1.7) is the minimum of the linear functionals all of whom are in the dual space of $C(M)$, which contains all the Radon measures on M. Furthermore, from standard theory it is known that Λ is concave and therefore is Hadamard differentiable.

Theorem 2

The sequence of random variables $\Lambda(F_N(\beta,\omega))$ converges in probability to $\mathbb{E}[F(\beta_0,\omega)]$. Furthermore $\Lambda(F_N(\beta,\omega)) - \mathbb{E}[F(\beta_0,\omega)] = 0_{\mathbb{P}}(1)$ and $\lim_{N\to\infty}\left[\arg\min_{\beta} F_N(\beta,\omega)\right] = \beta_0$.

Proof:

For some fixed ω, the function $F_N(\beta,\omega)$ is quadratic in terms of $\left(\int_0^\infty \varphi_r^x(t,\beta)w(t)dt\right)$ and $\left(\int_0^\infty \varphi_I^x(t,\beta)w(t)dt\right)$, thus having a minimum value when the former and latter take the values of $\left(\int_0^\infty X\cos(tX)w(t)dt\right)$ and $\left(\int_0^\infty X\sin(tX)w(t)dt\right)$ respectively.

Furthermore it is evident that $\mathbb{E}[F(\beta,\omega)]$ is also quadratic in terms of $\left(\int_0^\infty \varphi_r^x(t,\beta)w(t)dt\right)$ and $\left(\int_0^\infty \varphi_I^x(t,\beta)w(t)dt\right)$, hence attaining a minimum when $\beta = \beta_0$. Furthermore uniqueness of the characteristic function assures the uniqueness of the said minimum.

Next, we consider the subdifferential $\partial\Lambda$, which when evaluated at any $g \in \mathcal{C}(M)$ i.e. $\partial\Lambda(g)$, is a set of elements contained in the dual space of $\mathcal{C}(M)$. In particular this set contains all positive Radon measures of total mass 1 concentrated on the points where g attains its minima. Evaluated at $\mathbb{E}[F(\beta,\omega)]$ the said subdifferential is,

$$\left\langle \partial\Lambda(\mathbb{E}[F(\beta,\omega)]), h\right\rangle = \inf_{\beta\in M(\mathbb{E}[F(\beta,\omega)])} h(\beta)$$

This means that $\partial\Lambda(\mathbb{E}[F(\beta,\omega)])$ operating on h returns the minimum of h restricted to the points where $\mathbb{E}[F(\beta,\omega)])$ achieves its minimum, i.e. at β_0.

Finally applying the Delta method discussed in van der Vaart[10] with the Hadamard differentiability property of Λ operating on the series $F_N(\beta,\omega)$ we obtain convergence in probability. Finally $\lim_{N\to\infty}\left[\arg\min F_N(\beta,\omega)\right] = \beta_0$ follows from the fact that the minimum is unique.

□

Levin and Khramtsov[7] study the bias of estimators that are based on the minimization of a distance function involving the characteristic function and its empirical counterpart. Following in their footsteps, the next step in this section

will be that of examining the bias of the proposed estimator and in particular show that it is of order N^{-1}.
For simplicity in the following theorem we shall assume that we only have one parameter β to estimate. However, the following theorem can be easily adapted to the case when β is a vector of parameters.

Theorem 3
Given a set of iid increments X_j having a characteristic function $\varphi(t,\beta_0)$. If $\hat{\beta} = \arg\min_{\beta} F_N(\beta,\omega)$ is an estimator of β_0, then $\mathbb{E}(\hat{\beta}) - \beta_0 = O(N^{-1})$ and $\sqrt{N}(\hat{\beta} - \beta_0)$ converges to a normal distribution.

Proof
We focus on the objective function in (1.6) and call it $L(\beta)$. The first and second partial derivatives of $L(\beta)$ w.r.t. β are given below:

$$L'(\beta) = -\tfrac{1}{N}\sum_{j=1}^{N} 2\left\{\left[\int_0^{\infty} X_j \cos(tX_j)w(t)\,dt - \mathrm{Re}(\beta)\right]\mathrm{Re}'(\beta) + \left[\int_0^{\infty} X_j \sin(tX_j)w(t)\,dt - \mathrm{Im}(\beta)\right]\mathrm{Im}'(\beta)\right\}$$

and

$$L''(\beta) = \tfrac{1}{N}\sum_{j=1}^{N}\left\{-2\left(\int_0^{\infty} x_i \cos(tx_i)w(t)dt\right)\mathrm{Re}''(\beta) - 2\left(\int_0^{\infty} x_i \sin(tx_i)w(t)dt\right)\mathrm{Im}''(\beta)\right.$$
$$\left. +2[\mathrm{Re}'(\beta)]^2 + 2\,\mathrm{Re}(\beta)\,\mathrm{Re}''(\beta) + 2[\mathrm{Im}'(\beta)]^2 + 2I(\beta)\,\mathrm{Im}''(\beta)\right\}$$

where $\mathrm{Re}(B) = \int_0^{\infty} \varphi_r^x(t,\beta)w(t)dt$ and $\mathrm{Im}(B) = \int_0^{\infty} \varphi_I^x(t,\beta)w(t)dt$.

It is easy to see that $\mathbb{E}\left[L'(\beta)\right] = 0$ and $\mathbb{E}[L''(\beta)] > 0$ when $\beta = \beta_0$. In fact, $\mathbb{E}[L''(\beta_0)] = 2\left(R'(\beta_0)\right)^2 + 2\left(I'(\beta_0)\right)^2$. This means that with probability one, $L''(\beta_0) \to \mathbb{E}[L''(\beta_0)]$.
Furthermore, $\hat{\beta}$ satisfies the equation $L'(\hat{\beta}) = 0$ whose first order approximation is:

$$L'(\beta) + (\hat{\beta} - \beta)L''(\beta) = 0 \qquad (1.8)$$

From (1.8) it is evident that the ratio (1/*N*) which features in both $L'(\beta)$ and $L''(\beta)$ can be removed from the equation. Hence we will work with

$$\tilde{L}'(\beta) + (\hat{\beta} - \beta)\tilde{L}''(\beta) = 0 \qquad (1.9)$$

where $\tilde{L}'(\beta) = NL'(\beta)$ and similarly for $L''(\beta)$.

Next we let $U_i = \tilde{L}'(\beta, X_i),\ V_i = \tilde{L}''(\beta, X_i),\ I = -\sum_{j=1}^{N} \mathbb{E}(V_i),\ U = \sum_{j=1}^{N} U_j$ and $\rho = \mathbb{V}\mathrm{ar}[U]$.

Replacing V by its expectation and substituting these formulas into (1.9) we obtain:

$$\hat{\beta} - \beta = \frac{U}{I} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(\hat{\beta}) = \frac{\mathbb{V}\mathrm{ar}(U)}{I^2} = \frac{\rho}{I^2}.$$

It can be proven that $\rho < \infty$ and that ρ is of order N when $\beta = \beta_0$.

To obtain a more refined estimate we replace (1.9) by the second order approximation:

$$\tilde{L}'(\beta) + (\hat{\beta} - \beta)\tilde{L}''(\beta) + 0.5(\hat{\beta} - \beta)^2 \tilde{L}'''(\beta) = 0 \tag{1.10}$$

Taking expectations of (1.10) we obtain:

$$\begin{aligned} &\mathbb{E}\left(\tilde{L}'(\beta)\right) + \mathbb{E}\left((\hat{\beta} - \beta)\right)\mathbb{E}\left(\tilde{L}''(\beta)\right) + \mathrm{cov}\left((\hat{\beta} - \beta), \tilde{L}''(\beta)\right) + \ldots \\ &\ldots + 0.5\mathbb{E}\left((\hat{\beta} - \beta)^2\right)\mathbb{E}\left(\tilde{L}'''(\beta)\right) + \mathrm{cov}\left(0.5(\hat{\beta} - \beta)^2, \tilde{L}'''(\beta)\right) = 0 \end{aligned} \tag{1.11}$$

Where

$$\mathrm{cov}\left((\hat{\beta} - \beta), \tilde{L}''(\beta)\right) = \mathrm{cov}\left(\frac{U}{I}, V\right) = \frac{1}{I}\mathrm{cov}(U, V) = \frac{\mathbb{E}(U.V) - \mathbb{E}(U)\mathbb{E}(V)}{I} = \frac{J}{I},$$

In (1.10) we also use the $\tilde{L}'''(\beta)$ which is defined as follows:

$$\begin{aligned} \tilde{L}'''(\beta) = \sum_{i=1}^{N} \Bigg\{ &-2\left(\int_0^{\infty} x_i \cos(tx_i) w(t) dt\right)\mathrm{Re}'''(\beta) - 2\left(\int_0^{\infty} x_i \sin(tx_i) w(t) dt\right)\mathrm{Im}'''(\beta) \\ &\ldots + 6\,\mathrm{Re}(\beta)\,\mathrm{Re}''(\beta) + 6\,\mathrm{Im}(\beta)\,\mathrm{Im}''(\beta) + 2\,\mathrm{Re}(\beta)\,\mathrm{Re}'''(\beta) + 2\,\mathrm{Im}(\beta)\,\mathrm{Im}'''(\beta) \Bigg\} \end{aligned}$$

For simplicity we denote $\mathbb{E}[\tilde{L}'''(\beta)]$ by $K(\beta)$.

It is important to note that I, ρ, J and K are all of order N when $\beta = \beta_0$.

It can also be proven that $\mathrm{cov}(0.5(\hat{\beta} - \beta), \tilde{L}'''(\beta))$ is $O(N^{-1})$.

Thus substituting these results into (1.11) we obtain:

$$\mathbb{E}[U] - I\mathbb{E}(\hat{\beta} - \beta) + \frac{J(\beta)}{I(\beta)} + \frac{\rho(\beta)K(\beta)}{2I(\beta)^2} = 0 \tag{1.12}$$

rearranging (1.12) and letting $\beta = \beta_0$ we obtain the following result:

$$\text{bias} = \mathbb{E}(\hat{\beta} - \beta_0) = \frac{2J(\beta_0)I(\beta_0) + \rho(\beta_0)K}{2I(\beta_0)^3}$$

which is of order N^{-1} .

Finally, we prove convergence in distribution as follows:
It is easy to see that

$$\mathbb{V}\text{ar}\left\{\left[\int_0^\infty X\cos(tX)w(t)\,dt - \text{Re}(\beta_0)\right]\text{Re}'(\beta_0) + \left[\int_0^\infty X\sin(tX)w(t)\,dt - \text{Im}(\beta_0)\right]\text{Im}'(\beta)\right\}$$

equal to:

$$\mathbb{V}\text{ar}\left\{\left[\frac{X}{1+X^2}\right]\text{Re}'(\beta_0) + \left[\frac{X^2}{1+X^2}\right]\text{Im}'(\beta_0)]\right\} \tag{1.13}$$

and can be shown to be finite. For simplicity we shall denote (1.13) as $\sigma(\beta_0)$.

Using the Central Limit Theorem, we can say that at the true value β_0, $\sqrt{N}\tilde{L}(\beta_0) \to N(0,\sigma(\beta_0))$ .Using this fact, Taylor's expansion, and the fact that $L''(\beta_0) \to \mathbb{E}[L''(\beta_0)]$ we have that

$$\sqrt{N}(\hat{\beta} - \beta_0) \to N\left[0, \frac{4\sigma(\beta_0)}{(\mathbb{E}[L''(\beta_0)])^2}\right]$$

□

In the next section we shall briefly describe a method by which the stochastic program can be solved recursively as the sample size grows.

3 Solving the Stochastic Program using Kalman Filtering

Over the years, a number of methods have been devised to solve stochastic programs. Most of these algorithms assume that the objective function is convex. When this assumption fails, some algorithms exist which that still allow the user to solve the stochastic program. Sant and Caruana[9] used the Feasible Direction Interior Point Method (FDIPM) proposed by Karmsita, et al.[6] to solve a specific stochastic program. In this section we briefly discuss an algorithm which was proposed by Bertzekas[2] and is also discussed in Moriyama et al.[8] and Alessandri et al.[1] This algorithm is designed to solve (1.6) recursively. This clearly means that if additional data point had to be added to the original data set then the current estimate can be used together with the newly added data point to compute the new estimate. This property was completely absent in Sant and Caruana[9]. In fact, in the latter, if new data points are added to the data set, the new parameter estimates would not take

advantage of the current parameter estimates and would use the new data points together with the rest of the data set to estimate the parameters from scratch.

