İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 13, 2020, 2, 48-60

Geliş/Received:21.08.2020, Kabul/Accepted: 03.12.2020

www.istatistikciler.org

Araştırma Makalesi / Research Article

A multiple decrement life table model for orphan daughters in Turkey

İlker Şirin

Social Security Institution
Çankaya, Ankara, TURKEY
isirin@sgk.gov.tr

Abstract

The multiple decrement life table application in this study shows that, social insurances for surviving daughters in Turkey is considerably generous. When causes other than death are eliminated, expected benefit duration for surviving daughters, is higher than that is calculated for women pensioners who have earned their benefits after their career. Survivor pensioner daughters are encouraged to remain out of the labour force and receive their orphan's pensions, rather than start working. However, benefiting from the demographic dividend requires covering both men and women in the labour force.

Keywords: Multiple decrement life table, Survivor pensions, Orphan daughters.

Öz

Bu çalışmadaki çok-azalanlı yaşam tablosu uygulaması, ölüm aylığı alan yetim kız çocukları için sosyal sigortaların oldukça cömert olduğunu ortaya koymaktadır. Ölüm dışındaki ayrılış nedenleri analizden çıkarıldığında, Türkiye'de yetim kız çocukları için beklenen aylık ödeme süresi, kendi çalışmalarından kaynaklı emekliliğe hak kazanan kadın sigortalıların beklenen emekli aylığı alma sürelerinden fazladır. Yetim aylığı alan kız çocukları çalışmak yerine, iş gücü dışında kalarak yetim aylıklarını alma yönünde cesaretlendirilmektedir. Öte yandan, demografik fırsattan faydalanabilmek kadın ve erkeklerin birlikte iş gücüne dahil olmasından geçmektedir.

Anahtar sözcükler: *Çok azalanlı yaşam tablosu, Ölüm aylıkları, Yetim kız çocukları.*

1. Introduction

Life expectancies at all ages have risen due to the progress in quality of life, an increase in the number of people seeking health, and an increase in the availability and accessibility of sanitary and health services. Despite the difference in the phase of epidemiologic transition that countries are passing through, there is a worldwide trend in increased expected duration of life at old age, not only in the developed world but also in developing countries. According to the United Nations [1], life expectancy (LE) at birth has risen 10 to 25 years in different regions of the world. At individual level and for society as a whole, to live longer is a gift. However, if the prolonged life is in ill health, the process becomes burdensome for both the individual and the country. At the individual level, the quality of life is low, while at the country level it is the public pressure on finance, which makes the process difficult. Longevity leads to a rise at the durations of retirement benefits, thus putting more pressure on the national budget. The expenditure on public pensions when measured as percentage of Gross Domestic Product, have risen 2.5% on average since 1990 [2]. Therefore, increased life span and low levels of mortality during retirement is more frequently discussed in academic and policy oriented literature.

Pay As You Go financed public pension systems sustain the expenditures of current pensioners from today's active working population. Similarly, future contributors will finance future pension payments to today's active contributors. Nevertheless, population ageing disrupts the balance between contributors and beneficiaries. Pension reforms become indispensable instruments for tackling the burden of population ageing. If there is not enough number of contributors to pay for the liabilities of pensioners, public pension systems face financial unsustainability and budget deficits. In case, that the country manages the demographic gift appropriately, a cumulative burden on labour, sustainable pensions and economic growth is avoidable until the opportunity window closes [3]. In this respect, Turkey anticipates a window of opportunity until 2040 [4]. Turkish pension system has a challenge for the next 20 years to solve current problems regarding; early retirement, generous rights for surviving women and meanwhile has to prepare for the contemporary issues for creating jobs during the 'gift'.

The popular opinion about 'retirement' in Turkey covers benefits that are transferred from deceased family members. Common people call their survivor benefits as their 'retirement'. Besides, long term insurance branch in Turkey covers old age, disability and survivor insurance. The contribution load is 20% of declared wage. The person is under the insurance umbrella for the aforementioned risks by paying the contribution that is calculated on actual earnings. Survivors of two different insured groups can get survivor pensions. Firstly, the deceased person who was receiving (or was entitled to receive) an old age or disability pension before death is under coverage. Secondly, active working people who have paid at least 900 Days of Contribution (DoC) with at least five Years of Service before death are covered. Spouses are directly eligible to benefits regardless of their own pensions or active salaries. Children under 18 (20 if pre-university student and 25 if university student), disabled sons sor daughters at any age, unmarried, divorced or widowed daughters who are not insured or receive any sort of social security benefits are also eligible [5]. 50% of the pension amount of the deceased is paid to the spouse whereas orphans receive 25%. Apart from these, surviving daughters receive two years accumulated value of suvivor pensions as at marriage date, as a marriage grant.

In addition to the old-age pension which is paid to the retired individual, Social Security Institution (SSI) provides death benefits to survivors of the deceased according to special conditions. Surviving spouses, children, mother and father may be eligible to death benefits. Benefits are also stopped, not only by a single decrement "death" but also by marriage, by working or by reaching a specific age. Therefore a multiple decrement process is on duty.

In Turkey, major life table (LT) studies rely on indirect methods on account of the poor death registration in the country. Incompleteness of death counts was estimated to be around 17% in 2005 by Hoşgör [6]. The dataset, that is used, is either Turkey Demographic and Health Survey-TDHS (see [7-11]), population census data of Turkish Statistical Institute-TurkStat ([12-14]) or burial records [15]. All the authors who have used the TDHS data, estimate adult mortality with "orphanhood" method which is popular in case that the death registration is poor. Taylan and Yapar [16] use address based population register system data of Turkey and obtain period LTs for 2009, 2010 and 2011.

There are examples of LTs which are constructed via direct techniques with deaths records of pensioners or employees of SSI. In 2005, Tuzgöl [17] constructed a life table for pensioners beyond age 30, with four years (2000-2003) of SSI administrative data. Death rates were graduated by the method proposed by Whittaker [18]. In addition, she had to borrow the rates from Turkish national data of TurkStat or apply specific constant increment coefficients for crude death rates for females beyond 90 (and males 75+

respectively). Other LT technique applications with SSI data in recent native literature include, multiple incerement-decrement LTs for the actively insured employees ([19-20]) and the 2008 SSI table which was prepared under Life and Annuity Life Table Construction Project of the Turkish Treasury [21]. LTs were constructed with data of mixed groups of insured persons (retirees, invalidity pensioners, active contributors) as units of analysis.

The aim of this study is to construct multiple decrement life tables for survivor pensioner daughters. SSI data on withdrawals from survivor benefits and mid-year beneficiaries for 2012 to 2016 is used as input data. Units of analysis for the multiple decrement LT application in this study are daughters thar are receiving survivor pensions. In the following parts of the study, age specific expected duration of survivor benefits of daughters are analysed with multiple decrement LT application.

2. Data and Methods

In this section, the steps of calculations and the assumptions that is used for producing different columns of the LTs.are highlighted.

2.1. Data sources and description of data

SSI is the major information source in Turkey for the life events for the retired, as the institution pays the monthly benefits to the pensioners. Main events that requires the cessation of benefits is tracked. Therefore SSI has data on, in-year deaths, withdrawals and mid-year number of beneficiaries, which is the key material when constructing LTs. In this study, withdrawals and pensioner figures for 2012 to 2016 is used to depict average pension payment durations of surviving daughters. LTs are modelled with big data. In total, 140,740 withdrawals and 2,403,655 daughters that are exposed to risk are covered. Information for event and exposure is loaded to SSI server and data is prepared by SAS Enterprise Guide.

2.2. Description of age and 'daughter'

"Age at last birthday" which is also known as "age in completed years" or "actual age", takes the number of years completed alive by the person for a date of reference into account [22]. In this study, age is defined as "age at last birthday". Completed age at death is calculated by the exact date of death and the exact date of birth. Similarly, exact date of birth and mid-year data of pensioners (July 1) gives the age at mid-year for a specific year.

The daughter definition is made according to the current legislation. The surviving daughter of a deceased pensioner (or active contributor, with at least 1800 DoC -or 5 years of service with 900 DoC- for longterm insurances before death) is under coverage with specific additional conditions. The additional conditions are as follows:

- "The ones who have completed the age of 18, the age of 20 in case receiving education in high school or equivalent, or the age of 25 in case receiving higher education; or
- The ones who are found to be disabled by losing minimum 60% of working power based on Institution Health Committee decision (regardless of his or her marital status); or
- The daughters, whatever the ages are, not married, divorced or widow." (see [5]).

2.3. Multiple decrement processes

Let ${}_nD_x^i$ describe the number of decrements from cause i between age x to $x+n$, where ${}_nD_x$ is the number of deaths between age x and $x+n$ then, a "constant of proportionality" [23] reflecting decrements other than cause i between age x to $x+n$ is given by:

$$R^{-i} = \frac{{}_nD_x - {}_nD_x^i}{{}_nD_x}. \tag{1}$$

Probability of surviving ${}_np_x$, death probabilities ${}_nq_x$ and the average person years lived between age x and $x+n$, by those die in n years ${}_na_x$, the number of survivors l_x, are received from the all-cause LT. For the probability of surviving from x to $x+n$, where only decrements due to causes other than i between age x to $x+n$ is at force, is equal to:

$$ {}_n^*p_x^{-i} = \left[{}_np_x\right]^{R^{-i}}. \tag{2}$$

Similarly, one can calculate survivors at age $x+n$, with

$$ {}^*l_{x+n}^{-i} = {}^*l_x^{-i} \cdot {}_n^*p_x^{-i} \tag{3}$$

where the superscript $-i$ reflect the causes other than i that are in effect. In order to get the person years lived ${}_n^*L_x^{-i}$,

$$ {}_n^*L_x^{-i} = n.\ {}^*l_{x+n}^{-i} + {}_n^*a_x^{-i} \cdot {}_n^*d_x^{-i} \tag{4}$$

the average person years lived between age x and $x+n$, by those die due to causes other than i in the interval ${}_n^*a_x^{-i}$ and the number of pensioners that die between age x to $x+n$ is needed. For the first age group $x=45$ and the last age $x=84$ before the open ended interval, the approach proposed by Preston et al. [24] is used and ${}_n^*a_x^{-i}$ is obtained with the equation,

$$ {}_n^*a_x^{-i} = n + R^{-i}\,\frac{{}_nq_x}{{}_n^*q_x^{-i}}\left({}_na_x - n\right). \tag{5}$$

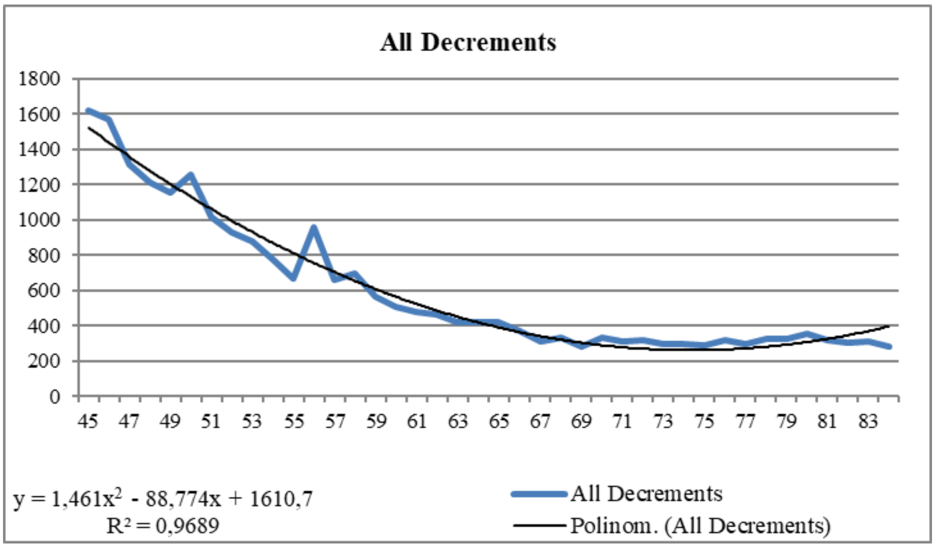


Figure 1. Distribution of all cause decrements that is needed to decide on the suitable ${}_{n}^{*}a_{x}^{-i}$.

Since all cause decrements support the assumption of second degree polynomial distribution with a R^2 of 97% (Figure 1), for the ages in between (x=46 to x=83), following Keyfitz [25], ${}_{n}^{*}a_{x}^{-i}$ is calculated by the equation;

$$ {}_{n}^{*}a_{x}^{-i} = \frac{\frac{-n}{24}\,{}_{n}^{*}d_{x-n}^{-i} + \frac{n}{2}\,{}_{n}^{*}d_{x}^{-i} + \frac{n}{24}\,{}_{n}^{*}d_{x+n}^{-i}}{{}_{n}^{*}d_{x+n}^{-i}}. \tag{6} $$

The LT functions for the open ended interval are as follows:

$$ {}_{\infty}^{*}a_{85}^{-i} = \frac{{}^{*}e_{85}^{0}}{R^{-i}} \tag{7} $$

Life expectancy at age x in the absence of cause i is, ${}^{*}e_{x}^{-i}$, and can be calculated by the following equations.

$$ {}^{*}T_{x}^{-i} = \sum_{\propto=x}^{\infty} {}_{n}^{*}L_{\propto}^{-i} \tag{8} $$

and

$$ {}^{*}e_{x}^{-i} = \frac{{}^{*}T_{x}^{-i}}{{}^{*}l_{x}^{-i}}. \tag{9} $$

3. Results

Figure 2 shows expected survivor benefit payment duration to daughters at selected ages when all causes are at force and when causes except death are eliminated. Life table details for all single ages are presented in Annex. Naturally, more decrements imply increased transition probability from the 'beneficiary' status and thus lower expected benefit duration. However, at each consecutive age, the contribution of causes other than death diminishes. As daughters get older, it is less likely that they work or get married. That is visible with shorter difference between the bars in Figure 2. After 70, the principal cause that affects expected years to be spent as survivor pensioner, is death. Finally, the survivor pension has to cease for some reason and that is death of the beneficiary daughter.

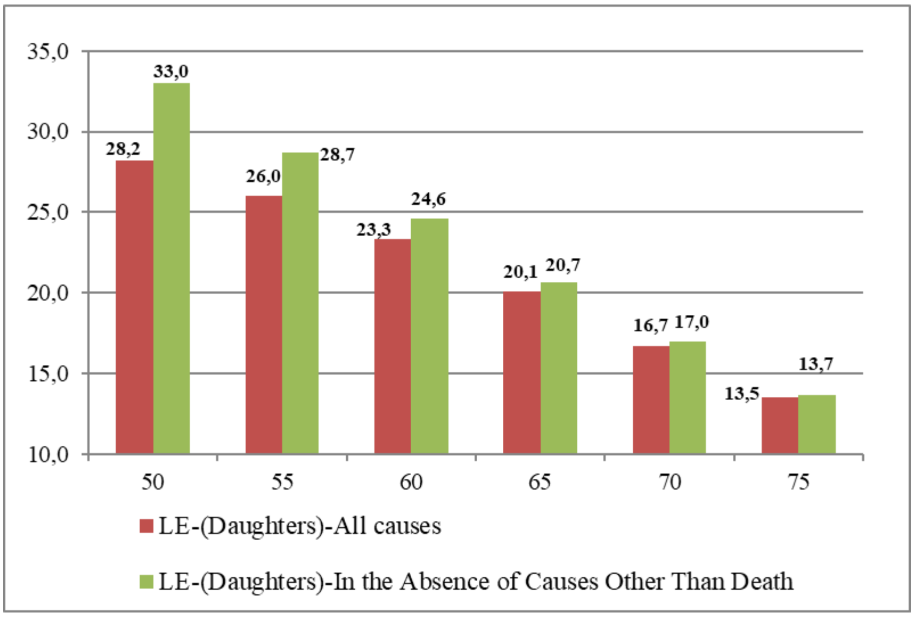


Figure 2. Expected duration of survivor benefits in years (Daughters).

In reality, all causes are at force according to the current legislation. The red bars in Figure 2 show that a 50 year surviving daughter is expected to reveice pension benefits during the remaining 28.2 years. Long benefit payment durations and generosity of the pension system in Turkey is visible. In other words, SSI is expected to pay pensions in the next 23.3 years to a 60 year old surviving daughter, under the assumption that daughters will be subject to transition rates during 2012-2016 in their remaning 'life'. But it is noteworthy that 'life' hereby represents duration that is spent as survivor beneficiary and life is not only continuable by being alive but also by not exiting to other causes.

Figure 3 is derived from Figure 2 and it represents the details of the difference between all cause and cause eliminated (causes other than death) LT application results. In Figure 3, the difference is decomposed by causes. Causes other than death have three categories; work, marriage and other. If deaths were the only cause, a 55 year old daughter would be expected to receive survivor benefits for the remaining 28.7 years. But work, marriage or other causes provide 2.7 years reduction in the average duration. Finding a job and starting to work and as a result losing the benefit is figured with 'Work'. It has the lowest contribution to the gap. 'Marriage' is the second reason in the ranking and finally the highest contribution to the gap is due to annulment, suspension, absence or other remaining reasons.

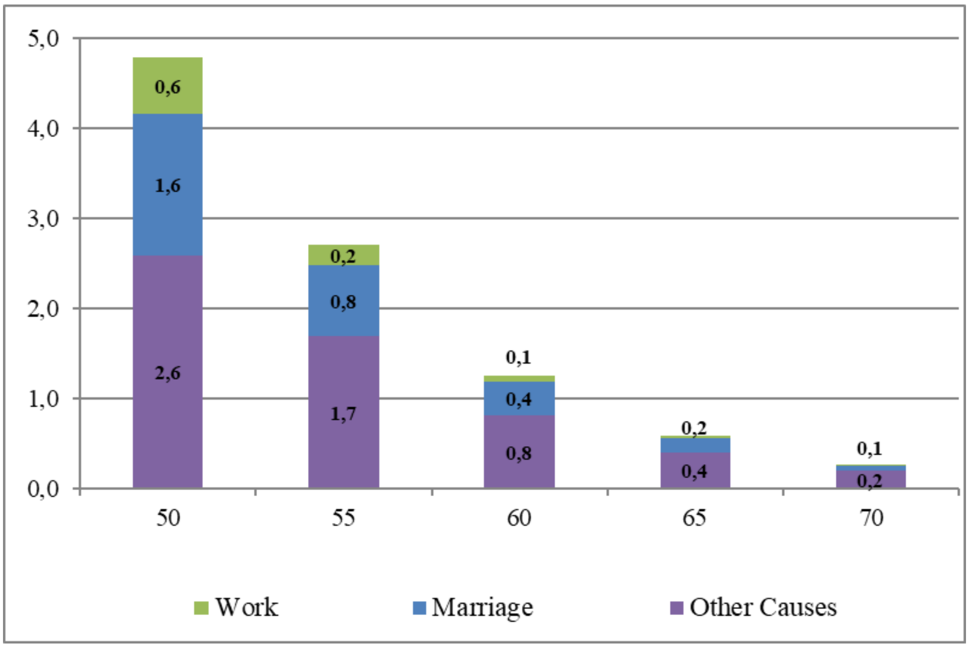


Figure 3. Relative contribution of causes other than death to the expected duration of survivor benefits.

The striking behaviour observed in recent years is the unregistered work of the surviving daughters in order not to lose the benefit that they get. Moreover, surviving daughters avoid marriage or if married, they get divorced and continue to live in the same household with the ex-husband. According to Figure 3, an additional duration can be gained by daughters (for instance 1.6+0.6=2.2 at age 50) if some life events are not declared to SSI. From the perspective of SSI, there is a potential abuse possibility which may increase the financial burden that is on the institution. Daughters, who get involved in such abuse are expected to continue being survivor pensioners considerably long.

The comparison of average benefit durations between orphan daughters and women old age pensioners is in Figure 4. Although survivor beneficiary daughters do not earn their own pension by working, but have a transfered pension, they are expected to get benefits less than max 6.5 years shorter (age 50) than an old age pensioner.

At all ages below 70, daughters receive benefits for durations that is equal to or considerably close to the average pension payment durations of women who have earned their benefits after life time working history. Weighted average durations are even 3.3 years higher for surviving daughters (26.9 years) when compared with women pensioners (26.6 years).

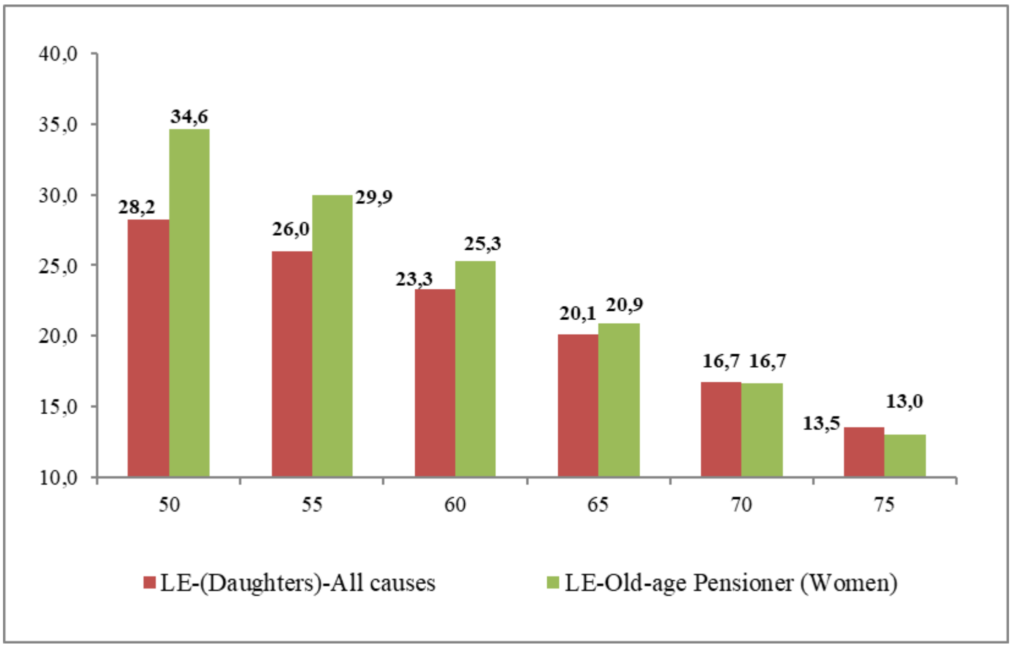


Figure 4. Expected duration of retirement benefits and survivor benefits.

4. Conclusion

The multiple decrement life table application in this study shows that, social insurances for surviving daughters in Turkey is generous. Weighted average of expected benefit payment duration to daughters, with number of beneficiaries as weights, is 26.9 years which is higher than the weighted average of 26.6 years calculated for women receiving old age pensions. Put it differently, the period SSI is expected to pay benefits to daughters that do not work, participate in the labour force, is more than the period that women pensioners are expected to receive as retirement pension.

Gruber and Wise [26] discuss whether social security has impacts on low labour force participation in the industiralized countries. They suggest that, social security contributed to the decline in participation rates via generous benefits at young ages, early retirement or easy pathways to disability insurance. In this sense, survivor benefit regime in Turkey is also generous that, non working daughters are encouraged to stay out of labour force and receive their pensions, rather than start working and lose their survivor benefits. However, the demographic window of opportunity is forecasted to end after 2040. To benefit from the dividend and develop sustainably, both men and women are needed in the economic production process. 2023 targets or benefiting from demographic dividend can not be possible by leaving half of the population outside the labour force. Therefore, encouraging survivor beneficiary daughters to earn their own pensions and offering incentives to those taking part in the labour force may be among suitable policy options.

An important posit hereby is the contradiction between government policies. Under main policy pillars of National Employment Strategy, policies to strenghten labour partcipation of women are discussed in the section “Increasing employment of those needing special policies” [27]. However, neither this section, nor the section on “strengthening links between employment and social protection” provides criticism for the existing generous survivor pensions system in the country.

References

[1] United Nations, 2017, *World Population Ageing 2017 Highlights*, Department of Economic and Social Affairs, Population Division (ST/ESA/SER.A/397). *https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf*

[2] OECD, 2017, *Pensions at a Glance 2017: OECD and G20 indicators*, OECD Publishing, Paris. *http://dx.doi.org/10.1787/pension_glance-2017-en*.

[3] AIV, 2009. *Demographic Changes and Development Cooperation*, Advisory Council On International Affairs, No.66, The Hague. *https://www.advisorycouncilinternationalaffairs.nl/documents/publications/2009/07/10/demographic-changes-and-development-cooperation*

[4] I., Koç, İ., M.A., Eryurt, T., Adalı, P., Seçkiner, 2009, *Türkiye'nin Demografik Dönüşümü: Demographic Transition of Turkey*. Hacettepe University Institute of Population Studies, Ankara.

[5] SSA, 2018, *Social Security Programs throughout the World: Europe, 2018,* Social Security Administration, SSA Publication No. 13-11801. *https://www.ssa.gov/policy/docs/progdesc/ssptw/2018-2019/europe/ssptw18europe.pdf*

[6] Ş. Hoşgör, 2005, Türkiye'de Devlet İstatistik Enstitüsünce Derlenen Ölüm Verilerindeki Eksik Kapsamın Boyut ve Nitelik Tahmini, *Nüfus Bilim Dergisi*, 27, 17-34.

[7] M.A., Eryurt, I., Koç, 2006, Türkiye için hayat tablolarının yetimlik tekniği ile oluşuturulması, *The Turkish Journal of Population Studies*, Vol. 28-29, 47-60.

