

A Stress-Strength Model with Dependent Variables to Measure Household Financial Fragility

Filippo Domma · Sabrina Giordano

Received: date / Accepted: date

Abstract The paper is inspired by the stress-strength models in the reliability literature, in which given the strength (Y) and the stress (X) of a component, its reliability is measured by $P(X < Y)$. In this literature, X and Y are typically modeled as independent. Since in many applications such an assumption might not be realistic, we propose a copula approach in order to take into account the dependence between X and Y . We then apply a copula-based approach to the measurement of household financial fragility. Specifically, we define as financially fragile those households whose yearly consumption (X) is higher than income (Y), so that $P(X > Y)$ is the measure of interest and X and Y are clearly not independent. Modeling income and consumption as non-identically Dagum distributed variables and their dependence by a Frank copula, we show that the proposed method improves the estimation of household financial fragility. Using data from the 2008 wave of the Bank of Italy's Survey on Household Income and Wealth we point out that neglecting the existing dependence in fact overestimates the actual household fragility.

Keywords Reliability · Dagum distribution · Copula · IFM · SHIW data

Mathematics Subject Classification (2000) 60E05 · 62H20 · 91B82

1 Introduction and motivation

The stress-strength term comes from a reliability problem: it describes the life of a component which has a random strength Y and is subject to a random stress X . If the stress exceeds the strength ($X > Y$) the component will fail, while the component works whenever $X < Y$. Thus, $\mathfrak{F} = P(X > Y)$ measures

Dipartimento di Economia e Statistica, Università della Calabria,
via Bucci, cubo 0C, 87036 Arcavacata di Rende (CS), Italy
E-mail: sabrina.giordano@unical.it

the likelihood of failure and $\mathfrak{R} = P(X < Y)$ is a measure of component reliability.

In the literature, stress-strength models refer to statistical methods associated with the evaluation of \mathfrak{R} or \mathfrak{F} . A comprehensive account of this topic is given by Kotz et al. (2003). Unlike the specific literature where \mathfrak{R} plays the leading role, we shall focus, without loss of generality, on the failure measure \mathfrak{F} .

Although the seminal use of the stress-strength models is to be found in problems of physics and engineering, interest in \mathfrak{R} is not confined to these contexts. In fact, it spreads across different disciplines, such as quality control, genetics, psychology, economics and in the last thirty years there have been numerous applications in medical problems and clinical trials (e.g. Gupta and Gupta 1990; Adimari and Chiogna 2006).

There is a well-developed theory for the case in which \mathfrak{R} is calculated under the assumption that X and Y are independent variables. By the end of the 1970s, inference on \mathfrak{R} was carried out for the majority of common distribution families, such as Gaussian, Pareto, Exponential, etc., to cite only the most recent contributes, see Krishnamoorthy and Lin (2010), Rezaei et al. (2010), Sengupta (2011) and Huang et al. (2012).

Yet little attention has been paid to the more realistic problem in which X and Y are dependent variables. Many situations require that X and Y are related in some way; for example, they may be influenced by a common factor, or may represent two measures pre and post treatment on the same subjects and so on.

Recently, estimators of \mathfrak{R} have been proposed assuming several bivariate distribution families for (X, Y) for describing the joint behavior of strength and stress variables, e.g. bivariate beta (Nadarajah 2005), bivariate exponential (Nadarajah and Kotz 2005) and bivariate log-normal (Gupta et al. 2010). Balakrishnan and Lai (2009) evaluated \mathfrak{R} for models in which X and Y are correlated. Nevertheless, the assumption on the bivariate distribution often admits only a certain specific form of dependence between margins and presupposes that both the marginal distributions are of the same type. Bivariate Normal distribution, for instance, restricts the type of association between margins to linear and the margins to be normally distributed, although these restrictions are not always realistic. In fact, empirical evidence of a non-linear form of dependence and non-normality of the data is widely unquestionable.

