Statistica & Applicazioni - Vol. XIX, n. 1, 2021, pp. 35-48
Published online november 2021

DISTRIBUTION OF THE RATIO OF FEMALE INCOME OVER MALE INCOME

Marcella Mazzoleni*
Angiola Pollastri*
Vanda Tulli*

SUMMARY

There is a great discussion concerning the differences between the income of females and males. The purpose of the present study is to estimate the distribution of the ratio of female income over male income. The methodology to study the ratio in exam is based on the distribution of the ratio of two Dagum random variables with three parameters proposed in 2010 by Pollastri and Zambruno. We will consider the official statistics of income and we will estimate the parameters of the Dagum distribution. The distribution of the ratio in question studied in two different situations can reveal the gender inequality concerning income in different times, regions, and age classes.

Keywords: Dagum Distribution, Distribution of the Ratio of two Dagum Random Variables, Distribution of the Ratio of Female and Male Income.
DOI: 10.26350/999999_000039
ISSN: 18246672 (print) 2283-6659 (digital)

1. INTRODUCTION

It is often a matter of discussion that, even with a higher educational level, women earn less than men do. The differences between men's and women's income on average are decreasing in recent years but income parity has not yet been achieved.

The purpose of this paper is to estimate the distribution of the ratio of females' income over males' income. The methodology used to study the ratio is based on the distribution of the ratio of two Dagum with three parameters (Pollastri and Zambruno, 2010).

Firstly, the parameters of the Dagum distribution are estimated using maximum likelihood estimation. Subsequently, the distribution of the ratio is analysed to reveal the gender inequality with applications to the income in different age classes, areas and times, and the retirement income in different times.

* Departimento di Statistica e Metodi Quantitativi - Università Milano-Bicocca - Via Bicocca degli Arcimboldi,8 - 20126 MILANO
(e-mail: ✉ marcella.mazzoleni@unimib.it; angiola.pollastri@unimib.it; vanda.tulli@unimib.it).

2. THE DAGUM DISTRIBUTION AND ESTIMATION METHOD

In literature we have many examples which confirm that the model proposed in 1977 by Camilo Dagum fits very well to many distributions of economic variables. In fact, the Dagum model fulfils many properties considered relevant to a model for the distribution of income in Italy such as the convergence to the Pareto law.

Supposing that X is a type I Dagum, then $X \sim D(\lambda, \beta, \delta)$ with $\lambda, \beta, \delta > 0$. The distribution function is defined as:

$$F_X(x) = \left(1 + \lambda x^{-\delta}\right)^{-\beta}$$

While the density function is:

$$f_X(x) = \beta\delta\lambda x^{-\delta-1}\left(1 + \lambda x^{-\delta}\right)^{-\beta-1} \tag{1}$$

For the estimation of the Dagum parameters the VGAM package in R is used. The function *dagum* gives maximum likelihood estimation of the 3-parameter Dagum distribution following Kleiber and Kotz (2003) definitions. Domanski and Jedrzejczak (1998) showed, through a simulation study, that the performance of maximum likelihood estimation is good for β and δ when $n > 2000$ or 3000, while for the scale parameter λ the bias tends to 0 when $n > 4000$.

3. THE DISTRIBUTION OF THE RATIO OF TWO DAGUM DISTRIBUTIONS

The purpose of this paper is to analyse the ratio:

$$U = \frac{X}{Y}$$

where

$$X \sim D(\lambda_1, \beta_1, \delta_1) \text{ and } Y \sim D(\lambda_2, \beta_2, \delta_2)$$

with X and Y independent.

