O P E R A T I O N S   R E S E A R C H   A N D   D E C I S I O N S

No. 3–4 2010

Angiola POLLASTRI*
Giovanni ZAMBRUNO*

DISTRIBUTION OF THE RATIO OF TWO INDEPENDENT DAGUM RANDOM VARIABLES

An estimation procedure of the distribution of the ratio of two independent Dagum random variables is proposed. Such an issue is of remarkable importance when analyzing the characteristics of ratios of economic variables which can be described by the Dagum model. The distribution and density functions are computed via numerical procedures; a numerical method is also proposed, in order to make the computation of the distribution easier and faster. Finally, some empirical investigations are reported, in order to establish the effectiveness of the model, and an application is presented concerning the estimation of the distribution of the ratio between the expenditures of a 2-member and a 1-member household, based on the Banca d'Italia 2006 survey.

Keywords: *Dagum distribution, independent Dagum r.v.'s, distribution functions*

1. Introduction

In the paper, some distributional characteristics of the ratio of two independent Dagum random variables (r.v.'s) with three parameters have been analysed. The model proposed by Dagum fulfils many properties considered relevant to a model of income distribution: the specifications of the model exploit the economic framework, convergence to the Pareto law and the economic significance of the parameters. In this paper, the choice of the Dagum model is also supported by the fact that it provides a good fit to both extreme sides of the observed distribution of income in Italy [12]. The model has also been successfully used to describe the distribution of the size of business firms [2].

*Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università di Milano Bicocca, Italia, e-mail addresses: angiola.pollastri@unimib.it, giovanni.zambruno@unimib.it

The subject of this study is the distribution of ratios between two economic variables. Some empirical studies confirm the validity of the model proposed. An application is also considered regarding the estimation of the distribution function of the ratio of r.v.'s describing the expenditures of families with different numbers of members.

2. Theoretical framework

Let X be an r.v. with a type I Dagum distribution [4], [5], [6]. Its cumulative distribution function (cdf) is

$$F_X(x) = \left(1 + \lambda x^{-\delta}\right)^{-\beta} \qquad \lambda, \beta, \delta > 0 .$$

Our purpose is to derive the distribution function of the r.v.

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

defined to be the ratio of two r.v.'s X and Y, which have Dagum distributions with the following parameters:

$$X \sim D(\lambda_1, \beta_1, \delta_1), \qquad Y \sim D(\lambda_2, \beta_2, \delta_2) .$$

We will develop our analysis under the assumption that these two r.v.'s are independent. Following Mood, Graybill and Boes ([13], p. 187), the density function of the variable U is

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

which, according to the assumption of independence and accounting for the support of Y, becomes

$$f_U(u) = \int_0^{+\infty} y f_X(uy) f_Y(y)\, dy .$$

Substituting in the expressions describing the Dagum densities yields

$$f_U(u) = \int_0^{+\infty} y \left(\beta_1 \delta_1 \lambda_1 (uy)^{-\delta_1 - 1} \left(1 + \lambda_1 (uy)^{-\delta_1}\right)^{-\beta_1 - 1}\right)$$
$$\times \left(\beta_2 \delta_2 \lambda_2 (y)^{-\delta_2 - 1} \left(1 + \lambda_2 (y)^{-\delta_2}\right)^{-\beta_2 - 1}\right) dy$$

and the cumulative probability distribution, after some minor rearrangements, takes the form

$$F_U(u) = Pr[U \le u] =$$
$$= \beta_1\beta_2\delta_1\delta_2\lambda_1\lambda_2 \int_0^u t^{-\delta_1 - 1} \int_0^{+\infty} y^{-\delta_1-\delta_2-1}\left(1+\lambda_1 (ty)^{-\delta_1}\right)^{-\beta_1-1}\left(1+\lambda_2 (y)^{-\delta_2}\right)^{-\beta_2-1} dydt.$$

In plain words, the above expression can be read as follows: Ignoring the constant of proportionality, we assign a fixed value t to the r.v. U and sum the probabilities over all the possible values of y and the corresponding $x = ty$ such that their ratio equals t. The outer integral then provides an estimate of the probability that the ratio takes a value not exceeding the upper bound u.

The expression in the inner integral does not match any form reported in currently available tables (e.g. [7]), nor does a change in variable appear to be suitable for reducing it to known forms: therefore, numeric quadrature is advisable. As is well-known, the goodness of an approximation depends largely on the appropriateness of the choice of the points where the integrand must be evaluated: on one hand, the number of points should be kept reasonably low for computational efficiency, on the other hand their location should span the whole range of integration, with a higher "density" around the most significant values.