Bertzekas' method is designed to solve mathematical programs of the form:

$$\arg\min_{\beta} \left\{ f(\beta) = \sum_{j=1}^{N} \left\| g_j(\beta) \right\|^2 \mid \beta \in \mathbb{R}^n \right\} \tag{1.14}$$

where $g_j : \mathbb{R}^n \to \mathbb{R}^r$ for $j = 1, 2, \ldots, N$ and are continuously differentiable functions. Here $||.||$ denotes the Euclidian norm of a vector. Furthermore it is evident that (1.6) can easily be re-written in the form of (1.14).
Bertzekas argues that (1.14) can be solved either by using Batch-Type algorithms (BTA) or the Incremental Type Algorithms (ITA). The former estimates β by using the whole data set, while the latter cycles through the data in sequence and updates the estimate of β after each data point is processes. Back Propagation Algorithms (BPA) are an example of ITA. Bertzekas outlines the fact that BPA are notoriously slow to converge. Other incremental methods are based on the Gauss-Newton method and have been proposed to increase the convergence speed. The Extended Kalman Filter (EKF) is one of these methods and was originally conceived as an iterative procedure to solve nonlinear least squares problems. In his paper, Bertzekas combines the advantages of the BPA for large data sets with the convergence rates of the Gauss-Newton method. Thus the algorithm in question cycles repeatedly through the data set. For this reason it, is sometimes called the Iterated Extended Kalman Filter. The author also studies the rates of convergence of this algorithm.

4 Results

The increments of two distinct Lévy processes were simulated using two different infinitely divisible distributions: the first being a Meixner distribution with location, scale, skew and shape parameters equal to 5, 2, 1 and 2 respectively. The second is a mixture of two exponential distributions one with parameter 4, the other with parameter 2. Moreover, each exponential distribution has equal weight, thus $\pi = 0.5$. Afterwards, the estimator discussed in section two together with the relative IEKF algorithm discussed in section three were implemented and the results are illustrated below. It must be said that the quality of the results obtained are comparable the to method proposed by Sant and Caruana [9].
The limits of integration which are present in (1.6) are taken from 0 to some constant *T*, and results are displayed below for different values of *T*.
In each case, 200 simulations were performed; each simulation contained 1000 data points.
The results obtained for the exponential mixture are displayed Table 1 below:

T	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\pi}$	$\widehat{\mathbb{V}\mathrm{ar}(\hat{\lambda}_1)}$	$\widehat{\mathbb{V}\mathrm{ar}(\hat{\lambda}_2)}$	$\widehat{\mathbb{V}\mathrm{ar}(\hat{\pi})}$
10	3.9880	2.0277	0.5008	0.1629	0.0177	0.0002
20	4.0201	2.0087	0.5001	0.1068	0.0150	0.0001

Table 1: estimates of exponential mixture

The results obtained for the Meixner distribution are displayed Table 2. The said table also contains simulation results obtained from the method of maximum likelihood (ML) and method of moments (MM). The results appear to be comparable.

T	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{d}$	$\widehat{\mathbb{V}\mathrm{ar}(a)}$	$\widehat{\mathbb{V}\mathrm{ar}(b)}$	$\widehat{\mathbb{V}\mathrm{ar}(c)}$	$\widehat{\mathbb{V}\mathrm{ar}(d)}$
10	2.00	1.002	4.981	2.016	0.005	0.001	0.021	0.019
20	1.99	1.001	4.999	1.998	0.004	0.001	0.009	0.002
ML	1.99	1.001	4.986	2.011	0.003	0.001	0.018	0.008
MM	1.98	1.003	5.004	1.996	0.004	0.001	0.023	0.014

Table2: estimates of Meixner distribution

Apart from the above results, the histograms illustrated below give a rough indication of the distribution of the difference between the parameter estimate and its actual value. From theorem 3 we know that this converges to a normal distribution. Figures 1 to 3 are related to the parameters of the exponential mixture, while Figures 4 to 7 are related to the parameters of the Meixner distribution.

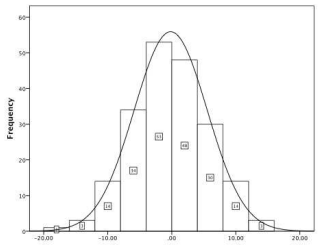


Fig.1. Simulated Distribution of $\sqrt{N}(\hat{\lambda}_1 - \lambda_1)$

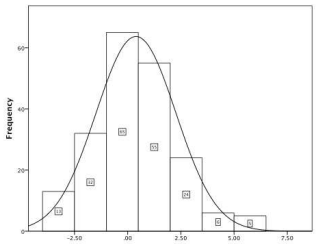


Fig.2. Simulated Distribution of $\sqrt{N}(\hat{\lambda}_2 - \lambda_2)$

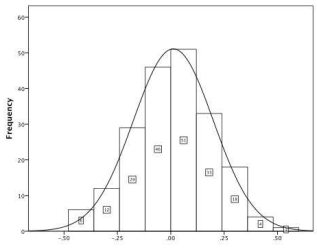


Fig.3. Simulated Distribution of $\sqrt{N}(\hat{\pi} - \pi)$

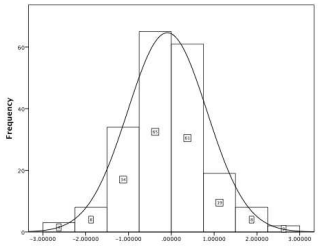


Fig.4. Simulated Distribution of $\sqrt{N}(\hat{a} - a)$

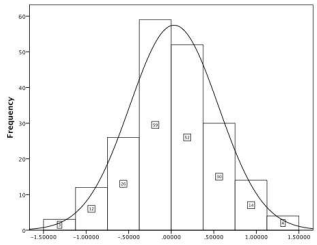


Fig.5. Simulated Distribution of $\sqrt{N}(\hat{b}-b)$

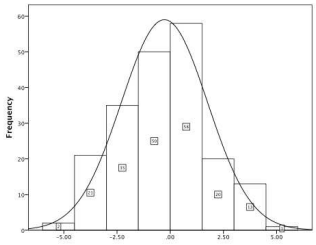


Fig.6. Simulated Distribution of $\sqrt{N}(\hat{c}-c)$

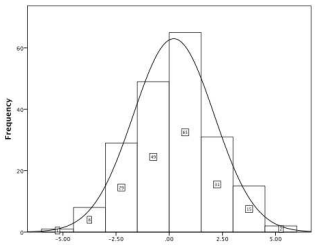


Fig.7. Simulated Distribution of $\sqrt{N}(\hat{d}-d)$

Finally, tests for normality were run using the simulation results. These tests were conducted to substantiate the asymptotic normality result of theorem 3. The p-values shown in table 3 below are all greater than 0.05, this means that the simulation results corroborate the asymptotic normality result.

	Shapiro-Wilk Test		
	Statistic	df	Sig.
$\sqrt{N}(\hat{\lambda}_1-\lambda_1)$	.996	200	.826
$\sqrt{N}(\hat{\lambda}_2-\lambda_2)$	.984	200	.210
$\sqrt{N}(\hat{\pi}-\pi)$	.996	200	.859
$\sqrt{N}(\hat{a}-a)$	.990	200	.203
$\sqrt{N}(\hat{b}-b)$	.991	200	.264
$\sqrt{N}(\hat{c}-c)$	.993	200	.425
$\sqrt{N}(\hat{d}-d)$	.997	200	.977

Table 3. Test for Normality

Conclusions

The estimation technique discussed in this paper combines the field of stochastic programming with the parameter estimation problem of Lévy process. It was shown that the proposed estimator enjoys some nice statistical properties that were discussed in theorem 3, these include, asymptotic unbiasedness and asymptotic normality. The latter was evidenced in the simulation results. Furthermore, the algorithm proposed by Bertzekas[2] allows the parameter estimation technique to be performed recursively. Moreover, as opposed to the method of maximum likelihood or the method of moments, the estimation technique presented in this paper does not require the user to know the functional form of the distribution function of the Lévy increments as it uses the characteristic function. Finally, the simulation results revealed that the estimators for the Meixner distribution were comparable to the estimators obtained from the method of maximum likelihood and the method of moments.

References

1. A. Alessandri, M. Cuneo, S. Pagnan and M. Sanguinetti. A recursive algorithm for nonlinear least-squares problems. Computational Optimization and Applications 38,195{216, 2007.
2. D. P. Bertsekas. Incremental least squares methods and the extended Kalman filter, Society for Industrial and Applied Mathematics 6, 3, 807{822, 1996.
3. F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes, Annales de l'Istitut Henri Poincare – Proabilites et Statistique, 46, 3, 595{617, 2010.
4. J.E. Figueroa-Lopez, Estimation methods for Lévy based model of asset prices, Financial Mathematics Seminar, 2006
5. N. C. Jain and M.B. Marcus. Central limit theorems for C(S)-valued random variables, Journal of Functional Analysis, 19, 3, 216{231, 1975.
6. N. Karmista, M.Tanaka and J. Herskovits. Nonconvex nonsmooth minimisation via cutting planes and feasible direction interior point method, Turku Centre for Computer Science, 2009.
7. A. Levin and V. Khramtsov. Effective Empirical Characteristic Function Methods for Estimation of Affine Diffusions Using the Realized Variance, Available at SSRN: http://ssrn.com/abstract=1993301 or http://dx.doi.org/10.2139/ssrn.1993301 , 2012.
8. H. Morijama, N. Yamashita and M. Fukushima. The incremental Gauss-Newton algorithm with adaptive stepsize rule, Computational Optimization and Applications, 26,107{141, 2003.
9. L. Sant and M. A. Caruana. Estimation of Lévy Processes through Stochastic Programming, Stochastic Modelling Techniques and Data Analysis, Conference Proceedings, 4, 43{50, 2014.
10. W. van der Vaart. Asymptototic Statistics, Cambridge University Press, pg 297, 2000.

Bowel Cancer Demographics

K.C. Chan[1], C.T. Lenard[2], T.M. Mills[3], and Ruth F.G. Williams[4]

[1] Computer Science and Information Technology, La Trobe University
(E-mail: ka.chan@latrobe.edu.au)
[2] Mathematics and Statistics, La Trobe University
(E-mail: c.lenard@latrobe.edu.au)
[3] Loddon Mallee Integrated Cancer Service
(E-mail: tmills@bendigohealth.org.au)
[4] Mathematics and Statistics, La Trobe University
(E-mail: ruth.williams@latrobe.edu.au)

Abstract. Bowel cancer is a major contributor to the burden of disease and injury in Australia. The purpose of this paper is to demonstrate the application of the cumulative rate to Australian bowel cancer statistics. In particular we will investigate the effects of sex, location and time of incidence and mortality of bowel cancer. A feature of this paper is that it uses the cumulative rate rather than the usual age-standardised rates as a basis for comparing incidence and mortality between populations. The cumulative rate gives the same sort of information as the age-standardised rate without the need to introduce a standard population. The results of this paper enhance our understanding of the impact of bowel cancer on Australia, and add to the methodology of descriptive epidemiology of cancer.
Keywords: Epidemiology, Demography, Cancer, Bowel, Colorectal, Cumulative rate, Incidence, Mortality, Australia.

1 Introduction

In this paper we will study the impact of bowel cancer on various subpopulations in Australia. We are interested in how the impact of the disease varies with sex, age, location and how the impact changes over time. Investigations such as this have many applications. They give us a better understanding of the effects of the disease on the population at large; they measure progress in dealing with the disease; and they can inform government on policies and priorities for screening, treatment, prevention, supportive care and research.

A key feature of this work is that our methods for assessing the impact of the disease are based on cumulative rates rather than age standardised rates.

1.1 Definitions

Diseases are classified according to *The International Statistical Classification of Diseases and Related Health Problems* (ICD-10-AM) [10]. Bowel cancer (also known as colorectal cancer) refers to a group of diseases in this classification system. For the purposes of this paper, we define bowel cancer as

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

those diseases classified as C18-C20 in [10]; this is in line with the definition of Australian Institute of Health and Welfare (AIHW) [1].

The *incidence* of a disease is the number of new cases reported in a given population in a given period of time, usually one year. In this paper the population of interest is the population of Australia. The *incidence rate* is the incidence per 100,000 head of population. The *age-specific incidence rate* is the incidence rate for a given age group. The *age-standardised incidence rate* is the incidence rate calculated when we assume that the age profile of the population matches some chosen standard population and the age-specific incidence rates match that of the original population.