With the aim of evaluating the role of dependence in the stress-strength models, we calculate the failure measure through a copula-based approach. To the best of our knowledge, until now a similar task has never been attempted by using copula functions (an attempt in this direction is in Domma and Giordano 2009 where the stress and strength variables are assumed to be non-identical Burr III distributed and their dependence is modeled through the Farlie-Gumbel-Morgenstern copula). This approach concerns modeling the bivariate distribution of stress and strength variables with univariate margins belonging to given parametric families and a copula function which summarizes the existing dependence structure.

To better understand the potential of this approach, consider that a copula function joins margins of any type (parametric, semi-parametric and non-parametric distributions) not necessarily belonging to the same family, and captures various forms of dependence (linear, non-linear, tail dependence etc).

The purpose of this work is to provide a contribution to the estimation of reliability measures in all practical problems where both the strength and stress variables are dependent and parametrically distributed. In this paper, X and Y are household consumption and income, respectively. The idea here is to estimate \mathfrak{F} as a measure of household financial fragility which occurs whenever expenses exceed household yearly income. To do this, we apply the Frank copula to model the dependence between margins whose non-identical distributions belong to the Dagum family. We show that neglecting the existing dependence overestimates the fragility measure.

The paper is organized as follows. Sect. 2 examines the issue of household financial fragility and Sect. 3 describes the data. Sect. 4 motivates the choice of Dagum models for the income and consumption distributions (Sect. 4.1) and Frank copula (Sect. 4.2) for the association to provide an estimate of the fragility measure (Sect. 4.3). Final developments are mentioned in Sect. 5.

2 Household financial fragility

The recent crisis has highlighted that understanding the ability of households to offset adverse changes in their financial circumstances is of great relevance for policy makers.

Attention to household financial stress has been rising in recent years. Several authors explore the determinants of measures of financial fragility to identify which households are potentially more vulnerable to unfavorable changes in the economic environment. Despite the increasing contributions, a highly debated, but still untangled point in the literature concerns exactly the definition of household financial fragility.

Brown and Taylor (2008) considered as measure of financial fragility the probability for a household to have negative net worth, that is being in a situation where total debts outweigh financial assets. Jappelli et al. (2008) investigate whether indebtedness is associated with increasing financial fragility measured as the sensitivity of household arrears to adverse shocks, such as unemployment and interest rate increase. They also explore institutional factors that can make households more willing to service scheduled loan payments. Moreover, Christelis et al. (2009) present a cross-country comparison of the capacity of elderly household heads to handle financial distress. They describe household vulnerability in 11 countries of Europe through four indicators: the net worth-income ratio, the proportion of households with low financial wealth, the proportion of households with debts other than housing and the reported difficulties in “making ends meet”. In a recent contribution, Lusardi et al. (2011) measure financial fragility by examining the household ability to come up with an unexpected expense of 2000 dollars in one month regardless of the

source of funds (i.e. savings, borrow from family/friends, traditional access to credit, work more etc). In all the cited studies, the financial fragility measures are compared across countries suggesting that the interest in this issue is not confined to the Italian experience.

In Italy, the recent increasing financial straits for households have been provoked by a drop in income mainly due to a rise in the rate of unemployment, to cuts in state transfers for welfare and increase in taxes in the attempt to reduce the budget deficit and prevent the growth of public debt, and, more generally, to the unsettled economic situation. On the other hand, there has been a rise in consumption mainly imputable to the growth of the financial sector which now offers households a broader range of products, such as consumer credit and loans. In fact, a wide variety of options to defer payments are now available. However, although most households are able to cope with a certain amount of debt, a number of households cannot really afford the loans and other debt they take on. They often do not limit debt payments to low percentages of their gross income.

Financial difficulties limit household borrowing capacity and ability to meet repayments. Moreover, income instability may affect the household liability. This issue has been much explored in the economic literature and the reader can refer for a better understanding of this specific topic to, among others, the recent contribution of Jappelli and Pistaferri (2010).