Following Mood, Graybill and Boes (1974) the density function of the ratio of two random variables is:

$$f_U(u) = \int_{-\infty}^{+\infty} |y| f_{X,Y}(uy, y)dy \tag{2}$$

Applying the independence of X and Y and the density function of a type I Dagum (1), it is possible to obtain from the (2) the density function for the ratio U:

$$f_U(u) = \int_0^{+\infty} y\{\beta_1\delta_1\lambda_1(uy)^{-\delta_1-1}[1 + \lambda_1(uy)^{-\delta_1}]^{-\beta_1-1}\}\times$$

$$\{\beta_2\delta_2\lambda_2(y)^{-\delta_2-1}[1 + \lambda_2(y)^{-\delta_2}]^{-\beta_2-1}\}dy$$

Using the definition of the cumulative distribution function, it is possible to obtain:

$$F_U(u) = \beta_1\delta_1\lambda_1\beta_2\delta_2\lambda_2 \int_0^u t^{-\delta_1-1} \int_0^u y^{-\delta_1-\delta_2-1}[1+\lambda_1(ty)^{-\delta_1}]^{-\beta_1-1}\times$$

$$[1+\lambda_2(y)^{-\delta_2}]^{-\beta_2-1}dydt$$

For the inner integral involved in the cumulative distribution function of the ratio a numeric quadrature is used (QDAGI subroutine from IMSL library). While for the outer integral, the argument is expressed in a tabular form. In Pollastri and Zambruno (2010) a graphical analysis of this method performance is exposed comparing the empirical and the computed density function, where the empirical one is created with the ratios of all the possible couples.

To understand the meaning of the ratio of the females' income over the males' income we consider a simple example. Let suppose that in a group composed by 5 males the incomes, indicated by Y, assume the following values: 2, 4, 5, 8, 16. And in a group composed by 5 females the incomes, indicated by X, are the following: 1, 2, 4, 5, 12.

In Table 1 we consider all possible ratios between the income of each female and the income of each male.

TABLE 1. - *Ratios between the females' income and the males' income*

X \ Y	2	4	5	8	16
1	0.5000	0.2500	0.2000	0.1250	0.0625
2	1.0000	0.5000	0.4000	0.2500	0.1250
4	2.0000	1.0000	0.8000	0.5000	0.2500
5	2.5000	1.2500	1.0000	0.6250	0.3125
12	6.0000	3.0000	2.4000	1.5000	0.7500

Each couple of a female and of a male has probability 0.04 to be selected. In Table 2 is reported the probability function of the random variable $U = X/Y$.

Now we compute the deciles of the random variable U. From Table 3 we can conclude, for example, that the ratio between the female income and male income is less than 0.125 in the 10% of possible couples and the ratio between the female income and male income is less than 0.625 in the 50% of possible couples.

TABLE 2. - *Probability function and Cumulative Distribution Function of the random variable U*

u	$p(u)$	$F(u)$
0.0625	0.04	0.04
0.1250	0.08	0.12
0.2000	0.04	0.16
0.2500	0.12	0.28
0.3125	0.04	0.32
0.4000	0.04	0.36
0.5000	0.12	0.48
0.6250	0.04	0.52
0.7500	0.04	0.56
0.8000	0.04	0.60
1.0000	0.12	0.72
1.2500	0.04	0.76
1.5000	0.04	0.80
2.0000	0.04	0.84
2.4000	0.04	0.88
2.5000	0.04	0.92
3.0000	0.04	0.96
6.0000	0.04	1.00
Total	1	

TABLE 3. - *Deciles of the ratio (Females' income over males' income)*

N	**DECILES**
1	0.1250
2	0.2500
3	0.3125
4	0.5000
5	0.6250
6	0.8000
7	1.0000
8	1.5000
9	2.5000

4. APPLICATIONS

We analyse two datasets. The first dataset is composed of the individual net incomes from the Bank of Italy Survey on Household Income and Wealth (SHIW). We compare the ratio of the females' and males' income in different groups:

- males and females in two different years 2016 and 1998
- males and females divided in three age classes: *young* $(age < 40)$, *adult* $(40 \leq age < 70)$ and *old* $(age \geq 70)$
- males and females divided in three areas: *N (North), C (Centre),* and *S&I (South and Islands)*

The dataset is composed of 11844 subjects. Of these 50.98% are males, and 49.02% are females. Concerning the division by ages 15.21% are aged less than 40 years, 54.41% are aged between 40 and 70 years, and 30.39% are aged equal or more than 70 years. Concerning the division by area, 43.90% of the subjects come from the north, 22.40% from the centre, and 33.70% from the south and the islands. Concerning the 1998 dataset, it is composed of 12616 subjects, of these 56.52% are males and 43.48% are females.