The *mortality* is the number of deaths attributed to the disease as the underlying cause of death, in a given population, and a given period of time, usually one year. Again, the population of interest is the population of Australia. The *mortality rate* is the mortality per 100,000 head of population. The *age-specific mortality rate* is the mortality rate for a given age group. The *age-standardised mortality rate* is the mortality rate calculated when we assume that the age profile of the population matches some chosen standard population and the age-specific mortality rates match that of the original population.

1.2 Incidence and mortality statistics

In Australia, cancer is a notifiable disease. Hence Australia has high quality data on the incidence and mortality of cancer over many years.

The incidence of bowel cancer in Australia has increased from 6,960 new cases in 1982 to 14,860 in 2010 as illustrated by the graph in figure 1.

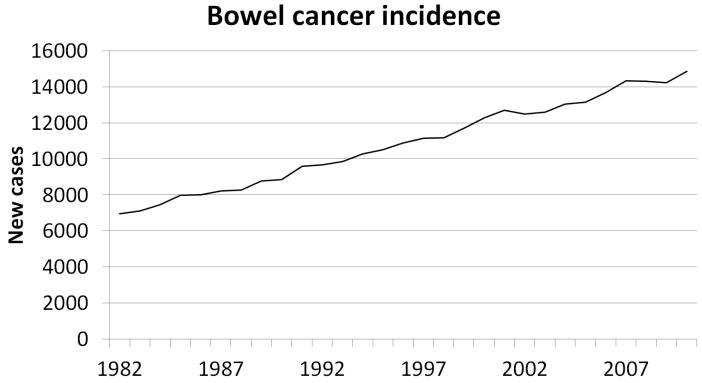


Fig. 1. Incidence of bowel cancer in Australia, 1982–2010

Bowel cancer is one of the leading causes of death attributed to cancer. However, over time, the pattern in mortality from bowel cancer in Australia

is dramatically different from the pattern in incidence. In 1968, bowel cancer was the cause of death for 2,500 Australians; in 2011, bowel cancer was the cause of death for 3,999 Australians [1]; the interesting graph in figure 2 raises many questions.

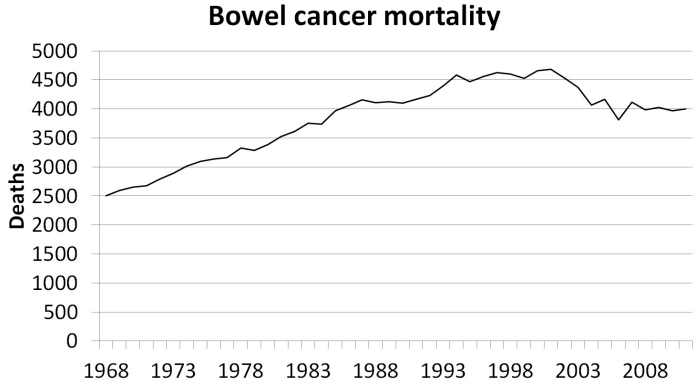


Fig. 2. Mortality from bowel cancer in Australia, 1968–2011

1.3 The cumulative rate

Because age is a major risk factor associated with cancer, one cannot simply compare the incidence or mortality rates of cancer between two populations without taking into account the age profiles of these populations. The traditional way in which one accounts for differences in age distributions is to use age-standardised incidence rates which involve selecting some standard population. In its summary of bowel cancer data, AIHW [1] gives the reader a choice of three standard populations. Yule [13, p.4] regards the introduction of a standard population as "superfluous".

In 1976, Nicholas Day [3] introduced the cumulative rate in the cancer literature as an alternative to the age-standardised rate. As its name suggests, the "cumulative rate" is the sum of age-specific rates for each 1-year age group up to a given age. The principal use of the cumulative rate is for comparing the impact of a disease on independent populations. The cumulative rate has many advantages [2, chapter 2]; it is easy to calculate from incidence (or mortality) data and population data; it does not rely on choosing, arbitrarily, some standard population; it is directly connected to the cumulative risk; its interpretation does not require knowledge of the choice of the standard population; when seen as a sum of age-specific rates, it is easy to interpret. Lancaster [5, p.79–80] states that the cumulative rate (which he calls the equivalent average rate) is "simpler" than the the usual age-standardised

rate; Lancaster reiterates this view in [6, p.34]. Despite these attractive features, the cumulative rate has not been as popular as one might expect.

There is an important link between cumulative rates and risks. The cumulative rate of being diagnosed with bowel cancer by age 75 is related to the risk of being diagnosed with the disease by age 75 [3]. Indeed there is a 1-1 transformation from one measure to the other. We eschew the use of risk as a measure because it is easily misunderstood, as we have explained elsewhere [7], [8]. An exposition of the theory on which the cumulative rate and risk are based can be found in [8]. The method for calculating cumulative rates, and testing for differences in two populations was outlined in [9] using hypothetical data.

In this paper we examine how the cumulative rate performs when applied to data on bowel cancer in Australia.

1.4 Research questions

We will address the following questions. How does the incidence of bowel cancer in Australia among men compare with the incidence among women? How has the incidence of the disease changed over time? How has the incidence among younger Australians changed over time? How does incidence in the state of Victoria compare with the incidence in the rest of Australia? We will address the same questions about mortality rather than incidence.

There are two distinctive statistical features of this paper. First, and foremost, our method for answering these questions will be based on the cumulative rate (rather than risk, or age-standardised rate). Second, we will use the methods of Dobson et al. [4] to calculate confidence intervals for cumulative rates. Hence we are interested in two statistical questions. How does the cumulative rate perform in terms of measuring inequalities between populations? How does the method of Dobson et al. for calculating confidence intervals perform in the context of cumulative rates?

2 Data and methods

2.1 Data

Data for Australia in this project were obtained from AIHW [1]. The analysis is based on three data sets: the incidence of bowel cancer in Australia for the years 1982–2010 classified by age, the mortality from bowel cancer for 1968–2011 classified by age, and the population of Australia for these years; all data are classified in five year age groups. These data were the most recent that were available when we started this project. Data for Victoria were obtained separately from Victorian Cancer Registry.

For the purposes of this research, only data for the age groups 0–4, 5–9, . . . , 70–74 were used because we will calculate cumulative rates to age 75.

2.2 Calculations

Although the methods for calculating cumulative rates were outlined in [9], we present them here in the context of incidence data for the sake of completeness. The calculations for mortality data are similar.

Group (i)	Age	Population	Observed incidence
1	0–4	n_1	x_1
2	5–9	n_2	x_2
⋮	⋮	⋮	
15	70–74	n_{15}	x_{15}

Table 1. Data in five-year age groups

Cumulative rate and standard deviation. Suppose that the incidence and population data are as in table 1. For each age group $i\,(=1,2,\ldots,15)$, assume that the incidence of the disease is a random variable X_i with a Poisson distribution with mean value θ_i, and that $X_1, X_2, \ldots, X_{15}$ are independent random variables. (In table 1, x_i is an observed value of X_i.) Then $E(X_i) = \text{Var } X_i = \theta_i$. Define the random variable

$$A(75) := 5\left(\sum_{i=1}^{15} \frac{X_i}{n_i}\right).$$

The expected value of $A(75)$ is

$$E(A(75)) = 5\left(\sum_{i=1}^{15} \frac{\theta_i}{n_i}\right)$$

and the variance of $A(75)$ is

$$\text{Var } A(75) = 25\left(\sum_{i=1}^{15} \frac{\theta_i}{n_i^2}\right).$$

We can estimate $E(A(75))$ by

$$\text{Cumulative incidence rate by age } 75 \;= a(75) := 5\left(\sum_{i=1}^{15} \frac{x_i}{n_i}\right).$$

We can estimate the standard deviation of $A(75)$ by

$$s(75) := 5\sqrt{\sum_{i=1}^{15} \frac{x_i}{n_i^2}}.$$

We note, in passing, that the estimated risk of being diagnosed with bowel cancer by age 75 is given by $1-\exp(-a(75))$. But we are not concerned with this measure for reasons mentioned above.

Normal approximation for confidence interval. One calculates an approximate $100(1-\alpha)\%$ confidence interval for the theoretical cumulative rate, $E(A(75))$, using a Normal approximation by

$$a(75) \pm z_{\alpha/2}\, s(75)$$

where, if Z has a standard Normal distribution, then $P(|Z| > z_{\alpha/2}) = \alpha$. In section 3 below, we refer to this method of calculating the CI as the Normal approximation.

Dobson approximation for confidence interval. In a delightful paper, Dobson et al. [4] noted that the above Normal approximation "may be poor" especially if the x_i are small—as will be the case in bowel cancer statistics. Dobson et al. propose an alternative approach based on the observation that $E(A(75))$ is a linear combination of Poisson parameters. Their approach can be applied to find approximate $100(1-\alpha)\%$ confidence intervals for the theoretical cumulative rates as follows.

Let $\chi^2(m, \alpha/2)$ satisfy $P(X < \chi^2(m, \alpha/2)) = \alpha/2$ where X is a random variable with a χ^2-distribution with m degrees of freedom. Let

$$\begin{aligned}
T &:= \sum_{i=1}^{15} x_i\,, \\
X_L &:= \frac{1}{2}\chi^2(2T, \alpha/2)\,, \\
X_U &:= \frac{1}{2}\chi^2(2T+2, 1-\alpha/2)\,, \\
T_L &:= a(75) - \sqrt{\frac{s(75)^2}{T}}\ (T - X_L)\,, \\
T_U &:= a(75) + \sqrt{\frac{s(75)^2}{T}}\ (X_U - T)\,.
\end{aligned}$$

Then the interval (T_L, T_U) is an approximate $100(1-\alpha)\%$ confidence interval for the theoretical cumulative rate. We will refer to this method of calculating the CI as the Dobson method for brevity without wishing to disregard the contributions of Dobson's co-authors.

Hypothesis testing. When comparing two populations, let us denote the cumulative rates by $a_1(75), a_2(75)$ respectively, and the standard deviations

by $s_1(75), s_2(75)$ respectively. Under the null hypothesis that the mean cumulative incidence rates in the two populations are equal, then the distribution of

$$z := \frac{a_1(75) - a_2(75)}{\sqrt{s_1(75)^2 + s_2(75)^2}}$$

is, approximately, a standard Normal distribution. This allows us to test the null hypothesis.

3 Results

We now summarise the results of our analyses of incidence and mortality data.

3.1 Incidence

Incidence and sex. In this subsection, we report the results from comparing the incidence of bowel cancer among men in Australia with the incidence of bowel cancer among women in the year 2010, which is the most recent data for incidence. The results are summarised in table 2.

	Male	Female
Cum. rate by age 75	0.0551	0.0367
Est. s.d. of cum. rate by age 75	0.0008	0.0006
95% CI for cum. rate (Normal)	[0.0536, 0.0566]	[0.0355, 0.0379]
95% CI for cum. rate (Dobson)	[0.0536, 0.0567]	[0.0355, 0.0380]
z (Normal approx.)	17.75	
$p = P(\|Z\| > z)$	$< .001$	

Table 2. Incidence of bowel cancer, Australia, 2010

The graph in figure 3 summarises the results from comparing incidence rates of bowel cancer among men in Australia with the incidence rates among women, for 1982–2010, using the cumulative incidence rate by age 75.

Incidence and location. In this subsection, we report the results from comparing the incidence of bowel cancer in Victoria with the incidence of bowel cancer in the rest of Australia in the year 2010. Data for Victoria was obtained from Thursfield and Farrugia [11]. The results are summarised in table 3.

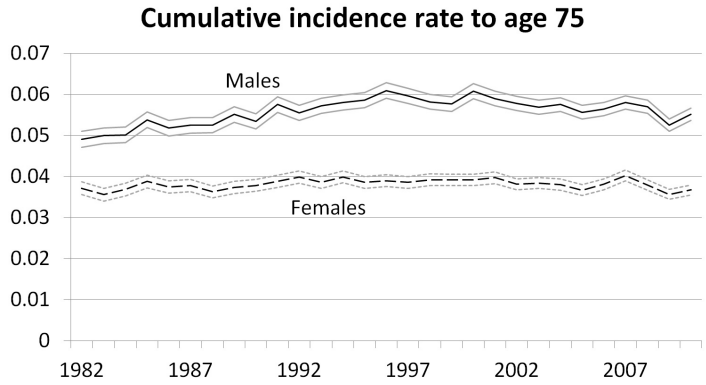


Fig. 3. Cumulative incidence rates to age 75 of bowel cancer in Australia, 1982–2010

	Victoria	Rest of Australia
Cum. rate by age 75	0.0442	0.0464
Est. s.d. of cum. rate by age 75	9.343E-07	3.298E-07
95% CI for cum. rate (Normal)	[0.0423, 0.0462]	[0.0453, 0.0475]
95% CI for cum. rate (Dobson)	[0.0442,0.0443]	[0.0452, 0.0475]
z (Normal approx.)	-1.89	
$p = P(\|Z\| > z)$	0.0581	

Table 3. Incidence of bowel cancer, Australia, 2010

Incidence among Australians under 40. Although bowel cancer is primarily a disease that affects older people, sadly it also affects younger people too, but not to the same extent. Young et al. [12] have recently published a review of the incidence of colorectal cancer among younger Australians. One of their conclusions is that incidence rates "are increasing in patients diagnosed under 40 years in Australia in contrast to stable or declining rates in older patients". This point is not widely appreciated. Hence, we investigated this matter using the cumulative rate up to age 40.