Our idea is to identify a stress-strength problem in the household budget management: the component is the household having at its disposal a random strength Y , income, which is subject to a stress X , consumption. If the stress exceeds the strength the component fails, so when expenses outpace the disposable income the household has a financial problem and needs to borrow or to decumulate its wealth and becomes vulnerable.

To put it bluntly, when consumption exceeds income, households face financial stress, so $\mathfrak{F} = P(X > Y)$ is directly interpretable as a measure of household financial fragility.

It is worth noting that we define as financial fragile those households whose yearly consumption is greater than yearly income. It may be the case, however, that households are reliant on their savings to support their consumption without difficulties in managing financially. But, our definition of fragility implicitly assumes that the household is also susceptible to financial stress when it needs to access its cumulated savings, as income does not suffice, to cope with expected or unexpected consumption within the year.

We believe it is important to estimate the propensity of households to exceed their yearly financial resources, especially as it is related to household attitude towards indebtedness which sometimes can lead to bankruptcy. It is not our task, however, to discuss the economic or institutional determinants of household financial fragility, our aim is rather to estimate the measure \mathfrak{F} taking into account the dependence underlying income and consumption variables. In the case at hand, in fact, the usual independence assumption of stress-strength variables is evidently untenable and will be relaxed in Section 4.

3 Income-consumption data

Income and consumption of Italian households are drawn from the 2008 wave of the Bank of Italy's Survey on Household Income and Wealth (SHIW08)¹.

The sample comprises 7977 households² (13266 income-earners) over about 350 Italian towns.

The variable Y indicates the net disposable household income obtained in the year 2008 as sum of income coming from payroll employment, pension and net transfers, self-employment and properties of all the income-earners in the household; X denotes the year's durable and non-durable consumption and in detail is obtained by summing up the following year's expenses:

- rental of dwelling or loans (mortgage instalments or personal loans) relating to the principal residence and other properties;
- extraordinary maintenance of dwelling and other properties owned by the household;
- total amount (if fully paid in 2008) or year's amount spent on instalment payments for the purchase of:
 - valuables (jewellery, gold, antiques, etc...);
 - means of transport (cars, motorbikes, caravans, motor boats, boats, bicycles);
 - furniture, furnishings, household appliances, sundry equipment;
 - non-durable goods (holidays, fur coats, etc...);
- fringe benefits³;
- maintenance, allowances, gifts, donations, loans to relatives or friends;
- life insurance premiums;
- contributions to supplementary pension schemes;
- all spending on average in 2008 in cash, by credit card, cheque or bancomat card for food and non-food (except previous items);

and subtracting the following year's received amounts:

- total value of objects (valuables and motor vehicles) sold in 2008;
- total amount deducted (36% or 41%) for renovation costs of all properties owned by the household.

¹ The data are freely available at www.bancaditalia.it/statistiche/indcamp/bilfait/dismicro. We refer the reader to the work of Brandolini and Cannari (1994) containing much detailed information on the survey; see also the supplements to the bulletin of the survey 2008 available from www.bancaitalia.it/statistiche/indcamp/bilfait/docum/ind_08.

² The sampled households actually involved in this study are 7955; as income or/and consumption for few households is negative or null they have been deleted from the original SHIW08 dataset.

³ Fringe benefits are included in both income (i.e. in payroll income) and consumption variables according to the Bank of Italy definition since this form of income is not cash, has an expire date, is usable only for the expenses it is devoted to and cannot be accumulated to increase household savings.

Unlike income Y , the variable X here considered as household consumption differs from that calculated by the Bank of Italy⁴ which also involves the total value of objects purchased by the household even if they have not been paid for in full during 2008. In particular, we focused only on the expenses the household incurred in 2008.

Summary statistics pertaining to income and consumption data are reported in Table 1.

Table 1 Descriptive statistics for Income and Consumption (in euro)

	<i>Income</i>	<i>Consumption</i>
min	65.68066	1200
max	629339.7	278900
median	26790.56	17030
mean	32429.08	20157.44
std error	24333.98	14008.44
skewness	5.24	4.32
kurtosis	78.23	47.82

4 \mathfrak{F} for dependent variables

Dependence between variables is completely described by the joint distribution function which can be specified by modelling the univariate margins and the dependence structure separately. A flexible tool to model different kinds of dependence which has received growing recognition in several applications is the copula function. The monographs by Joe (1997) and Nelsen (2006) give a comprehensive introduction to the most popular copula families and their properties.