For the second dataset, we compare the ratio of the females' and males' retirement income in different times, 2016 and 1997, using the data given by the Central register of retirees (INPS, the National Social Welfare Institution). The 2016 dataset is composed of 16064508 retirees, 47.32% are males and 52.68% are females. Whilst the 1997 dataset is composed of 9483681 retirees, 55.31% are males and 44.69% are females.

4.1 *Results of the ratio distribution for Bank of Italy dataset*

Firstly, in Table 4 we compare the mean value in each sub-group. We observe that the income mean value for females is lower than that for males. For young subjects, the females have a higher income mean value but increasing the age it is possible to observe an inversion and an increase in this difference at the expense of the females. We observe that in the partition by area the income mean value for the females is always lower than the males' income, but in the centre, we observe that the mean values are close. On the contrary, for the other areas we observe a bigger difference between gender, especially for the south and islands. Comparing the income mean value for females and males in 2016 and in 1998, we observe a bigger difference for the 1998 dataset. This mean values analysis suggests us what to aspect analysing and comparing the distribution of the ratio of the income of females over males.

Using the function *dagum* of the VGAM package in R, we obtain for the Dagum income distribution the following parameters: $\delta = 4.2384$, $\beta = 0.3298$, $\lambda = 1033064$.

While the parameters of Dagum distribution for the income of the females are

TABLE 4. - *Summaries of Bank of Italy dataset divided by gender in each age, area, and time (the income is divided by 1000 and for the 1998 it is in millions of Italian Lira)*

	MEAN	MEDIAN	STANDARD DEVIATION
TOTAL 2016	**19.103**	**16.908**	**15.152**
FEMALES	18.580	16.414	14.385
MALES	19.607	17.491	15.839
TOTAL 1998	**28.792**	**23.509**	**29.289**
FEMALES	20.011	18.000	15.920
MALES	35.547	29.330	34.898
YOUNG	**17.692**	**15.761**	**13.431**
FEMALES	17.819	15.902	13.677
MALES	17.597	15.628	13.251
ADULT	**19.360**	**17.090**	**14.648**
FEMALES	18.860	16.665	14.723
MALES	19.812	17.700	14.567
OLD	**19.350**	**17.000**	**16.734**
FEMALES	18.445	16.373	14.117
MALES	20.458	18.014	19.412
NORTH	**20.826**	**18.026**	**16.902**
FEMALES	20.190	18.000	14.837
MALES	21.457	18.500	18.708
CENTRE	**19.668**	**17.498**	**13.415**
FEMALES	19.649	17.000	14.538
MALES	19.687	17.820	12.205
SOUTH & ISLANDS	**16.484**	**14.802**	**13.364**
FEMALES	15.639	13.856	13.153
MALES	17.249	15.500	13.511

$\delta = 4.2307,\ \beta = 0.3124,\ \lambda = 999772.4$ and the parameters of Dagum distribution for the income of the males are $\delta = 4.2307,\ \beta = 0.3124,\ \lambda = 999772.4$. We accept that the data follows these distributions using Anderson-Darling and Cramer-von Mises test of goodness-of-fit with Braun's adjustment. Braun's adjustment (1980) is used as these tests assume that the parameters of the null distribution are known. On the contrary, in our case the parameters were estimated (calculated from the data), then this method is used to adjust for the effect of estimating the parameters from the data. Braun's method (1980) involves randomly dividing the data into two equally sized subsets. This technique is expected to work well when the number of observations is large, such as our case.

After having estimated the cumulative distribution function of the ratio of the females' income over the males' income, it is possible to analyse the results of the ratio divided by age, area, and time.