However, there is a substantial difference between the two integrals which are to be evaluated. The inner integrand has an analytical representation: therefore, one can efficiently use any of the numerical quadrature subroutines available. In this case, we used the QDAGI subroutine from the IMSL library, which is designed for integration over unbounded intervals. In contrast, the outer integral's argument is expressed in a tabular form, although the user has some choice regarding the most effective values of t. For this purpose, we observe that values close to either boundary of the support have an extremely low probability of occurrence (especially in the case of the upper boundary), while we expect that a considerable probability is concentrated around the value $\nu = E(X)/E(Y)$ (of course, this value is only used with the purpose of providing a very rough approximation for the mean of the ratio, whose value we do not know yet).

Therefore, a good choice for the sequence of points

$$\{t_1, t_2, \ldots, t_n\}$$

at which to evaluate the inner integral would be, for instance, to have them fairly densely clustered around the value v and more and more scattered the further we move away from it. One way to produce such a result would be, for instance, to consider a continuous and strictly increasing function defined on (0,1), with vertical asymptotes at both boundaries of the domain and one inflexion point: a good example of such a function is $x: \to g(x) = -(\ln(x))^{-1}$. Then take a suitable sequence of equally spaced values p_i, $i = 1, \dots, n$ in (0,1) with p_1 and p_n as close as possible to 0 and 1, respectively. Finally, define

$$t_i = g(\kappa p_i)$$

where the scale coefficient κ is such that the inflexion point of g corresponds to the value v.

These values are then fed into each inner integral, whose value is in turn an input for the outer integral, which is computed using the standard trapezium rule.

An example of this procedure is presented in the next section.

3. Some applications

The purpose of this section is to illustrate, by way of simple applications based on real data, how the method proposed here performs. Before starting, we ascertain whether the method works in replicating the true distribution of observations: in other words, in this preliminary step we want to avoid any problem related to the estimation of parameters. For this purpose, we consider two Dagum distributions with fixed parameters $\lambda_1 = 0.7$, $\beta_1 = 1.2$, $\delta_1 = 4$ and $\lambda_2 = 0.9$, $\beta_2 = 1.4$, $\delta_2 = 6$, respectively.

We draw at random two series, each of 800 numbers, from the uniform distribution on (0,1). Call these series p_i and q_j ($i, j = 1, \dots, 800$). Then we compute the numbers x_{p_i}, y_{q_j} defined to be the p_i and q_j quantiles of the two Dagum distributions, respectively, i.e.

$$x_{p_i} = \left(\frac{p_i^{-1/\beta_1} - 1}{\lambda_1} \right)^{-1/\delta_1}, \qquad y_{q_j} = \left(\frac{p_j^{-1/\beta_2} - 1}{\lambda_2} \right)^{-1/\delta_2}$$

Next, we form all the possible ratios that can be obtained by taking any x_{p_i} as the numerator and any y_{q_j} as the denominator. In this way, we obtain a series of 800×800 = 640 000 observations from which we can draw a frequency distribution. This is plotted in Fig. 1 together with the distribution obtained by running the procedure presented earlier, with the same set of parameters. The close match is quite apparent.

Then two corresponding cumulative distribution functions were computed (illustrated in Fig. 2). They are very similar as well.

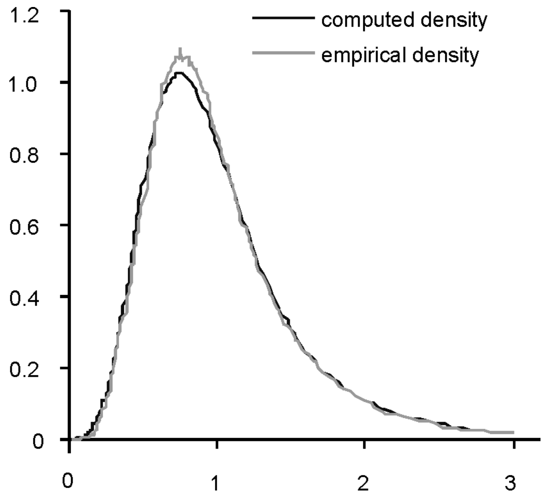

Fig. 1. Empirical and computed density functions of ratios between numbers drawn from Dagum distributions with parameters: $\lambda_1 = 0.7$, $\beta_1 = 1.2$, $\delta_1 = 4$ and $\lambda_2 = 0.9$, $\beta_2 = 1.4$, $\delta_2 = 6$