[8] E., Kırkbeşoğu, İ. Koç, 2010, Mortality table problems in the life insurance sector: An advisory alternative solution for Turkey, *The Turkish Journal of Population Studies*, Vol. 31-32, 5-30.

[9] E., Kırkbeşoğu, 2006, *Construction of mortality tables for life insurance sector from the 2003 Turkey demographic and health survey*, Unpublished master's thesis, Hacettepe University Institute of Population Studies, Ankara.

[10] Y. Coşkun, 2002, *Estimation of Adult Mortality by Using the Orphanhood Method from the 1993 and 1998 Turkish Demographic and Health Surveys*, Unpublished masters thesis, Hacettepe University Institute of Population Studies, Ankara.

[11] A. Hancıoğlu, 1991, *Estimation of Levels and Trends in Mortality from Information on the Survival Statusof a Close Relative: Turkey 1970-1985*, Unpublished Doctoral Dissertation, Hacettepe University Institute of Population Studies, Ankara.

[12] Ş. Hoşgör, 1992, *Estimation of Post-Childhood Life Tables Using Age and Sex Distributions and Intercensal Growth Rates in Turkey, (1930-1990),* Unpublished master's thesis, Hacettepe University Institute of Population Studies, Ankara.

[13] Ş. Hoşgör, 1997, *Estimation of Post-Childhood Life Tables of Provinces and Regions in Turkey, by Using Age and Sex Distributions and Intercensal Growth Rates (1985-1990)*, Unpublished PhD thesis, Hacettepe University Institute of Population Studies, Ankara.

[14] A. Gjonça, 2006, *Training on preparing a life table for Turkey*, Project Completion Report. Hifab International.

[15] D. Demirbüken, 2001, *An Evaluation of Burial Records of Ankara City Cemeteries*, Unpublished masters thesis, Hacettepe University Institute of Population Studies, Ankara.

[16] H., Taylan, G., Yapar, 2013, Construction of a Life Table by Using the Turkish General Death Data, *Journal of Statistical Research*,10 (2), 1-24.

[17] H. Tuzgöl, 2005, *SSK Ölüm İstatistiklerinin İncelenmesi ve Farklı Gruplar İçin Yaşam Tablolarının Oluşturulması*, Unpublished expertise thesis, Social Security Institution, Ankara.

[18] E. Whittaker, 1923, On a new method of graduation, *Proceedings of Edinburgh Mathematical Society*, 41, 63-75.

[19] H., Tuzgöl, M., Sucu, Ş., Hoşgör, 2010, Çoklu Artan-Azalan Hayat Tablosu ve Türkiye Sosyal Güvenlik Sistemine Bir Uygulaması. *The Turkish Journal of Population Studies*, Vol. 31-32, 31-44.

[20] Y., G., Ündemir, H., Özuysal, Ş., Hoşgör, 2010, Sosyal güvenliğe kayıtlı 4/1-a bendi kapsamındaki zorunlu sigortalılara ilişkin çoklu azalan hayat tabloları, *The Turkish Journal of Population Studies*, Vol. 31-32, 83-102.

[21] Hacettepe University, 2018, 'Türkiye Hayat ve Hayat Annüite Tablolarının Oluşturulması Projesi', http://www.aktuerya.hacettepe.edu.tr/TurkiyeHayatTablolari.php

[22] Europan Commission, 2003, *Demographic statistics: Definitions and methods of collection in 31 European Countries*, 3/2003/E/no25.

[23] C. L. Chiang, 1991, Competing risks in mortality analysis. *Annu Rev Public Health*, 1991; 12: 281-307.

[24] S., H., Preston, P., Heuveline, M., Guillot, 2001, *Demography. Measuring and Modelling Population Processes,* Blackwell Publishers: Oxford.

[25] N. Keyfitz, 1966, A Life Table that Agrees with the Data, *Journal of the American Statistical Association*, 61(314): 305-12.

[26] J., Gruber, D., Wise, 1998, Social Security and Retirement: An International Comparison. *The American Economic Review*, Vol 88, No.2., 158-163.

[27] Ministry of Labour and Social Security-MoLSS (2014), *National Employment Strategy & Action Plans*, General Directorate of Labour. *https://www.resmigazete.gov.tr/eskiler/2017/07/20170707M1-1.pdf*

ANNEX. Multiple Decrement LTs

LT in the absence of causes other than death (Daughters)

Age	q	l	d	L	T	e	p	n	a	$_nq^i_x$	$_nq^j_x$	l_x	l^i_x	l^j_x	$_nd_x$	$_nd^i_x$	$_nd^j_x$	$_nL_x$	T_x	e_x	R^{-i}	$_np^{-i}_x$	$_na^{-i}_x$	$_nl^{-i}_x$	$_nd^{-i}_x$	$_nL^{-i}_x$	T^{-i}_x	e^{-i}_x
45	0,03312	100000	3312	98344	2901969	29,02	0,96688	1	0,5	0,03083	0,00229	100000	27931	72069	3312	3083	229	98344	2901969	29,02	0,06909	0,99768	0,5078	100000	232	99886	3749725	37,50
46	0,03092	96688	2990	95193	2803625	29,00	0,96908	1	0,5	0,02852	0,00241	96688	24847	71841	2990	2757	233	95193	2803625	29,00	0,07786	0,99756	0,5072	99768	244	99647	3649839	36,58
47	0,02868	93698	2687	92354	2708432	28,91	0,97132	1	0,5	0,02536	0,00332	93698	22090	71608	2687	2376	311	92354	2708432	28,91	0,11577	0,99664	0,5064	99524	335	99359	3550192	35,67
48	0,02650	91011	2412	89805	2616078	28,74	0,97350	1	0,5	0,02326	0,00324	91011	19714	71297	2412	2117	294	89805	2616078	28,74	0,12211	0,99673	0,4972	99189	325	99026	3450833	34,79
49	0,02447	88599	2168	87515	2526273	28,51	0,97553	1	0,5	0,02134	0,00313	88599	17597	71002	2168	1891	277	87515	2526273	28,51	0,12792	0,99684	0,5035	98864	313	98709	3351807	33,90
50	0,02267	86431	1959	85451	2438758	28,22	0,97733	1	0,5	0,01913	0,00353	86431	15706	70725	1959	1654	305	85451	2438758	28,22	0,15580	0,99643	0,5152	98552	351	98381	3253098	33,01
51	0,02112	84472	1784	83580	2353307	27,86	0,97888	1	0,5	0,01667	0,00445	84472	14052	70420	1784	1408	376	83580	2353307	27,86	0,21084	0,99551	0,5099	98200	441	97984	3154717	32,13
52	0,01987	82688	1643	81866	2269727	27,45	0,98013	1	0,5	0,01523	0,00463	82688	12644	70044	1643	1259	383	81866	2269727	27,45	0,23326	0,99533	0,5031	97759	456	97532	3056733	31,27
53	0,01891	81045	1533	80279	2187860	27,00	0,98109	1	0,5	0,01406	0,00485	81045	11385	69660	1533	1139	393	80279	2187860	27,00	0,25653	0,99511	0,5007	97303	475	97065	2959200	30,41
54	0,01827	79513	1453	78786	2107582	26,51	0,98173	1	0,5	0,01351	0,00476	79513	10245	69267	1453	1074	379	78786	2107582	26,51	0,26059	0,99521	0,5097	96827	464	96600	2862135	29,56
55	0,01861	78060	1453	77333	2028796	25,99	0,98139	1	0,5	0,01259	0,00602	78060	9171	68889	1453	983	470	77333	2028796	25,99	0,32340	0,99394	0,5043	96363	584	96074	2765535	28,70
56	0,01965	76607	1505	75854	1951462	25,47	0,98035	1	0,5	0,01421	0,00543	76607	8188	68419	1505	1089	416	75854	1951462	25,47	0,27651	0,99453	0,5127	95779	524	95524	2669462	27,87
57	0,01950	75102	1464	74370	1875608	24,97	0,98050	1	0,5	0,01173	0,00776	75102	7099	68003	1464	881	583	74370	1875608	24,97	0,39818	0,99219	0,5038	95255	744	94886	2573938	27,02
58	0,01866	73637	1374	72950	1801239	24,46	0,98134	1	0,5	0,01244	0,00622	73637	6218	67420	1374	916	458	72950	1801239	24,46	0,33333	0,99374	0,5012	94511	592	94216	2479051	26,23
59	0,01757	72263	1270	71629	1728288	23,92	0,98243	1	0,5	0,00951	0,00806	72263	5302	66962	1270	687	582	71629	1728288	23,92	0,45870	0,99190	0,5142	93920	760	93550	2384835	25,39
60	0,01657	70994	1177	70406	1656660	23,34	0,98343	1	0,5	0,00748	0,00909	70994	4615	66379	1177	531	645	70406	1656660	23,34	0,54851	0,99088	0,5050	93159	850	92739	2291285	24,60
61	0,01588	69817	1109	69263	1586254	22,72	0,98412	1	0,5	0,00657	0,00931	69817	4083	65734	1109	459	650	69263	1586254	22,72	0,58613	0,99066	0,4993	92309	862	91878	2198546	23,82
62	0,01556	68709	1069	68174	1516991	22,08	0,98444	1	0,5	0,00645	0,00911	68709	3625	65084	1069	443	626	68174	1516991	22,08	0,58568	0,99086	0,5006	91447	836	91030	2106668	23,04
63	0,01558	67639	1054	67112	1448817	21,42	0,98442	1	0,5	0,00595	0,00963	67639	3182	64458	1054	402	651	67112	1448817	21,42	0,61814	0,99034	0,5029	90611	875	90176	2015638	22,24
64	0,01584	66586	1054	66058	1381705	20,75	0,98416	1	0,5	0,00586	0,00997	66586	2779	63806	1054	390	664	66058	1381705	20,75	0,62972	0,99000	0,5075	89736	898	89294	1925462	21,46
65	0,01626	65531	1065	64999	1315646	20,08	0,98374	1	0,5	0,00462	0,01164	65531	2389	63142	1065	303	763	64999	1315646	20,08	0,71599	0,98833	0,5090	88839	1037	88330	1836168	20,67
66	0,01680	64466	1083	63924	1250648	19,40	0,98320	1	0,5	0,00405	0,01274	64466	2086	62380	1083	261	822	63924	1250648	19,40	0,75871	0,98723	0,5065	87802	1121	87249	1747839	19,91
67	0,01746	63383	1107	62830	1186723	18,72	0,98254	1	0,5	0,00351	0,01395	63383	1825	61558	1107	223	884	62830	1186723	18,72	0,79872	0,98603	0,5047	86681	1211	86081	1660590	19,16
68	0,01829	62276	1139	61707	1123894	18,05	0,98171	1	0,5	0,00359	0,01470	62276	1602	60674	1139	224	915	61707	1123894	18,05	0,80363	0,98528	0,5055	85470	1258	84847	1574509	18,42
69	0,01932	61137	1181	60547	1062187	17,37	0,98068	1	0,5	0,00300	0,01631	61137	1379	59759	1181	184	997	60547	1062187	17,37	0,84452	0,98366	0,5084	84211	1376	83535	1489662	17,69
70	0,02059	59956	1234	59339	1001640	16,71	0,97941	1	0,5	0,00205	0,01853	59956	1195	58761	1234	123	1111	59339	1001640	16,71	0,90030	0,98145	0,5060	82835	1537	82076	1406127	16,97
71	0,02212	58722	1299	58073	942301	16,05	0,97788	1	0,5	0,00249	0,01963	58722	1072	57650	1299	146	1153	58073	942301	16,05	0,88746	0,98035	0,5048	81298	1598	80507	1324051	16,29
72	0,02393	57423	1374	56736	884228	15,40	0,97607	1	0,5	0,00234	0,02159	57423	926	56498	1374	134	1240	56736	884228	15,40	0,90221	0,97839	0,5068	79701	1722	78851	1243544	15,60
73	0,02603	56049	1459	55320	827492	14,76	0,97397	1	0,5	0,00194	0,02409	56049	791	55258	1459	109	1350	55320	827492	14,76	0,92542	0,97589	0,5062	77978	1880	77050	1164692	14,94
74	0,02845	54590	1553	53814	772172	14,14	0,97155	1	0,5	0,00217	0,02627	54590	683	53908	1553	119	1434	53814	772172	14,14	0,92359	0,97370	0,5060	76098	2002	75109	1087642	14,29
75	0,03117	53037	1653	52211	718358	13,54	0,96883	1	0,5	0,00195	0,02923	53037	564	52474	1653	103	1550	52211	718358	13,54	0,93750	0,97075	0,5074	74096	2168	73029	1012533	13,67
76	0,03420	51384	1757	50505	666148	12,96	0,96580	1	0,5	0,00108	0,03313	51384	461	50924	1757	55	1702	50505	666148	12,96	0,96855	0,96685	0,5067	71929	2384	70753	939505	13,06
77	0,03753	49627	1862	48695	615642	12,41	0,96247	1	0,5	0,00088	0,03664	49627	405	49221	1862	44	1818	48695	615642	12,41	0,97643	0,96334	0,5053	69545	2549	68284	868752	12,49
78	0,04118	47764	1967	46781	566947	11,87	0,95882	1	0,5	0,00075	0,04043	47764	361	47403	1967	36	1931	46781	566947	11,87	0,98182	0,95956	0,5042	66995	2710	65652	800468	11,95
79	0,04527	45798	2073	44761	520166	11,36	0,95473	1	0,5	0,00139	0,04388	45798	326	45472	2073	64	2010	44761	520166	11,36	0,96933	0,95609	0,5050	64286	2823	62889	734816	11,43
80	0,05003	43724	2188	42631	475405	10,87	0,94997	1	0,5	0,00042	0,04961	43724	262	43462	2188	18	2169	42631	475405	10,87	0,99160	0,95038	0,5050	61463	3050	59954	671928	10,93
81	0,05584	41537	2319	40377	432774	10,42	0,94416	1	0,5	0,00122	0,05462	41537	244	41293	2319	51	2269	40377	432774	10,42	0,97813	0,94535	0,5049	58413	3192	56833	611974	10,48
82	0,06319	39218	2478	37978	392397	10,01	0,93681	1	0,5	0,00124	0,06196	39218	193	39025	2478	48	2430	37978	392397	10,01	0,98046	0,93800	0,5061	55221	3424	53530	555141	10,05
83	0,07269	36739	2671	35404	354419	9,65	0,92731	1	0,5	0,00141	0,07128	36739	144	36595	2671	52	2619	35404	354419	9,65	0,98058	0,92867	0,5072	51798	3695	49977	501611	9,68
84	0,08499	34069	2895	32621	319015	9,36	0,91501	1	0,5	0,00060	0,08439	34069	93	33976	2895	21	2875	32621	319015	9,36	0,99291	0,91559	0,5002	48103	4060	46073	451634	9,39
85	1,00000	31173	31173	286394	286394	9,19	0,00000			0,00231	0,99769	31173	72	31101	31173	72	31101	286394	286394	9,19	0,99769	0,00000	9,2085	44042	44042	405561	405561	9,21

LT in the absence of causes other than deaths&marriages (Daughters)

Age	q	l	d	L	T	e	p	n	a	$_nq^i_x$	$_nq^j_x$	l_x	l^i_x	l^j_x	$_nd_x$	$_nd^i_x$	$_nd^j_x$	$_nL_x$	T_x	e_x	R^{-i}	$_np^{-i}_x$	$_na^{-i}_x$	$_nl^{-i}_x$	$_nd^{-i}_x$	$_nL^{-i}_x$	T^{-i}_x	e^{-i}_x
45	0,03312	100000	3312	98344	2901969	29,02	0,96688	1	0,5	0,01990	0,01322	100000	18771	81229	3312	1990	1322	98344	2901969	29,02	0,39914	0,98665	0,5050	100000	1335	99339	3439340	34,39
46	0,03092	96688	2990	95193	2803625	29,00	0,96908	1	0,5	0,01818	0,01275	96688	16781	79907	2990	1757	1233	95193	2803625	29,00	0,41225	0,98713	0,5046	98665	1269	98036	3340001	33,85
47	0,02868	93698	2687	92354	2708432	28,91	0,97132	1	0,5	0,01645	0,01223	93698	15024	78674	2687	1541	1146	92354	2708432	28,91	0,42650	0,98766	0,5041	97395	1201	96800	3241965	33,29
48	0,02650	91011	2412	89805	2616078	28,74	0,97350	1	0,5	0,01463	0,01187	91011	13483	77528	2412	1331	1080	89805	2616078	28,74	0,44802	0,98804	0,4949	96194	1150	95613	3145166	32,70
49	0,02447	88599	2168	87515	2526273	28,51	0,97553	1	0,5	0,01339	0,01108	88599	12151	76448	2168	1186	982	87515	2526273	28,51	0,45290	0,98884	0,4906	95043	1060	94503	3049553	32,09
50	0,02267	86431	1959	85451	2438758	28,22	0,97733	1	0,5	0,01304	0,00962	86431	10965	75466	1959	1127	832	85451	2438758	28,22	0,42448	0,99032	0,4969	93983	910	93525	2955050	31,44
51	0,02112	84472	1784	83580	2353307	27,86	0,97888	1	0,5	0,01051	0,01061	84472	9838	74634	1784	888	897	83580	2353307	27,86	0,50246	0,98933	0,5005	93073	993	92577	2861525	30,75
52	0,01987	82688	1643	81866	2269727	27,45	0,98013	1	0,5	0,00991	0,00995	82688	8950	73738	1643	820	823	81866	2269727	27,45	0,50108	0,99000	0,4959	92080	921	91615	2768948	30,07
53	0,01891	81045	1533	80279	2187860	27,00	0,98109	1	0,5	0,00906	0,00985	81045	8131	72915	1533	734	799	80279	2187860	27,00	0,52100	0,99010	0,4954	91159	902	90703	2677333	29,37
54	0,01827	79513	1453	78786	2107582	26,51	0,98173	1	0,5	0,00922	0,00905	79513	7397	72116	1453	733	720	78786	2107582	26,51	0,49551	0,99090	0,5014	90256	821	89847	2586629	28,66
55	0,01861	78060	1453	77333	2028796	25,99	0,98139	1	0,5	0,00827	0,01035	78060	6663	71396	1453	645	808	77333	2028796	25,99	0,55589	0,98961	0,4951	89435	929	88966	2496782	27,92
56	0,01965	76607	1505	75854	1951462	25,47	0,98035	1	0,5	0,01166	0,00798	76607	6018	70588	1505	893	612	75854	1951462	25,47	0,40644	0,99197	0,5054	88506	711	88155	2407816	27,21
57	0,01950	75102	1464	74370	1875608	24,97	0,98050	1	0,5	0,00791	0,01159	75102	5125	69977	1464	594	870	74370	1875608	24,97	0,59422	0,98837	0,5032	87795	1021	87288	2319662	26,42
58	0,01866	73637	1374	72950	1801239	24,46	0,98134	1	0,5	0,00961	0,00905	73637	4531	69107	1374	708	666	72950	1801239	24,46	0,48498	0,99091	0,4963	86774	789	86377	2232374	25,73
59	0,01757	72263	1270	71629	1728288	23,92	0,98243	1	0,5	0,00655	0,01102	72263	3823	68440	1270	473	797	71629	1728288	23,92	0,62742	0,98894	0,5082	85985	951	85517	2145997	24,96
60	0,01657	70994	1177	70406	1656660	23,34	0,98343	1	0,5	0,00512	0,01145	70994	3350	67644	1177	363	813	70406	1656660	23,34	0,69109	0,98852	0,4994	85034	976	84545	2060480	24,23
61	0,01588	69817	1109	69263	1586254	22,72	0,98412	1	0,5	0,00477	0,01111	69817	2987	66831	1109	333	776	69263	1586254	22,72	0,69958	0,98886	0,4977	84058	936	83587	1975935	23,51
62	0,01556	68709	1069	68174	1516991	22,08	0,98444	1	0,5	0,00446	0,01111	68709	2654	66055	1069	306	763	68174	1516991	22,08	0,71367	0,98887	0,5009	83121	925	82660	1892347	22,77
63	0,01558	67639	1054	67112	1448817	21,42	0,98442	1	0,5	0,00398	0,01160	67639	2347	65292	1054	269	785	67112	1448817	21,42	0,74463	0,98838	0,5002	82196	955	81719	1809688	22,02
64	0,01584	66586	1054	66058	1381705	20,75	0,98416	1	0,5	0,00441	0,01143	66586	2078	64507	1054	293	761	66058	1381705	20,75	0,72170	0,98855	0,5036	81241	931	80779	1727969	21,27
65	0,01626	65531	1065	64999	1315646	20,08	0,98374	1	0,5	0,00338	0,01288	65531	1785	63746	1065	221	844	64999	1315646	20,08	0,79236	0,98710	0,5076	80310	1036	79800	1647190	20,51
66	0,01680	64466	1083	63924	1250648	19,40	0,98320	1	0,5	0,00270	0,01410	64466	1564	62902	1083	174	909	63924	1250648	19,40	0,83914	0,98589	0,5055	79274	1119	78721	1567390	19,77
67	0,01746	63383	1107	62830	1186723	18,72	0,98254	1	0,5	0,00234	0,01512	63383	1390	61993	1107	149	958	62830	1186723	18,72	0,86581	0,98486	0,5036	78155	1183	77568	1488669	19,05
68	0,01829	62276	1139	61707	1123894	18,05	0,98171	1	0,5	0,00243	0,01586	62276	1241	61035	1139	151	988	61707	1123894	18,05	0,86707	0,98412	0,5048	76972	1222	76367	1411101	18,33
69	0,01932	61137	1181	60547	1062187	17,37	0,98068	1	0,5	0,00184	0,01747	61137	1090	60048	1181	113	1068	60547	1062187	17,37	0,90459	0,98251	0,5061	75750	1325	75096	1334734	17,62
70	0,02059	59956	1234	59339	1001640	16,71	0,97941	1	0,5	0,00155	0,01903	59956	977	58979	1234	93	1141	59339	1001640	16,71	0,92447	0,98095	0,5044	74425	1418	73723	1259638	16,92
71	0,02212	58722	1299	58073	942301	16,05	0,97788	1	0,5	0,00192	0,02020	58722	884	57838	1299	113	1186	58073	942301	16,05	0,91318	0,97978	0,5044	73008	1476	72276	1185915	16,24
72	0,02393	57423	1374	56736	884228	15,40	0,97607	1	0,5	0,00196	0,02196	57423	771	56652	1374	113	1261	56736	884228	15,40	0,91798	0,97802	0,5059	71532	1573	70755	1113639	15,57
73	0,02603	56049	1459	55320	827492	14,76	0,97397	1	0,5	0,00176	0,02427	56049	658	55391	1459	99	1360	55320	827492	14,76	0,93220	0,97571	0,5062	69959	1699	69120	1042884	14,91
74	0,02845	54590	1553	53814	772172	14,14	0,97155	1	0,5	0,00170	0,02675	54590	559	54031	1553	93	1460	53814	772172	14,14	0,94020	0,97323	0,5061	68260	1827	67357	973764	14,27
75	0,03117	53037	1653	52211	718358	13,54	0,96883	1	0,5	0,00162	0,02955	53037	466	52571	1653	86	1567	52211	718358	13,54	0,94792	0,97043	0,5067	66433	1965	65464	906407	13,64
76	0,03420	51384	1757	50505	666148	12,96	0,96580	1	0,5	0,00097	0,03323	51384	380	51004	1757	50	1708	50505	666148	12,96	0,97170	0,96675	0,5065	64468	2144	63410	840943	13,04
77	0,03753	49627	1862	48695	615642	12,41	0,96247	1	0,5	0,00063	0,03689	49627	331	49296	1862	31	1831	48695	615642	12,41	0,98316	0,96309	0,5053	62324	2300	61186	777533	12,48
78	0,04118	47764	1967	46781	566947	11,87	0,95882	1	0,5	0,00062	0,04055	47764	299	47465	1967	30	1937	46781	566947	11,87	0,98485	0,95943	0,5044	60024	2435	58818	716347	11,93
79	0,04527	45798	2073	44761	520166	11,36	0,95473	1	0,5	0,00083	0,04443	45798	269	45528	2073	38	2035	44761	520166	11,36	0,98160	0,95555	0,5049	57589	2560	56322	657529	11,42
80	0,05003	43724	2188	42631	475405	10,87	0,94997	1	0,5	0,00028	0,04975	43724	231	43493	2188	12	2175	42631	475405	10,87	0,99440	0,95024	0,5047	55029	2738	53673	601207	10,93
81	0,05584	41537	2319	40377	432774	10,42	0,94416	1	0,5	0,00105	0,05479	41537	219	41318	2319	43	2276	40377	432774	10,42	0,98125	0,94518	0,5047	52291	2867	50872	547534	10,47
82	0,06319	39218	2478	37978	392397	10,01	0,93681	1	0,5	0,00124	0,06196	39218	176	39042	2478	48	2430	37978	392397	10,01	0,98046	0,93800	0,5063	49425	3064	47912	496662	10,05
83	0,07269	36739	2671	35404	354419	9,65	0,92731	1	0,5	0,00094	0,07175	36739	127	36612	2671	35	2636	35404	354419	9,65	0,98706	0,92821	0,5071	46360	3328	44720	448751	9,68
84	0,08499	34069	2895	32621	319015	9,36	0,91501	1	0,5	0,00060	0,08439	34069	93	33976	2895	21	2875	32621	319015	9,36	0,99291	0,91559	0,5002	43032	3632	41217	404030	9,39
85	1,00000	31173	31173	286394	286394	9,19	0,00000			0,00231	0,99769	31173	72	31101	31173	72	31101	286394	286394	9,19	0,99769	0,00000	9,2085	39400	39400	362814	362814	9,21