In Table 5 we observe that the deciles of the ratio of the females' income over males' income are higher in 2016 with respect to the ratio in 1998.

TABLE 5. - *Deciles of the ratio (Females' over males' income in 2016 and 1998)*

DECILES	**2016**	**1998**
1	0.2293	0.1504
2	0.3946	0.2516
3	0.5554	0.3514
4	0.7274	0.4596
5	0.9258	0.5860
6	1.1741	0.7450
7	1.5186	0.9654
8	2.0802	1.3216
9	3.3469	2.1134

These results are confirmed also by the density function graph of the ratio in Figure 1, where we observe that the density function of the ratio of the females' income over males' income in 1998 has higher value for lower value, compared with the density function of the ratio of the females' income over males' income in 2016.

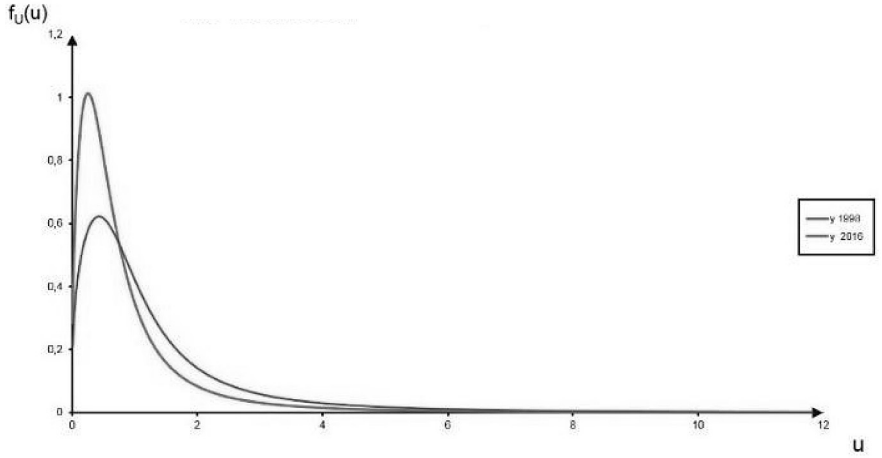

FIGURE 1. - *Density function of the ratios of the income of females over the income of males in 1998 and 2016*

The cumulative distribution function graph of the in Figure 2 endorses these conclusions as the cumulative distribution function of the ratio of the females' income over males' income in 1998 is higher for lower value compared with the cumulative distribution function of the ratio of the females' income over males' income in 2016.

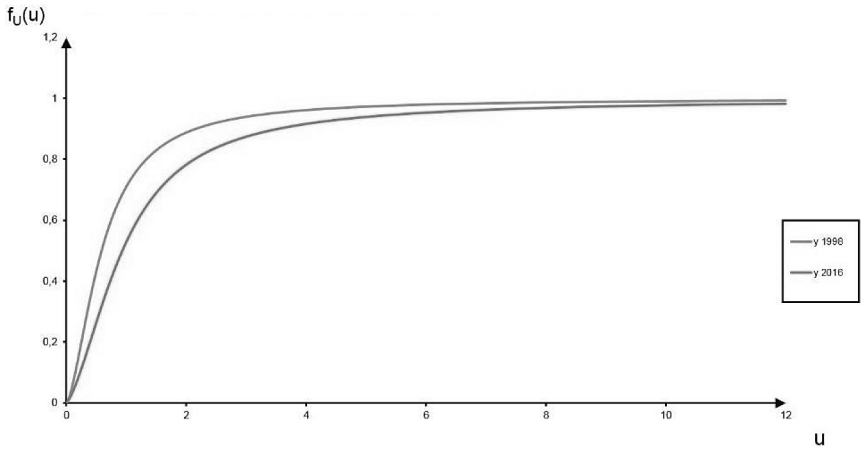

FIGURE 2. - *Cumulative distribution function of the ratios of the income of females over the income of males in 1998 and 2016*

Comparing the ratio of the females' income over males' income in different age classes in Table 6, we observe higher value of deciles for younger subjects, lower for the adult group, and even lower for the old group, except for the first three deciles that are higher for the old group.