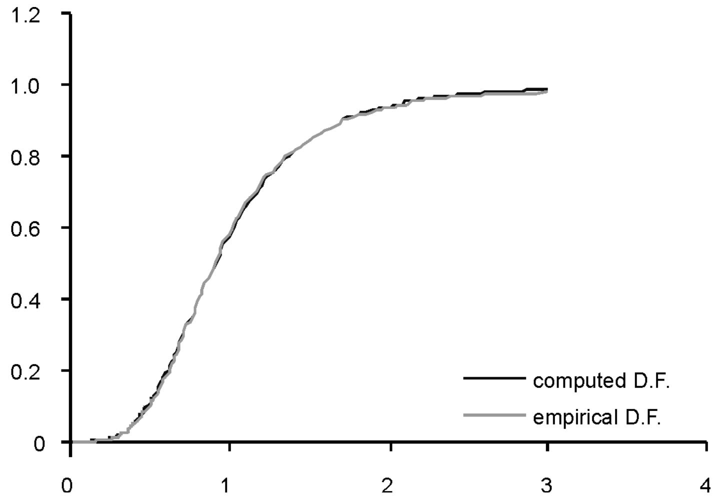

Fig. 2. Empirical and computed cumulative distribution functions of ratios between numbers drawn from Dagum distributions with parameters: $\lambda_1 = 0.7$, $\beta_1 = 1.2$, $\delta_1 = 4$ and $\lambda_2 = 0.9$, $\beta_2 = 1.4$, $\delta_2 = 6$

3.1. The data set

A very important source of microdata about expenditure, income and wealth in Italy is provided by the Banca d'Italia survey, which consists of a series of interviews.

The sampling unit is one household and the survey population is the whole set of households in Italy. In the 2006 survey, the sample size was 7768 households. Each family is randomly drawn in a two stage procedure. In the questionnaire there is also an item regarding the average monthly expenditure on all kinds of consumption. Our goal is to analyze such data. Indeed, for each household size we have collected data on income, whence we can fit a Dagum distribution by estimating the relevant parameters.

Let the r.v. X_r (r = 1, 2, …) describe the income of a household with r members. Assume that each X_r has a Dagum distribution. Therefore, we can estimate the parameters λ_r, β_r, δ_r.

Now consider the r.v. X_r/X_1 as the ratio of two Dagum r.v.'s It is possible to numerically find the distribution function of this r.v. which represents the result of the experiment of selecting at random a household of one member and a household of r members. The r.v. X_r is independent of the r.v. X_1. Subsequently, we generated the distribution of all the possible ratios between the expenditure of a household with r members and the expenditure of a household with one member.

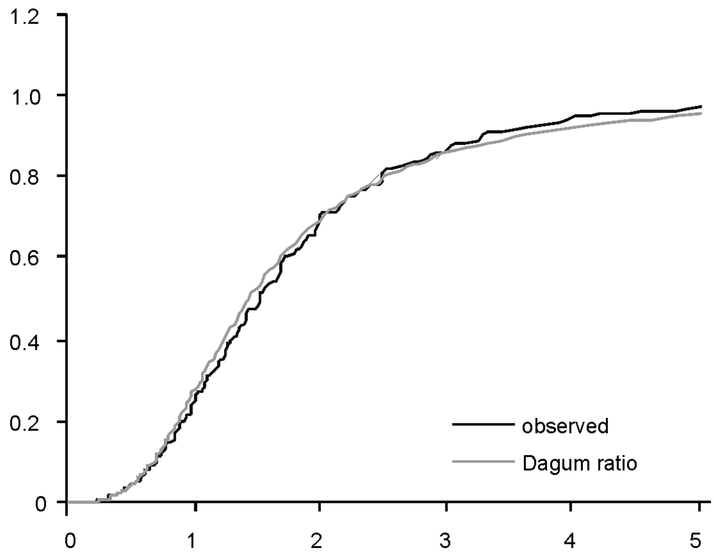

Fig. 3. Cumulative distribution function of X_2/X_1 as the ratio of two Dagum r.v.'s and as estimated from the observed ratios

In the present study, we estimate the distribution function of the ratio of the expenditure of a household with two members divided by the expenditure of a household with one member. In 2006, the sample size of households with 2 members was $n_2 = 2366$ and the sample size of households with a single member was $n_1 = 1327$.