LT in the absence of causes other than deaths&marriages&work (Daughters)

Age	q	l	d	L	T	e	p	n	a	$_nq^i_x$	$_nq^j_x$	l_x	l^i_x	l^j_x	$_nd_x$	$_nd^i_x$	$_nd^j_x$	$_nL_x$	T_x	e_x	R^{-i}	$_np^{-i}_x$	$_na^{-i}_x$	$_nl^{-i}_x$	$_nd^{-i}_x$	$_nL^{-i}_x$	T^{-i}_x	e^{-i}_x
45	0,03312	100000	3312	98344	2901969	29,02	0,96688	1	0,5	0,01126	0,02186	100000	13630	86370	3312	1126	2186	98344	2901969	29,02	0,66009	0,97801	0,5028	100000	2199	98907	3272779	32,73
46	0,03092	96688	2990	95193	2803625	29,00	0,96908	1	0,5	0,01103	0,01989	96688	12504	84184	2990	1067	1923	95193	2803625	29,00	0,64327	0,98000	0,5028	97801	1956	96829	3173872	32,45
47	0,02868	93698	2687	92354	2708432	28,91	0,97132	1	0,5	0,00946	0,01922	93698	11437	82261	2687	886	1801	92354	2708432	28,91	0,67022	0,98068	0,5024	95845	1851	94924	3077043	32,10
48	0,02650	91011	2412	89805	2616078	28,74	0,97350	1	0,5	0,00905	0,01745	91011	10551	80460	2412	824	1588	89805	2616078	28,74	0,65842	0,98247	0,4925	93994	1647	93158	2982120	31,73
49	0,02447	88599	2168	87515	2526273	28,51	0,97553	1	0,5	0,00772	0,01675	88599	9727	78872	2168	684	1484	87515	2526273	28,51	0,68453	0,98318	0,4888	92346	1553	91552	2888962	31,28
50	0,02267	86431	1959	85451	2438758	28,22	0,97733	1	0,5	0,00917	0,01349	86431	9043	77388	1959	793	1166	85451	2438758	28,22	0,59539	0,98644	0,4890	90793	1231	90164	2797410	30,81
51	0,02112	84472	1784	83580	2353307	27,86	0,97888	1	0,5	0,00747	0,01365	84472	8251	76221	1784	631	1153	83580	2353307	27,86	0,64631	0,98630	0,4976	89562	1227	88946	2707245	30,23
52	0,01987	82688	1643	81866	2269727	27,45	0,98013	1	0,5	0,00678	0,01309	82688	7620	75068	1643	561	1082	81866	2269727	27,45	0,65875	0,98687	0,4938	88335	1160	87748	2618299	29,64
53	0,01891	81045	1533	80279	2187860	27,00	0,98109	1	0,5	0,00685	0,01206	81045	7059	73986	1533	555	978	80279	2187860	27,00	0,63791	0,98789	0,4913	87175	1055	86638	2530551	29,03
54	0,01827	79513	1453	78786	2107582	26,51	0,98173	1	0,5	0,00741	0,01086	79513	6504	73008	1453	589	864	78786	2107582	26,51	0,59435	0,98910	0,4993	86120	939	85650	2443913	28,38
55	0,01861	78060	1453	77333	2028796	25,99	0,98139	1	0,5	0,00643	0,01218	78060	5915	72145	1453	502	950	77333	2028796	25,99	0,65425	0,98778	0,4938	85181	1041	84654	2358263	27,69
56	0,01965	76607	1505	75854	1951462	25,47	0,98035	1	0,5	0,01037	0,00927	76607	5412	71194	1505	795	710	75854	1951462	25,47	0,47193	0,99068	0,5013	84141	784	83750	2273608	27,02
57	0,01950	75102	1464	74370	1875608	24,97	0,98050	1	0,5	0,00676	0,01274	75102	4618	70484	1464	507	957	74370	1875608	24,97	0,65350	0,98721	0,5020	83356	1066	82826	2189859	26,27
58	0,01866	73637	1374	72950	1801239	24,46	0,98134	1	0,5	0,00854	0,01012	73637	4110	69527	1374	629	745	72950	1801239	24,46	0,54220	0,98984	0,4949	82291	836	81868	2107033	25,60
59	0,01757	72263	1270	71629	1728288	23,92	0,98243	1	0,5	0,00577	0,01179	72263	3481	68782	1270	417	852	71629	1728288	23,92	0,67135	0,98817	0,5069	81455	964	80979	2025165	24,86
60	0,01657	70994	1177	70406	1656660	23,34	0,98343	1	0,5	0,00423	0,01234	70994	3064	67930	1177	301	876	70406	1656660	23,34	0,74455	0,98763	0,4982	80491	995	79992	1944185	24,15
61	0,01588	69817	1109	69263	1586254	22,72	0,98412	1	0,5	0,00434	0,01154	69817	2763	67054	1109	303	806	69263	1586254	22,72	0,72689	0,98843	0,4954	79496	920	79032	1864194	23,45
62	0,01556	68709	1069	68174	1516991	22,08	0,98444	1	0,5	0,00422	0,01134	68709	2461	66248	1069	290	779	68174	1516991	22,08	0,72885	0,98863	0,5008	78576	893	78130	1785162	22,72
63	0,01558	67639	1054	67112	1448817	21,42	0,98442	1	0,5	0,00353	0,01204	67639	2171	65469	1054	239	815	67112	1448817	21,42	0,77327	0,98793	0,5001	77683	937	77214	1707032	21,97
64	0,01584	66586	1054	66058	1381705	20,75	0,98416	1	0,5	0,00418	0,01165	66586	1932	64654	1054	279	776	66058	1381705	20,75	0,73585	0,98832	0,5029	76746	896	76300	1629818	21,24
65	0,01626	65531	1065	64999	1315646	20,08	0,98374	1	0,5	0,00310	0,01315	65531	1653	63878	1065	203	862	64999	1315646	20,08	0,80907	0,98683	0,5072	75849	999	75357	1553517	20,48
66	0,01680	64466	1083	63924	1250648	19,40	0,98320	1	0,5	0,00252	0,01428	64466	1450	63016	1083	163	920	63924	1250648	19,40	0,84987	0,98571	0,5047	74850	1070	74320	1478160	19,75
67	0,01746	63383	1107	62830	1186723	18,72	0,98254	1	0,5	0,00229	0,01517	63383	1287	62096	1107	145	962	62830	1186723	18,72	0,86901	0,98481	0,5036	73780	1121	73224	1403840	19,03
68	0,01829	62276	1139	61707	1123894	18,05	0,98171	1	0,5	0,00227	0,01602	62276	1142	61134	1139	141	998	61707	1123894	18,05	0,87613	0,98396	0,5048	72659	1166	72082	1330616	18,31
69	0,01932	61137	1181	60547	1062187	17,37	0,98068	1	0,5	0,00177	0,01754	61137	1001	60136	1181	109	1073	60547	1062187	17,37	0,90813	0,98244	0,5059	71494	1255	70874	1258534	17,60
70	0,02059	59956	1234	59339	1001640	16,71	0,97941	1	0,5	0,00149	0,01909	59956	893	59064	1234	89	1145	59339	1001640	16,71	0,92749	0,98089	0,5044	70238	1342	69573	1187661	16,91
71	0,02212	58722	1299	58073	942301	16,05	0,97788	1	0,5	0,00185	0,02027	58722	803	57919	1299	109	1190	58073	942301	16,05	0,91640	0,97971	0,5042	68896	1398	68203	1118087	16,23
72	0,02393	57423	1374	56736	884228	15,40	0,97607	1	0,5	0,00196	0,02196	57423	695	56729	1374	113	1261	56736	884228	15,40	0,91798	0,97802	0,5059	67499	1484	66766	1049884	15,55
73	0,02603	56049	1459	55320	827492	14,76	0,97397	1	0,5	0,00168	0,02435	56049	582	55467	1459	94	1365	55320	827492	14,76	0,93559	0,97562	0,5064	66015	1609	65221	983118	14,89
74	0,02845	54590	1553	53814	772172	14,14	0,97155	1	0,5	0,00161	0,02684	54590	488	54102	1553	88	1465	53814	772172	14,14	0,94352	0,97314	0,5060	64406	1730	63551	917898	14,25
75	0,03117	53037	1653	52211	718358	13,54	0,96883	1	0,5	0,00152	0,02966	53037	400	52637	1653	80	1573	52211	718358	13,54	0,95139	0,97032	0,5065	62676	1860	61758	854347	13,63
76	0,03420	51384	1757	50505	666148	12,96	0,96580	1	0,5	0,00097	0,03323	51384	320	51064	1757	50	1708	50505	666148	12,96	0,97170	0,96675	0,5064	60815	2022	59817	792589	13,03
77	0,03753	49627	1862	48695	615642	12,41	0,96247	1	0,5	0,00063	0,03689	49627	270	49356	1862	31	1831	48695	615642	12,41	0,98316	0,96309	0,5053	58793	2170	57720	732772	12,46
78	0,04118	47764	1967	46781	566947	11,87	0,95882	1	0,5	0,00062	0,04055	47764	239	47525	1967	30	1937	46781	566947	11,87	0,98485	0,95943	0,5046	56623	2297	55485	675052	11,92
79	0,04527	45798	2073	44761	520166	11,36	0,95473	1	0,5	0,00069	0,04457	45798	209	45588	2073	32	2041	44761	520166	11,36	0,98466	0,95541	0,5050	54326	2422	53127	619567	11,40
80	0,05003	43724	2188	42631	475405	10,87	0,94997	1	0,5	0,00014	0,04989	43724	177	43547	2188	6	2181	42631	475405	10,87	0,99720	0,95011	0,5047	51904	2590	50621	566440	10,91
81	0,05584	41537	2319	40377	432774	10,42	0,94416	1	0,5	0,00087	0,05496	41537	171	41366	2319	36	2283	40377	432774	10,42	0,98438	0,94501	0,5047	49314	2712	47971	515818	10,46
82	0,06319	39218	2478	37978	392397	10,01	0,93681	1	0,5	0,00103	0,06216	39218	135	39083	2478	40	2438	37978	392397	10,01	0,98371	0,93780	0,5063	46603	2899	45172	467847	10,04
83	0,07269	36739	2671	35404	354419	9,65	0,92731	1	0,5	0,00071	0,07199	36739	94	36645	2671	26	2645	35404	354419	9,65	0,99029	0,92799	0,5069	43704	3147	42152	422675	9,67
84	0,08499	34069	2895	32621	319015	9,36	0,91501	1	0,5	0,00060	0,08439	34069	69	34000	2895	21	2875	32621	319015	9,36	0,99291	0,91559	0,5002	40557	3423	38846	380523	9,38
85	1,00000	31173	31173	286394	286394	9,19	0,00000			0,00154	0,99846	31173	48	31125	31173	48	31125	286394	286394	9,19	0,99846	0,00000	9,2014	37133	37133	341678	341678	9,20

İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 13, 2020, 2, 61-77

Geliş/Received:20.11.2020, Kabul/Accepted: 10.12.2020

www.istatistikciler.org

Araştırma Makalesi / Research Article

Eliptik sözde-kopulalar ile esnek bağımlılık modellemesi

Övgücan Karadağ ERDEMİR

Hacettepe Üniversitesi
Aktüerya Bilimleri Bölümü
06800-Çankaya, Ankara, Turkey
ovgucan@hacettepe.edu.tr
0000-0002-4725-3588

Meral SUCU

Hacettepe Üniversitesi
Aktüerya Bilimleri Bölümü
06800-Çankaya, Ankara, Turkey
msucu@hacettepe.edu.tr
0000-0002-7991-1792

Öz

Finansal ve aktüeryal bağımlılık modellemesi çalışmalarında sıklıkla tercih edilen eliptik kopulalar, dinamik bağımlılık modellemesi elde etmek ve kuyruk bağımlılığı ile eliptik bağımlılığın modellenebilmesi gibi çeşitli nedenler ile düzenlenebilir. Düzenlenmiş sözde-kopula ile esnek bir bağımlılık modellemesi elde edilir. Bu çalışmada, düzenlenmiş sözde-kopula fonkiyonlarının düzenlemeden sonra da sözde-kopula fonksiyonu özelliğini koruduğu gösterilmiş ve bu fonksiyonların elde edilme aşamaları verilmiştir. Uygulama bölümünde, sözde-kopula fonksiyonlarının düzenlenmesinin etkinliği perspektif ve izohips eğrileri ile incelenmiş ve düzenlemenin sağladığı fayda, düzenlenmiş sözde-kopula regresyon modelleri yardımıyla gösterilmiştir.

Anahtar sözcükler: *Bağımlılık, Düzenlenmiş sözde-Gauss kopula, Eliptik kopula, Kuyruk bağımlılığı, Sözde-gözlemler.*

Abstract

Flexible dependence modelling with elliptical pseudo-copulas

Elliptical copulas, which are frequently preferred in financial and actuarial dependency modeling studies, can be modified for various reasons such as obtaining dynamic dependency modeling and modeling tail dependency and elliptic dependence. With the modified pseudo-copula, a flexible dependency modeling is obtained. In this study, it is shown that the modified pseudo-copula functions preserve pseudo-copula function feature after the modification and the stages of obtaining these functions are given. In the application part, the efficiency of the modification of the pseudo-copula functions is examined with perpective curves and contour lines, and the benefit provided by the modification is demonstrated with the help of modified pseudo-copula regression models.

Keywords: *Dependence, Modified Gaussian pseudo-copula, Elliptical copula,, Tail dependence, Pseudo-observation.*

1. Giriş

Bağımlılık modellemesi denince akla ilk gelen modeller kuşkusuz kopula fonksiyonlarıdır. Ortak olasılık dağılım fonksiyonlarının, marjinal olasılık dağılım fonksiyonları ile ifade edilmesini sağlayan Sklar'ın Teoremi, birçok bağımlılık çalışmasının temelini oluşturur [1]. Bağımlılık çalışmalarında kopula, daha çok finansal ve istatistiksel veriler için kullanılmış, aktüerya bilimlerinde ise bağımlılığın kopula ile modellenmesi yakın geçmişte çalışılmıştır.

Aktüeryal çalışmalarda kopula ile bağımlılığın modellenmesinde, Frees ve Valdez'in [2] çalışması önemli ve temel kaynak niteliğindedir. Hayat sigortalarında çoklu azalım modellerinde ve birleşik yaşam ürünlerinde [3, 4], sağlık sigortalarında hasar tutarı ile sayısı arasındaki [5, 6] bağımlılık kopula ile modellenmiştir. Hayat dışı sigortalarda ise, hasar tutarı ile sayısı arasındaki bağımlılığın [7, 8, 9], hasar sayıları arasındaki bağımlılığın [10, 11], hasar türleri arasındaki bağımlılığın [12-14], hasar nedenleri arasındaki bağımlılığın [15, 16], hasar süresi ve hasar tutarı arasındaki bağımlığın [17] ve mekansal bağımlılığın [18, 19] modellemesinde kopula fonksiyonundan yararlanılmaktadır. Hayat dışı sigortaların rezerv hesaplamalarında [20-23] ve aktüeryal verilerle yapılan finansal çalışmalarda da [24-26] bağımlılığın modellenmesinde kopula kullanılmıştır.

Eliptik dağılımların toplamı yine bir eliptik dağılım olduğundan, çok değişkenli eliptik dağılımlardan elde edilen eliptik kopula fonksiyonları, finansal çalışmalarda varlık portföylerindeki bağımlığın ve aktüerya bilimlerinde toplam hasardaki bağımlılığın modellenmesinde tercih edilir [27]. Eliptik dağılım ailesinin en sık kullanılan üyeleri Gauss kopula ve t-kopula fonksiyonlarıdır. Gauss kopula fonksiyonu, çok değişkenli normal dağılıma bağlı olarak tanımlandığından ve değişken sayısının çok olması durumunda, diğer parametrik kopula fonksiyonlarına göre çalışılması daha kolay olduğundan tercih edilir [28, 29]. Gauss kopula fonksiyonu sahip olduğu bu avantajların yanında, asimetrik bağımlılığın, eliptik bağımlılığın, dairesel simetriye göre bağımlılığın ve kuyruk bağımlılığının modellenmesinde yetersiz kalmaktadır [28, 30, 31, 32].

Bağımlılık çalışmalarında, var olan kopula fonksiyonları daha dinamik bağımlılık yapıları elde etmek ya da kuyruk bağımlılığı ve eliptik bağımlılığın modellenebilmesi gibi çeşitli sebeplerle düzenlenmiştir. Patton [33, 34] zamana bağlı olarak değişen koşullu bağımlılığın ve çok değişkenli bağımlılığın modellenebilmesi amacıyla koşullu kopula tanımını geliştirmiştir. Fermanian ve Wegkamp [35, 36], koşullu kopula fonksiyonlarını geliştirerek sözde-kopula fonksiyonlarını tanımlamıştır. Sözde-kopula fonksiyonu ile kopulanın durağan yapısından uzaklaşarak, dinamik yani zamana göre değişen bir yapı kazanması amaçlanmaktadır. Sözde-kopula fonksiyonunun ilişki katsayısı, çeşitli fonksiyonlar yardımıyla değiştirilerek, düzenlenmiş sözde-Gauss kopula geliştirilmiştir [30, 31]. Kopula fonksiyonlarında yapılan değişikliklerin etkisinin test edilmesi ve farklı kopula fonksiyonlarının karşılaştırılaştırılarak uygun kopulanın seçilmesi ile ilgili çalışmalar da yapılmıştır [37, 38].

Bağımlılık ve ilişkili olma durumu farklı olmakla birlikte, birbiri ile bağlantılıdır. Gauss kopula gibi bazı bağımlılık modelleme yöntemlerinde parametrenin direkt ilişki katsayısı olmasından dolayı bağımlılık ilişki katsayısı ile doğrudan ilişkilendirilebilir. Değişkenler arasındaki ilişki; kovaryans, varyans-kovaryans matrisi, ilişki katsayısı ve ilişki matrisi ile ölçülebilir. İlişki katsayısı ve ilişki matrisi; kovaryans ve varyans-kovaryans matrisinin standartlaştırılmış halidir. Rastlantı değişkenleri arasındaki ilişki; Pearson doğrusal ilişki katsayısı, sıra ilişki katsayılarından Spearman'ın ρ ile Kendall'ın τ katsayıları ve kuyruk bağımlılığı katsayısı olarak sınıflandırılan bağımlılık ölçümleri ile incelenebilir.

Bu çalışmada, sözde-Gauss kopula fonksiyonlarının düzenlenmesiyle elde edilen düzenlenmiş sözde-kopula fonksiyonları ile esnek bir bağımlılık modellemesi sunulmuştur. Düzenleme yardımıyla verideki bağımlılığa uygun kopula fonksiyonu ile doğru bir bağımlılık modellemesi elde edilerek, bağımlılık hesaplamalara dahil edilebilir ve gerçeğe daha yakın hesaplamalar yapılabilir.

Bu çalışmada İkinci Bölüm'de genel olarak kopula fonksiyonları, özellikleri ile eliptik kopulalardan Gauss kopula ve sözde-Gauss kopula fonksiyonları hakkında kısaca bilgi verilmiştir. Üçüncü Bölüm'de düzenlenmiş sözde-kopula fonkiyonlarının elde edilme aşamaları ve düzenlemeden sonra da kopula fonksiyonunun sözde-kopula fonksiyonu özelliğini koruduğu gösterilmiştir. Dördüncü Bölüm'de düzenlenmiş sözde-kopula regresyon model yardımıyla bağımlılık varsayımı altında hasar tutarı ile sayısının ortak dağılım ve ortak olasılık yoğunluk fonksiyonlarının çıkarımına kısaca yer verilmiştir. Beşinci Bölüm'de Sigorta Bilgi Merkezi (SBM)'den alınan verinin kullanıldığı uygulamada, sözde-kopula fonksiyonlarında düzenleme yapmanın etkinliği perspektif ve izohips eğileri ile incelenmiş ve düzenlemenin sağladığı esneklik düzenlenmiş sözde-kopula regresyon modeli yardımıyla gösterilmiştir. Çalışmanın sonucunda elde edilen sonuç ve bulgulara ise Altıncı Bölüm'de yer verilmiştir.

2. Yöntem

2.1. Kopula

Bağımlılık modellemesi denince akla ilk gelen kopula fonksiyonu 1959 yılında Sklar tarafından ortaya atılmış ve hala birçok çalışmanın temelini oluşturmaktadır [1]. İstatistik, ekonomi, ekonometri ve finans gibi birçok bilim dalında değişkenler arasındaki bağımlılığın modellenmesinde kopula fonksiyonlarından yararlanılmaktadır. Aktüerya Bilimleri alanında kopulanın ilk kullanıldığı çalışmalar ağırlıklı olarak finansal çalışmalar olmakla birlikte, zaman içinde hayat ve hayat dışı sigortalar matematiğinde bağımlılık modellemesinde tercih edilmiştir.

Bağ, ilişki ve birlikte hareket eden anlamlarına gelen kopula, çok değişkenli ve normal olmayan dağılımların tanımlanmasında ve bağımlılığın modellenmesinde kullanılır. Kopula fonksiyonu, marjinal dağılımları birleştiren çok değişkenli dağılım fonksiyonu olarak tanımlanabilir [39]. Dağılım fonksiyonlarının tanım aralığı **I**=[0,1] olduğu için kopula; marjinal uniform dağılımlı rastlantı değişkenlerini birleştiren çok değişkenli bir dağılım fonksiyonu olarak da tanımlanabilir.

Kopula fonksiyonlarının daha iyi anlaşılabilmesi için reel eksen, genişletilmiş reel eksen, artan ve azalan fonksiyonların özellikleri, bir fonksiyonun hacmi gibi temel matematiksel bilgilerin verilmesinde fayda vardır. $(-\infty,\infty)$ açık aralığı $\boldsymbol{R}$ reel eksenini, $[-\infty,\infty]$ kapalı aralığı ise $\overline{\boldsymbol{R}}$ genişletilmiş reel ekseni ifade eder. Benzer biçimde; $\boldsymbol{R}^2=\boldsymbol{R}x\boldsymbol{R}$ reel düzlem iken, $\overline{\boldsymbol{R}}^{\,2}=\overline{\boldsymbol{R}}\,x\overline{\boldsymbol{R}}$ genişletilmiş reel düzlemdir. Genişletilmiş reel düzlem üzerinde iki kapalı aralığın kartezyen çarpımı; köşeleri (x_1,y_1), (x_1,y_2), (x_2,y_1) ve (x_2,y_2) olan $B=[x_1,x_2]x[y_1,y_2]$ dikdörtgenidir. İki değişkenli $H(.,.)$ fonksiyonunun tanım kümesi $DomH$, $\overline{\boldsymbol{R}}^{\,2}$ genişletilmiş reel düzlemin alt kümesidir ve aynı fonksiyonun görüntü kümesi $RanH$ fonksiyonu, $\boldsymbol{R}$ reel eksenin alt kümesidir [39].

S_1 ve S_2, $\overline{\boldsymbol{R}}$ genişletilmiş reel ekseninin boş olmayan alt kümeleri ve $H(.,.)$ tanım kümesi $DomH=S_1xS_2$ olan bir fonksiyon olsun. $B=[x_1,x_2]x[y_1,y_2]$, $H(.,.)$'nin tanım bölgesindeki tüm noktaları içeren bir bölge olsun. B bölgesinin *H-hacmi*,

$$V_H(B)=H(x_2,y_2)-H(x_2,y_1)-H(x_1,y_2)-H(x_1,y_1) \quad (1)$$

biçimindedir.

Köşe noktaları, $H(.,.)$ fonksiyonunun tanım kümesi $DomH$ üzerinde olan tüm B dikdörtgenleri için $V_H(B) \geq 0$ ise, iki değişkenli $H(.,.)$ fonksiyonu 2-artandır.

$\mathbf{I}^2$=[0,1]x[0,1]'den **I**=[0,1]'e tanımlanan ve $C(.,.)$ ile gösterilen iki-değişkenli kopula fonksiyonu çalışmalarda kısaca C kopula fonksiyonu olarak kullanılmaktadır. $\boldsymbol{u} = (u_1, u_2)$ ve $\boldsymbol{v} = (v_1, v_2)$ olmak üzere $C: \boldsymbol{I}^2 \rightarrow \boldsymbol{I}$ 2-değişkenli kopula fonksiyonunun sağlaması gereken özellikler aşağıda belirtilmiştir [39]:

1. Her $\boldsymbol{u} \epsilon [0,1]$ için, $\boldsymbol{u}$'nun üyelerinden biri 0 iken, kopula fonksiyonu 0 değerini alır.

   $C(u_1, 0) = C(0, u_2) = 0$

2. Her $\boldsymbol{u} \epsilon [0,1]$ için aşağıdaki eşitlikler sağlanmalıdır.

   $C(u_1, 1) = u_1$ ve $C(1, u_2) = u_2$

3. Her $\boldsymbol{u} \epsilon [0,1]$ ve $\boldsymbol{v} \epsilon [0,1]$ için $u_1 \leq u_2$ ve $v_1 \leq v_2$ ise kopula hacmi $V_C = V_{[u_1,v_1]x[u_2,v_2]}$ pozitiftir.