TABLE 6. - *Deciles of the ratio (Females' over males' income divided by age classes)*

DECILES	YOUNG	ADULT	OLD
1	0.2055	0.2160	0.2602
2	0.3801	0.3816	0.4168
3	0.5563	0.5445	0.5647
4	0.7477	0.7192	0.7230
5	0.9706	0.9211	0.9015
6	1.2516	1.1736	1.1224
7	1.6450	1.5245	1.4249
8	2.2938	2.0974	1.9079
9	3.7788	3.3948	2.9660

Analysing the density function of the ratio of the females' income over males' income in different age classes in Figure 3, the results of the deciles are confirmed, as

we observe higher value of density function for lower value of the ratio for the old group with respect to the other two groups, and higher value of density function for lower value for the adult group than the young group.

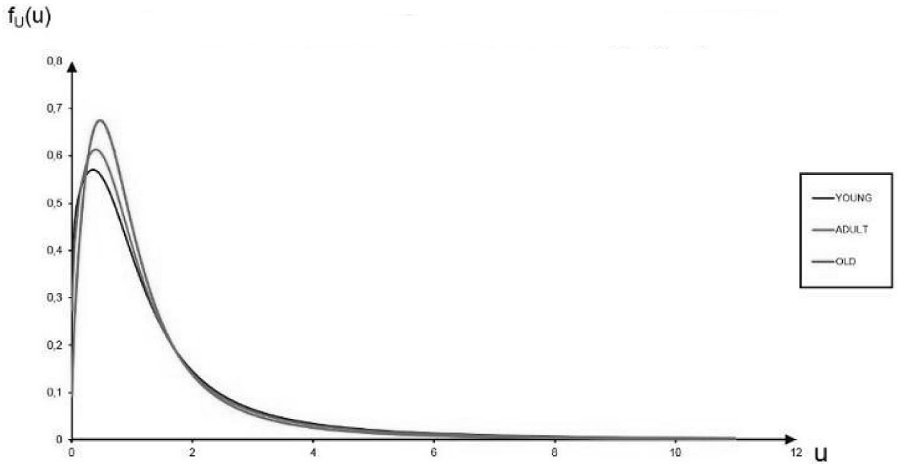

FIGURE 3. - *Density function of the ratios of the income of females over the income of males for three age groups*

This is endorsed also by the cumulative distribution function graph in Figure 4, as we observe higher value of cumulative distribution function for lower value of the ratio for the old group with respect to the other two groups, and higher values of cumulative distribution function for lower value for the adult group with respect to the young group.

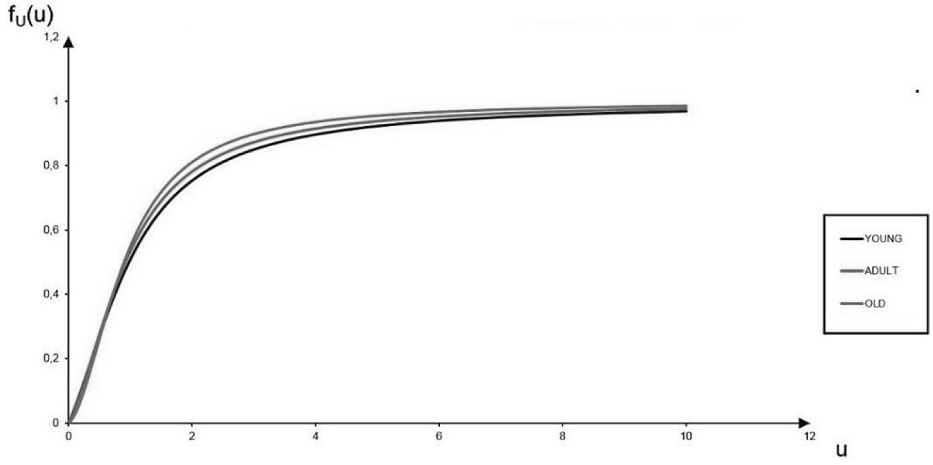

FIGURE 4. - *Cumulative distribution function of the ratios of the income of females over the income of males for three age groups*

In Table 7 it is possible to observe the deciles of ratio of the females' income over males' income comparing different areas, we observe close and higher value for the subjects that live in north and centre of Italy, and lower value for the subjects that live in south and islands, except for the ninth decile.