The graph in figure 4 summarises the results from comparing incidence rates of bowel cancer among men in Australia with the incidence rates among women, for 1982–2010, using the cumulative incidence rate by age 40.

We investigated the cumulative incidence rate to various ages, not only 75 and 40. The graphs in figure 5 represent the cumulative incidence rates for all persons across a wide range of ages for selected years.

3.2 Mortality

Mortality and sex. In this subsection, we report the results from comparing the mortality from bowel cancer among men in Australia with the mortality

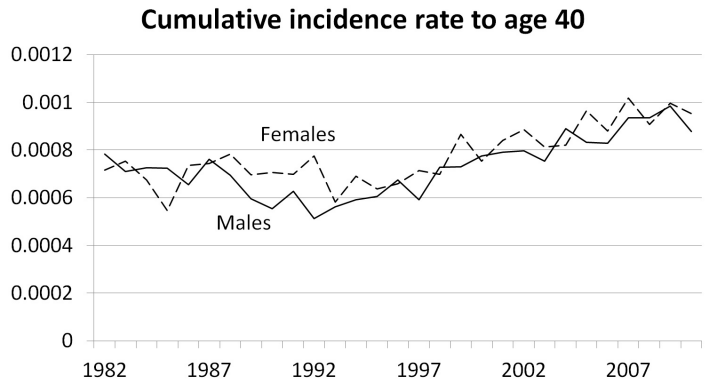


Fig. 4. Cumulative incidence rates to age 40 of bowel cancer in Australia, 1982–2010

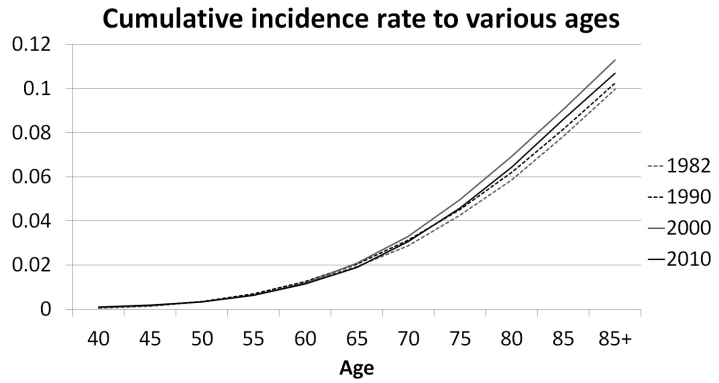


Fig. 5. Cumulative incidence rates for various ages of bowel cancer in Australia for selected years

from bowel cancer among women in the year 2011, which is the most recent data for mortality. The results are summarised in table 4.

	Male	Female
Cum. rate by age 75	0.0122	0.0068
Est. s.d. of cum. rate by age 75	0.0004	0.0003
95% CI for cum. rate (Normal)	[0.0115, 0.0130]	[0.0062, 0.0073]
95% CI for cum. rate (Dobson)	[0.0115, 0.0126]	[0.0063, 0.0073]
z (Normal approx.)	11.69	
$p = P(\|Z\| > z)$	$< .001$	

Table 4. Mortality from bowel cancer, Australia, 2011

Mortality over time. The graph in figure 6 summarises the results from comparing mortality rates of bowel cancer among men in Australia with the mortality rates among women, for 1968–2011, using the cumulative incidence rate by age 75.

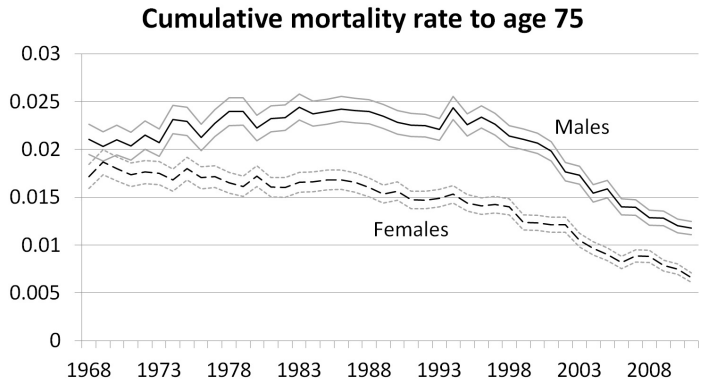


Fig. 6. Cumulative mortality rates to age 75 for bowel cancer in Australia, 1968–2011

Mortality among Australians under 40. The graph in figure 7 summarises the results from comparing mortality rates of bowel cancer among men in Australia with the mortality rates among women, for 1968–2011, using the cumulative incidence rate by age 40.

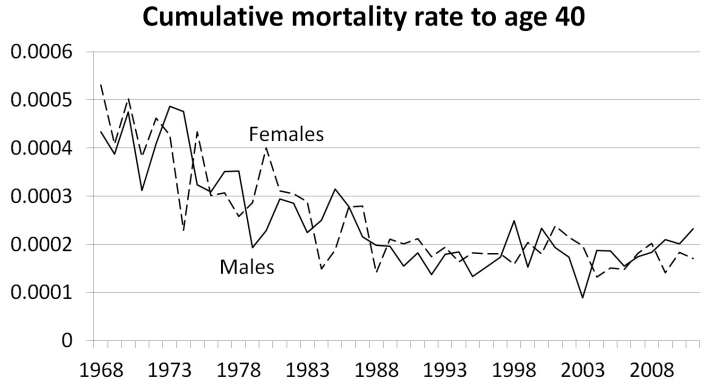


Fig. 7. Cumulative mortality rates to age 40 for bowel cancer in Australia, 1968–2011

We investigated the cumulative mortality rates to various ages, not only 75 and 40. The graphs in figure 8 represent the cumulative mortality rates for all persons across a wide range of ages for selected years.

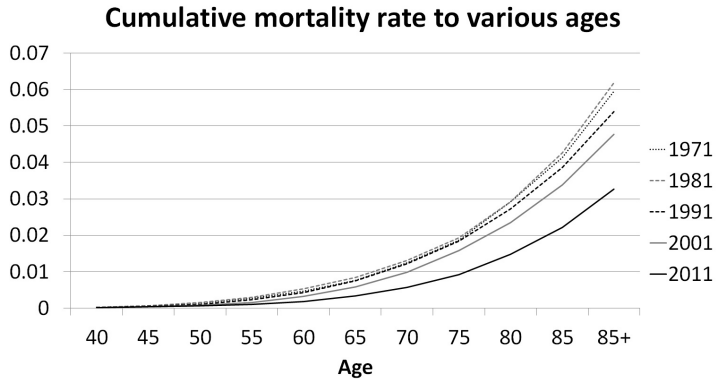


Fig. 8. Cumulative mortality rates for various ages of bowel cancer in Australia for selected years.

Mortality and location. In this subsection, we report the results from comparing the mortality from bowel cancer in Victoria with the mortality from bowel cancer in the rest of Australia in the year 2010. Data for Victoria was kindly provided by Victorian Cancer Registry. The results are summarised in table 5,

	Victoria	Rest of Australia
Cum. rate by age 75	0.0122	0.0089
Est. s.d. of cum. rate by age 75	0.0005	0.0003
95% CI for cum. rate (Normal)	[0.0111, 0.0132]	[0.0084, 0.0094]
95% CI for cum. rate (Dobson)	[0.0112, 0.0132]	[0.0084, 0.0094]
z (Normal approx.)	5.6382	
$p = P(\|Z\| > z)$	< 0.001	

Table 5. Mortality from bowel cancer, Australia, 2010

4 Discussion

We now discuss the results presented in the previous section.

4.1 Bowel cancer in Australia

Bowel cancer has a bigger impact on men than on women. From table 2 and figure 3, it is clear that the cumulative incidence rate to age 75 has been higher among men than among women in Australia since 1982.

The graphs in figure 1 and figure 3 illustrate a key point about the cumulative rate. Although the incidence of bowel cancer has been steadily increasing over time, the graph of the cumulative incidence rate is similar to the graph of the age-standardised incidence rate reported by AIHW [1] which uses the Australian 2001 standard population. Thus the cumulative rate gives the same sort of information as the age standardised rate without the need to introduce a standard population.

From table 4, it is clear that the cumulative mortality rate of bowel is higher among men than among women in Australia ($p<.001$). Figure 6 makes it clear that this inequality has been evident for a long time. The decrease in mortality rates for men and women in recent years is a notable feature of these graphs.

From table 3, we see that the cumulative incidence rate of bowel cancer in Victoria is slightly lower than in the rest of Australia but not significantly so ($p=0.0581$). By contrast, from table 5, we see that the cumulative mortality rate of bowel cancer in Victoria is significantly higher than in the rest of Australia ($p<.001$). This is a surprising discovery. It would be interesting to us to compare incidence and mortality across all the states of Australia.

The graphs in figure 5 and figure 8 raise several questions that deserve further consideration. For example, at what ages do the differences in incidence and mortality between males and females appear to emerge? And then there is the harder question: why?

4.2 Statistical issues

A limitation of this study is that we have used many 95% confidence intervals. Thus the probability that some of our intervals do not contain the parameter of interest will be much larger than 5%.

Overall, in the context of our study, the Normal approximation for calculating confidence intervals seemed to be as effective as the method proposed by Dobson et al. [4] in nearly every case. (The results in table 3 are exceptional.)

4.3 Measuring inequality

Cancer can affect anyone. Yet, the impact is not uniform with respect to age, sex, geographical location, and year of diagnosis. These inequalities raise serious questions about setting priorities and allocating resources to deal with the disease. There are many ways in which the cumulative rate can be used to identify inequalities in the impact of cancer on different populations.

5 Acknowledgements

We thank the organisers of 16th conference of the Applied Stochastic Models and Data Analysis (ASMDA) International Society, 30 June–4 July 2015 at the University of Piraeus for the opportunity to present this work. We thank Victorian Cancer Registry for providing data on bowel cancer in Victoria.

References

1. Australian Institute of Health and Welfare (AIHW). ACIM (Australian Cancer Incidence and Mortality) Books [monograph on the internet]. Canberra: AIHW. (2014) [cited 12 Nov 2014]. Available from: http://www.aihw.gov.au/cancer/
2. N.E. Breslow and N.E. Day. *Statistical Methods in Cancer Research, Vol. II: The Design and Analysis of Cohort Studies.* International Agency for Research on Cancer, Lyon, 1987.
3. N. Day. A new measure of age standardized incidence, the cumulative rate. In: R. Payne and J. Waterhouse (eds) *Cancer incidence in five continents.* Vol. III. IARC Scientific Publications, No. 15. International Agency for Research on Cancer, Lyon, 443–452, 1976.
4. A.J. Dobson, K. Kuulasmaa, E. Eberle, and J. Scherer. Confidence intervals for weighted sums of Poisson parameters. *Statistics in Medicine*, 10, **3**, 457–462, 1991.
5. H.O. Lancaster. *An Introduction to Medical Statistics.* John Wiley & Sons, New York, 1974.
6. H.O. Lancaster. *Expectations of Life. A Study in the Demography, Statistics, and History of World Mortality.* Springer-Verlag, New York, 1990.
7. C.T. Lenard, T.M. Mills, and R.F.G. Williams. The risk of being diagnosed with cancer. *Australian & New Zealand J. Public Health*, 37, **5**, 489, 2013.
8. C.T. Lenard, T.M. Mills, and R.F.G. Williams. Cumulative incidence rates of cancer. *The Mathematical Scientist*, 39, **2**, 83–89, 2014.
9. C.T. Lenard, T.M. Mills, and R.F.G. Williams. Comparing the cumulative rates of cancer in two regions. In: C. H. Skiadas (ed.) *Proc. SMTDA2014, 3rd Stochastic Modeling Techniques and Data Analysis International Conference*, Lisbon, 11-14 June, 399–403, 2014.
10. National Centre for Classification in Health (Sydney). *The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification*, Fifth edition, Sydney: National Centre for Classification in Health, Faculty of Health Sciences, The University of Sydney, 2006.
11. V. Thursfield and H. Farrugia. *Cancer in Victoria: Statistics and Trends 2010.* Cancer Council Victoria, Melbourne, 2011.
12. J. P. Young, A. K. Win, C. Rosty, I. Flight, D. Roder, G. P. Young, O. Frank, G. K. Suthers, P. J. Hewett, A. Ruszkiewicz, Ehud Hauben, B-A. Adelstein, S. Parry, A. Townsend, J. E. Hardingham, T. J. Price. Rising incidence of early-onset colorectal cancer in Australia over two decades: Report and review. *Journal of Gastroenterology and Hepatology*, (to appear).
13. G. Udny Yule. On some points relating to vital statistics, more precisely statistics of occupational mortality. *Journal of the Royal Statistical Society*, 97, **1**, 1–84, 1934.