   $V_C = C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) - C(u_1, v_1) \geq 0$

Kopula fonksiyonu, *u* ve *v* değişkenleri kullanılarak tanımlanabildiği gibi, Sklar'ın Teoremi yardımıyla dağılım fonksiyonları kullanılarak da tanımlanabilir [1]. Dağılım fonksiyonları $F_{X_1}(x_1)$, $F_{X_2}(x_2)$,…,$F_{X_n}(x_n)$ olan n rastlantı değişkeni olsun. Sklar'ın Teoremi'ne göre,

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = C[F_{X_1}(x_1), \dots, F_{X_n}(x_n)] \tag{2}$$

eşitliğini sağlayan bir C kopula fonksiyonu mevcuttur.

$F_{X_1}(x_1), \dots, F_{X_n}(x_n)$ marjinal dağılım fonksiyonları sürekli ise, C kopula fonksiyonu tektir ve değer kümeleri üzerinde tek olarak tanımlanır. Diğer taraftan, C bir kopula fonksiyonu ve $F_{X_1}(x_1), \dots, F_{X_n}(x_n)$ marjinal dağılım fonksiyonları ise $C[F_{X_1}(x_1), \dots, F_{X_n}(x_n)]$ fonksiyonu n rastlantı değişkeninin ortak olasılık dağılım fonksiyonudur. Sklar'ın Teoremi, kopula fonksiyonlarının varlığını gösteren bir teoremdir. Kopula fonksiyonunun elemanları kesikli veya karma ise C fonksiyonu tek değildir ancak bir kopula fonksiyonudur.

Kopulalar parametrik ve parametrik olmayan kopulalar olarak ayrılabilir. Parametrik kopulalar genel olarak eliptik kopula (t-kopula (Student) ve Gauss kopula) ve Arşimet kopula (Clayton kopula, Gumbel kopula, Frank kopula) olarak ayrılır [40]. Parametrik olmayan kopulalar Bernstein, grid-tip ve kernel kopulalar olarak sınıflandırılabilir. Bu kopula fonksiyonlarının dışında birçok kopula fonksiyonu mevcut olup, farklı bağımlılık yapılarının modellenmesinde kullanılırlar. Eliptik kopula türlerinden biri olan Gauss kopula fonksiyonu oldukça yaygın kullanılan bir kopula türüdür.

2.1.1. Gauss Kopula

$\Phi(.)$ tek değişkenli standart normal dağılıma ait dağılım fonksiyonunu, $\Phi_n(.|\boldsymbol{\Gamma})$ $\boldsymbol{\Gamma}$ ilişki matrisine sahip n değişkenli standart normal dağılıma ait dağılım fonksiyonunu göstersin. $\boldsymbol{u} = (u_1, \dots, u_n) \in \boldsymbol{I}^n$ olmak üzere $C: \boldsymbol{I}^n \rightarrow \boldsymbol{I}$ n-değişkenli Gauss kopula fonksiyonu aşağıdaki gibi yazılır:

$$C(u_1, \dots, u_n|\boldsymbol{\Gamma}) = \Phi_n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)|\boldsymbol{\Gamma}) \tag{3}$$

n=2 durumunda elde edilecek 2-değişkenli Gauss kopula fonksiyonu aşağıdaki gibi elde edilir:

$$C(u_1, u_2|\mathbf{\Gamma}) = \Phi_n(\Phi^{-1}(u_1), \Phi^{-1}(u_2)|\mathbf{\Gamma}) \quad (4)$$

$$C(u_1, u_2|\mathbf{\Gamma}) = \frac{\partial}{\partial u_1}\int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_{i2})}\frac{1}{2\pi\sqrt{1-(\rho_{12})^2}}exp\left\{-\frac{s^2-2\rho_{12}st+t^2}{2(1-(\rho_{12})^2)}\right\}dsdt \quad (5)$$

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}$$

Bağımlılık modellemesi çalışmalarında Gauss kopula diğer kopula türlerine göre daha çok tercih edilir. Gauss Kopula fonksiyonunun avantajları Brigo ve diğerleri [28] ile Song [29] tarafından aşağıdaki gibi sıralanmıştır:

- Kopula fonksiyonundaki değişken sayısı çok olduğunda, Gauss kopula fonksiyonu ile çalışmak diğer parametrik kopula fonksiyonlarına göre daha kolaydır.
- Çok değişkenli normal dağılıma bağlı olarak tanımlanan Gauss kopula fonksiyonundaki ilişki ve bağımlılık ölçümleri kolay tanımlanır.
- Gauss kopula fonksiyonunun tüm bileşenleri klasik doğrusal regresyon modeli olduğunda, vektör genelleştirilmiş doğrusal model (VGDM), klasik çok değişkenli doğrusal modele dönüşür. Ancak diğer parametrik kopula türleri ile oluşturulan VGDM'ler için aynı durum söz konusu değildir.

Gauss kopula fonksiyonunun avantajları olmasına karşın; asimetrik bağımlılığın, eliptik bağımlılığın, kuyruk bağımlılığının ve dairesel simetriye göre bağımlılığın modellenmesinde yetersiz kalmaktadır [28, 30, 31, 32].

Dağılım fonksiyonları $F_{X_1}(x_1)$ ve $F_{X_2}(x_2)$ olan iki rastlantı değişkeni olsun. Eğer,

$$\lambda_U = \lim_{u \to 1} P\left(X_2 > F_{X_2}^{-1}(u) \middle| X_1 > F_{X_1}^{-1}(u)\right) \quad (6)$$

ise X_1 ve X_2 rastlantı değişkenleri üst kuyruk bağımlılığına sahiptir ve $\lambda_U \in (0,1]$'dir. Eğer,

$$\lambda_L = \lim_{u \to 0} P\left(X_2 \leq F_{X_2}^{-1}(u) \middle| X_1 \leq F_{X_1}^{-1}(u)\right) \quad (7)$$

ise X_1 ve X_2 rastlantı değişkenleri alt kuyruk bağımlılığına sahiptir ve $\lambda_L \in (0,1]$'dir. Eğer X_1 ve X_2 rastlantı değişkenleri bağımsız ise $\lambda_U = \lambda_L = 0$'dır [28]. Gauss kopula fonksiyonunda üst ve alt kuyruk bağımlılığı olmadığı için $\lambda_U = \lambda_L = 0$'dır.

Finansal ve aktüeryal verilerde asimetrik bağımlılık ve kuyruk bağımlılığı ile karşılaşma olasılığı yüksektir. Bu nedenle özellikle gerçek veri ile yapılan uygulamalarda, bağımlılığın modellenmesinde kopula fonksiyonlarında belirli düzenlemeler yapılabilmektedir [30, 31].

2.2. Sözde-Kopula

Kopula fonksiyonları ile bağımlılık durağan olarak modellenmektedir, ancak finansal ve aktüeryal verilerdeki bağımlılık dinamik bir yapıya sahiptir. Fermanian ve Wegkamp [35, 36] çalışmalarında daha dinamik bir bağımlılık yapısı elde etmek amacıyla Patton [33, 34] tarafından öne sürülen koşullu kopulayı geliştirerek sözde-kopula fonksiyonunu tanımlamışlardır. Patton'un [34] koşullu kopula tanımı, Fermanian ve Wegkamp [36] tarafından ölçüm kuramı ve Sklar'ın Teoremi'ne ait özellikleri sağlayacak şekilde geliştirilmiştir. Koşullu ve sözde-kopulalar, Cherubini ve diğerleri [41] tarafından parametreleri ve şekli zamana göre değişen dinamik kopulalar olarak adlandırılmıştır.

$\boldsymbol{u} = (u_1, \dots, u_n)$ ve $\boldsymbol{v} = (v_1, \dots, v_n)$ olmak üzere $PC: \boldsymbol{I}^n \to \boldsymbol{I}$ n-değişkenli sözde-kopula fonksiyonun sağlaması gereken özellikler aşağıda belirtilmiştir:

1. Her $\boldsymbol{u}\epsilon[0,1]^n$ için, $\boldsymbol{u}$'nun üyelerinden en az biri 0 iken $i = 1, \dots, n$ için $PC(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_n) = 0$'dır.
2. $PC(1, \dots, 1) = 1$'dir.
3. Her $\boldsymbol{u}\epsilon[0,1]^n$ ve $\boldsymbol{v}\epsilon[0,1]^n$ için, $\boldsymbol{u} \le \boldsymbol{v}$ $(u_1 \le v_1, \dots, u_n \le v_n)$ iken, $[u, v]$'nin kopula hacmi olan $V_{[u_1,v_1]x\dots x[u_n,v_n]}$ pozitiftir [35, 41].

Sözde-kopulalar, bir özellik dışında kopula fonksiyonu özelliklerini sağlar. Sözde-kopulalara has olan bu özellik, $PC(u_1, \dots, u_n)$ çok değişkenli kopula fonksiyonunda $1 \le k \le n$ için u_k dışında tüm elemanların 1'e eşit olma koşulunun aranmamasıdır.

Fermanian ve Wegkamp [35] ile Cherubini ve diğerleri [41] sözde-kopulanın diğer özelliklerini tanımlamak amacıyla bazı teoremlerden yararlanmışlardır. $H(.)$ R^n'de tanımlı n-değişkenli bir dağılım fonksiyonu ve $F_1(.), \dots, F_n(.)$ R'de tanımlı tek değişkenli n adet dağılım fonksiyonu olsun. Her $\boldsymbol{x} = (x_1, \dots, x_n)$, $\widetilde{\boldsymbol{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$ için $1 \le j \le n$ iken $F_j(x_j) = F_j(\tilde{x}_j)$ ise $H(\boldsymbol{x}) = H(\widetilde{\boldsymbol{x}})$ olur. Her $\boldsymbol{x} = (x_1, \dots, x_n)$ için $H(\boldsymbol{x}) = C(F_1(x_1), \dots, F_n(x_n))$ eşitliğinin sağlandığı bir $C(.)$ sözde-kopula fonksiyonu mevcuttur. Ayrıca $C(.)$ eğer bir sözde-kopula fonksiyonu ve $F_1(.), \dots, F_n(.)$ fonksiyonları tek değişkenli dağılım fonksiyonları ise, $H(.)$ fonksiyonu n-değişkenli bir dağılım fonksiyonudur [35]. $C(.)$ sözde-kopulası her $j = 1, \dots, n$ ve $\boldsymbol{x} = (x_1, \dots, x_n) \in R^n$ için ancak ve ancak $H(\infty, \dots, x_j, \dots, \infty) = F_j(x_j)$ eşitliğini sağlıyorsa bir kopula fonksiyonudur [41].

Sözde-Gauss kopula, en yaygın kullanılan sözde-kopula fonksiyonudur ve bir tür eliptik kopuladır. Eliptik kopulalar, eliptik dağılımlar için önerilmiş kopulalardır. Gauss dağılımı dışında finansal çalışmalara uygunluk sağlayan t-dağılımının kullanıldığı kopula fonksiyonları da vardır. Eliptik sözde-kopulalar ile daha dinamik bir yapı elde edilmektedir. Ancak dairesel simetriden dolayı alt ve üst kuyruğu aynı algılandığından, alt ve üst kuyruk bağımlılığını modelleme de yetersiz kalırlar. Zaman içinde bu sorunları gidermek için Arşimet ve sözde-Arşimet kopula fonksiyonları tanımlanmıştır [39]. Arşimet kopulalarla çok değişkenli çalışmalar yapmak zor olduğundan ve gerçek veriye uyum sağlamada sorun yaşandığından bağımlılığın modellenmesinde yetersiz kalmıştır. Kopula fonksiyonları belirli düzenlemelerle değiştirilerek [30, 31] veya farklı kopula fonksiyonları ağırlıklı ortalama veya doğrusal kombinasyon gibi belirli fonksiyonlarla bir araya getirilerek [42] bu sorunlar giderilmeye çalışılmaktadır.

2.3. Düzenlenmiş Sözde-Kopula

Kuyruk bağımlılığının olmaması, özellikle gerçek veri uygulamalarında oldukça kullanışlı olan Gauss kopula fonksiyonunun eksik yönlerinden biridir. Standart Gauss kopula yerine sözde-Gauss kopula fonksiyonunu kullanan Fang [30] ile Fang ve Madsen [31] çalışmalarında eliptik ve kuyruk bağımlılığını daha iyi modellemek amacıyla bu kopula fonksiyonlarını düzenlemişlerdir. Gerçek sigorta verileri ile kullanılabilecek dinamik yapıda, eliptik bağımlılığı ve kuyruk bağımlılığını modelleyebilen düzenlenmiş sözde-kopula fonksiyonları önerilmiştir. Sözde-kopula fonksiyonları yeniden tanımlanırken ilişki matrisi düzenlenmiştir. $\boldsymbol{\Gamma}$ ilişki matrisinin elemanları olan ilişki katsayılarını; modelin kuyruğa yakınsama hızını kontrol eden a_{ij} parametresi, modelin kuyruk şeklini kontrol eden b_{ij} parametresi ile kopulanın elemanlarının $(u_1, \dots, u_n)$ bir fonksiyonu olarak düzenleyerek $\boldsymbol{\Gamma}^*$ düzenlenmiş ilişki matrisi elde edilmiştir. Yapılan düzenlemeler aşağıda kısaca özetlenmiştir:

- $C(u_1, \dots, u_n)$ standart bir kopula fonksiyonudur. Kesim 2.1'de verilen tüm özellikleri sağladığı varsayılır.
- $PC(u_1, \dots, u_n)$ bir sözde-kopula ise, standart kopulanın sağlaması gereken özelliklerden, $C(1, \dots, u_k, \dots 1) = u_k,\ 1 \le k \le p$ eşitliğini sağlamak zorunda değildir.
- $PC(u_1, \dots, u_n)$ bir düzenlenmiş sözde-kopula ise, standart kopulanın sağlaması gereken özelliklerinden biri olan; $C(1, \dots, u_k, \dots 1) = u_k,\ 1 \le k \le p$ eşitliğinin sağlanması gerekmez. $\mathbf{\Gamma}^*$ düzenlenmiş ilişki matrisinin elemanları olan ρ_{ij} ilişki katsayıları, modelin kuyruğa yakınsama hızını kontrol eden a_{ij} parametresi ve modelin kuyruk şeklini kontrol eden b_{ij} parametresi ile kopula fonksiyonunun elemanları olan $\boldsymbol{u} = (u_1, \dots, u_n)$'nun bir fonksiyonu olarak $\rho_{ij}^* = f(u_i, u_j; a_{ij}, b_{ij})$ biçimindedir. $\mathbf{\Gamma}^*$ düzenlenmiş ilişki matrisi aşağıda verilmiştir.

$$\mathbf{\Gamma}^* = \begin{bmatrix} 1 & \rho_{12}^* = f(a_{12}, b_{12}; u_1, u_2) & \cdots & \rho_{1n}^* = f(a_{1n}, b_{1n}; u_1, u_n) \\ \rho_{12}^* = f(a_{12}, b_{12}; u_1, u_2) & 1 & \cdots & \cdots \\ \vdots & \cdots & \ddots & \cdots \\ \rho_{1n}^* = f(a_{1n}, b_{1n}; u_1, u_n) & \cdots & \cdots & 1 \end{bmatrix}$$

Fang ve Madsen [31] ilişki katsayısını tanımlamak için beş farklı fonksiyon ele almıştır. Bu fonksiyonların seçiminde eliptik yapıdan dolayı trigonometrik fonksiyonlar tercih edilmiştir. Fang ve Madsen [31], $\rho_{ij} = f(a_{ij}, b_{ij}, u_i, u_j)$ fonksiyonunu kısaca $\rho(u, v; a, b)$ olarak ele alıp, tanımlama yapmıştır. İlişki katsayısını tanımlamak için ele aldıkları fonksiyonlar $b(1 - auv)$, $b\,cos\left(\frac{\pi}{2}a(1 - uv)\right)$, $b\,sin\left(\frac{\pi}{2}a(1 - uv)\right)$, $b\,tan\left(\frac{\pi}{4}a(1 - uv)\right)$ ve $b\exp\{-a(1 - uv)\}$'dır. İlk dört fonksiyon için $a \in [0,1]$, son fonksiyon için $a \in [0, \infty)$'dır. Tüm fonksiyonlar için $b \in [-1,1]$ alınmıştır.

Bu fonksiyonlar 0 ve 1 aralığında değer alan fonksiyonlardır. $a = 0$ ve $b = 0$ olduğunda, düzenlenen kopula fonksiyonları bilinen kopula fonksiyonlarına dönüşmektedir. Tanımlanan beş fonksiyonda, $b = 0$ ise $\rho = 0$ olacağından kopula modeli $C(u, v) = uv$ biçiminde bağımsız kopula modeline dönüşmektedir. $a = 0$ iken, Fonksiyon 3 ve Fonksiyon 4 için, $\rho = 0$ olacağından kopula modeli bağımsız kopula modeline dönüşmektedir. $a = 0$ iken; Fonksiyon 1, Fonksiyon 2 ve Fonksiyon 5 kopula modeli sözde- Gauss kopula modelidir [31].

Fang ve Madsen [31] tarafından önerilen düzenlenmiş kopula fonksiyonunun Gauss kopula fonksiyonuna göre daha kullanışlı olmasını sağlayan özellikler aşağıdaki gibi sıralanabilir:

- Kalın kuyruklu, alt/üst kuyruklu ve eliptik dağılımlar ile iyi sonuçlar verir.
- Simetriklik ve asimetriklik sorununa çözüm sunar.
- Pozitif ve negatif bağımlılık sorununa çözüm sunar.
- Sözde-kopula, olasılık yoğunluk fonksiyonu için kapalı form tanımlayabilir.

3. Sözde-Kopula Fonksiyonlarının Düzenlenmesi

Sözde-kopula fonksiyonları, Kesim 2.3'te verilen düzenlemeden sonra da sözde-kopula fonksiyonu olma özelliklerini korur. Bu özelliklerin sağlandığı aşağıda gösterilmiştir. $PC(u_{i1}, u_{i2}|\mathbf{\Gamma}^*)$ eşitliğinde q_{i1} ve q_{i2} normal skorlarının $\Phi^{-1}(u_{i1})$ ve $\Phi^{-1}(u_{i2})$ değerlerine eşit olduğu bilgisiyle, değişkenler y ve z olarak değiştirilirse aşağıdaki eşitlik elde edilir:

$$PC(u_{i1},u_{i2}|\mathbf{\Gamma}^*) = \frac{1}{N}\int_0^{u_{i1}}\int_0^{u_{i2}} \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz$$

$$N = \int_0^1\int_0^1 \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz$$

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & \rho_{12}^* \\ \rho_{12}^* & 1 \end{bmatrix}$$

1. **Özellik:** u_{i1} veya u_{i2} değerlerinden herhangi biri 0 olduğunda $PC(u_{i1},u_{i2}|\mathbf{\Gamma}^*)$ kopula fonksiyonu 0 değerini alır.

$$u_{i1} = 0$$

$$PC(0,u_{i2}|\mathbf{\Gamma}^*) = \frac{1}{N}\int_0^{0}\int_0^{u_{i2}} \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz = \frac{1}{N}*0 = 0$$

$$u_{i2} = 0$$

$$PC(u_{i1},0|\mathbf{\Gamma}^*) = \frac{1}{N}\int_0^{u_{i1}}\int_0^{0} \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz = \frac{1}{N}\int_0^{u_{i1}} 0\,dz = \frac{1}{N}*0 = 0$$

2. **Özellik:** u_{i1} ve u_{i2} değerlerinden ikisi de 1 olduğunda $PC(u_{i1},u_{i2}|\mathbf{\Gamma}^*)$ kopula fonksiyonunu 1 değerini alır.

$u_{i1} = 1$ ve $u_{i2} = 1$

$$PC(1,1|\mathbf{\Gamma}^*) = \frac{1}{N}\int_0^1\int_0^1 \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz$$

$N = \int_0^1\int_0^1 \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz$ biçiminde tanımlandığından,

$$PC(1,1|\mathbf{\Gamma}^*) = \frac{1}{N}*N = 1$$

3. **Özellik:** Kopula hacmi $V_{[u_{i1},v_{i1}]x[u_{i2},v_{i2}]} \geq 0$'dır.

N ve dolayısıyla $\frac{1}{N}$ pozitif değerlidir.

$$V_{[u_{i1},v_{i1}]x[u_{i2},v_{i2}]} = \frac{1}{N}\int_{u_{i1}}^{v_{i1}}\int_{u_{i2}}^{v_{i2}} \frac{\phi_2(\Phi^{-1}(y),\Phi^{-1}(z)|\mathbf{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))}dydz$$

$$N = \int_0^1 \int_0^1 \frac{\phi_2(\Phi^{-1}(y), \Phi^{-1}(z)|\boldsymbol{\Gamma}^*)}{\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z))} > 0$$

$$\frac{1}{N} > 0$$

İntegralin içindeki tüm değerler pozitif değerlidir.

$0 \leq \phi_2(\Phi^{-1}(y), \Phi^{-1}(z)|\boldsymbol{\Gamma}^*) \leq 1$ (iki değişkenli olasılık yoğunluk fonksiyonu)

$0 \leq \phi(\Phi^{-1}(y)) \leq 1$ (tek değişkenli olasılık yoğunluk fonksiyonu)

$0 \leq \phi(\Phi^{-1}(z)) \leq 1$ (tek değişkenli olasılık yoğunluk fonksiyonu)

$\phi(\Phi^{-1}(y))\phi(\Phi^{-1}(z)) > 0$ (tek değişkenli olasılık yoğunluk fonksiyonu)

İntegrali oluşturan tüm değerler 0 veya 0'dan büyük olduğundan kopula hacmi 0 veya 0'dan büyüktür.

$$V_{[u_{i1},v_{i1}]x[u_{i2},v_{i2}]} \geq 0$$

İki değişkenli kopula fonksiyonu ile ilgili özel bir durum söz konusudur. İki değişkenli kopula fonksiyonunda kopula hacmi, $0 \leq u_{i1} < u_{i2} \leq 1$ ve $0 \leq v_{i1} < v_{i2} \leq 1$ olmak üzere,

$$V_{[u_{i1},v_{i1}]x[u_{i2},v_{i2}]} = C(u_{i2}, v_{i2}) - C(u_{i1}, v_{i2}) - C(u_{i2}, v_{i1}) + C(u_{i1}, v_{i1})$$

eşitliği ile ifade edilir. Bu fonksiyon, $[u_{i1}, v_{i1}]x[u_{i2}, v_{i2}]$ bölgesinde,

$P(u_{i1} \leq u \leq u_{i2}, v_{i1} \leq v \leq v_{i2})$ bir dağılım fonksiyonudur. Bir dağılım fonksiyonu olduğu için $P(u_{i1} \leq u \leq u_{i2}, v_{i1} \leq v \leq v_{i2}) \geq 0$'dır. Diğer bir ifade ile, $V_{[u_{i1},v_{i1}]x[u_{i2},v_{i2}]} \geq 0$ olur ve üçüncü özelliğin de sağlandığı görülür.

Fang ve Madsen [30, 31]çalışmalarında önerdikleri düzenlenmiş sözde-Gauss kopula fonksiyonlarının, standart Gauss kopula fonksiyonuna göre daha iyi olduğunu uyum iyiliği testleri ile incelemişlerdir. Ayrıca düzenlenmiş sözde-kopula fonksiyonun diğer sözde-kopulara göre daha iyi performans gösterdiği ve daha esnek bir model olduğu sigorta ve finans verisini içeren gerçek veri analizi ve Kendall'ın τ yaklaşımına göre yapılan benzetim çalışması ile gösterilmiştir. Düzenlenmiş sözde-Gauss kopula fonksiyonun, sözde-Gauss kopula fonksiyona göre kuyruk bağımlılığını daha iyi modellediği ve yine düzenlenmiş sözde Gauss kopula fonksiyonun, parametrik kopula türlerinden biri olan Arşimet kopulalara göre dairesel simetriye göre bağımlılığı daha iyi modellediği gösterilmiştir.

4. Bağımlılık Varsayımı Altında Hasar Tutarı ile Sayısının Ortak Dağılım ve Ortak Olasılık Yoğunluk Fonksiyonları

Bu çalışmada sözde-kopula fonksiyonları üzerindeki düzenleme, düzenlenmiş sözde-kopula regresyon modelleri [43] yardımıyla bağımlılık varsayımı altında elde edilen hasar tutarı ile sayısının ortak dağılım ve ortak olasılık yoğunluk fonksiyonları ile gösterilmiştir. Bu nedenle kopula regresyon modelleri [7, 8] ile düzenlenmiş sözde-kopula regresyon modellerine [43] kısaca değinilmiştir.