In order to obtain the c.d.f. in question, we estimated the parameters of the Dagum distributions describing the expenditures of households with one member and the expenditures of households with two members. We obtained the estimates of the sets of parameters for the variables X_1 and X_2 using the minimum Chi-Square method (see, e.g., [8]). Numerical computations give the following: $(\hat{\lambda}_N = 0.9, \hat{\beta}_N = 1.24, \hat{\delta}_N = 3.98)$ are the minimum χ^2 estimates of the parameters for the r.v. X_2 and $(\hat{\lambda}_D = 0.93, \hat{\beta}_D = 0.4, \hat{\delta}_D = 5.59)$ the estimates of the parameters for the r.v. X_1.

We estimated the c.d.f. of the variable X_2/X_1 using the method described above. Then, we construct the empirical c.d.f. of the ratios x_{2i}/x_{1j} by taking all the possible ratios between the observed expenditure of a household with two members, denoted by x_{2i} (i = 1, …, 2366), and the observed expenditure of a household with a single component, denoted by x_{1j} (j = 1, …, 1327). The empirical and theoretical c.d.f. are shown in Fig. 3. They overlap quite closely.

Table 1. Deciles for the ratio between the expenditures of households of two and one members

I	1	2	3	4	5	6	7	8	9
Deciles	0.68799	0.88385	1.05890	1.22824	1.44257	1.68913	2.01885	2.52895	3.55029

It is also possible to estimate the deciles. They are given in Table 1. The deciles describe many characteristics of the ratio. For instance, we can establish that the estimate of the median of the ratio between the expenditure of a household with two members and the expenditure of a household with only one member is 1.44257.

4. Conclusions

The present study is the first proposal for a method of estimating the distribution of the ratio of two independent r.v.'s with Dagum distributions. The main purpose is to study the distribution of the ratio of two economic variables, often used in economic indexes.

An efficient numerical integration procedure has been proposed, consisting of a rule constructing the gridpoints adapted specifically to this particular situation. This offered the possibility of estimating the percentiles quite easily, and also comparing the estimated distribution with the empirical distribution computed directly by drawing a sample from actual data.

This technique may prove useful in carrying out inference based on a number of economic and financial indexes, whenever they are defined as ratios of two independent random variables with Dagum distributions.

References

[1] Banca d'Italia, I bilanci delle famiglie italiane nell'anno 2006, *Supplementi al Bollettino Statistico*, XVII, Centro Stampa Banca d'Italia, Roma, 2008.

[2] BISANTE E., FIORI A.M., F*irm size distribution e modello di Dagum: un'indagine empirica sull'industria meccanica italiana*, Working paper n. 181, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università di Milano Bicocca, Milano, 2009.

[3] BURR I.W., *Cumulative frequency functions*, Annals of Mathematical Statistics, 1942, 13, 215–232.

[4] DAGUM C., *A new model for personal income distribution: specification and estimation,* Economie Appliquée, 1977, 30, 413– 437.

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

[6] DANCELLI L., *Tendenza alla massima ed alla minima concentrazione nel modello di distribuzione del reddito personale di Dagum, Scritti in onore di Francesco Brambilla*, Vol. 1, Edizioni di Bocconi Comunicazione, Milano, 1986.

[7] GRADSTHTEYN I.S., RYZHIK I.M., *Table of integrals, series, and products*, Academic Press, Boston, 1994.

[8] KENDALL M.G., STUART A., *The advanced theory of statistics*, C. Griffin and Co., London, 1973.

[9] KOT S.M., T*he estimation of the social welfare functions, Inequality aversion, and equivalence scale*, International Workshop Income Distribution and Welfare, May 30th–June 1st, Università Bocconi, Milano, Italy, 2002.

[10] KOT S.M., *On the estimation and calibration of the social welfare function*, [In:] W. Ostasiewicz (Ed.), *Quality of life research*, Chapter 4, Yang's Scientific Press, Tucson , USA, 2002, 61–71.

[11] LATORRE G., P*roprietà Campionarie del Modello di Dagum per la distribuzione dei redditi*, Statistica, 1988, 48 (1–2), 15–27.

[12] LATORRE G., *Asymptotic distributions of indices of concentration: Empirical Verification and application*, [In:] *Studies in contemporary economics. Income and wealth distribution, inequality and poverty*, C. Dagum, M. Zenga (Eds.), Springer, Berlin, 1989.

[13] MOOD A.M., GRAYBILL F.G., BOES D.C., I*ntroduction to the theory of statistics*, Wiley, New York, 1974.

[14] POLLASTRI A., *Scale di equivalenza tramite l'impiego della distribuzione di Dagum*, Working paper No. 62, Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali, Università di Milano-Bicocca, 2003.

[15] POLLASTRI, A., *Estimation of equivalence scales in Italy based on income distribution*, Statistica et Applicazioni, 2007, 2, 131–140.