TABLE 7. - *Deciles of the ratio (females' over males' income divided by areas)*

DECILES	NORTH	CENTRE	SOUTH & ISLANDS
1	0.2670	0.2507	0.1830
2	0.4331	0.4169	0.3406
3	0.5910	0.5757	0.4997
4	0.7587	0.7437	0.6720
5	0.9516	0.9356	0.8719
6	1.1921	1.1731	1.1235
7	1.5244	1.4978	1.4759
8	2.0621	2.0163	2.0592
9	3.2608	3.1509	3.4096

Analysing the density function of the ratio of the females' income over males' income in different areas in Figure 5, the results of deciles are confirmed, as we observe higher value of density function for lower value of the ratio for the subjects that come from the south and islands with respect to the other two groups, while for the subjects that come from north and centre the density function are almost overlapped.

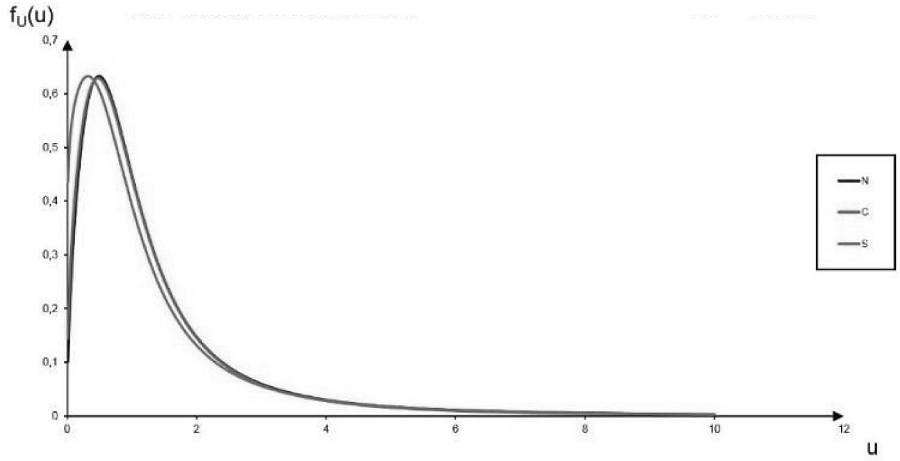

FIGURE 5. - *Density function of the ratios of the income of females over income of males relating to the Italian territorial division*

This is endorsed by the cumulative distribution function graph of the ratio of the females' income over males' income in different areas in Figure 6, as we observe higher value of cumulative distribution function for lower value of the ratio for the subjects that come from the south and islands with respect to the other two groups, while for the subjects that come from north and centre the cumulative distribution function are almost overlapped.