Investigating the structure of Schwartz's Human Values Scale

Anastasia Charalampi[1], Catherine Michalopoulou[2] and Clive Richardson[3]

[1] Ph.D. Candidate, Department of Social Policy, Panteion University of Political and Social Sciences, Athens, Greece
(E-mail: natashaharal@yahoo.gr)
[2] Associate Professor of Statistics, Department of Social Policy, Panteion University of Political and Social Sciences
(E-mail: kmichal@panteion.gr)
[3] Professor of Social Statistics, Department of Economic and Regional Development, Panteion University of Political and Social Sciences
(E-mail: crichard@panteion.gr)

Abstract. *Purpose:* To investigate the dimensionality and psychometric properties of the Schwartz Human Values Scale Short Form included in the European Social Survey (ESS) questionnaire.
Method: Greek and Slovenian ESS data of 2002 were used. Each sample was split randomly into two halves and Exploratory Factor Analysis (EFA) was performed on one half-sample in order to assess the construct validity of the scales. The structure was validated by carrying out Confirmatory Factor Analysis (CFA) on the second half.
Results: In both countries, EFA resulted in a three-factor solution but the third subscale was found not to be reliable. In the case of Greece, CFA showed acceptable fit for a 16-item model defined by two first-order correlated factors, Openness to change and Self-transcendence. In the case of Slovenia, CFA showed best fit for a 13-item model defined by three first-order correlated factors, Openness to change, Self-transcendence and Self-enhancement.
Conclusions: Although our analysis did not confirm the dimensionality of the Schwartz Human Values Scale as proposed in the literature, it did produce two subscales for Greece and three subscales for Slovenia that were reliable and valid. Our results suggest that further research is necessary in each country in order to provide subscales suitable for use in analyses.
Keywords: Schwartz's human values scale (PVQ-21); Reliability; Construct validity; Exploratory Factor Analysis; Confirmatory Factor Analysis

1 Introduction

The European Social Survey aims to measure attitudes, beliefs and behaviour patterns of populations across Europe. It has been conducted every two years since 2002 and 36 countries have taken part in one or more rounds. One component that has been present in every round is Schwartz's Human Values Scale, which is designed to classify respondents according to their value orientation. As with any scale, it is important to carry out detailed analysis of its psychometric properties and, in the case of a scale to be used in cross-national research, to establish that these are the same in each country. The present paper contributes to this investigation of the structure and properties of the Human Values Scale.

In 1992, Schwartz developed his theory of basic human values which has been widely used by social and cross-cultural psychologists in order to study differences in values among individuals [8]. This theory includes the following ten motivationally distinct basic values which encompass the major value orientations recognized cross-culturally: Self-direction (SD), universalism (UN), benevolence (BE), tradition (TR), conformity (CO), security (SEC), power (PO), achievement (AC), hedonism (HE) and stimulation (ST) [6-7]. Schwartz derived these values from three universal requirements of the human condition: needs of individuals as biological organisms, requisites of coordinated social interaction and requirement for the survival and welfare needs of group [6-7, 12].

Schwartz presented the ten basic values defined as sub-dimensions in a circular structure based on the relations of conflict and congruity among the types of values [7]. More similar value types are close to each other in either direction around the circle and consequently share more similar underlying motivations. On the other hand, conflicting value types appear on opposite sides of the circle and have more antagonistic underlying motivations [7-8]. Moreover, the circular structure also summarizes two dimensions of relations between these values: the self-enhancement versus self-transcendence dimension opposes power and achievement values to universalism and benevolence values, and the openness to change versus conservation dimension opposes self-direction and

stimulation values to security, conformity and traditional values; hedonism shares elements of both openness to change and self-enhancement [7-8].

Lilleoja and Saris [14] pointed out that Schwartz first used a 57-item questionnaire in his survey (Schwartz's value survey, SVS), which was later replaced by the 40-item Portrait Value Questionnaire (PVQ). The European Social Survey (ESS) Human Values Scale was derived from the earlier 40-item PVQ, but because of space limitations, the number of items was reduced to 21 (PVQ-21) [7]. According to Knoppen and Saris [12], ESS selected Schwartz's value scale because it was considered to be one of the most comprehensive models and also had been widely validated across cultures.

The ESS PVQ-21 questionnaire is administered as a self-completion questionnaire after the end of the interview. Each item presents a verbal portrait of a different person, worded according to the respondent's gender, describing that person's goals, aspirations or wishes showing implicitly the importance of a value [7]. Each of the ten values listed above is represented by two items, apart from the universalism value, which is expressed by three items. There are six possible responses to each item as follows: 1 (very much like me), 2 (like me), 3 (somewhat like me), 4 (a little like me), 5 (not like me) and 6 (not like me at all). The score for each respondent is calculated by averaging his or her responses on the items defining each value, i.e. ten subscale scores are constructed by computing the mean of items that measure each one [7].

Davidov, Schmidt and Schwartz [7: 440-441] showed that "the scale failed to exhibit scalar invariance across the 20 countries. Hence, one should not compare the mean importance of the values across all 20 countries simultaneously. However, as illustrated for Denmark and Spain, one can compare means for values across subsets of countries where scalar invariance or partial scalar invariance are found." In each country, they found that there were at least two pairs of values which were dependent on each other that could not be separated. In order to solve the problem of non-positive definite covariance matrices of the constructs, Davidov, Schmidt and Schwartz [7] unified in pairs the strongly associated values. Their results showed that there were between five and eight distinct values in the different countries. They found that 69 out of 71 pairs of unified values across the 20 countries were adjacent in the circular structure of the Schwartz theory of values. Their analyses for Greece and Slovenia resulted in a similar structure containing the following five unified values: PO/AC, CO/TR, UN/BE, HE/ST and ST/SD. The purpose of the present paper is to investigate the dimensionality of the Schwartz scale for these two countries by applying both Exploratory and Confirmatory Factor Analysis and to indicate how the values should be treated in country-level analyses.

2 Method

The items comprising Schwartz's scale of human values are Likert-type with six response categories and, therefore their level of measurement is ordinal. In applications where the number of response categories used for each item is at least five, ordinal categories can be understood as being interval and one may perform statistical analyses using these pseudo-interval variables [2].

In the first step of the analysis, the sample in each country was randomly split into two halves. For the data of the first split-half sample, item analysis was carried out to examine their distributional properties and decide on the items to be included in the analysis. For the construct validity of the scale, Exploratory Factor Analysis was performed using IBM SPSS Statistics Version 20.

In the second step of analysis, the structure was validated by carrying out Confirmatory Factor Analysis on the second split-half sample using IBM SPSS AMOS Version 21.

2.1 Exploratory Factor Analysis (EFA)

The size of the first split-half samples for Greece (N = 1,283) and Slovenia (N = 759) was considered adequate for factor analysis (KMO = 0.901 for Greece and KMO=0.831 for Slovenia; see also Tabachnick and Fidell [23]). First, Principal Axis Factoring (PAF) was performed to define factors as subscales and factor loadings were reported [9]. In performing PAF, the following sequence of decisions was required [4, 23-24]:

1. Initially, univariate statistics were computed for each item and their distributional properties were inspected (testing for normality) to decide on the appropriateness of the methods to be used. Also, corrected item-total correlations were computed and items meeting the criteria of correlations greater than .30 and extraction communalities greater than .40 were included in the analysis [15, 22].

2. Missing data was replaced by the mean values (which for most items coincided with the median). Data screening for unengaged responses (standard deviation = .000) in the Greek and Slovenian data sets identified only three and five cases, respectively, and it was decided not to reject them from analysis. Data screening for outliers was based on the following background variables: gender (dichotomy), age (ratio), education (pseudo-interval). In the case of Greece, only four outlying

cases with Higher Education degree were detected and it was decided not to reject them from analysis. In the case of Slovenia, outliers were not detected.

3. The decision on the number of factors to be extracted was based on the eigenvalue greater than 1.0 rule, scree test, parallel analysis and interpretability [11, 22]. Parallel analysis [13, 16, 19] was performed using the parallel analysis engine provided by Patil, Singh, Mishra and Donavan [17-18].

4. Factor rotation method: Promax (oblique) rotation was applied [9]. Items with loadings greater than .30 on one factor and greater than .22 on another factor [1, 21] were considered as "cross-loading" items, i.e. items that loaded on multiple factors.

5. Subscales were constructed. Descriptive statistics, Cronbach's alpha and split-half reliability coefficients of the subscales were computed. A subscale was considered reliable if the coefficients were ≥.70 [15]. Average inter-item correlations in the recommended range of .15-.5 that clustered near their mean value were used as an indication of the unidimensionality of the subscales [5]. To determine whether or not subscales were warranted, the condition of average correlation between subscale items "significantly greater than zero but substantially less than the average within-subscale values (say, .20)" [5: 318] was used for justifying subscales. As Clark and Watson [5: 318] pointed out, "if this condition cannot be met, then the subscales should be abandoned in favor of a single overall score".

2.2 Confirmatory Factor Analysis (CFA)

The size of the second split-half samples for Greece (N = 1,283) and Slovenia (N = 760) was considered adequate for factor analysis [23]. In performing CFA, the following sequence of decisions was required [3, 4, 10, 24]:

1. The decision on the items to be included in the analysis was based on the item analysis results carried out before performing EFA. Missing data was replaced by the mean or median values as in the case of EFA. Data screening for unengaged responses in the Greek and Slovenian data sets resulted in only eleven and one cases-respondents, respectively, and it was decided not to reject them from analysis. In the case of Greece, only four outlying cases with Higher Education degree were detected and it was decided not to reject them from analysis. In the case of Slovenia, outliers were not detected.

2. CFA was performed using the covariance matrix of associations and using maximum likelihood for estimation.

3. Rival models: It was decided to consider the following models: one first-order factor (model 1); two first-order correlated factors employing all items (model 2a); two first-order correlated factors employing all items (model 2b) with cross-loadings; two first-order correlated factors based on the solution obtained from EFA with consideration of the subscales' reliability (model 2c); two first-order correlated factors based on the solution obtained from EFA with consideration of the subscales' reliability (model 2d) with cross-loadings; three first-order correlated factors based on the EFA results (model 3a); three first-order correlated factors based on the EFA results (model 3b) with cross-loadings and the five first-order correlated factors model based on Davidov and Schwartz's [4] results (model 4).

Lilleoja and Saris [14: 157] pointed out that "Schwartz has criticized CFA approach, because it contradicts the view of values as arrayed on a continuum, as it seeks to confirm relatively pure factors and each item ideally loads on only one factor [20]. The latter remark is not true because cross-loadings are in principle allowed in CFA, but in that case they have to be specified in the model. If they are ignored, the misspecification leads to improper estimates, like correlations larger than 1.0." Therefore, the presentation of cross-loadings in CFA is required and models 2, 3 and 4 were run again by considering the respective cross-loadings resulting from EFA. Where necessary, error variances were correlated.

4. Model Fit statistics: In CFA, model fit was considered adequate when χ^2/df was smaller than 3; Standardized Root Mean Square Residual (SRMR) was lower than 0.05; the Comparative Fit Index (CFI) and the Adjusted Goodness-of-Fit Index (AGFI) were greater than 0.95, the Normed Fit Index (NNFI) was greater than 0.95 and the Root-Mean-Square Error Approximation (RMSEA) was lower than 0.06 [3, 19, 24].