Hasar tutarının gamma ve hasar sayısının Poisson dağılımlı olduğu varsayımı altında, $\boldsymbol{\Gamma}$ ilişki matrisi ve $C(.,.|\boldsymbol{\Gamma})$ Gauss kopula fonksiyonunu göstermek üzere, kopula regresyon modeli,

C(Gamma GDM, Poisson GDM|**Γ**)

biçiminde oluşturulmaktadır. Düzenlenmiş sözde-kopula regresyon modelinin elde edilmesi amacıyla Fang [30] tarafından önerilen düzenlenmiş sözde-kopula fonksiyonu, Czado ve diğerleri [7] tarafından önerilen karma regresyon modeli içine dahil edilmiş ve özellikle hayat-dışı sigortalarda hasar tutarı ile sayısı arasındaki bağımlılığın modellenmesinde kullanılan modeller önerilmiştir. $\boldsymbol{\Gamma}^*$ düzenlenmiş ilişki matrisi ve $PC(.,.|\boldsymbol{\Gamma}^*)$ düzenlenmiş sözde-Gauss kopula fonksiyonu olmak üzere, düzenlenmiş sözde-kopula regresyon modeli,

PC(Gamma GDM, Poisson GDM|$\boldsymbol{\Gamma}$*)

biçiminde oluşturulabilir [43]. Kastenmeirer [8] ile Czado ve arkadaşlarının [7] çalışmasına benzer şekilde kopula fonksiyonun türevi, kısmi artık test istatistiği yardımıyla ifade edilirse,

$$PC_1'(u_1, u_{i2}|\boldsymbol{\Gamma}^*) = \Phi\left(\frac{\Phi^{-1}(u_{i2}) - \rho_{12}^*\Phi^{-1}(u_1)}{\sqrt{(1-(\rho_{12}^*)^2)}}\right) := D_{\rho_{12}^*}(u_1, u_{i2}) \tag{8}$$

ilişkisi elde edilir. Hasar sayısına ilişkin kesikli kısmın kopula türevi için

$$PC_1'(u_1, u_{i2}^-|\boldsymbol{\Gamma}^*) := \Phi\left(\frac{\Phi^{-1}(u_{i2}^-) - \rho_{12}^*\Phi^{-1}(u_1)}{\sqrt{(1-(\rho_{12}^*)^2)}}\right) := D_{\rho_{12}^*}(u_1, u_{i2}^-) \tag{9}$$

ilişkisi kurulmuştur.

$PC_1'(u_1, u_{i2}|\boldsymbol{\Gamma}^*)$ kopula türevi, ortak olasılık yoğunluk fonksiyonuna geçmek amacıyla bulunmuştur. Kopula fonksiyonu ortak olasılık dağılım fonksiyonu ile kopula türevi de ortak olasılık yoğunluk fonksiyonu ile ilişkilidir. Bağımlılık varsayımı altında hasar tutarı ile sayısına ait ortak olasılık dağılım fonksiyonu Skalar'ın Teoremi yardımıyla Eşitlik (10) ile ve ortak olasılık yoğunluk fonksiyonu karma kopula yaklaşımı [7], Radon-Nikodym Türevi [29] yardımıyla Eşitlik (11) ile aşağıdaki gibi elde edilmiştir:

$$F_{X_1X_2}(x_{i1}, x_{i2}) = PC(u_{i1}, u_{i2}|\boldsymbol{\Gamma}^*) \tag{10}$$

$$f_{X_1X_2}(x_{i1}, x_{i2}) = f_{X_1}(x_{i1})[PC_1'(u_{i1}, u_{i2}|\boldsymbol{\Gamma}^*) - PC_1'(u_{i1}, u_{i2}^-|\boldsymbol{\Gamma}^*)] \tag{11}$$

$PC_1'(.,.,|\boldsymbol{\Gamma}^*) := D_{\rho_{12}^*}(.,.)$ olduğundan Eşitlik (11),

$$f_{X_1X_2}(x_{i1}, x_{i2}) = \begin{cases} f_{X_1}(x_{i1})\left[D_{\rho_{12}^*}\left(F_{X_1}(x_{i1}), F_{X_2}(x_{i2})\right)\right], & , x_{i2} = 0 \\ f_{X_1}(x_{i1})\left[D_{\rho_{12}^*}\left(F_{X_1}(x_{i1}), F_{X_2}(x_{i2})\right) - D_{\rho_{12}^*}\left(F_{X_1}(x_{i1}), F_{X_2}(x_{i2}-1)\right)\right], & x_{i2} \geq 1 \end{cases} \tag{12}$$

biçiminde Eşitlik (12) ile ifade edilebilir. Burada Karadağ Erdemir [43] tarafından önerilen düzenlenmiş sözde-kopula regresyon modeli yardımıyla elde edilmiş hasar tutarı ile sayısına ait ortak olasılık dağılım ve ortak olsılık yoğunluk fonksiyonları en sade şekliyle özetlenmiştir. Model varsayımları, fonksiyonların elde edilme aşamaları hakkında detaylı bilgi için Karadağ Erdemir'in [43] çalışması incelenebilir.

5. Uygulama

Sözde-kopula fonksiyonunun uygulamalarında sözde-gözlemlerden yararlanılır [30, 31]. R_{ij}, X_{ij} rastlantı değişkeninin rankı ve $\hat{u}_{ij} = \frac{R_{ij}}{(n+1)}$ olmak üzere, $\hat{\boldsymbol{u}}_{\boldsymbol{i}} = (\hat{u}_{i1}, \dots, \hat{u}_{id})^T$ sözde-gözlemlerini göstermektedir. $\boldsymbol{X_i} = (X_{i1}, \dots, X_{id})^T$ rastlantı değişkenlerinden rank yardımıyla hesaplanır (i=1,…n; j=1,…d). $\hat{F}_j(.)$

dağılım fonksiyonu ve $\frac{n}{(n+1)}$ ölçekleme katsayısı olmak üzere, bir sözde-gözlem $\hat{u}_{ij} = \frac{n\hat{F}_j(X_{ij})}{(n+1)}$ biçiminde dağılım fonksiyonu yardımıyla da ifade edilir [44]. Gauss kopula fonksiyonu içinde $\hat{\boldsymbol{u}}_{\boldsymbol{i}} = (\hat{u}_{i1}, \hat{u}_{i2}) \epsilon [0,1]^2$ sözde-gözlemlerinden yararlanılmıştır. Sözde-gözlemler ile çalışmanın avantajı, kitleyi en iyi şekilde temsil edecek örneklemle çalışma olanağı sağlamasıdır [45]. Sözde-gözlemler R'da copula paketi altındaki *pobs()* fonksiyonundan yararlanarak, Sigorta Bilgi Merkezi'nden (SBM) alınan 2017 yılı Türkiye kasko sigortalarına ait hasar tutarı ve sayılarından üretilmiştir. SBM'den elde edilen hasar değişkenlerine ait betimleyici istatistikler Çizelge 1'de verilmiştir.

Çizelge 1. Hasar değişkenlerinin betimleyici istatistikleri

	Minimum	1.Çeyrek	Medyan	Ortalama	3.Çeyrek	Maximum
Toplam Hasar Tutarı	51,67	425	1.000	2.128,48	2.350	174.206
Ortalama Hasar Tutarı	51,67	429	966,43	2.068,57	2.249,37	130.000
Hasar Sayısı	1	1	1	1,035	1	8

Hasar tutarı değişkenini modellemek için Gamma GDM ve hasar sayısı değişkenini modellemek için Poisson GDM'den yararlanılmıştır. Bu iki marjial GDM, düzenlenmiş sözde-Gauss kopula fonksiyonu içine dahil edilerek Dördüncü Bölüm 'de bahsedilen düzenlenmiş sözde-kopula regresyon modelleri elde edilmiştir.

Sözde-Gauss kopula fonksiyonu, kopulanın parametresi olan ilişki katsayısı yardımıyla düzenlendiğinden, düzenlemenin faydasını analiz etmek amacıyla, sabit ilişki katsayılı sözde-Gauss kopula fonksiyonu ve çeşitli fonksiyonlarla düzenlenmiş sözde-Gauss kopula fonksiyonlarının perspektif ve izohips grafikleri elde edilmiştir. Düzenlemede Fang ve Madsen'in [31] çalışmasındaki üç fonksiyon ele alınmıştır. a_{12}, modelin kuyruğa yakınsama hızını ve b_{12}, modelin kuyruk şeklini kontrol eden parametreler, $(\hat{u}_1, \hat{u}_2)$ sözde-gözlemler olmak üzere bu fonksiyonlar;

$$\rho_{12}^{*1} = b_{12} cos\left(\frac{\pi}{2} a_{12}(1 - \hat{u}_1 \hat{u}_2)\right), \ a_{12} \in [0, \infty) \text{ ve } b_{12} \in [-1,1] \quad (13)$$

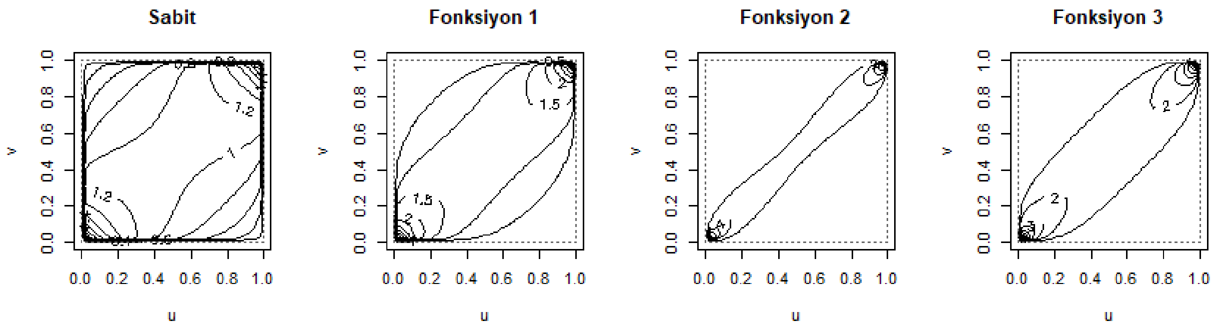
$$\rho_{12}^{*2} = b_{12} sin\left(\frac{\pi}{2} a_{12}(1 - \hat{u}_1 \hat{u}_2)\right), \ a_{12} \in [0, \infty) \text{ ve } b_{12} \in [-1,1] \quad (14)$$

$$\rho_{12}^{*3} = b_{12} tan\left(\frac{\pi}{4} a_{12}(1 - \hat{u}_1 \hat{u}_2)\right), \ a_{12} \in [0, \infty) \text{ ve } b_{12} \in [-1,1] \quad (15)$$

biçimindedir. Fonksiyon 1, 2 ve 3 sırasıyla ρ_{12}^{*1}, ρ_{12}^{*2} ve ρ_{12}^{*3} ile gösterilmiştir.

Düzenlenmiş Gauss kopula fonksiyonlarının persfektif ve izohips grafiklerini vermeden önce kısaca bu grafiklerin tanımının ve özelliklerinin verilmesi faydalıdır. Perspektif grafikleri değişkenler arasındaki ilişkileri incelemek amacıyla kullanılmaktadır. Üç boyutlu grafik yorumları zor olduğundan, perspektif grafikleri yerine genellikle iki boyutlu izohips eğrileri yorumlanır. Başka bir ifade ile izohips eğrileri üç boyutlu şekilleri iki boyutlu olarak ifade eder. Gauss kopula fonksiyonu ile çalışıldığından çok değişkenli normal dağılımdan yararlanılır. İzohips eğrileri özdeğerler ve varyans-kovaryans matrisinin özvektörleri yardımıyla çizilir. İzohips eğrileri grafiğinde çizgiler eşit olasılıklı olayları gösterir ve olasılık izohipsleri eliptik bir şekle sahiptir. Elipslerin yönü özdeğerlerin yönüne, elipslerin uzunluğu ise özvektör çarpımlarına bağlıdır [46].

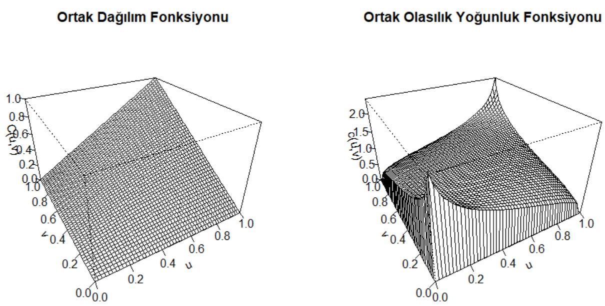
$(\hat{u}_1, \hat{u}_2)$ sözde-gözlem çifti yardımıyla sabit ilişki katsayısı, fonksiyon 1, 2 ve 3'e göre düzenlenmiş ilişki katsayılarına göre izohips eğrileri verilmiştir. Çok sayıda farklı ρ , a_{12} ve b_{12} değerleri ile grafikler çizdirilmiştir. Kopula fonksiyonları üzerinde ilişki matrisi yardımıyla düzenleme yapmanın avantajının daha iyi anlaşılması için sabit ve düzenlenmiş ilişki katsayılı fonksiyonlar için ρ=0,30, a_{12}=0,75 ve b_{12}=0,75 alındığında elde edilen grafik çizimleri Şekil 1'de verilmiştir.



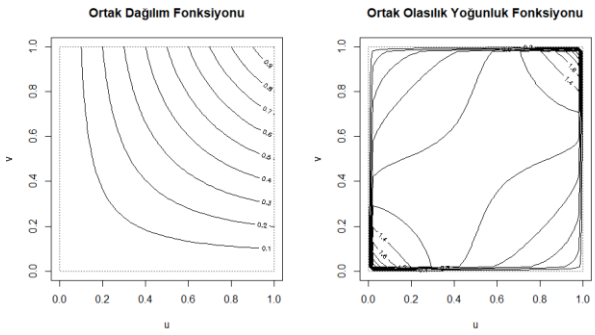
Şekil 1. Sabit ve düzenlenmiş ilişki katsayılı sözde-Gauss kopula fonksiyonlara göre ortak olasılık yoğunluk fonksiyonlarının izohips eğrileri

Sabit ve her bir düzenlenmiş ilişki katsayısı için ortak olasılık yoğunluk fonksiyonlarının grafiklerine bakılırsa farklı kuyruk bölgesi olasılıkları görülmüştür. İlişki katsayısının sabit bir değer yerine, fonksiyon şeklinde düzenlenmesi daha esnek ve dinamik bir bağımlılık modellemesi sağlamaktadır. Ayrıca kuyruğun şekli fonksiyondan fonksiyona değiştiği için farklı kuyruk bağımlılık yapıları elde edilebilmektedir. Bu düzenleme risk grubunun özelliğine göre farklı ilişki katsayıları ve dolayısıyla farklı ilişki matrisleri ile çalışılmasına olanak tanımaktadır.

Düzenlemenin etkisi incelendikten sonra, düzenlenmiş sözde-kopula regresyon modeli ile bulunan hasar tutarı ve sayısının ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonlarının üç boyutlu perspektif ve iki boyutu izohips eğrileri sabit ρ_{12}=0,25 ilişki katsayısı için sırasıyla Sekil 2'de ve Şekil 3'te verilmiştir.



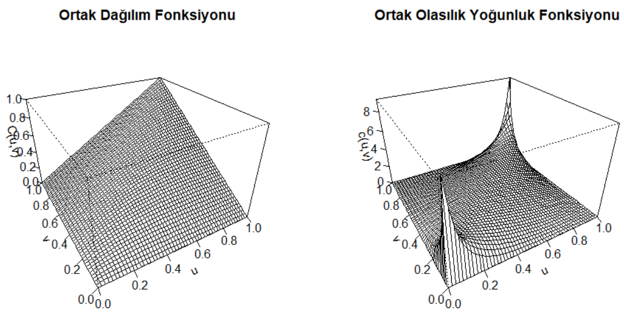
Şekil 2. Ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonunun perspektif grafiği (sözde-Gauss kopula)



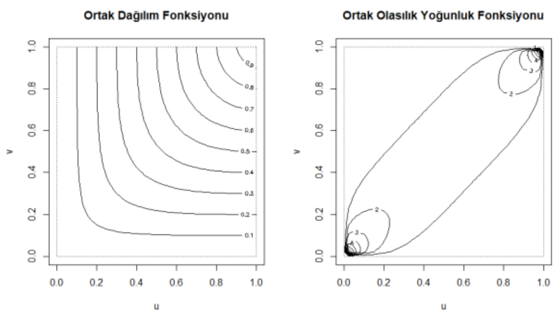
Şekil 3. Hasar tutarı ve sayısının ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonunun izohips eğrileri (sözde-Gauss kopula)

Sabit ilişki katsayılı sözde-Gauss kopulanın kullanıldığı durumda, Şekil 2'de üç boyutlu olarak verilen ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonları, Şekil 3'te iki boyutlu ifade edilmiştir. Ortak olasılık yoğunluk fonksiyonuna ait izohips eğrisinden, pozitif yönlü ve düşük düzeyli bir ilişki gözlemlenmektedir ve ρ=0,25 gibi pozitif küçük bir değer olduğundan bu beklenen bir sonuçtur.

Düzenlenmiş fonksiyonlardan ρ_{12}^{*3} için, a_{12}=0,75 ve b_{12}=0,75 iken hasar tutarı ve sayısının ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonlarının üç boyutlu perspektif ve iki boyutu izohips eğrilerini sırasıyla Şekil 4'te ve Şekil 5'te verilmiştir.



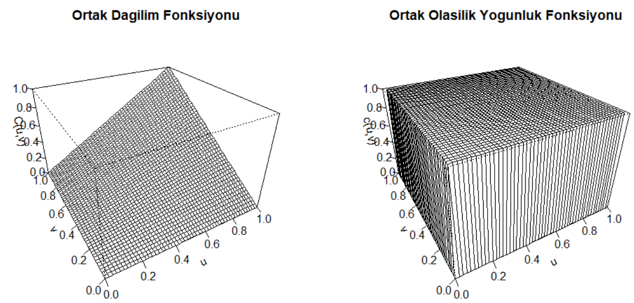
Şekil 4. Ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonunun perspektif grafiği (düzenlenmiş sözde-Gauss kopula)



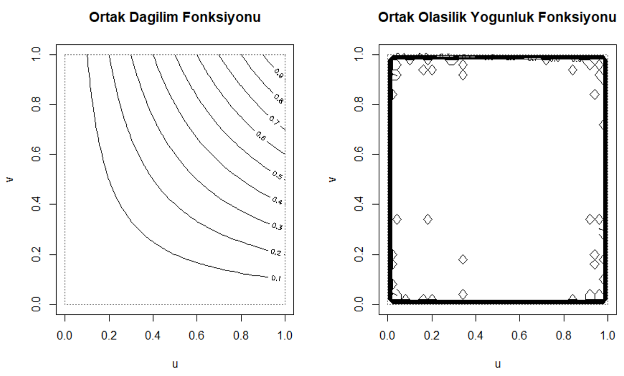
Şekil 5. Ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonunun izohips eğrileri (düzenlenmiş sözde-Gauss kopula)

Üçüncü fonksiyona göre düzenlenmiş ilişki katsayılı sözde-Gauss kopulanın kullanıldığı durumda, Şekil 4'te üç boyutlu olarak verilen ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonları, Şekil 5'te iki boyuta indirgenmiştir. Ortak olasılık yoğunluk fonksiyonuna ait izohips eğrisinden pozitif yönlü ve yüksek düzeyli bir ilişki gözlemlenmektedir. a_{12}=0,75 ve b_{12}=0,75 katsayıları pozitif ve bulunduğu değer aralığına göre yüksek değerler olduğundan, izohips eğrisinden gözlenen ilişki beklenen bir sonuçtur.

Yukarıda Şekil 1, Şekil 3 ve Şekil 5 ile incelenen tüm ortak olasılık yoğunluk fonksiyonlarının grafiklerine bakıldığında, farklı kuyruk bölgesi olasılıkları görülmüştür. Standart Gauss kopula fonksiyonunda kuyruk bağımlılığı yoktur. Düzenleme ile Gauss kopulanın bu eksikliği esnetilir ve farklı kuyruk bağımlılık yapıları elde edilebilir [30, 31]. Normal dağılım simetrik bir dağılım olduğundan, alt ve üst kuyruk bölgesindeki olasılık değerleri aynı ve simetriktir. $\rho = 0$ olduğunda, sözde-Gauss kopula fonksiyonu bağımsız sözde-Gauss kopula fonksiyonu $C(\hat{u}_1, \hat{u}_2) = \hat{u}_1\hat{u}_2$'ye dönüşür. Gauss kopula fonksiyonunda ilişki durumu bağımlılık ile doğrudan ilişkilendirilebildiğinden, değişkenler arasında ilişki olmadığı durumda diğer bir ifade ile bağımsızlık durumunda (ρ_{12}=0) grafikler, Şekil 6 ve Şekil 7 ile verilmiştir.



Şekil 6. Ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonunun izohips eğrileri (bağımsız sözde-Gauss kopula)



Şekil 7. Ortak olasılık dağılım ve ortak olasılık yoğunluk fonksiyonunun izohips eğrileri (bağımsız sözde-Gauss kopula)

Ortak olasılık yoğunluk fonksiyonunun perspektif grafiği, ilişki olmaması diğer bir ifade ile bağımsızlık durumunda dikdörtgensel bir şekil almıştır. Ortak olasılık yoğunluk fonksiyonunun izohips grafiğinden ise bağımsızlık durumu açıkça görülmektedir.

6. Sonuç ve öneriler

İstatistiksel ve finansal bağımlılık modellemesi çalışmalarında sıklıkla kullanılan kopula fonksiyonları zaman içinde aktüerya bilimlerinde de yoğun olarak kullanılmaya başlamıştır. Hayat sigortalarından sağlık sigortalarına, sağlık sigortalarından hayat dışı sigortalara kadar aktüerya bilimlerinin birçok alt branşında ve finansal aktüeryal verilerdeki bağımlılık çalışmalarında kopula fonksiyonlarından yararlanılmaktadır.

Özellikle eliptik sözde-kopula fonksiyonlarından sözde-Gauss kopula fonksiyonları eliptik dağılım özellikleri nedeniyle diğer kopula türlerine göre daha çok tercih edilmektedir. Avantajlarının yanında Gauss kopula fonksiyonlarının bağımlılık modellemesinde yetersiz kaldığı noktalarda vardır. Gauss kopula fonksiyonları, özellikle gerçek sigorta ve finans verilerinde karşılaşılma olasılığı yüksek olan kuyruk bağımlılığı, asimetrik bağımlılık, eliptik bağımlılık ve dairesel simetriye göre bağımlılık gibi bazı bağımlılık türlerini modellemede eksik kalmaktadır. Bu soruna çözüm üretmek amacıyla Gauss kopula fonksiyonunun parametresi olan ilişki katsayısı, belirli fonksiyonlarla düzenlenerek esnek bir bağımlılık modellemesi elde edilmiştir.

Fang ve Madsen [31] tarafından önerilen düzenleme yaklaşımı, Karadağ Erdemir [43] tarafından sözde-kopula regresyon modellerine dâhil edilmiştir. Bu kombinasyon sonucunda düzenlenmiş sözde-kopula regresyon modelleri tanımlanmış ve bağımlılık varsayımı altında hasar tutarı ile sayısının ortak dağılım ve ortak olasılık yoğunluk fonksiyonlarının kapalı formları elde edilmiştir.

Kopula fonksiyonları üzerinde düzenleme yapılması sonucunda esnek bir bağımlılık modellemesi elde edilmiştir. Düzenlemenin sağladığı etkinlik, bağımlılık varsayımı altında hasar tutarı ve sayısının ortak dağılım ve ortak olasılık yoğunluk fonksiyonlarının perspektif ve izohips eğrilerinin grafikleri ile incelenmiştir. Çalışılan verideki hasar tutarı ve sayısı arasındaki ilişkiye göre belirlenebilecek fonksiyonlar ve yine fonksiyonlarda kullanılacak kuyruğa yakınsama hızını ve kuyruğun şeklini kontrol eden parametreler yardımıyla kullanıcı tarafından kontrol edilebilen esnek bir bağımlılık modellemesi elde edilmiştir. Sözde fonksiyonların kullanılmasıyla birlikte daha dinamik ve farklı fonksiyonların kullanılabilmesiyle daha esnek bir bağımlılık modellemesi sunulmuştur.

Kaynaklar

[1] A. Sklar, 1959, Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229.
[2] E. Frees, E. Valdez, 1998, Understanding Relationships Using Copulas, *North Ameracan Actuarial Journal*, 2, 1.
[3] E. S. Sarıdaş, 2012, Bağımlı Yaşam Sürelerinin Modellenmesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
[4] Ö. Bakar, 2018, Bağımlı Çoklu Hayat Anüitelerinde Uzun Ömürlülük Riskinin Stokastik Analizi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
[5] L. Hua, 2015, Tail Negative Dependence and Its Applications for Aggregate Loss Modeling, *Insurance: Mathematics and Economics*, 61, 135.
[6] A. Şentürk Acar, 2016, Sağlık Sigortalarında Toplam Hasar Üzerinde Heterojenliğin Etkisi, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
[7] C. Czado, R. Kastenmeier, E.C. Brechmann, A. Min, 2012, A Mixed Copula Model for Insurance Claims and Claim Sizes, *Scandinavian Actuarial Journal*, 4, 278.
[8] R. Kastenmeier, 2008, Joint Regression Analysis of Insurance Claims and Claim Sizes, Diploma Thesis, Technische Universitat München, Mathematical Sciences.
[9] N. Krämer, E. C. Brechmann, D. Silvestrini, C. Czado, 2013, Total Loss Estimation Using Copula-Based Regression Models, *Insurance: Mathematics and Economics*, 53, 829.
[10] K. Wang, A. H. Lee, K. K. W. Yau, P. J. W. Carrivick, 2003, A Semisupervised Regression Model for Mixed Numerical and Categorical Variables, *Accident Analysis and Prevention*, 35 (4), 625.
[11] L. Madsen, Y. Fang, 2011, Joint Regression Analysis for Discrete Longitudinal Data, *Biometrics*, 67 (3), 1171.

[12] E. W. Frees, E. A. Valdez, 2008, Hierarchical Insurance Claims Modelling, *Journal of the American Statistical Association*, 5, 41.