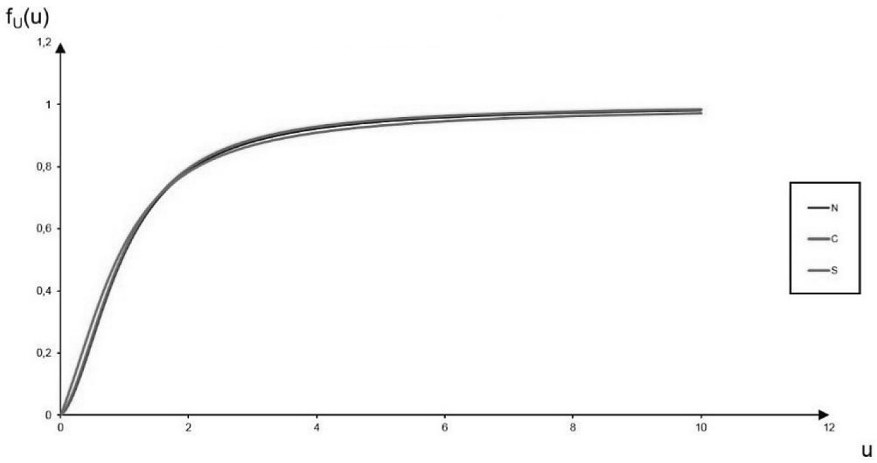

FIGURE 6. - *Cumulative distribution function of the ratios of the income of females over income of males relating to the Italian territorial division*

4.2 *Results of the ratio distribution for Retirement dataset*

In 1997 the mean value for the retirement income was 22.577 million of Italian Lira, with a mean for females which was 18.182 million of Italian Lira and a mean value for males which was 26.129 million of Italian Lira. Concerning the 2016 dataset, we observe a mean value for the retirement income, which was 17580 Euros, with a mean for females which was 14780 Euros and a mean value for males which was 20697 Euros.

The deciles of the ratio of the females' retirement income over males' retirement income in 1998 and 2016 are shown in Table 8. It is possible to observe that the deciles of the ratio of the females' retirement income over males' retirement income are higher in 2016 with respect to the ratio in 1998.

Analysing the density function of the ratio of the females' retirement income over males' retirement income in different years in Figure 7, the results of deciles are confirmed, as we observe higher value of density function for lower value of the ratio for the 1997 dataset than the 2016 dataset.

This is endorsed also by the cumulative distribution function graph in Figure 8, as we observe higher value of cumulative distribution function for the ratio for the older dataset with respect to the younger dataset.

TABLE 8. - *Deciles of the ratio of retirement income (Females' over males' retirement income in 1997 and 2016)*

DECILES	1997	2016
1	0.1053	0.2159
2	0.2276	0.3465
3	0.3628	0.4671
4	0.5114	0.5927
5	0.6787	0.7362
6	0.8770	0.9165
7	1.1314	1.1706
8	1.5060	1.5941
9	2.2411	2.5627

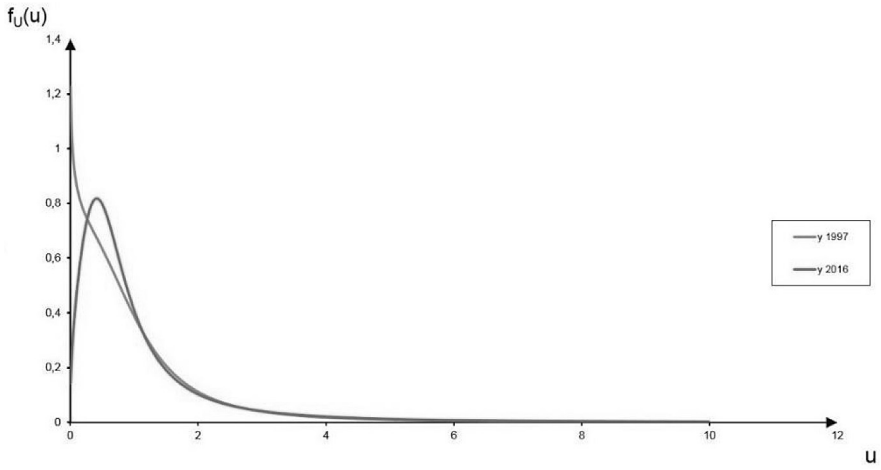

FIGURE 7. - *Density function of the ratios of the retirement income of females over the retirement income of males compared in 1997 ad 2016*