5. Model misspecification searches: searches for modification indices [3, 24].

2.3 Subscale construction and reliability assessment

Subscales were constructed and assessed for the total sample as in step 5 of the EFA sequence of decisions.

3 Results

3.1 Results of EFA

3.1.1 Greece. ESS Round 1 (2002)

The majority of the responses were clustered at the lower end of the scale in the first two response categories (Table 1). Low mean responses were found for items defining Security (SEC5, SEC14), Tradition (TR20), Universalism (UN3, UN19) and Benevolence (BE12, BE18). Relatively high mean responses were found for items defining Stimulation (ST15), Power (PO2), Achievement (AC4, AC13) and Hedonism (HE10).

As shown in Table 1, the proportion of missing values for all the items was negligible, ranging from 0.2 to 0.6%. Testing for randomness indicated that the data was probably missing completely at random (Little's MCAR test: χ^2 = 146.5, *df* = 132, *p* = 0.18). Non-normality was not severe for any item (skewness>2; kurtosis>7). Based on the criteria of corrected item-total correlations and extraction communalities, the following three items were rejected from analysis: TR9 ("It is important to him to be humble and modest. He tries not to draw attention to himself"); TR20 ("Tradition is important to him. He tries to follow the customs handed down by his religion or his family"); CO7 ("He believes that people should do what they are told. He thinks people should follow rules at all time, even when no-one is watching").

Table 1. Item analysis of Schwartz scale values for ESS – 2002: Greece (*n*=1,283)

Item	Mean (SD)	95% CI for mean	Frequency percent of response categories							Skew	Kurt.	CC
			1	2	3	4	5	6	NA			
SD1	2.34 (1.10)	2.27-2.40	22.1	41.7	23.1	6.9	5.1	0.9	0.3	0.92	0.69	0.47
SD11	2.04 (0.98)	1.99-2.10	31.0	44.5	16.5	4.8	2.1	0.7	0.4	1.19	1.88	0.52
UN3	1.79 (0.89)	1.74-1.84	42.8	42.2	10.5	2.5	1.5	0.4	0.2	1.50	3.20	0.37
UN8	2.11 (0.96)	2.06-2.16	26.0	47.5	18.8	4.5	1.9	0.9	0.4	1.18	2.12	0.48
UN19	1.81 (0.85)	1.77-1.86	40.8	41.5	13.8	2.7	0.6	0.3	0.3	1.17	2.10	0.47
BE12	2.00 (0.91)	1.95-2.05	30.2	46.6	17.7	3.4	1.0	0.7	0.4	1.17	2.43	0.40
BE18	1.81 (0.85)	1.76-1.85	39.8	44.6	11.9	2.0	0.7	0.6	0.4	1.46	3.76	0.41
TR9	2.27 (1.08)	2.21-2.33	24.9	40.7	22.1	8.0	3.3	1.0	0.2	0.93	0.80	0.11
TR20	1.73 (0.89)	1.68-1.78	49.2	34.9	11.0	3.5	0.8	0.4	0.2	1.45	2.57	0.22
CO7	2.52 (1.28)	2.45-2.59	20.7	38.9	20.4	10.0	6.8	3.0	0.3	0.90	0.24	0.23
CO16	2.13 (1.02)	2.07-2.18	26.8	46.5	17.6	4.9	2.6	1.1	0.5	1.24	2.05	0.38
SEC5	1.71 (0.89)	1.66-1.76	49.6	35.6	9.2	3.7	1.0	0.3	0.5	1.52	2.75	0.37
SEC14	1.71 (0.92)	1.66-1.76	50.4	35.5	8.6	2.9	1.9	0.4	0.3	1.69	3.44	0.43
PO2	3.58 (1.39)	3.50-3.66	6.4	13.4	36.1	15.4	16.9	11.6	0.2	0.17	-0.75	0.33
PO17	2.47 (1.25)	2.40-2.54	23.9	33.8	24.2	8.7	7.2	1.8	0.3	0.79	0.08	0.40
AC4	2.75 (1.33)	2.68-2.82	17.8	30.6	26.0	13.2	8.7	3.6	0.3	0.60	-0.30	0.51
AC13	2.84 (1.28)	2.77-2.91	13.9	31.2	27.0	14.8	9.7	2.8	0.6	0.51	-0.39	0.55
HE10	2.92 (1.38)	2.84-2.99	14.1	29.3	29.0	10.9	10.7	5.6	0.4	0.59	-0.38	0.47
HE21	2.53 (1.32)	2.46-2.60	22.9	34.1	23.8	8.9	5.6	4.4	0.3	0.91	0.33	0.52
ST6	2.68 (1.33)	2.61-2.76	19.5	31.7	24.4	12.9	7.6	3.7	0.3	0.67	-0.20	0.53
ST15	3.76 (1.59)	3.68-3.85	9.1	16.5	19.1	17.4	19.6	18.1	0.2	-0.13	-1.14	0.34

Notes: SD = standard deviation; CI = confidence interval; NA = no answer (missing values); Kurt. = kurtosis; CC = corrected item-total correlation. The values were defined as follows: SD=self-direction; UN=universalism; BE=benevolence; TR=tradition; CO=conformity; SEC=security; PO=power; AC=achievement; HE=hedonism; ST=stimulation. Item number indicates order as presented in questionnaire. Items were assigned the following response categories: 1 (very much like me), 2 (like me), 3 (somewhat like me), 4 (a little like me), 5 (not like me) and 6 (not like me at all). Standard errors for skewness and kurtosis were 0.068 and 0.137, respectively.

The eigenvalue rule >1 and scree test suggested the retention of a three-factor solution that best explained the variance when eigenvalues from the target data set were compared to the average and the 95th percentile of the random data sets. Parallel analysis confirmed this result as actual eigenvalues (7.472, 2.969, 1.980) were greater than the randomly generated ones for both the average (1.210, 1.172, 1.141) and the 95th percentile (1.248, 1.199, 1.166) eigenvalue criteria.

Table 2 presents the structure of the three-factor solution obtained by Principal axis factoring. The first factor refers mostly to Openness to change, the second factor expresses Self-transcendence and the third factor is defined by Self-enhancement. These three factors explain 31.595, 12.556 and 8.372% of the variance, respectively. Almost all of the 18 items possessed strong factor loadings ≥0.45 on at least one factor.

Subscales were constructed by averaging the defining items of each factor. As shown in Table 3, Cronbach's alpha reliability coefficients for the subscales Openness to change, Self-transcendence and Self-enhancement were 0.819, 0.806 and 0.467, respectively, indicating that the third factor was not reliable. Split-half reliabilities were 0.821, 0.792 and 0.467, respectively. Average inter-item correlations were 0.368, 0.344 and 0.306 within subscales and 0.239, 0.334 and 0.271 between subscales, indicating that the values were within the recommended range.

Table 2. Factor loadings from Exploratory Factor Analysis (Principal axis factoring) with promax rotation (3 factors): Greece 2002 (*n*=1,283)

Item	Factor I Openness to change	Factor II Self-transcendence	Factor II Self-enhancement	Unique variance
SD1	**.598**	.140	-.099	.621
SD11	**.479**	**.263**	.025	.616
UN3	-.014	**.589**	-.062	.671
UN8	**.280**	**.507**	-.073	.612
UN19	.124	**.650**	-.099	.547
BE12	.031	**.592**	.003	.637
BE18	-.003	**.567**	.070	.656
CO16	-.188	**.478**	**.356**	.637
SEC5	-.129	**.564**	.167	.653
SEC14	-.055	**.620**	.089	.602
PO2	**.422**	-.124	.289	.671
PO17	-.068	.132	**.595**	.626
AC4	**.407**	.046	**.361**	.561
AC13	**.408**	.022	**.396**	.532
HE10	**.396**	-.089	**.411**	.567
HE21	**.659**	.043	-.006	.552
ST6	**.794**	.066	-.161	.421
ST15	**.678**	**-.224**	.046	.553
	Correlations between factors			
Factor I	—			
Factor II	.285	—		
Factor III	.418	.233	—	

Notes: Factor loadings >.22 are in boldface.

Table 3. Descriptive statistics, reliability coefficients and internal consistencies of the subscales: Greece 2002 (*n*=1,283)

	Subscale		
	Openness to change	Self-transcendence	Self-enhancement
Range (number of items)	8	8	2
Mean (standard error)	2.81 (0.024)	1.88 (0.017)	2.69 (0.030)
95% Confidence interval	2.77-2.86	1.85-1.92	2.63-2.75
Standard deviation	0.865	0.594	1.060
Skewness (standard error)	0.528 (0.068)	1.311 (0.068)	0.573 (0.068)
Kurtosis (standard error)	0.018 (0.137)	4.079 (0.137)	0.010 (0.137)
Cronbach's alpha reliability coefficient	0.819	0.806	0.467
Split-half reliability coefficient	0.821	0.792	0.467
Average inter-item correlations	0.368	0.344	0.306
Minimum-maximum correlations	0.178-0.497	0.243-0.476	0.306-0.306
Range of correlations	0.319	0.233	0.000
	Average inter-item correlations between subscales		
Openness to change	—		
Self-transcendence	0.239	—	
Self-enhancement	0.334	0.271	—

3.1.2 Slovenia. ESS Round 1 (2002)

The majority of the responses were clustered at the lower end of the scale in the first two response categories (Table 4). Low mean responses were found for items defining Universalism (UN3, UN19), Self-Direction (SD11) and Security (SEC5, SEC14) and high mean responses for items defining Stimulation (ST15), Power (PO2) and Conformity (CO7).

As shown in Table 4, the proportion of missing values for all the items was greater in Slovenia than in Greece, ranging from 2.4 to 3.2%. Testing for randomness indicated that the data was probably not missing completely at random (Little's MCAR test: $\chi^2 = 472.7$, $df = 305$, $p < 0.001$). Non-normality was not severe for any item (skewness>2; kurtosis>7). Based on the criteria of corrected item-total correlations and extraction communalities, the following five items were rejected from analysis: TR9 ("It is important to him to be humble and modest. He tries not to draw attention to himself"); TR20 ("Tradition is important to him. He tries to follow the customs handed down by his religion or his family"); CO7 ("He believes that people should do what they are told. He thinks people should follow rules at all time, even when no-one is watching"); PO2 ("It is important to him to be rich. He wants to have a lot of money and expensive things"); ST15 ("He looks for adventure and likes to take risks. He wants to have an exciting life").

Table 4. Item analysis of Schwartz scale values for ESS – 2002: Slovenia (*n*=759)

Item	Mean (SD)	95% CI for mean	Frequency percent of response categories							Skew	Kurt.	CC
			1	2	3	4	5	6	NA			
SD1	2.52 (1.12)	2.44-2.60	14.8	40.7	25.8	8.0	7.0	1.1	2.6	0.86	0.46	0.38
SD11	2.06 (1.03)	1.99-2.13	30.6	44.1	13.4	5.3	3.3	0.7	2.6	1.28	1.82	0.45
UN3	2.01 (0.95)	1.95-2.08	28.6	50.1	11.5	3.6	2.9	0.7	2.8	1.48	3.00	0.30
UN8	2.44 (1.07)	2.37-2.52	13.3	48.7	21.1	7.9	5.0	1.4	2.5	1.12	1.24	0.35
UN19	2.09 (0.94)	2.02-2.16	26.7	44.8	18.4	5.4	1.3	0.7	2.6	1.07	1.76	0.39
BE12	2.30 (0.96)	2.23-2.37	18.1	44.8	24.8	7.1	2.5	0.4	2.4	0.83	0.92	0.46
BE18	2.35 (1.04)	2.28-2.43	16.7	46.5	22.1	7.2	2.9	1.7	2.8	1.16	1.74	0.45
TR9	2.28 (1.04)	2.21-2.36	19.6	47.8	17.1	8.2	4.0	0.7	2.6	1.06	1.08	0.28
TR20	2.62 (1.21)	2.53-2.70	14.8	39.0	22.0	12.0	7.4	2.1	2.8	0.80	0.10	0.27
CO7	3.11 (1.39)	3.01-3.21	9.4	31.6	21.7	13.8	15.3	5.5	2.6	0.44	-0.78	0.25
CO16	2.52 (1.17)	2.43-2.60	15.2	43.6	20.7	8.6	7.9	1.3	2.8	0.94	0.38	0.41
SEC5	2.21 (1.06)	2.13-2.28	23.3	47.0	15.8	6.1	4.2	0.9	2.6	1.21	1.54	0.35
SEC14	2.21 (1.08)	2.14-2.29	24.2	46.2	15.5	5.5	5.3	0.8	2.4	1.20	1.33	0.40
PO2	4.08 (1.30)	3.99-4.18	2.0	11.3	20.0	17.3	34.3	12.6	2.5	-0.34	-0.85	0.23
PO17	2.84 (1.21)	2.75-2.93	9.4	35.2	26.9	12.9	10.9	2.0	2.8	0.63	-0.30	0.46
AC4	2.82 (1.28)	2.73-2.91	12.1	33.5	26.5	10.4	12.4	2.4	2.8	0.63	-0.38	0.48
AC13	2.69 (1.18)	2.61-2.78	12.3	39.4	24.9	10.8	8.4	1.8	2.4	0.74	0.07	0.55
HE10	2.90 (1.30)	2.80-2.99	10.3	34.4	24.5	12.0	13.3	2.9	2.6	0.59	-0.51	0.40
HE21	2.64 (1.35)	2.54-2.73	20.4	33.9	18.2	11.6	10.9	2.5	2.5	0.69	-0.44	0.44
ST6	2.51 (1.20)	2.42-2.59	17.8	41.4	18.2	10.9	8.3	0.9	2.5	0.81	-0.04	0.48
ST15	3.81 (1.46)	3.71-3.91	6.5	15.3	18.6	14.8	31.0	10.8	3.2	-0.26	-1.02	0.26

Notes: SD = standard deviation; CI = confidence interval; NA = no answer (missing values); Kurt. = kurtosis; CC = corrected item-total correlation. Standard errors for skewness and kurtosis were 0.068 and 0.137, respectively.