[13] L. Bermúdez, D. Karlis, 2011, Bayesian multivariate Poisson models for insurance ratemaking, *Insurance: Mathematics and Economics*, 48 (2), 226.

[14] J. Ren, 2012, A Multivariate Aggregate Loss Model, *Insurance: Mathematics and Economics*, 51, 402.

[15] J. D. Cummins, L. J. Wiltbank, 1983, Estimating The Total Claims Distribution Using Multivariate Frequency and Severity Distributions, *Journal of Risk and Insurance*, 50 (3), 377.

[16] E. W. Frees, G. Myers, C. David, 2010, Dependent Multi-peril Ratemaking Models, *Astin Bulletin*, 40, 699.

[17] M. Ayuso, L. Bermúdez, M. Santolino, 2016, Copula-Based Regression Modeling of Bivariate Severity of Temporary Disability And Permanent Motor Injuries, *Accident Analysis & Prevention*, 89, 142.

[18] Y. Fang, 2014, A Bayesian Approach to Inference and Prediction for Spatially Correlated Count Data Based on Gaussian Copula Model, *International Journal of Applied Mathematics*, 44(3), 126.

[19] P. Shi, K. Shi, 2017, Territorial Risk Classification Using Spatially Dependent Frequency-Severity Models, *ASTIN Bulletin: The Journal of the IAA*, 47 (2), 437.

[20] G. Pettere, T. Kollo, 2006, Modelling Claim Size in Time via Copulas, in Transactions of 28th International Congress of Actuaries.

[21] P. Weke, C. Ratemo, 2013, Estimating IBNR Claims Reserves for General Insurance Using Archimedean Copulas, *Applied Mathematical Sciences*, 7 (25), 1223.

[22] E. Usta, 2016, Risk Premium Estimation in MTPL Insurance Using Copula: Turkey Case, Master of Science Thesis, The Graduate Scohool of Applied Mathematics of Middle East Technical University, Ankara.

[23] E. Usta, 2016, The Estimation of IBNR Reserve Using Copula, *Journal of Insurance Research*, 12 (10), 3.

[24] E. Kole, K. Koedijk, M. Verbeek, 2007, Selecting Copulas for Risk Management, *Journal of Banking & Finance*, 31 (8), 2405.

[25] B. Z. Karagül, 2013, Hayat Dışı Sigortalarda Doğrusal Olmayan Bağımlılığın Kopulalar ile Dinamik Finansal Analizi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Yüksek Lisans Tezi, Ankara.

[26] U. Karabey, 2015, Importance of Modelling the Dependence for Risk Capital Allocation, *Journal of Statisticians: Statistics and Actuarial Sciences*, 8 (1), 1.

[27] Z. M. Landsman, E. A. Valdez, 2003, Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7 (4), 55-71.

[28] D. Brigo, A. Pallavicini, R. Torresetti, 2010, *Credit Models and The Crisis: A Journey Into Cdos, Copulas, Correlations And Dynamic Models*, John Wiley & Sons.

[29] P. X.-K. Song, 2007, Correlated Data Analysis: Modeling, Analytics, And Applications. Springer Science & Business Media, Ontario, Canada.

[30] Y. Fang, 2012, Extensions to Gaussian Copula Models, Doctoral Dissertation, Oregon State University.

[31] Y. Fang, L. Madsen, 2013, Modified Gaussian Pseudo-Copula: Application in Insurance and Finance, *Insurance: Mathematics and Economics*, 53, 292.

[32] G. Masarotto, C. Varin, 2017, Gaussian Copula Regression in R, *Journal of Statistical Software*, 77 (8).

[33] A. J. Patton, 2001, Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula, UCSD Discussion Paper No. 01-09, SSRN.

[34] A. J. Patton, 2006, Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, 47 (2), 527.

[35] J. D. Fermanian, M. Wegkamp, 2004, Time Dependent Copulas. Preprint INSEE, Paris, France.

[36] J. D. Fermanian, M. Wegkamp, 2012, Time-Dependent Copulas. Journal of Multivariate Analysis, 110, 19.

[37] Y. Fang, L. Madsen, L. Liu, 2014, Comparison of Two Methods to Check Copula Fitting, *International Journal of Applied Mathematics*, 44 (1).

[38] M. A. Boateng, A. Y. Omari-Sasu, R. K. Avuglah, N. K. Frempong, 2017, On Two Random Variables and Archimedean Copulas, *International Journal of Statistics and Applications*, 7 (4), 228.

[39] R. B. Nelsen, 2006, *An Introduction to Copulas*, Springer Science & Business Media, Portland, Oregon, USA.

[40] R. A. Parsa, S. A. Klugman, 2011, Copula Regression, *Variance Advancing and Science of Risks*, 5, 45.

[41] U. Cherubini, S. Mulinacci, F. Gobbi, S. Romagnoli, 2011, *Dynamic Copula Methods in Finance*, John Wiley & Sons.

[42] Z. S. Ouyang, H. Liao, X. Q. Yang, 2009, Modeling Dependence Based on Mixture Copulas and Its Application in Risk Management, *Applied Mathematics-A Journal of Chinese Universities*, 24 (4) 393.

[43] Ö. G. Erdemir, 2020, Düzenlenmiş Sözde-Kopula Regresyon Modeli, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

[44] I. Kojadinovic, J. Yan, 2010, Modeling Multivariate Distributions with Continuous Margins Using The Copula R Package, *Journal of Statistical Software*, 34(9) 1.

[45] R. N. Mortensen, 2013, Pseudo-Observations in Survival Analysis, Master of Science Thesis, Aalborg University.
[46] Steorts, R. C., 2015, Visualizing the Multivariate Normal, Lecture 9, http://www2.stat.duke.edu/~rcs46/lectures_2015/02-multivar2/02-multivar2.pdf.

İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 13, 2020, 2, 78-100

Geliş/Received:05.12.2020, Kabul/Accepted: 26.12.2020

www.istatistikciler.org

Araştırma Makalesi / Research Article

A new combination method in mathematical theory of evidence "Analytic Fusion Process"

Murat Büyükyazıcı

Hacettepe Üniversitesi
Fen Fakültesi, Aktüerya Bilimleri Bölümü
06800 Beytepe, Ankara
muratby@hacettepe.edu.tr
orcid.org/*0000-0002-8622-4659*

Meral Sucu

Hacettepe Üniversitesi
Fen Fakültesi, Aktüerya Bilimleri Bölümü
06800 Beytepe, Ankara
msucu@hacettepe.edu.tr
orcid.org/*0000-0002-7991-1792*

Abstract

In this paper, we are interested in a consensus generator which combine two belief functions obtained from equally reliable and independent sources of information. The independence mentioned here is between occurrences of the sources of information. We propose a new consensus generator called "Analytic Fusion Process" which satisfy the idempotent and commutative law. Furthermore, this method also produces a measure of conflict shows whether the original beliefs were in harmony or in conflict. Another advantage is that the measure of conflict produced by this method reflects both qualitative and quantitative conflict.

Keywords: *Mathematical theory of evidence; Dempster-Shafer theory; Belief function; Data fusion; Consensus generator; Analytic fusion process; Integer programming.*

Özet

Kanıt kuramında yeni bir birleştirme yöntemi "Analitik Birleştirme Süreci"

Bu çalışmada, eşit derecede güvenilir ve bağımsız bilgi kaynaklarından elde edilen iki kanaat fonksiyonunu birleştiren bir uzlaşma oluşturucu ile ilgileniyoruz. Burada bahsedilen bağımsızlık, bilgi kaynaklarının oluşumları arasındadır. Eş kuvvetlilik ve değişme özelliklerini sağlayan "Analitik Birleştirme Süreci" adlı yeni bir uzlaşma oluşturucu öneriyoruz. Bu yöntem aynı zamanda orijinal kanaatlerin uyum içinde mi yoksa çelişki halinde mi olduğunu gösteren bir çelişki ölçüsü üretir. Diğer bir avantaj, bu yöntemle üretilen çelişki ölçüsünün hem nitel hem de nicel çelişkiyi yansıtmasıdır.

Keywords: *Kanıt kuramı, Dempter-Shafer kuramı, Kanaat fonksiyonu, Veri birleştirme, Uzlaşı üretici, Analitik birleştirme süreci, Tam sayılı programlama.*

1. Introduction

The mathematical theory of evidence (MTE), also known as Dempster-Shafer (DS) theory of evidence, has gained relatively wide acceptance as a reasonable tool for the representation, and combination such as revision, updating and conditioning of uncertain knowledge, evidence or information. The seminal work [1] on this subject is carried out by G. Shafer in 1976. This work was an expansion of A. Dempster's study

[2] about upper and lower probabilities dated 1967. In a finite discrete space, MTE can be interpreted as a generalization of probability theory where probabilities are assigned to sets as opposed to mutually exclusive singletons [3]. On the other hand, there are also some theoritical contributions which can be considered as an extension of the classical MTE but includes fundamental differences [4, 5, 6, 7].

Aggregation or fusion of information are basic concerns for all kinds of knowledge-based systems from image processing to decision making, from pattern recognition to machine learning [8]. The basic rule of combination to aggregate or fusion of information from distinct bodies of evidence in the framework of MTE is Dempster's rule of combination (DRC). However, after L. A. Zadeh presented an example for which this method gives counter-intuitive results especially in the case of existence of high conflict between bodies of evidence [9, 10], the success story of DS theory was abruptly slowed down [11]. In literature, also other examples claiming that DRC gives counter-intuitive results can be found [12, 13, 14, 15, 16]. Although R. Haenni and many others state that the counter-intuitive results are not a problem of DRC, but rather a problem of misunderstandings and misapplications [17, 11]; many alternative combination rules have been proposed in the framework of the MTE [18, 19, 20, 21, 13, 14, 22, 23, 24, 25, 26] for situations where DRC is not applicable, where its assumptions are not satisfied. D. Dubois and H. Prade have stated that many alternative rules can potentially occupy a continuum between conjunction (AND-based on set intersection) and disjunction (OR-based on set union) [27]. DRC is a *conjunctive operation,* because in the pairwise combination phase, the intersections of focal elements from distinct bodies of evidence are used. In this phase of some combination rules, the union of focal elements are used. These rules are called *disjunctive operation*. There is also a third type of combination rules called *tradeoff operation* in which task is carried out by using both intersection and union operators [27].

There are some alternative combination rules that have different assumptions from MTE on dependency of the sources of information or on reliability of the sources. R. Haenni and S. Hartmann's paper approaches the problem of independent and partially reliable information source from a very general perspective by using the *theory of probabilistic argumentation* [6] as modeling and DS theory of evidence as the underlying mathematical mechanism [28]. If the independency of the sources of information assumption is questionable, M. E. G. V. Cattaneo suggests using the least specific combination minimizing the conflict among the ones allowed by a simple generalization of Dempster's rule [29], does not propose a new method. In MTE, if the requirement of independence between sources of information is not satisfied, it is prevented a direct application of Dempster's rule. So, in [30] P. A. Monney and M. Chan proposed a method for dealing with this situation. The method relies on the ability of the *theory of hints* [5] to explicitly represent variables for which probabilistic information is avalibale. The fundamental limitation of DRC lies in the assumption that the belief functions combined be based on distinct bodies of evidence. So, in the case of nondistinct bodies of evidence, T. Denœux proposed two new commutative, associative and idempotent operators for belief functions; the *cautious conjunctive rule*, and the *bold disjunctive rule* [31]. These operators rely on the *transferable belief model* [4]. Beside the distinctness assumption, the choice of one operator between two depends on assumptions regarding reliability of the sources. If all are reliable then the cautious conjunctive rule, or if at least one is reliable then the bold disjunctive rule was recommended. In the recently paper, K. Yamada proposed a new model of combination and a new rule of combination called *combination by compromise* as a consensus generator [32]. This combination model for consensus generation is based on the assumptions that the information sources are independent in the sense of occurrance but may collide with each other over their contents, and that the information sources may not be totally reliable.

Experts and researchers in the fields today have no agreement on the superiority on one combination method over the others [33, 17, 34, 15]. To prevent this confusion about the belief combination, it should be clear what it means to combine belief functions; conditioning, revisioning, updating, consensus generating or anything else. Then, it should be stated expressly what kind of dependence or independence assumed. There are many concepts of dependence. In [30], P. A. Monney and M. Chan discriminate between dependence at the sources level, namely at the level of the assumption variables and dependence at the Θ level, namely *distinctness* [15] of bodies of evidence. In two papers [35, 36], B. B. Yaghlane et al. introduce the concept of *doxastic independency*. This concept of independence is once again represented at the Θ level [30]. K. Yamada asserts that the problem comes from confusion of independence between occurences of bodies of evidence with *consistency between contents of the information* [32]. It also should be stated expressly how reliable the sources of information are; partially or totally? Are the reliabilities of the sources of information equally or not?

A set of experts or other sources of information can provide more information than a single expert. Although it is sometimes reasonable to provide a decision maker with only the individual experts' opinions seperately, it is often necessary to combine the opinions into a single one [37]. In many cases, a single belief function is needed for input into a decision model. Even if this is not the case, it can also be informative and effective to generate a *combined belief function* as a summary of the available information. So, a *consensus generator* in the framework of the MTE is needed. A consensus generator is expected to be idempotent and commutative, but associativity is dispensable such as in the case of simple aritmethic mean operator. Also, a *measure of conflict* is useful to show whether the original beliefs were in harmony or in conflict. A consensus generator and measure of conflict are more informative if they are given together as in the case mean and variance. Recently, K. Yamada's rule of combination called combination by compromise as a consensus generator is commutative; however, it does not satisfy the idempotent law nor the associative law [32]. The *weighted average operator* [38] also can be tought as a consensus generator. The weighted average operator also does not satisfy the idempotent law nor the associative law. Furthermore, neither of both does not generate any measure of conflict.

In MTE, there are some rules which also produce a measure of conflict. However, there are no mechanisms to measure the degree of conflict other than using the mass of the combined belief assigned to the emptyset before normalization [39]. Is the mass of the combined belief to the emptyset before normalization real conflict? Under this consideration W. Liu has proposed a method to measure the conflict among beliefs using a pair of values, and then investigated the effect of these measures on deciding when Dempster's rule can be applied [39]. Still, there is not any real single measure of conflict which measures how the original beliefs were in conflict.

In this paper, we are interested in a consensus generator which combines two belief functions obtained from equally reliable and independent sources of information. The independence mentioned here is between occurrences of the sources of information. We propose a new consensus generator called *Analytic Fusion Process* (AFP) which satisfies the idempotent and commutative law. Furthermore, this method also produces a measure of conflict which shows how the original beliefs were in conflict. Another advantage of the AFP is that the measure of conflict produced by AFP reflects both qualitative and quantitative conflict. Section 2 presents a background on MTE and DRC. Some definitions and principles of the AFP are given in Section 3, and Section 4 introduces the three stages of the AFP: definition, matching, and combination stages. Section 5 gives two examples, one of them is for a general case example and the other for a special case example. A comparison among the combination by compromise, the weighted average operator and the AFP is provided in Section 6, and the conclusion is provided at Section 7.

2. Mathematical Theory of Evidence

In MTE, knowledge is represented by *basic belief assignment (bba)* or *belief function*. These two functions have one-to-one correspondence [1]. The set called *sample space* in the traditional probability theory is called *frame of discernment* in the MTE. The frame of discernment is a finite set of mutually exclusive elements, denoted by Θ hereafter. The set of all subsets of Θ is called the power set of Θ and is denoted by 2^{Θ}. The *bba* is defined as follows:

Definition 1. Let's assume Θ is a frame of discernment, then a function $m : 2^{\Theta} \to [0,1]$ is called a *bba* whenever it satisfies the following conditions:

(1) $m(\varnothing) = 0$

(2) $\sum_{A \subseteq \Theta} m(A) = 1$.

The quantity of $m(A)$ is called *basic belief mass* (*bbm*), and it is meant to be the measure of the belief that is committed exactly to *A* (exactly in A in nothing smaller). Each subset $A \subseteq \Theta$ with $m(A) > 0$ is called a *focal element* of m and their set is represented by ε. In this case *body of evidence* is represented by $E(\varepsilon, m)$. If the true state of interest is denoted by θ and the set of its possible values by Θ, then the *propositions* of interest are precisely those of the form "the true value of θ is exactly in A" where A is a subset of Θ [1]. In this paper we will use the word "proposition" to refer not only to set A but also to its *bbm* $m(A)$. In this case, if we denote the true state (or value) of interest by θ and the set of its possible values by Θ, any focal element A of m and its *bbm* $m(A)$ can also thought as a proposition which asserts that "the true state of θ is exactly in A with a quantity of belief $m(A)$". Thus, a proposition is composed of two parts: a focal element, which we call the *qualitative part* of the proposition because it gives the content of the proposition, and a *bbm*, which we call the *quantitative part* of the proposition because it gives the quantity of the proposition.

The belief function measures how much the information from a body of evidence supports the belief in a set specified elements as the right answer [18]. A definition of the belief function which can be equivalently represented by m is given as follows:

Definition 2. A function $Bel : 2^{\Theta} \to [0,1]$ is called a belief function over Θ if it is given for some *bba* $m : 2^{\Theta} \to [0,1]$ as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B).$$

The quantity $Bel(A)$ is called *A*'s *degree of belief*, and it is meant to be the measure of the total belief that is committed to *A*. If all the focal elements are singletons than this is the *Bayesian belief function* [1].

Let Bel_1 and Bel_2 be two belief functions from two distinct bodies of evidence $E_1(\varepsilon_1, m_1)$ and $E_2(\varepsilon_2, m_2)$ over the same frame of discernment Θ. If Bel_1's focal elements are denoted by $\varepsilon_1 = \{A_1, ..., A_n\}$ and Bel_2's focal elements by $\varepsilon_2 = \{B_1, ..., B_m\}$, then the combination result (denoted here by index *D*) is given by the following equation.

$$m_D(X) = \frac{\sum\limits_{\substack{i,j \\ A_i \cap B_j = X}} m_1(A_i) \times m_2(B_j)}{1 - \sum\limits_{\substack{i,j \\ A_i \cap B_j = \varnothing}} m_1(A_i) \times m_2(B_j)} \quad , \quad X \neq \varnothing \text{ and } X \subseteq \Theta$$

$m_D(X)$ is a proper *bba*, for all non-empty $X \subset \Theta$, if and only if the denominator in above equation is non-zero. The *degree of conflict* between the bodies of evidence $E_1(\varepsilon_1, m_1)$ and $E_2(\varepsilon_2, m_2)$ is defined by

$$\kappa_D = \sum_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) \times m_2(B_j).$$

Shafer explained in [1] at pg. 66 what the combination means if it is obtained by DRC as follows:

> "*Dempster's rule of combination permits a simple description of how the assimilation of new evidence should change our beliefs: if our initial beliefs are expressed by a belief function* Bel_1 *over* Θ*, and the new evidence alone determines a belief function* Bel_2 *over* Θ*, then after assimilating the new evidence we should have the beliefs given by* $Bel_1 \oplus Bel_2$*. This description avoids the doctrine that a body of evidence can always be cast in the form of a single proposition known with certainty.*"

Shafer revealed that if the effect of the new evidence on the frame of discernment Θ is to establish any particular subset with certainty then the result of DRC is similar to the result of Bayes' rule of conditioning. When the new evidence occurs in the form of a certainty, he called this special case of DRC as *Dempster's rule of conditioning* [1]. So it is obvious to say that DRC is a conditioning operation which generalize Bayes' rule of conditioning to the case where it is not necessary for the new evidence to occur in the form of certainty.

3. Foundations of the Analytic Fusion Process

Mathematical aggregation methods range from simple summary measures such as arithmetic or geometric means of probabilities to procedures based on axiomatic approaches or on various models of the information aggregation process requiring inputs regarding characteristics such as the quality of and dependence among the experts' probabilities [37]. C. Genest and J. V. Zidek declared that the "logarithmic opinion pool" involves many advantages over the "linear opinion pool" when finding the consensus distribution of the subjective probability distributions. For details, you can see [40]. If all the weights are equal, as in the case of equally reliable sources of information, the consensus distribution is proportional to the geometric mean of the individual distributions. So, as a consensus generation process, AFP is a geometric mean based analytical method that operates on the individual belief functions to produce a single combined belief function when the individual belief functions comes from equally reliable and occurrence independent sources of information.

It is now appropriate to give some definitions of AFP before it is explained in detail. In any context in this paper even it is not denoted, belief functions which will be combined are from equally reliable and occurrence independent sources of information.

3.1. Pairwise combination

Let $E_1(e_1, m_1)$ and $E_2(e_2, m_2)$ be two occurrence independent bodies of evidence defined over the same frame of discernment Θ. In this way, a belief function Bel_1 with a focal element set e_1 and *bba* m_1 and another belief function Bel_2 with a focal element set e_2 and *bba* m_2 are given. If the number of focal elements of the Bel_1 is taken as n, each focal element of this function is shown by $e_{1i},\ i = 1,...,n$. Similarly, if the number of focal elements of the Bel_2 is taken as m, each focal element of this function is shown by $e_{2j},\ j = 1,...,m$. In other words, there is n propositions for Bel_1 and m for Bel_2. So, it is obvious that reciprocal propositions of two *bba*s defined above can produce $n \times m$ pairs. In a consensus generation process, since we are looking the answer of how can we generate a new *compromised proposition* which embody the joint effect of the two propositions, one needs a kind of summarizing process for each one of the $n \times m$ pairs. The summarizing process of a pair of propositions is called *pairwise combination*. A pairwise combination process is expected to generate a compromised proposition which has its own qualitative parts and quantitative parts. In Section 3.1.1 generating qualitative parts and in Section 3.1.3 calculating quantitative parts of the compromised proposition will be given.

3.1.1. Compromised focal element

We called the qualitative parts of compromised proposition as *compromised focal element (cfe)*. According to the concept of reliability of the information sources, three ideas might be possible to generate *cfe*s if information sources are equally reliable. If information sources are completely reliable $e_{1i} \cap e_{2j}$ should be chosen as in DRC [1], if at least an unknown one of the sources reliable $e_{1i} \cup e_{2j}$ should be chosen as in Dubois and Prade's combination rule [20]. K. Yamada proposes a third approach, combination by compromise, as a natural consensus [32]. In this paper we use this third approach to generate a *cfe.* According to this approach *cfe* composed of three subsets of $e_{1i} \cup e_{2j}$: $e_{1i} \cap e_{2j}$, $e_{i\backslash j}^{1\backslash 2}$, and $e_{i\backslash j}^{2\backslash 1}$. It is obvious to see that the first subset $e_{1i} \cap e_{2j}$ is just an intersection set. The second subset $e_{i\backslash j}^{1\backslash 2}$ is the set of all elements which are members of e_{1i}, but not members of e_{2j}. Similarly, the third subset $e_{i\backslash j}^{2\backslash 1}$ is the set of all elements which are members of e_{2j}, but not members of e_{1i}. The reason why *1\2* is used as a superscript is that so as to exterminate the confusion which may occur under the condition that *i* and *j* focal element numbers given in subscript are the same. In AFP additionally, we called the $e_{1i} \cap e_{2j}$ as *agreement set*, $e_{i\backslash j}^{1\backslash 2}$ as *conflict set one,* and $e_{i\backslash j}^{2\backslash 1}$ as *conflict set two*. Consequently, the *cfe* can be represented as follows:

$$(e_{i\backslash j}^{1\backslash 2})((e_{1i} \cap e_{2j}))(e_{i\backslash j}^{2\backslash 1}) . \qquad (1)$$

3.1.2. Region sharing of compromised focal element

The focal elements, namely qualitative parts of two propositions, e_{1i}, e_{2j} and the *bbm*s of the focal elements, namely quantitative parts of the same propositions, $m_1(e_{1i})$, $m_2(e_{2j})$ are given. A pairwise combination of two propositions includes two issues: how to make *qualitative summarizing*, and how to make *quantitative summarizing*. Producing the *cfe* releases the first issue. The second issue can be rewritten as how to share the region of $m_1(e_{1i}) \times m_2(e_{2j})$, namely the size of *pairwise combination region,* among subsets $e_{1i} \cap e_{2j}$, $e_{i\backslash j}^{1\backslash 2}$, and $e_{i\backslash j}^{2\backslash 1}$. The justification of using the word "region" is simple: When the *bbm*s of m_1 and m_2 are depicted graphically on line segments on y-axis and x-axis respectively with both sizes one, the size of $m_1(e_{1i}) \times m_2(e_{2j})$ of any pairwise combination is equal to a region in the total area equal to one (See Fig. 2 in Section 3.2). In each pairwise combination region, we called the parts as *agreement region* belonging to agreement set elements $e_{1i} \cap e_{2j}$ and the remaining parts as *conflict region one* and *conflict region two* belonging to conflict set elements $e_{i\backslash j}^{1\backslash 2}$ and $e_{i\backslash j}^{2\backslash 1}$, respectively. The region sharing can be made proportionally with respect to both their *bba*s assigned by the information sources and the cardinalities of the subsets. Thus, if the size of agreement region is given as r_{ij}^{a} whereas the size of conflict regions is given as r_{ij}^{b}, and r_{ij}^{d} respectively, the pairwise combination region sharing can be made as in Fig. 1.