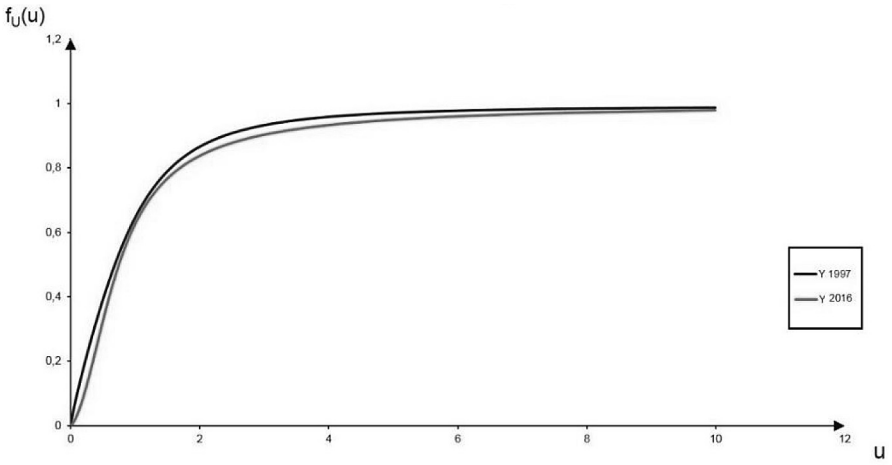

FIGURE 8. - *Cumulative distribution function of the ratios of the retirement income of females over the retirement income of males compared in 1997 ad 2016*

5. CONCLUSIONS

In this paper we propose to use the ratio of two type I Dagum random variables for analysing the difference of the income of females and males. We observe that this method is good as it is possible to compare the ratio of females' over males' income and retirement income in different age, territorial groups and different times. Analysing the results, it is possible to obtain interesting conclusions concerning the difference of the income of females and males.

As a matter of fact, in the applications we observe less diversity for females' and males' income in the younger group and the diversity rises increasing the subjects age, passing then to adult and old group. In the division by areas, the deciles of the ratio of females' income over males' income for the north and centre are close, while for the south and islands a bigger income difference between gender is observed. The difference of the income for males' and females' is decreasing over the years (1998-2016) and it is possible to obtain the same conclusion for the retirement income as the difference is decreasing over the years (1997-2016). In fact, the deciles of the ratio of females' income over males' income are higher for 2016 datasets than the deciles of the previous dataset considered.

REFERENCES

Alleva G. (2017). *Indagine conoscitiva sulle politiche in materia di parità tra uomini e donne*. Istat, Rome.

Baldacci E., Inglese L. (2000). *Le caratteristiche socio-economiche dei pensionati in Italia, Analisi della distribuzione dei redditi da pensione*. No. 03. Center for Research on Pensions and Welfare Policies, Turin (Italy).

Bank of Italy (2016). *I bilanci delle famiglie italiane nell'anno 2016. Supplementi al Bollettino Statistico,* XVII. Centro Stampa Banca d'Italia, Roma.

Bank of Italy (1998). *I bilanci delle famiglie italiane nell'anno 1998. Supplementi al Bollettino Statistico*, XVII. Centro Stampa Banca d'Italia, Roma.

Braun H. (1980). A simple method for testing goodness-of-fit in the presence of nuisance parameters. *Journal of the Royal Statistical Society* **42**, 53-63.

Dagum C. (1977). A new model of personal income distribution: specification and estimation. *Economie Appliquée*, **30**(3), 413-437.

Dagum C. (1990). Generation and properties of income distribution functions. In C. Dagum and M. Zenga (Eds.), *Studies in Contemporary economics. Income and wealth distribution, inequality and poverty*, Springer, Berlin.

Domanski, C., Jedrzejczak A. (1998). Maximum likelihood estimation of the Dagum model parameters. *International Advances in Economic Research*, **4**, 243-252.

Kleiber, C., Kotz S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, NJ, USA, Wiley-Interscience.

Mood A., Graybill F., Boes D. (1974). *Introduction to the theory of statistics*. Wiley, New York.

Pollastri A., Zambruno G. (2010) Distribution of the ratio of two independent Dagum random variables. *Operations Research and Decisions*, **3**(20), 95-102.

Yee T. (2019). VGAM: Vector Generalized Linear and Additive Models. R package.