The eigenvalue rule >1 and scree test suggested the retention of a three-factor solution that best explained the variance when eigenvalues from the target data set were compared to the average and the 95th percentile of the random data sets. Parallel analysis confirmed this result as the actual eigenvalues (5.775, 2.466, 1.603) were greater than randomly generated ones for both the average (1.254, 1.202, 1.162) and the 95th percentile (1.308, 1.239, 1.193) eigenvalue criteria.

Table 5. Factor loadings of Exploratory Factor Analysis (Principal axis factoring) with Promax rotation (3 factors): Slovenia 2002 (*n*=759)

Item	Factor I Openness to change	Factor II Self-transcendence	Factor III Self-enhancement	Unique variance
SD1	**.479**	.117	-.035	.729
SD11	**.452**	.207	.003	.683
UN3	**.233**	**.442**	**-.246**	.726
UN8	**.221**	**.359**	-.108	.794
UN19	.085	**.538**	-.077	.695
BE12	.027	**.606**	.031	.607
BE18	**.239**	**.383**	.020	.721
CO16	**-.230**	**.572**	**.257**	.600
SEC5	-.145	**.501**	.176	.713
SEC14	-.075	**.496**	.180	.697
PO17	.055	.089	**.540**	.641
AC4	**.380**	-.028	**.417**	.581
AC13	**.383**	.013	**.501**	.455
HE10	**.561**	-.140	.162	.646
HE21	**.708**	-.145	.073	.517
ST6	**.556**	.123	.015	.618
	Correlations between factors			
Factor I	—			
Factor II	.367	—		
Factor III	.364	.318	—	

Note: Factor loadings >.22 are in boldface.

Table 5 presents the structure of the three-factor solution obtained by Principal axis factoring. The first factor refers mostly to Openness to change, the second factor expresses Self-transcendence and the third factor is defined by Self-enhancement. These three factors explain 28.367, 12.115 and 7.873% of the variance, respectively.

Almost all of the 16 items demonstrated strong factor loadings ≥0.45 on at least one factor.

Table 6. Descriptive statistics, reliability coefficients and internal consistencies of the subscales: Slovenia 2002 (*n*=759)

	Subscale		
	Openness to change	Self-transcendence	Self-enhancement
Number of items	5	8	3
Mean (standard error)	2.53 (0.030)	2.27 (0.022)	2.78 (0.035)
95% Confidence interval	2.47-2.58	2.22-2.31	2.72-2.85
Standard deviation	0.822	0.615	0.958
Skewness (standard error)	0.623 (0.089)	0.671 (0.089)	0.577 (0.089)
Kurtosis (standard error)	0.301 (0.177)	1.222 (0.177)	0.035 (0.177)
Cronbach's alpha reliability coefficient	0.742	0.737	0.682
Split-half reliability coefficient	0.860	0.716	0.651
Average inter-item correlations	0.333	0.261	0.418
Minimum-maximum correlations	0.198-0.501	0.140-0.391	0.345-0.521
Range of correlations	0.304	0.250	0.176
	Average inter-item correlations between subscales		
Openness to change	—		
Self-transcendence	0.216	—	
Self-enhancement	0.318	0.230	—

Subscales were constructed by averaging the defining items of each factor. As shown in Table 6, Cronbach's alpha reliability coefficients for the subscales Openness to change, Self-transcendence and Self-enhancement were 0.742, 0.737 and 0.682, respectively, indicating that the third factor was not reliable. Split-half reliabilities were 0.860, 0.716 and 0.651, respectively. Average inter-item correlations were 0.333, 0.261 and 0.418 within subscales and 0.216, 0.318 and 0.230 between subscales; these values lie within the recommended range.

3. 2 Results of CFA

3.2.1 Greece. ESS Round 1 (2002)

Using CFA, eight different models were tested: one first-order uncorrelated factor based on the 18 observed variables (model 1); two first-order correlated factors based on the 18 observed variables (model 2a); two first-order correlated factors based on the 18 observed variables with cross-loadings (model 2b); two first-order correlated factors based on the 16 observed variables as indicated by the subscale reliability analysis results (model 2c); two first-order correlated factors based on the 16 observed variables with cross-loadings (model 2d); three first-order correlated factors based on the EFA results (model 3a); three first-order correlated factors with cross-loadings (model 3b); and five first-order correlated factors of unified values (model 4) as proposed by Davidov et al. [7].

For the justification of the models 2b and 2d, PAF was performed (Table 7). The two factors of model 2b explained 31.595 and 12.556% of the variance, respectively. Parallel analysis confirmed this result as the actual eigenvalues (7.472, 2.969) were greater than randomly generated ones for both the average (1.210, 1.172) and the 95th percentile (1.248, 1.199) eigenvalue criteria. The two factors of model 2d explained 32.637 and 14.284% of the variance, respectively. This also was confirmed by parallel analysis, as the actual eigenvalues (6.594, 2.886) were greater than the randomly generated ones for both the average (1.194, 1.154) and the 95th percentile (1.232, 1.182) eigenvalue criteria.

Table 7. Factor loadings of Exploratory Factor Analysis (Principal axis factoring) with Promax rotation (2 factors): Greece 2002 (*n*=1,283)

Item	Factor I Openness to change	Factor II Self-transcendence	Unique variance
SD1	**.547**	.107	.647
SD11	**.473**	**.253**	.625
UN3	-.089	**.592**	.679
UN8	.214	**.494**	.634
UN19	.048	**.632**	.576
BE12	.005	**.601**	.637
BE18	-.002	**.584**	.660
CO16	-.020	**.507**	.750
SEC5	-.086	**.616**	.652
SEC14	-.049	**.661**	.585
PO2	**.545**	-.164	.741
AC4	**.577**	.084	.625
AC13	**.602**	.058	.609
HE21	**.647**	.022	.570
ST6	**.695**	.017	.508
ST15	**.743**	**-.266**	.521
	Correlations between factors		
Factor I	—		
Factor II	.362	—	

Note: Factor loadings >.22 are in boldface.

As shown in Table 8, the fit of model 1 was not adequate; model 2a also had poor fit to the data and using all cross-loadings indicated by EFA (model 2b) improved model fit; model 2c presented inadequate fit to the data and using all cross-loadings indicated by EFA (model 2d) improved model fit; model 3a had also a poor fit to the data and using all cross-loadings indicated by EFA (model 3b) improved model fit; model 4 resulted in a non-positive definite matrix. Therefore, model 2d (Figure 1) provided a better fit to the data than all other models.

Table 8. Confirmatory Factor Analysis (maximum likelihood), goodness-of-fit indices of eight models: Greece 2002 (*n*=1,283)

Models tested	Factor structure	χ^2/df	SRMR[b]	NFI[a]	CFI[a]	TLI[a]	RMSEA[b] (95 % CI)
1	1 first-order uncorrelated factor	10.83	0.089	0.826	0.839	0.793	0.088 (0.083-0.092)
2a	2 first-order correlated factors (18 items)	6.70	0.065	0.888	0.903	0.880	0.067 (0.062-0.071)
2b	2 first-order correlated factors (18 items) with cross-loadings	5.55	0.054	0.909	0.924	0.904	0.060 (0.055-0.064)
2c	2 first-order correlated factors (16 items)	6.39	0.066	0.903	0.917	0.897	0.065 (0.060-0.070)
2d	2 first-order correlated factors (16 items) with cross-loadings	4.87	0.052	0.927	0.941	0.926	0.055 (0.050-0.060)
3a	3 first-order correlated factors	8.05	0.071	0.864	0.879	0.852	0.074 (0.070-0.078)
3b	3 first-order correlated factors with cross-loadings	5.28	0.047	0.916	0.931	0.910	0.058 (0.053-0.062)
4	5 first-order correlated factors of unified values	—	0.066	0.863	0.880	0.855	0.068 (0.064-0.072)

Notes: *df* = degrees of freedom; SRMR = standardized root mean square residual; CI = confidence interval; NFI = normed fit index; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation. The covariance matrix of the 5 first-order correlated factors of unified values was not positive definite.
a Higher values indicate better model fit
b Lower values indicate better model fit

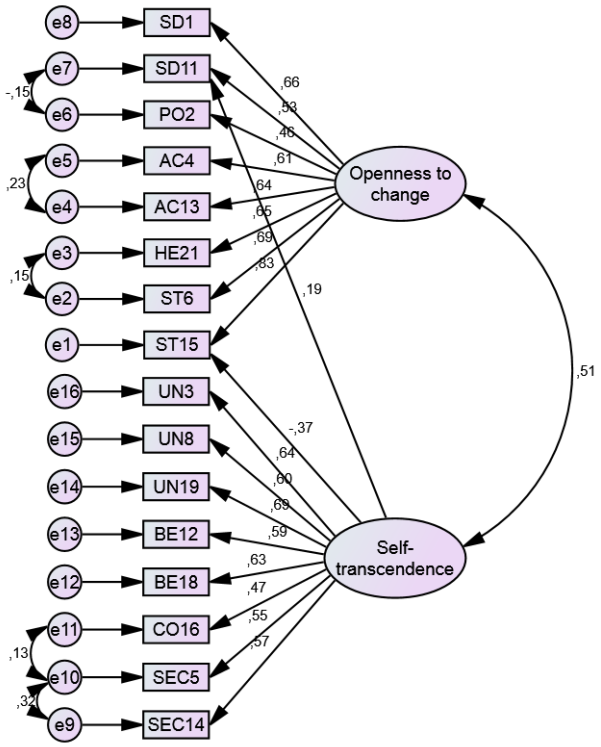


Figure 1. Standardized solution for the 2 first-order correlated factors (model 2d; 16 items) with cross-loadings based on CFA analysis (*n* = 1,283). Observed variables are represented by rectangles and latent variables are enclosed in ellipses: Greece 2002.

3.2.2 Slovenia, ESS Round 1 (2002)

Using CFA, eight different models were tested: one first-order uncorrelated factor based on the 16 observed variables (model 1); two first-order correlated factors based on the 16 observed variables (model 2a); two first-order correlated factors based on the 16 observed variables with cross-loadings (model 2b); two first-order correlated factors based on the 13 observed variables as indicated by the subscale reliability analysis results (model 2c); two first-order correlated factors based on the 13 observed variables with cross-loadings (model 2d); three first-order correlated factors based on the EFA results (model 3a); five first-order correlated factors with cross-loadings (model 3b); and five-order correlated factors of unified values (model 4) as proposed by Davidov et al. [7].

For the justification of the models 2b and 2d PAF was performed. The two factors of model 2b explained 28.367 and 12.115% of the variance, respectively. Parallel analysis confirmed this result as the actual eigenvalues (5.775, 2.466) were greater than randomly generated ones for both the average (1.254, 1.202) and the 95th percentile (1.308, 1.239) eigenvalue criteria. The two factors of model 2d explained 28.002 and 15.245% of the variance, respectively. Parallel analysis again confirmed this result, as the actual eigenvalues (4.438, 2.416) were greater than the randomly generated ones for both the average (1.219, 1.168) and the 95th percentile (1.266, 1.205) eigenvalue criteria.