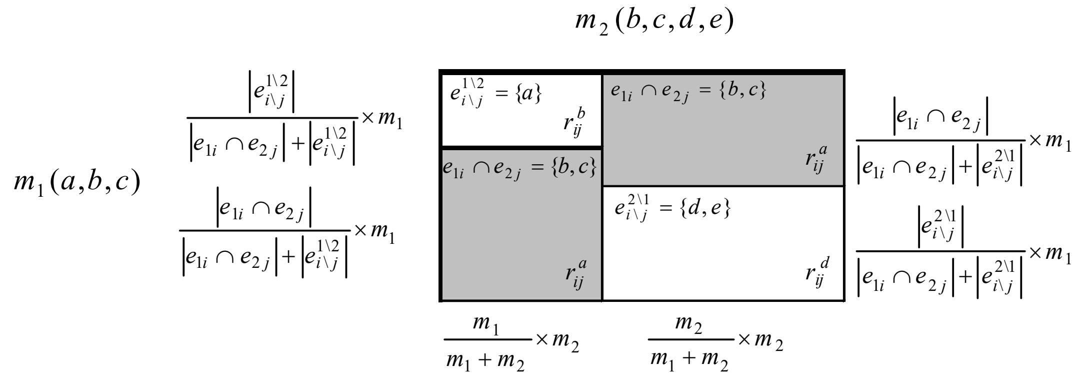


Fig. 1. Illustration of region sharing.

In Fig. 1, four small rectangles represent how a pairwise combination region can be allocated to the subsets of *cfe* proportionally with respect to both their *bba*s and the cardinalities of the subsets. Firstly, In Fig.1, the vertical line shares the biggest rectangle two parts proportionally with respect to their *bba*s. Secondly, the two horizontal lines share the two separated parts proportionally with respect to their subset's cardinalities. So, the region sharing can be made by the following equations:

$$r_{ij}^a=\left(\frac{m_1(e_{1i})}{m_1(e_{1i})+m_2(e_{2j})}\times\frac{\left|e_{1i}\cap e_{2j}\right|}{\left|e_{1i}\cap e_{2j}\right|+\left|e_{i\setminus j}^{1\setminus 2}\right|}+\frac{m_2(e_{1i})}{m_1(e_{1i})+m_2(e_{2j})}\times\frac{\left|e_{1i}\cap e_{2j}\right|}{\left|e_{1i}\cap e_{2j}\right|+\left|e_{i\setminus j}^{2\setminus 1}\right|}\right)\times m_1(e_{1i})\times m_2(e_{2j}),$$

$$r_{ij}^b=\frac{m_1(e_{1i})}{m_1(e_{1i})+m_2(e_{2j})}\times\frac{\left|e_{i\setminus j}^{1\setminus 2}\right|}{\left|e_{1i}\cap e_{2j}\right|+\left|e_{i\setminus j}^{1\setminus 2}\right|}\times m_1(e_{1i})\times m_2(e_{2j}),$$

$$r_{ij}^d=\frac{m_1(e_{1i})}{m_1(e_{1i})+m_2(e_{2j})}\times\frac{\left|e_{i\setminus j}^{2\setminus 1}\right|}{\left|e_{1i}\cap e_{2j}\right|+\left|e_{i\setminus j}^{2\setminus 1}\right|}\times m_1(e_{1i})\times m_2(e_{2j}).$$

3.1.3. Mass sharing of compromised focal element

It is important in the case of consensus generation to emphasize that the geometric mean of *bbm*s should be used for quantitative summarizing not just the area of the pairwise combination region as in the case of conditioning with DRC. In conditioning case, multiplying of two *bbm*s is reasonable, but in the case of consensus generation averaging is expected. So, after sharing out of the region $m_1(e_{1i})\times m_2(e_{2j})$ between the agreement region, the conflict region one and the conflict region two, it is need to take square roots of distributed region as in the case of two probabilities' geometric mean. We called the parts as *mass of agreement* belonging to agreement set elements $e_{1i}\cap e_{2j}$, *mass of conflict one* belonging to conflict set $e_{i\setminus j}^{1\setminus 2}$, and *mass of conflict two* belonging to conflict set $e_{i\setminus j}^{2\setminus 1}$. If the mass of agreement is given by $a_{ij}(e_{1i}\cap e_{2j})$ or shortly a_{ij}, the mass of conflict one and two are given by $b_{ij}(e_{i\setminus j}^{1\setminus 2})$ or shortly b_{ij}, and $d_{ij}(e_{i\setminus j}^{2\setminus 1})$ or shortly d_{ij} respectively, the mass sharing can be made by the following equations:

$$a_{ij}=\sqrt{r_{ij}^a}\ , \tag{2}$$

$$b_{ij} = \sqrt{r_{ij}^{b}}, \quad (3)$$

$$d_{ij} = \sqrt{r_{ij}^{d}}, \quad (4)$$

Sum of the mass of conflict one and the mass of conflict two, that is $b_{ij} + d_{ij}$, is called *mass of conflict.*

3.1.4. Qualitative and quantitative conflicts

In the phase of pairwise combination, the elements which are involved in one of the focal elements which will be combined but not involved in the other one cause a *qualitative conflict*. As a measure of qualitative conflict, we use the proportion of the mass of conflict in sum of the mass of agreement and the mass of conflict for each pairwise combination. Since this is a proportion, it takes value on $[0, 1]$. So, if the size of qualitative conflict is given by c_{ij}, it can be formulated as follows:

$$c_{ij} = (b_{ij} + d_{ij})/(a_{ij} + b_{ij} + d_{ij}) \quad (5)$$

Whether there is a qualitative conflict or not, there can be a quantitative conflict. *Quantitative conflict* is a conflict which stems from inequality of $m_1(e_{1i})$ and $m_2(e_{2j})$ *bbms*. It has been stated that mass of agreement and mass of conflicts are geometric means of parts of $m_1(e_{1i})$ and $m_2(e_{2j})$ *bbms* which belong to agreement set elements and conflict sets elements. As a measure of quantitative conflict we use the measure of deviation from geometric mean g_{ij} of $m_1(e_{1i})$ and $m_2(e_{2j})$ *bbms*, namely geometric variance v_{ij}. When $n = 2$, geometric mean and geometric variance values can be obtained as in Eq. (6) and Eq. (7).

$$g_{ij} = \sqrt{m_{1i}(e_{1i}) \times m_{2j}(e_{2j})} \quad (6)$$

$$v_{ij} = 1 - \frac{\max\left\{\frac{m_{1i}(e_{1i})}{g_{ij}}, \frac{g_{ij}}{m_{1i}(e_{1i})}\right\}}{\frac{\max\{m_{1i}(e_{1i}), m_{2j}(e_{2j})\}}{\min\{m_{1i}(e_{1i}), m_{2j}(e_{2j})\}}} \quad (7)$$

In Eq. (7) geometric variance is developed for AFP as being defined on $[0, 1]$. It represents the size of quantitative conflict. This value is closer to 0 in situations where the difference between *bbms* is low and it is closer to 1 in situations where the difference between *bbms* is high.

3.2. Matching the most suitable propositions

In this section, to determine the matching the most suitable propositions principle of the AFP, we will start making analogy between the simple case of averaging two Bayesian belief functions and the general case of averaging two belief functions.

Example 1. Given two Bayesian belief functions, Bel_1 and Bel_2, over the same frame of discernment $\Theta = \{a, b, c\}$ from two bodies of evidence $E_1(\varepsilon_1, m_1)$ and $E_2(\varepsilon_2, m_2)$ as follows:

$$E_1(\varepsilon_1, m_1) = \{ m_1(a) = 0.1 ; \ m_1(b) = 0.2 ; \ m_1(c) = 0.7 \}$$

$$E_2(\varepsilon_2, m_2) = \{ m_2(a) = 0.2 \,;\; m_2(b) = 0.3 \,;\; m_2(c) = 0.5 \}.$$

In this simple case, to obtain a single combined belief function from given these two Bayesian belief functions, we simply match the propositions which have the same focal elements reciprocally and average the *bbm*s of matched propositions. Since we use the geometric mean as an averaging operator, it is needed normalization. Normalization is carried out by calculating the portions of these geometric means in their sum. The combined belief function obtained with matching, averaging and normalization process is as follows: $m_{geo}(a) = 0.1446$, $m_{geo}(b) = 0.2505$, $m_{geo}(c) = 0.6049$.

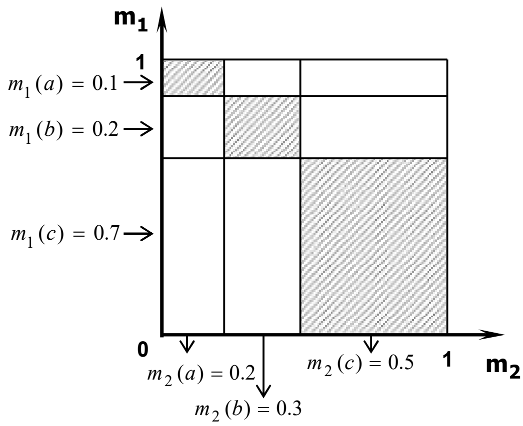


Fig. 2. Illustration of pairwise combinations of propositions.

The matching situations can be seen graphically in Fig. 2. The propositions $m_1(a) = 0.1$, $m_1(b) = 0.2$, and $m_1(c) = 0.7$ of Bel_1 were matched and pairwise combined with the propositions $m_2(a) = 0.2$, $m_2(b) = 0.3$, and $m_2(c) = 0.5$ of Bel_2, respectively. In the figure, when the shaded regions correspond to *matched pairwise combinations*, the clear regions correspond to the unmatched pairwise combinations.

It is important in the case of consensus generation to emphasize that the clear regions in this example neither represent the conflict mass nor have any other sense. So, the right way of matching a proposition from a body of evidence would be to match it with only one proposition in the other body of evidence. In other words, proposition $m_1(a) = 0.1$ of Bel_1 has to be matched with most suitable one $m_2(a) = 0.2$ of Bel_2. The propositions $m_1(a) = 0.1$ and $m_2(a) = 0.2$ must not be pairwise combined with others. Therefore, in Fig. 2, only one matching must be made in each row and column. So, a proposition from a body of evidence must be matched and pairwise combined with the most suitable proposition from the other body of evidence. The size of qualitative conflict, and the size of quantitative conflict lead determination of the most agreeable matching among propositions. This will be explained in Section 4.2.

4. Stages of the Analytic Fusion Process

AFP is composed of three stages. These three stages are called *definition*, *matching*, and *combination* stages respectively. In the definition stage, a general combination table which lays outs the outputs produced by pairwise combinations the propositions of the *bba*s which will be combined. In the matching stage, the propositions are matched so as to minimize the conflict by an integer programming problem for conflict minimization. In the combination stage, the combined *bba* is obtained based on the matches of the previous stage.

4.1. Definition stage

A *bba* having n focal elements and a *bba* having *m* focal elements are to be combined. In the definition stage of the process, the pairwise combinations are performed $n \times m$ times; then *cfe*, mass of agreement a_{ij}, mass of conflicts b_{ij} and d_{ij}, the size of qualitative conflict c_{ij}, and the size of quantitative conflict v_{ij} are obtained in the result of each combination as explained in Section 3.1. So, all results are achieved for each of $n \times m$ combinations. The table in which the results obtained through $n \times m$ numbered pairwise combinations of propositions which belong to two *bba*s is called *general combination table* (See Table 1 in Section 5.1). In the i th row and the j th column of this table, pairwise combinations results of the the i th proposition of the first *bba* and the j th proposition of the second *bba* are given.

4.2. Matching stage

A *linear programming* [41] problem is an optimization problem for which we attempt to maximize or minimize a linear function of the decision variables. The function that is to be maximized or minimized is called the *objective function*. The values of the decision variables must satisfy a set of *constraints*. Each constraint must be a linear equation or linear inequality. In any linear programming model, the *decision variables* should completely describe the decisions to be made [41]. As a special case of the linear programming problem, if all decision variables are restricted to be integers the problem is called integer programming, if all decision variables are restricted to 0 or 1 values the problem is called *0-1 integer programming* [41, 42].

In Section 3.2 with Example 1 we explained why it is needed to match the most suitable propositions of the bbas. So, the propositions of two *bba*s which are closest to each other in the sense of quality and quantity aspects should be matched. The combined *bba* should be calculated based on the results of these matching. In matching stage, we propose to make the matching with the assistance of a 0-1 integer programming problem, shortly we said *matching problem*. Decision variables, constraints and objective function of matching problem will be given in Section 4.2.1, Section 4.2.2, and Section 4.2.3, respectively. In Section 4.2.4 the matching problem will be given as a whole.

4.2.1. Decision variables

Taking the general combination table with *n* row and *m* column into consideration, the matching of the propositions is made by valuing the matching decision variable of the cells as *1* where the matching is agreed, and valuing the matching decision variable of the cells as *0* where the matching is not agreed. Thus, x_{ij} the *matching decision variable* is defined as:

$$x_{ij} = \begin{cases} 1, & \textit{if the } i\,\text{th}\,\textit{focal element of the first bba is matched} \\ & \textit{with the } j\,\text{th}\,\textit{focal element of the second bba} \\ 0, & \textit{otherwise} \end{cases} \tag{8}$$

After the definition of the decision variables about the matching of the propositions, the constraints and the objective function should be determined using the decision variables.

4.2.2. Constraints

In the case that the numbers of propositions of the two *bba*s to be combined, *n* and *m,* are equal, each proposition *bba* should have been matched with a proposition of the other *bba* or vice versa. However, in the case that the numbers of propositions, *n* and *m* are not equal, it is natural that some of the propositions will be out of matching. When a generalization is made by considering this situation, the sum of matching decision variables in each row and also each column of the general combination table can be at most *1*. So, the constraints related to each row cells are as follows,

$$\sum_{j=1}^{m} x_{ij} \leq 1 \qquad i = 1,...,n. \tag{9}$$

The constraints related to each column cells are as follows,

$$\sum_{i=1}^{n} x_{ij} \leq 1 \qquad j = 1,...,m. \tag{10}$$

The sum of matching decision variables in each row and each column does not exceed *1* and this has been acquired with the constraints given in Eq. (9) and Eq. (10).

However, in such a case, all the matching decision variables would be *0* in a conflict minimization problem. By adding the constraint that the total of all the matching decision variables will be $\min\{n, m\}$, the smallest value of the numbers of row or column is matched. To this end, the constraint

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} = \min\{n, m\} \tag{11}$$

is added to the problem. Lastly, the constraints showing that the matching decision variables can be only 0 or 1,

$$x_{ij} = \{0,1\} \qquad i = 1,...,n, \; j = 1,...,m \tag{12}$$

are added to the problem. These constraints have the generalized forms for the situations that the numbers of propositions, *n* and *m,* are equal or not.

4.2.3. Objective function

As the matching decision variables and the related constraints have been determined, the fact that under which circumstances the matching decision variables will take the value 0 or 1 based on constraints Eq. (9), Eq. (10), Eq. (11) and Eq. (12) should be determined. Following the pairwise combination phase a_{ij} mass of agreement, c_{ij} size of qualitative conflict and v_{ij} size of quantitative conflict for each of the pairwise combinations have been calculated and all of these results have been shown in the general combination table.

Taking only the size of qualitative conflict or the size of quantitative conflict in the objective function into account would be not enough. So, the objective function should be a form of minimizing a function of both. Such an objective function can be given as shown in Eq. (13) in order to minimize the total of the *generalized means* [37] of the size of qualitative conflict c_{ij}, and the size of quantitative conflict v_{ij}.

$$\min \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\left(\frac{c_{ij}^2 + v_{ij}^2}{2}\right)^{1/2} \times x_{ij}\right) \tag{13}$$

Either the arithmetic or the geometric mean of the c_{ij} and v_{ij} could have been used in Eq. (13). However, the generalized mean has been preferred here as it always yields in a bigger value than the arithmetic and geometric mean. So, it has been made harder to make matching between propositions have bigger qualitative and/or quantitative conflict.

4.2.4. Matching problem

The problem as how to match the $\min\{n,m\}$ pair from $n \times m$ probable pairwise combinations under constraints given in Eq. (9), Eq. (10), Eq. (11) and Eq. (12) with objective function given in Eq. (13) can be overcome through optimal solution of 0-1 integer programming problem. Thus; matching problem of the AFP given as follows in Eq. (14):

$$
\begin{aligned}
&\min \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{c_{ij}^2+v_{ij}^2}{2}\right)^{1/2} \times x_{ij} \\
&\qquad \sum_{j=1}^{m} x_{ij} \le 1, \qquad i=1,\dots,n \\
&\qquad \sum_{i=1}^{n} x_{ij} \le 1, \qquad j=1,\dots,m \\
&\qquad \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} = \min\{n,m\} \\
&\qquad x_{ij} = 0,1 \quad i=1,\dots,n,\ j=1,\dots,m .
\end{aligned} \tag{14}
$$

The $n \times m$ cells of the general combination table had $n \times m$ matching decision variables x_{ij}, in a matching problem. In linear programming, a *feasible solution* is a choice of values for decision variables that satisfies all constraints. For a minimization problem as in the case matching problem, the *optimal solution* is a feasible solution that gives the smallest objective function value. If an optimization problem has more than one optimal solution, it is said that it has *alternative optimal solutions* [41]. In an optimal solution of a model of Eq. (14) created in matching stage; $\min\{n,m\}$ of them get *1* and the rest get 0. As stated in Section 4.2.1, matching decision variables which have value *1* in optimal solution determine the matched pairwise combinations. The optimal solution method of Eq. (14) is not subject of this paper. Detailed information about solving this mathematical model can be found at [41, 42]. There are also many computer programs solving this kind of optimization problem.

4.3. Combination stage

In the calculation of *combined bba*; the matched pairwise combinations will be used as base and if necessary, the other non-base pairwise combinations will be considered with a lower weight than the weight of the matched pairwise combinations. A pairwise combination of matched propositions is called *base of pairwise combination*. The cell corresponding to this pairwise combination in the general combination table will be called *base cell of pairwise combination*. In other words, the cells corresponding to the matching decision variables which get *1* in the general combination table are called base cells of pairwise combination. The weights of the cells of pairwise combination on the general combination result are shown with w_{ij} at the first row of every cell of pairwise combination in the general combination table. The weight of base cell of pairwise combination will be shown with w_{ij}^b, the size of qualitative conflict with c_{ij}^b and the size of quantitative conflict with v_{ij}^b. The general combination table in which the weights are also shown is called *weighted general combination table*. Table 2 in Section 5.1 gives a weighted general combination table for Example 2.

The weights of the base cells of pairwise combination on the general combination result are calculated as shown in Eq. (15).

$$w_{ij}^{b}=1-\left(\frac{(c_{ij}^{b})^{2}+(v_{ij}^{b})^{2}}{2}\right)^{1/2} \tag{15}$$

Then, firstly the one with the biggest weight of the base cells of pairwise combination is considered. The weight of the non-base cells of pairwise combination that is positioned in the same row or column with the base cell of pairwise combination with the biggest weight is calculated as shown in Eq. (16).

$$w_{ij}=1-w_{ij}^{b} \tag{16}$$

If there is neither qualitative nor quantitative conflict ($c_{ij}^{b}=0$, $v_{ij}^{b}=0$) between two *bba*s of the base cell of pairwise combination; Eq. (15) and Eq. (16) provide that the combination weight of the base cell of pairwise combination is $w_{ij}^{b}=1$. Thus, the combination weight of the other cells of pairwise combination that are in the same row or column as this base cell of pairwise combination are zero as in the case of Example 1. If there is any measure of qualitative or quantitative conflict, Eq. (15) and Eq. (16) provide that the weight of the base cell of pairwise combination on the combination is measured by subtracting the generalized mean of this qualitative or quantitative measure from *1*. In such a case, the combination weight of other pairwise combinations that are in the same row or column with this base cell of pairwise combination equals to the generalized mean of conflicts in the base cell of pairwise combination.

A non-base cell of pairwise combination can be in the same row with a base cell of pairwise combination and in the same column with another base cell of pairwise combination. If this is the case, how can the combination weight for a non-base cell of pairwise combination be achieved? Which one of the base cells of pairwise combination is affecting the non-base cell of pairwise combination? It will be wise firstly to calculate combination weights of the non-base cells of pairwise combination, which are in the same row or column with the base cell of pairwise combination with the biggest combination weight of all base cells of pairwise combination. Once the weight of a non-base cell of pairwise combination is determined, it will not change then.

In the general combination table, after the weights for all the cells of pairwise combination are calculated the general combination results are calculated as in Eq. (17),

$$m_{AFP}(X)=\frac{\sum\limits_{\substack{i,j\\ e_{i\cap j}=X}}\sum w_{ij}\times a_{ij}(X)+\sum\limits_{\substack{i,j\\ e_{i\setminus j}^{1\setminus 2}=X}}\sum w_{ij}\times b_{ij}(X)+\sum\limits_{\substack{i,j\\ e_{i\setminus j}^{2\setminus 1}=X}}\sum w_{ij}\times d_{ij}(X)}{\sum\limits_{\substack{i,j\\ A\in 2^{\Theta}}}\sum w_{ij}\times a_{ij}(A)+\sum\limits_{\substack{i,j\\ A\in 2^{\Theta}}}\sum w_{ij}\times b_{ij}(A)+\sum\limits_{\substack{i,j\\ A\in 2^{\Theta}}}\sum w_{ij}\times d_{ij}(A)},\quad X\neq\varnothing \tag{17}$$

and then the *combined bba* and so the *combined belief function* is achieved. Eq. (17) shows the combined *bba* $m_{AFP}(X)$, which is given for any of the subsets $X\in 2^{\Theta}$ products of agreement or conflict sets achieved through pairwise combination phases. It can be seen in the numerator of Eq. (17) that the mass of agreement or conflict (a_{ij}, b_{ij}, d_{ij}) committed to the sets with *X* agreement or conflict sets ($e_{i\cap j}=X$, $e_{i\setminus j}^{1\setminus 2}=X$, or $e_{i\setminus j}^{2\setminus 1}=X$) of all the cells of pairwise combination are added by multiplying with their weights (w_{ij}). In the denominator the mass of agreement and conflict committed to all the agreement and conflict sets ($A \in 2^{\Theta}$) are added by multiplying with their weights. Indeed, this is normalization. So, the value N in Eq. (18) is called *normalization constant*.

$$N = \sum_{\substack{i,j \\ A\in 2^{\Theta}}}\sum w_{ij} \times a_{ij}(A) + \sum_{\substack{i,j \\ A\in 2^{\Theta}}}\sum w_{ij} \times b_{ij}(A) + \sum_{\substack{i,j \\ A\in 2^{\Theta}}}\sum w_{ij} \times d_{ij}(A) \tag{18}$$

Then, given Eq. (18), Eq. (17) can be rewritten as Eq. (19).

$$m_{AFP}(X) = \left(\sum_{\substack{i,j \\ e_{i\cap j}=X}}\sum w_{ij} \times a_{ij}(X) + \sum_{\substack{i,j \\ e_{i\setminus j}^{1\setminus 2}=X}}\sum w_{ij} \times b_{ij}(X) + \sum_{\substack{i,j \\ e_{i\setminus j}^{2\setminus 1}=X}}\sum w_{ij} \times d_{ij}(X) \right) \Big/ N, \quad X \neq \varnothing \tag{19}$$

The normalization term N given in Eq. (18) of the combined *bba* is composed of three terms. In the first term the mass of agreement committed to all of the agreement sets are added by multiplying with their weights. In the second and third terms the mass of conflict committed to all of the conflict sets are added by multiplying with their weights. In the Eq. (18) if the first term is shown by N_a, and the total of the second term and the third term is shown by N_c Eq. (18) can be rewritten as Eq. (20). The total *measure of conflict* κ as an indicator of the effect of the conflict sets over the general combination results is achieved by the Eq. (21).

$$N = N_a + N_c \tag{20}$$

$$\kappa = N_c / (N_a + N_c) = N_c / N \tag{21}$$

It is said that as the total measure of conflict approaches *1* the conflict between the *bba*s which will be combined is high and as it approaches *0* the conflict between the *bba*s which will be combined is low. The fact that the conflict is high does not make the result meaningless but it stresses the existence of the high conflict between the data from two evidence sources.

5. Numerical examples

In this section we give two numerical examples. In these examples, we are interested in two individual belief functions to produce a single combined belief function when the individual belief functions come from equally reliable and occurrence independent sources of information. In Section 5.1, an example is illustrated to show the use of the AFP for general case. On the other hand, in Section 5.2, we chosen a special case example which have alternative optimal solution for its matching problem.

5.1. General case example

Example 2 has an optimal solution in the matching stage, so it serves showing the use of the AFP for general case.

Example 2. Given two belief functions, Bel_1 and Bel_2, over the same frame of discernment $\Theta = \{a,b,c\}$ from two bodies of evidence $E_1(\varepsilon_1, m_1)$ and $E_2(\varepsilon_2, m_2)$ as follows:

$$E_1(\varepsilon_1, m_1) = \{\, m_1(a,b,c) = 0.3\,; m_1(c) = 0.1\,;\; m_1(c,d) = 0.2\,;\; m_1(a,b,c,d) = 0.4 \,\}$$

$$E_2(\varepsilon_2, m_2) = \{\, m_2(a,b) = 0.4\,;\; m_2(c,d) = 0.6 \,\}.$$

Definition stage

The focal element set of the first *bba* is e_1 and the number of focal elements is $n = |e_1| = 4$. The focal element set of the second *bba* is e_2 and the number of focal elements is $m = |e_2| = 2$.

The agreement set of the *cfe* set to be obtained as a result of pairwise combination of the first focal element (*i=1*) of the first *bba,* namely $e_{11} = (a,b,c)$ and the first focal element (*j=1*) of the second *bba*, namely $e_{21} = (a,b)$; is

$$e_{1\cap 1} = e_{11} \cap e_{21} = \{a,b,c\} \cap \{a,b\} = \{a,b\} .$$

The conflict sets of the *cfe*, are obtained as;

$$e_{1\backslash 1}^{1\backslash 2} = e_{11} - \mathrm{e}_{21} = \{a,b,c\} - \{\mathrm{a,b}\} = \{\mathrm{c}\}$$

$$e_{1\backslash 1}^{2\backslash 1} = e_{21} - \mathrm{e}_{11} = \{a,b\} - \{\mathrm{a,b,c}\} = \{\varnothing\}$$

So with respect to the conflict set one is $\{\mathrm{c}\}$, the agreement set is $\{a,b\}$, and the conflict set two is $\{\varnothing\}$; by using the Eq. (1), the *cfe* of the first pairwise combination can be represented in the first cell of Table 1 as follows:

$$(c)\,((ab))\,(\ ) .$$

By using the Eq. (2), Eq. (3), and Eq. (4) for mass sharing, the mass of agreement, mass of conflict one and mass of conflict two are calculated as follows:

$$a_{11} = a_{11}(a,b) = \sqrt{r_{11}^{a}} = \sqrt{0.103} \cong 0.321 ,\ b_{11} = b_{11}(c) = \sqrt{r_{11}^{b}} = \sqrt{0.017} \cong 0.131 \text{, and}$$

$$d_{11} = d_{11}(\varnothing) = \sqrt{r_{11}^{d}} = 0 .$$

By using the Eq. (5) the size of qualitative conflict is calculated as,

$$c_{11} = \frac{b_{11} + d_{11}}{a_{11} + b_{11} + d_{11}} = \frac{0.131 + 0}{0.321 + 0.131 + 0} = 0.290 .$$

By using the Eq. (6) and Eq. (7) the size of quantitative conflict is calculated as,

$$g_{11} = \sqrt{0.3 * 0.4} = \sqrt{0.12} \cong 0.346 ,$$

$$v_{11} = 1 - \frac{\max\left\{\dfrac{m_{11}(e_{11})}{g_{11}}, \dfrac{g_{11}}{m_{11}(e_{11})}\right\}}{\dfrac{\max\{m_{11}(e_{11}), m_{21}(e_{21})\}}{\min\{m_{11}(e_{11}), m_{21}(e_{21})\}}} = 1 - \frac{\max\left\{\dfrac{0.3}{0.346}, \dfrac{0.346}{0.3}\right\}}{\dfrac{\max\{0.3, 0.4\}}{\min\{0.3, 0.4\}}} \cong 0.134 .$$

Up to here, only the results of the first pairwise combination have been obtained but the rest of the *4*2-1=7* pairwise combinations will not be given here in detail. The results of all pairwise combinations are given in Table 1.

Table 1. General combination table for Example 2

	m2(ab)=0.400	m2(cd)=0.600
m1 (abc) 0.300	(c)((ab))() b11 a11 d11 0.131 0.321 0.000 c11=0.290 v11=0.134	(ab)((c))(d) b12 a12 d12 0.200 0.283 0.245 c12=0.611 v12=0.293
m1 (c) 0.100	(c)(())(ab) b21 a21 d21 0.089 0.000 0.179 c21=1.000 v21=0.500	()((c))(d) b22 a22 d22 0.000 0.185 0.160 c22=0.464 v22=0.592
m1 (cd) 0.200	(cd)(())(ab) b31 a31 d31 0.163 0.000 0.231 c31=1.000 v31=0.293	()((cd))() b32 a32 d32 0.000 0.346 0.000 c32=0.000 v32=0.423
m1 (abcd) 0.400	(cd)((ab))() b41 a41 d41 0.200 0.346 0.000 c41=0.366 v41=0.000	(ab)((cd))() b42 a42 d42 0.219 0.438 0.000 c42=0.333 v42=0.184

Matching stage

In the matching stage for this example, the matching problem formulation by utilizing Eq. (14) is obtained as:

$$\min \quad 0.226x_{11} + 0.479x_{12} + 0.791x_{21} + 0.532x_{22} + 0.737x_{31} + 0.299x_{32} + 0.259x_{41} + 0.269x_{42}$$

$$x_{11} + x_{12} \le 1,\ x_{21} + x_{22} \le 1,\ x_{31} + x_{32} \le 1,\ x_{41} + x_{42} \le 1$$

$$x_{11} + x_{21} + x_{31} + x_{41} \le 1, \quad x_{12} + x_{22} + x_{32} + x_{42} \le 1$$

$$x_{11} + x_{12} + x_{21} + x_{22} + x_{31} + x_{32} + x_{41} + x_{42} = 2$$

$$x_{ij} = 0,1 \ , \qquad i = 1,\dots,4 \quad j = 1,2$$

In the optimal solution of the matching problem stated above, the values of the two matching decision variables are obtained as $x^*_{11} = 1$, $x^*_{42} = 1$ and the others are obtained as *0*. Thus, the *1*st proposition of the first *bba* and the *1*st proposition of the second *bba* are matched and also the *4*th proposition of the first *bba* and the *2*nd proposition of the second *bba* are matched.

Combination stage

In conformity with the above stated matching, the cells *(1,1)* and *(4,2)* of the general combination table become base cells of pairwise combination. Thus, the weights of the base cells of pairwise combination from the Eq: (15) are:

$$w^b_{11} \cong 1 - 0.226 = 0.774, \qquad w^b_{42} \cong 1 - 0.269 = 0.731$$

As the largest weight between the base cells of pairwise combination is $w^b_{11} = 0.774$, the weights of the cells *(1,2)* existing in the same row with the base cell *(1,1)* and the weights of the cells *(2,1)*, *(3,1)* and *(4,1)* existing in the same column with the base cell *(1,1)* are calculated by the Eq (16) as:

$$w_{21} = w_{31} = w_{42} = 0.226$$

As the second largest weight between the base cells of pairwise combination is $w_{42}^{b} = 0.731$ the weights of the cells *(2,2)* and *(3,2)*, existing in the same column with the base cell *(4,2)* are (using the Eq (16)) obtained as:

$$w_{22} = w_{32} = 0.269$$

If these obtained weight values are placed in the general combination table, the weighted general combination table will be obtained as given in Table 2.

Table 2. Weighted general combination table for Example 2

	m2(ab)=0.400	m2(cd)=0.600
m1 (abc) 0.300	w11=0.774 (c)((ab))() b11 a11 d11 0.131 0.321 0.000 c11=0.290 v11=0.134	w12=0.226 (ab)((c))(d) b12 a12 d12 0.200 0.283 0.245 c12=0.611 v12=0.293
m1 (c) 0.100	w21=0.226 (c)(())(ab) b21 a21 d21 0.089 0.000 0.179 c21=1.000 v21=0.500	w22=0.269 ()((c))(d) b22 a22 d22 0.000 0.185 0.160 c22=0.464 v22=0.592
m1 (cd) 0.200	w31=0.226 (cd)(())(ab) b31 a31 d31 0.163 0.000 0.231 c31=1.000 v31=0.293	w32=0.269 ()((cd))() b32 a32 d32 0.000 0.346 0.000 c32=0.000 v32=0.423
m1 (abcd) 0.400	w41=0.226 (cd)((ab))() b41 a41 d41 0.200 0.346 0.000 c41=0.366 v41=0.000	w42=0.731 (ab)((cd))() b42 a42 d42 0.219 0.438 0.000 c42=0.333 v42=0.184

By using the Table 2, Eq. (18), and Eq. (20) the normalization coefficient N is obtained as:

$$N_a = 0.854\,,\; N_c = 0.600\,,\; N = 0.854 + 0.600 = 1.454\,.$$

By using the Table 2 and the Eq. (19) the combined *bba* is obtained as:

$$m_{AFP}(c) = \left(w_{12}a_{12} + w_{22}a_{22} + w_{11}b_{11} + w_{21}b_{21}\right)/N = 0.162$$

$$m_{AFP}(d) = \left(w_{12}d_{12} + w_{22}d_{22}\right)/N = 0.068$$

$$m_{AFP}(a,b) = \left(w_{11}a_{11} + w_{41}a_{41} + w_{12}b_{12} + w_{42}b_{42} + w_{21}d_{21} + w_{31}d_{31}\right)/N = 0.429$$

$$m_{AFP}(c,d) = \left(w_{32}a_{32} + w_{42}a_{42} + w_{31}b_{31} + w_{41}b_{41}\right)/N = 0.341\,.$$

The total measure of conflict from the Eq. (21) is obtained as,

$$\kappa = 0.510/1.330 \cong 0.413\,.$$

5.2. Special case example

As a special case, some of the matching problems may have alternative optimal solutions. Example 3 has alternative optimal solutions in the matching stage, so it serves showing how can be obtained a final combined *bba* from alternative combined *bba*s.

Example 3. Given two belief functions, Bel_1 and Bel_2, over the same frame of discernment $\Theta = \{a,b,c\}$ from two bodies of evidence $E_1(\varepsilon_1, m_1)$ and $E_2(\varepsilon_2, m_2)$ as follows:

$$E_1(\varepsilon_1, m_1) = \{\, m_1(a,b,c) = 1 \,\}$$

$$E_2(\varepsilon_2, m_2) = \{\, m_2(a,b) = 0.5 \,;\; m_2(b,c) = 0.5 \,\}.$$

Definition stage

The general combination table in which the results obtained through $1\times 2 = 2$ numbered pairwise combinations of propositions which belong to two *bba*s is given in Table 3.

Table 3. General combination table for Example 3

	m2(ab)=0.500	m2(bc)=0.500
m1 (abc) 1.000	(c)((ab))() b11 a11 d11 0.333 0.624 0.000 c11=0.348 v11=0.293	(a)((bc))() b12 a12 d12 0.333 0.624 0.000 c12=0.348 v12=0.293

Matching stage

In the matching stage for this example, the matching problem formulation by utilizing Eq. (14) is obtained as:

$$\min \quad 0.308x_{11} + 0.308x_{12}$$

$$x_{11} + x_{12} \le 1,\; x_{11} \le 1,\; x_{12} \le 1$$

$$x_{11} + x_{12} = 1$$

$$x_{1j} = 0,1 \;, \qquad j = 1,2$$

The matching problem stated above has alternative optimal solutions. One of the alternative solutions is $x_{11}^* = 1$, $x_{12}^* = 0$, and the other is $x_{11}^* = 0$, $x_{12}^* = 1$.

Combination stage

The weighted general combination tables according to the two alternative optimal solutions are given in Table 4a and Table 4b, respectively.

Table 4a. Weighted general combination table for first alternative optimal solution of Example 3

```
           |     m2(ab)=0.500      |     m2(bc)=0.500      |
----------- --------------------- ---------------------
           |      w11=0.678        |      w12=0.322        |
   m1      |     (c)((ab))()       |     (a)((bc))()       |
  (abc)    |    b11  a11  d11      |    b12  a12  d12      |
  1.000    | 0.333  0.624  0.000   | 0.333  0.624  0.000   |
           | c11=0.348  v11=0.293  | c12=0.348  v12=0.293  |
----------- --------------------- ---------------------
```

Table 4b. Weighted general combination table for second alternative optimal solution of Example 3

```
           |     m2(ab)=0.500      |     m2(bc)=0.500      |
----------- --------------------- ---------------------
           |      w11=0.322        |      w12=0.678        |
   m1      |     (c)((ab))()       |     (a)((bc))()       |
  (abc)    |    b11  a11  d11      |    b12  a12  d12      |
  1.000    | 0.333  0.624  0.000   | 0.333  0.624  0.000   |
           | c11=0.348  v11=0.293  | c12=0.348  v12=0.293  |
----------- --------------------- ---------------------
```

So according to two weighted general combination tables given in Table 4a and 4b, two alternatives combined bba are obtained. Then, as doing in Example 1 in Section 3.2, the geometric mean of these two alternatives combined bba can be used as the final combined bba. The results are given as follows:

The first alternative combined bba	*The second alternative combined bba*	*The final combined bba with geometric mean*
$m_{AFP}(a) = 0.112$	$m_{AFP}(a) = 0.236$	$m_{AFP}(a) = 0.174$
$m_{AFP}(c) = 0.236$	$m_{AFP}(c) = 0.112$	$m_{AFP}(c) = 0.174$
$m_{AFP}(a,b) = 0.442$	$m_{AFP}(a,b) = 0.210$	$m_{AFP}(a,b) = 0.326$
$m_{AFP}(b,c) = 0.210$	$m_{AFP}(b,c) = 0.442$	$m_{AFP}(b,c) = 0.326$
$\kappa = 0.348$,	$\kappa = 0.348$,	$\kappa = 0.348$.

6. Comparison of three methods

This section describes four examples that compare weighted average operator (WAO), combination by compromise (CBC) and the AFP. The definitions of the WAO and the CBC can be seen in [38] and [32], respectively.

Example 4: Zadeh's example

In the literature, Zadeh's example appears in different but essentially equivalent versions of disagreeing experts. We will present Zadeh's example by the story of that a patient examined by two doctors [9, 10]. Assume that the first doctor diagnosis is that patient has either meningitis, with probability 0.99, or brain tumor, with probability 0.01. The second doctor agrees with the first one that the probability of brain tumor is 0.01, but believes that it is the probability of concussion rather than meningitis that is 0.99. So, they provide the following diagnosis:

$$m_1(m) = 0.99 \quad m_1(t) = 0.01 \text{ and } m_2(c) = 0.99 \quad m_2(t) = 0.01 .$$

The combination results are given in Table 5. In the columns for the WAO and AFP, it is observed that the *bbm* of the agreed *t* increases and the *bbm* of conflicted *m* and *c* decreases. However, with CBC, the *bbm* of the agreed *t* decreases. According to the AFP, the measure of conflict for Zadeh's example is *0.976*. The other methods did not give a measure of conflict.

Table 5. Comparison of methods in Zadeh's example

	m_1	m_2	WAO	CBC	AFP
m	0.99	0.00	0.494951	0.499851	0.488097
t	0.01	0.01	0.010099	0.000298	0.023805
c	0.00	0.99	0.494951	0.499851	0.488097
conflict, κ			-	-	0.976000

Example 5: Zadeh's modified example

When introducing a small amount of uncertainty in the doctor's opinions, the *bbm*s and the results of applying the methods are given in Table 6. In this case, only in the column for the AFP, it is observed that the *bbm* of the agreed *t* and the *bbm* of the agreed Θ increases. According to the AFP, the measure of conflict for Zadeh's modified example is *0.953*. When comparing with the measure of conflict of Zadeh's example, there is a decrease in the measure of conflict. This decrease is also a consistent result.

Table 6. Comparison of methods in Zadeh's modified example

	m_1	m_2	WAO	CBC	AFP
m	0.98	0.00	0.490000	0.494801	0.476522
t	0.01	0.01	0.010100	0.000398	0.023478
c	0.00	0.98	0.490000	0.494801	0.476522
m,t	0.00	0.00	0.000000	0.004900	0.000000
m,c	0.00	0.00	0.000000	0.000100	0.000000
t,c	0.00	0.00	0.000000	0.004900	0.000000
Θ	0.01	0.01	0.009900	0.000100	0.023478
conflict, κ			-	-	0.953000

Example 6: Bayesian belief functions

Let m_1 and m_2 belong to two Bayesian belief functions over the $\Theta = \{a, b, c\}$. The *bbm*s and the results of applying the methods are given in Table 7. In this example, the results are expected to be between $(0.10, 0.20)$ for singleton a, $(0.20, 0.30)$ for singleton b, and $(0.50, 0.70)$ for singleton c. The results produced by CBC are not corresponding to these expectations. The WAO produce results in expected intervals, however the results are very close to the borders. In the columns for the AFP the results are consistent with expectations. Furthermore, the results are consistent with geometric means between m_1's and m_2's *bbm*s of same singletons. According to the AFP, the measure of conflict for this example is *0.242*. It is reasonable since the existence of the quantitative conflict between *bbm*s.

Table 7. Comparison of methods in Bayesian belief functions

	m_1	m_2	WAO	CBC	AFP
a	0.10	0.20	0.105500	0.086944	0.152085
b	0.20	0.30	0.202500	0.194071	0.264078
c	0.70	0.50	0.692000	0.718984	0.583837
conflict, κ			-	-	0.242000

Example 7: Identical Bayesian belief functions

Let m_1 and m_2 belong to two identical Bayesian belief functions over the $\Theta = \{a, b, c\}$. The *bbm*s and the results of applying the methods are given in Table 8. In this case, according to the idempotency rule, the results are expected to be the same with these identical *bba*s. Only the AFP produced the expected results. According to the AFP, the measure of conflict for this example is zero. It is reasonable since there is no conflict between two bbas with respect both qualitative and quantitative.

Table 8. Comparison of methods in identical Bayesian belief functions

	m_1	m_2	WAO	CBC	AFP
a	0.10	0.10	0.056000	0.040833	0.100000
b	0.20	0.20	0.132000	0.128889	0.200000
c	0.70	0.70	0.812000	0.830278	0.700000
conflict, κ			-	-	0.000000

7. Conclusions

We propose a new consensus generator called "Analytic Fusion Process" in the framework of mathematical theory of evidence. The proposed method is a geometric mean based analytical method that operate on the individual belief functions to produce a single combined belief function when the individual belief functions comes from equally reliable and occurrence independent sources of information. This method satisfies the idempotent and commutative law. Furthermore, this method also produces a measure of conflict shows whether the original beliefs were in harmony or in conflict. Another advantage is that the measure of conflict produced by this method reflects both qualitative and quantitative conflict. Unfortunately, Analytic Fusion Process is not associative. However, an *n*-ary version of the method can be developed, and combining *n* basic belief assignments simultaneously can be a practical substitute for associativity in many real world application. The other disadvantage of the proposed method is it needs many calculation process. However, using a computer program to accomplish the whole calculation process of the methods simplifies the task. We have written such a program that accomplish the whole calculation process of the Analytic Fusion Process.

The proposed method, the weighted average operator, and the combination by compromise method are compared with each other using four examples. The results show that the Analytic Fusion Process produce results that are much more convincing than the others. Furthermore, it is seen that measure of conflict produced by Analytic Fusion Process is really reasonable. To our knowledge it is the only real single measure of conflict which measures how the original beliefs were in conflict in the framework of mathematical theory of evidence.

Acknowledgements

The authors thank the anonymous reviewer for providing valuable remarks and suggestions which helped to improve quality of the paper. We also thank Prof. Glen Meeden, Yasemin Gençtürk, and Gürkan Özel for providing language help.

References

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, London, 1976.

[2] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, Ann. Mathematic Statistics, (1967) Vol. 38, pp. 325-339.

[3] K. Sentz, S. Ferson, Combination of Evidence in Dempster-Shafer Theory, Sandia National Laboratories Report, SAND2002-0835, California, 2002.

[4] Ph. Smets, R. Kennes, The Transferable Belief Model, Artificial Intelligence, Vol. 66 (1994), pp. 191-234.

[5] J. Kohlas, P.A. Monney, *A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence*, Vol. 425 of Lecture Notes in Economics and Mathematical Systems, Springer –Verlag, 1995.

[6] R. Haenni, J. Kohlas, N. Lehman, Probabilistic argumentation systems, J. Kohlas, S. Moral (Eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, Vol. 5 of Algorithms for Uncertainty and Defeasible Reasoning, pp. 221-288, Kluwer Academic Publishers, 2000.

[7] F. Smarandache, J. Dezert (Eds.), *Advances and Applications of DSmT for Information Fusion, Smarandache*, Vol 2. American Research Press, Rehoboth, 2006.

[8] M. Detyniecki, Fundamentals on Aggregation Operators, http://www.cs.berkeley.edu/~marcin/agop.pdf , 2001.

[9] L.A. Zadeh, On the validty of Dempster's rule of combination of evidence. Technical Report 79/24, University of California, Berkely, 1979.

[10] L.A. Zadeh, A mathematical theory of evidence (book review), AI Mag. 5(3) (1984), pp. 81-83.

[11] R. Haenni, Shedding new light on Zadeh's Criticism of Dempster's rule of Combination, Proceedings of Information Fusion 2005, Philadelphia, July 2005.

[12] H.Y. Hau, R.L. Kashyap, On the Robustness of Dempster' s Rule of Combination, IEEE International Workshop on Tools for Artificial Intelligence, Fairfax, VA, Oct. (1989), pp. 578-582.

[13] E. Lefevre, O. Colot, P. Vannoorenberghe, Belief Function Combination and Conflict Management, Information Fusion, Vol. 3 (2002), pp. 149-162.

[14] C.K. Murphy, Combining Belief Functions when Evidence Conflicts, Decision Support Systems, Vol. 29 (2000), pp. 1-9.

[15] Ph. Smets, Analyzing the Combination of Conflicting Belief Functions, Information Fusion, Vol. 8, (2007), pp. 387-412.

[16] F. Voorbraak, On the justification of Dempster's rule of combination, Artificial Intelligence, 48, (1991), pp. 171-197.

[17] R. Haenni, Are alternatives to Dempster's rule of combination real alternatives? Comments on "About the belief function combination and the conflict management problem"-Lefevre et al, Information Fusion, Vol. 3 (2002), pp. 237-239.

[18] F. Campos, S. Cavalcante, An Extended Approach for Dempster-Shafer Theory, IEEE International Conference on Information Reuse and Integration (IRI 2003), Las Vegas, USA, October, 2003.

[19] M. Daniel, Associativity in Combination of belief functions; a derivation of minC combination, Soft Computing, 7(5), (2003), pp. 288–296.

[20] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence, Volume 4, Issue 3, (1988), pp. 244-264.

[21] A. Josang, The consensus operator for combining beliefs, Artificial Intelligence, 141 (2002), pp. 157-170.

[22] R. R. Yager, On the Dempster-Shafer framework and new combination rules, Information Sciences, (1987), 41 (2), p.93-137.

[23] Fuyuan Xiao, Evidence combination based on prospect theory for multi-sensor data fusion, ISA Transactions, Volume 106, (2020), Pages 253-261,

[24] Jiang W., Zhan J., A modified combination rule in generalized evidence theory, Appl. Intell., (2017), 46 (3) pp. 630-640.

[25] Jian W., Kuoyuan Q., Zhiyong Z., An improvement for combination rule in evidence theory, Future Generation Computer Systems, (2019), Volume 91, Pages 1-9,

[26] Wenjun M., Yuncheng J., Xudong L., A flexible rule for evidential combination in Dempster–Shafer theory of evidence, Applied Soft Computing, (2019), Volume 85.

[27] D. Dubois, H. Prade, On the combination of evidence in various mathematical frameworks, J. Flamm and T. Luisi (Eds.), *Reliability Data Collection and Analysis*, Brussels, ECSC, EEC, EAFC: pp. 213-241, 1992.

[28] R. Haenni, S. Hartmann, Modeling partially reliable information source: A general approach based on Dempster-Shafer theory, Information Fusion, Vol. 7 (2006), pp. 361-379.

[29] M.E.G.V. Cattaneo, Combining belief functions issued from dependent sources, In: J.M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications (ISIPTA'03), 2003, Carleton Scientific, Lugano, Switzerland. pp. 133-147.

[30] P.A. Monney, M. Chan, Modelling dependency in Dempster-Shafer theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 15, No. 1 (2007), pp. 93-114.

[31] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial Intelligence, 172, (2008), 234-264.

[32] K. Yamada, A new combination of evidence based on compromise, Fuzzy Sets and Systems, 159, (2008), 1689-1708.

[33] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, K. Sentz, Constructing Probability Boxes and Dempster-Shafer Structures, Sandia National Laboratories Report, SAND2002-4015, California, 2003.

[34] F. Smarandache, J. Dezert, Proportional Conflict Redistribution Rules for Information Fusion, arXiv Archives, Los Alamos National Laboratory; the Abstract and the whole paper are available at http://arxiv.org/PS_cache/cs/pdf/0408/0408064.pdf , 2005.

[35] B.B. Yaghlane, P. Smets, K. Mellouli, Belief function independence: I. The marginal case, International Journal of Approximate Reasoning, Vol. 29, (2002), pp. 47-70.

[36] B.B. Yaghlane, P. Smets, K. Mellouli, Belief function independence: II. The conditional case, International Journal of Approximate Reasoning, Vol. 31, (2002), pp. 31-75.

[37] R.T. Clemen, R.L. Winkler, Combining Probability Distributions From Experts in Risk Analysis, Risk Analysis, (1999) Vol. 19, No. 2 pp. 187-203.

[38] A. Josang, M. Daniel, P. Vannoorenberghe, Strategies for combining conflicting dogmatic beliefs, Applications of plausable, paradoxical, and neutrosophical reasoning for information fusion (The sixth international conference on information fusion), Cairns, Queensland, Australia, July 8-11, 2003.

[39] W. Liu, Analyzing the degree of conflict among belief functions, Artificial Intelligence, Vol. 170 (11) (2006), pp. 909-924.

[40] C. Genest, J.V. Zidek, Combining probability distributions: A critique and a annotated bibliography, Statistical Science, Vol. 1, No. 1, (1986), pp.114-148.

[41] W.L. Winston, *Operations Research Aplications and Algorithms*, Duxbury Press, Belmont, 1994.

[42] L.A. Wolsey, *Integer Programming*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience Publication, NewYork, 1998.