As shown in Table 9, the fit of model 1 was not adequate; model 2a had also a poor fit to the data and using all cross-loadings indicated by EFA (model 2b) improved model fit; model 2c presented inadequate fit and using all cross-loadings indicated by EFA (model 2d) improved model fit; model 3a had also a poor fit to the data and using all cross-loadings indicated by EFA (model 3b) improved model fit; model 4 also resulted in poor fit. Therefore, model 3b (Figure 2) provided a better fit to the data than all other models.

Table 9. Confirmatory Factor Analysis (maximum likelihood), goodness-of-fit indices of eight models: Slovenia 2002 (*n*=760)

Models tested	Factor structure	χ^2/df	SRMR[b]	NFI[a]	CFI[a]	TLI[a]	RMSEA[b] (95 % CI)
1	1 first-order uncorrelated factor	5.12	0.063	0.832	0.859	0.810	0.074 (0.067-0.080)
2a	2 first-order correlated factors (16 items)	4.84	0.059	0.829	0.858	0.822	0.071 (0.065-0.078)
2b	2 first-order correlated factors (16 items) with cross-loadings	4.78	0.057	0.833	0.862	0.825	0.071 (0.064-0.077)
2c	2 first-order correlated factors (13 items)	4.74	0.057	0.845	0.873	0.837	0.070 (0.062-0.078)
2d	2 first-order correlated factors (13 items) with cross-loadings	4.62	0.055	0.851	0.878	0.842	0.069 (0.061-0.077)
3a	3 first-order correlated factors	5.54	0.066	0.796	0.825	0.790	0.077 (0.071-0.084)
3b	3 first-order correlated factors with cross-loadings	4.48	0.051	0.847	0.876	0.839	0.068 (0.061-0.074)
4	5 first-order correlated factors of unified values	5.45	0.072	0.777	0.808	0.769	0.077 (0.071-0.082)

Notes: *df* = degrees of freedom; SRMR = standardized root mean square residual; CI = confidence interval; NFI = normed fit index; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation.
a Higher values indicate better model fit
b Lower values indicate better model fit

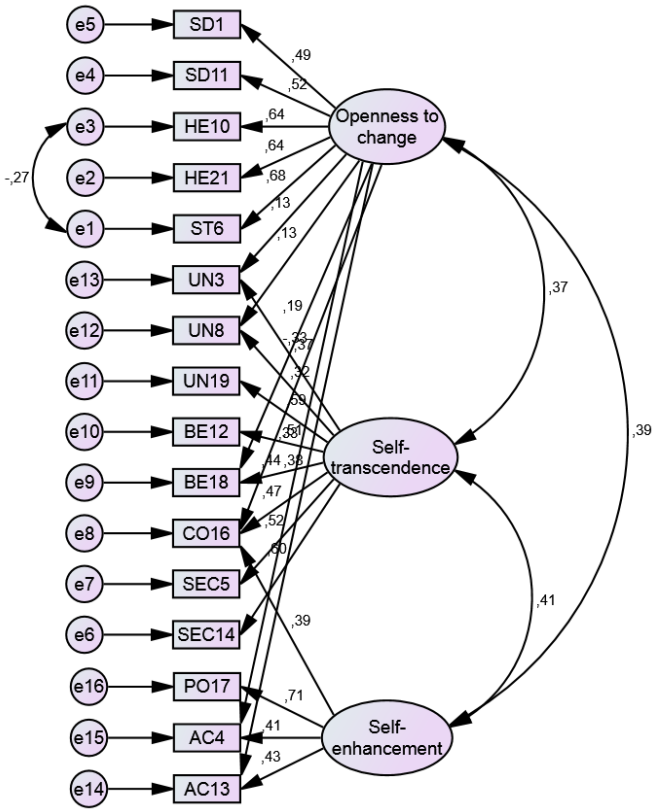


Figure 2. Standardized solution for the 3 first-order correlated factors (model 3b) with cross-loadings based on CFA analysis (n = 760). Observed variables are represented by rectangles and latent variables are enclosed in ellipses: Slovenia 2002.

3.3 Descriptive statistics, reliability coefficients and internal consistencies of the subscales for Greece and Slovenia

In Table 10, descriptive statistics, reliability coefficients and internal consistencies of the subscales based on the two-factor solution (16 items) are presented for Greece for the full sample.

Table 10. Descriptive statistics, reliability coefficients and internal consistencies of the subscales: Greece 2002 (N = 2,566)

	Subscale	
	Openness to change	Self-transcendence
Number of items	8	8
Mean (standard error)	2.82 (0.017)	1.90 (0.012)
95% Confidence interval	2.79-2.85	1.88-1.92
Standard deviation	0.861	0.601
Skewness (standard error)	0.521 (0.048)	1.447 (0.048)
Kurtosis (standard error)	0.077 (0.097)	5.176 (0.097)
Cronbach's alpha reliability coefficient	0.825	0.812
Split-half reliability coefficient	0.828	0.773
Average inter-item correlations	0.377	0.354
Minimum-maximum correlations	0.166-0.512	0.279-0.507
Range of correlations	0.346	0.228
	Average inter-item correlations between subscales	
Openness to change	—	
Self-transcendence	0.251	—

As shown, Cronbach's alpha reliability coefficients for the subscales Openness to change and Self-transcendence were 0.825 and 0.812, respectively (Table 10). Split-half reliabilities were 0.828

and 0.773, respectively. Average inter-item correlations within subscales were 0.377 and 0.354 and between subscales 0.251, indicating that the values were within the recommended range.

In Table 11, descriptive statistics, reliability coefficients and internal consistencies of the subscales based on the three-factor solution are presented for Slovenia, this time for the full sample. Cronbach's alpha reliability coefficients for the subscales Openness to change, Self-transcendence and Self-enhancement were 0.713, 0.729 and 0.683, respectively. Split-half reliabilities were 0.726, 0.684 and 0.547, respectively. Average inter-item correlations within subscales were 0.332, 0.255 and 0.419 and between subscales 0.207, 0.310 and 0.227, indicating that the values were within the recommended range.

Table 11. Descriptive statistics, reliability coefficients and internal consistencies of the subscales: Slovenia 2002 (N = 1,519)

	Subscale		
	Openness to change	Self-transcendence	Self-enhancement
Number of items	5	8	3
Mean (standard error)	2.57 (0.021)	2.27 (0.015)	2.80 (0.024)
95% Confidence interval	2.53-2.62	2.24-2.30	2.75-2.84
Standard deviation	0.821	0.602	0.945
Skewness (standard error)	0.585 (0.063)	0.485 (0.063)	0.515 (0.063)
Kurtosis (standard error)	0.330 (0.125)	0.856 (0.125)	0.076 (0.125)
Cronbach's alpha reliability coefficient	0.713	0.729	0.683
Split-half reliability coefficient	0.726	0.684	0.547
Average inter-item correlations	0.332	0.255	0.419
Minimum-maximum correlations	0.214-0.467	0.137-0.375	0.356-0.501
Range of correlations	0.253	0.238	0.145
	Average inter-item correlations between subscales		
Openness to change	—		
Self-transcendence	0.207	—	
Self-enhancement	0.310	0.227	—

Conclusions

In this paper, the investigation of the dimensionality of the ESS Schwartz values scale by applying the traditional approaches of EFA and CFA to randomly split half-samples indicated a different structure from the five unified values that had been proposed by Davidov et al. [7] for Greece and Slovenia in order to solve the problem of non-positive definite matrices of the constructs in single-country CFAs. In the case of Greece, when the proposed solution was tested on the second split-half sample, CFA showed that this problem was still present. In both countries, item analyses carried out on the first split-half samples indicated that a number of items had first to be excluded from further analysis. Eight models were tested. A model with two first-order correlated factors based on 16 items (model 2d) in Greece and one with three first-order correlated factors (model 3b) in Slovenia, provided the best fit to the data. The fit was improved by considering cross-loadings as suggested by Lilleoja and Saris [14]. In the case of Greece, the resulting two underlying dimensions were defined as Openness to change and Self-transcendence values. In the case of Slovenia, the three underlying dimensions were defined as Openness to change, Self-transcendence and Self-enhancement. These solutions provided reliable subscales for both countries that could be used in further analyses.

The ESS has included Schwartz's Short Form of the Human Values Scale in all its rounds and therefore this work could be extended to cover all participating countries in every round. In this way, researchers would be provided with valid and reliable subscales for their analyses.

References

1. F. Anagnostopoulos, J. Yfantopoulos, I. Moustaki and D. Niakas. Psychometric and factor analytic evaluation of the 15D health-related quality of life instrument: the case of Greece. Quality of Life Research, 22, 1973-1986, 2013.
2. D. J. Bartholomew, F. Steele, I. Moustaki and J. Galbraith. Analysis of multivariate social science data, Chapman & Hall/CRC, London, 2008.

3. T. A. Brown. Confirmatory factor analysis for applied research, The Guilford Press, New York, 2006.
4. P. Cabrera-Nguyen. Author guidelines for reporting scale development and validation results. Journal of the Society for Social Work and Research, 1, 99-103, 2010.
5. L. A. Clark and D. Watson. Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7, 309-319, 1995.
6. G. Datler, W. Jagodzinski and P. Schmidt. Two theories on the test bench: Internal and external validity of the theories of Ronald Inglehart and Shalom Schwartz. Social Science Research, 42, 906-925, 2013.
7. E. Davidov, P. Schmidt and S. H. Schwartz. Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. Public Opinion Quarterly, 72, 420-445, 2008.
8. European Social Survey (n.d.). Core Questionnaire. Retrieved from http://www.europeansocialsurvey.org/docs/methodology/core_ess_questionnaire/ESS_core_questionnaire_human_ values. pdf
9. L. R. Fabrigar, D. T. Wegener, R. C. MacCallum and E. J. Strahan. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4, 272-299, 1999.
10. J. A. Gillapsy Jr., D. L. Jackson and R. Purc-Stephenson. Reporting practices in confirmatory factor analysis: An overview and some recommendations. Psychological Methods, 14, 6-23, 2009.
11. J. C. Hayton, D. G. Allen and V. Scarpello. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational Research Methods, 7, 191-205, 2004.
12. D. Knoppen and W. Saris. Do we have to combine values in the Schwartz' human values scale? A comment on the Davidov studies. Survey Research Methods, 3, 91-103, 2009.
13. R. D. Ledesman and P. Valero-Mora. Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out parallel analysis. Practical Assessment, Research & Evaluation, 12, 1-11, 2007.
14. L. Lilleoja and W. E. Saris. Testing a new operationalization of the basic values on Estonian-and Russian-speaking subpopulations in Estonia. Social Indicators Research, 116, 153-172, 2014.
15. J. C. Nunnally and I. H. Bernstein. Psychometric theory, McGraw-Hill, New York, 1994.
16. B. P. O'Connor. SPSS and SAS programs for determing the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments & Computers, 32, 396-402, 2000.
17. V. H. Patil, S. N.Singh, S. Mishra and T. Donavan. Parallel analysis engine to aid determining number of factors to retain [Computer software], 2007. Available from Instruction and Research Server, University of Kansas website: http://ires.ku.edu/~smishra/parallelengine.htm.
18. V. H. Patil, S. N. Singh, S. Mishra and T. Donavan. Efficient theory development and factor retention criteria: A case for abandoning the "eigenvalue greater than one" criterion. Journal of Business Research, 61, 162-170, 2008.
19. T. A. Schmitt. Current methodological considerations in exploratory and confirmatory factor analysis. Journal of Psychoeducational Assessment, 29, 304-322, 2011.
20. S. H. Schwartz. Studying values: Personal adventure, future directions. Journal of Cross-Cultural Psychology, 42, 307-319, 2011.
21. J. Stevens. Applied multivariate statistics for the social sciences, Lawrence Erlbaum Associates, New Jersey, 2002.
22. M. Symeonaki, C. Michalopoulou and A. Kazani. A fuzzy set theory solution to combining Likert items into a single overall scale (or subscales). Quality & Quantity, 49, 739-762, 2015.
23. B. G. Tabachnick and L. S. Fidell. Using multivariate statistics, Pearson International Edition, 2007.
24. B. Thompson. Exploratory and confirmatory factor analysis: Understanding concepts and applications, American Psychological Association, Washington DC, 2005.