A two-dimensional copula is merely a bivariate distribution function with Uniform $(0, 1)$ margins. Sklar's theorem clarifies how copulas join the multivariate distribution functions with their univariate margins. According to Sklar's theorem any bivariate distribution $H(x, y) = P(X \leq x, Y \leq y)$ with continuous margins $F(x)$ and $G(y)$ can be written as $H(x, y) = C(F(x), G(y))$, where C is a unique copula.

The joint density function is denoted by $h(x, y) = c(F(x), G(y)) f(x)g(y)$ where $c(F(x), G(y)) = \frac{\partial^2 C(F(x), G(y))}{\partial F(x) \partial G(y)}$ is the copula density, and $f(x), g(y)$ indicate the marginal density functions.

⁴ The variables used by the Bank of Italy are defined in the legend downloadable from www.bancaditalia.it/statistiche/indcamp/bilfait/docum/ind_08/descrizione_arc/eng_Legen08.pdf.

Consequently, the measure \mathfrak{F} allowing for the dependence between X and Y , with positive values for simplicity, turns out to be

$$\begin{aligned}\mathfrak{F} = P(X > Y) &= \int_0^{+\infty} \int_0^x h(x, y) dy dx \\ &= \int_0^{+\infty} \int_0^x c(F(x), G(y)) f(x)g(y) dy dx.\end{aligned}\quad (1)$$

Hereafter, the argument of functions will be omitted when necessary for the sake of simplicity, so that $F(x), G(y)$ will be denoted in short form by F, G .

Copulas depend on one or more parameters, while the marginal distributions may belong to either non-parametric and parametric families. It is easy to imagine that there is an uncountable variety of possible combinations of copula and marginal distributions. In the parametric approach, we can choose the parametric distributions which best fit the marginal behavior of the two variables and capture the existing dependence through a suitable copula function.

In the rest of the section we concentrate on two steps: 1. to find appropriate marginal parametric distributions (Sect. 4.1) and a copula (Sect. 4.2) which serve the need to suitably model the joint distribution of income-consumption data; 2. to estimate the measure \mathfrak{F} of Italian household fragility accounting dependent income and consumption variables (Sect. 4.3).

4.1 Distributions for margins

The origins of the Dagum (Dagum 1977) model stem from an attempt to overcome the limits of the classical Log-Normal and Pareto distributions by accommodating both the heavy tails and the mode of the empirical income density function. In fact, the Dagum distribution⁵ has been shown to provide an excellent fit to income distribution data. Consequently, the Dagum model has been appreciated by economists and is often preferred to its parametric competitors to model income data as highlighted in several empirical applications (e.g. Azzalini et al. 2003; Bandourian et al. 2003).

For this reason, we assume that the marginal densities of household income and consumption belong to the three-parameter Dagum family.

First, we introduce some notation. The cumulative distribution function of the Dagum variable is

$$F(x; \gamma) = (1 + \lambda x^{-\delta})^{-\beta} \quad (2)$$

which leads to the probability density function

$$f(x; \gamma) = \beta \lambda \delta x^{-\delta-1} (1 + \lambda x^{-\delta})^{-\beta-1} \quad (3)$$

⁵ Kleiber (2008) provides an exhaustive overview on the genesis of the Dagum distribution and its developments in the literature on applied statistics.

where $x \in \mathbb{R}^+$, and $\gamma = (\beta, \lambda, \delta)$, positive. The scale parameter is λ , while β and δ are shape parameters.

Here, consumption and income data are modeled by non-identical Dagum distributions, henceforth denoted by $F(x; \gamma_c)$ and $G(y; \gamma_i)$, where the subscripts c and i clearly refer to consumption and income, respectively.

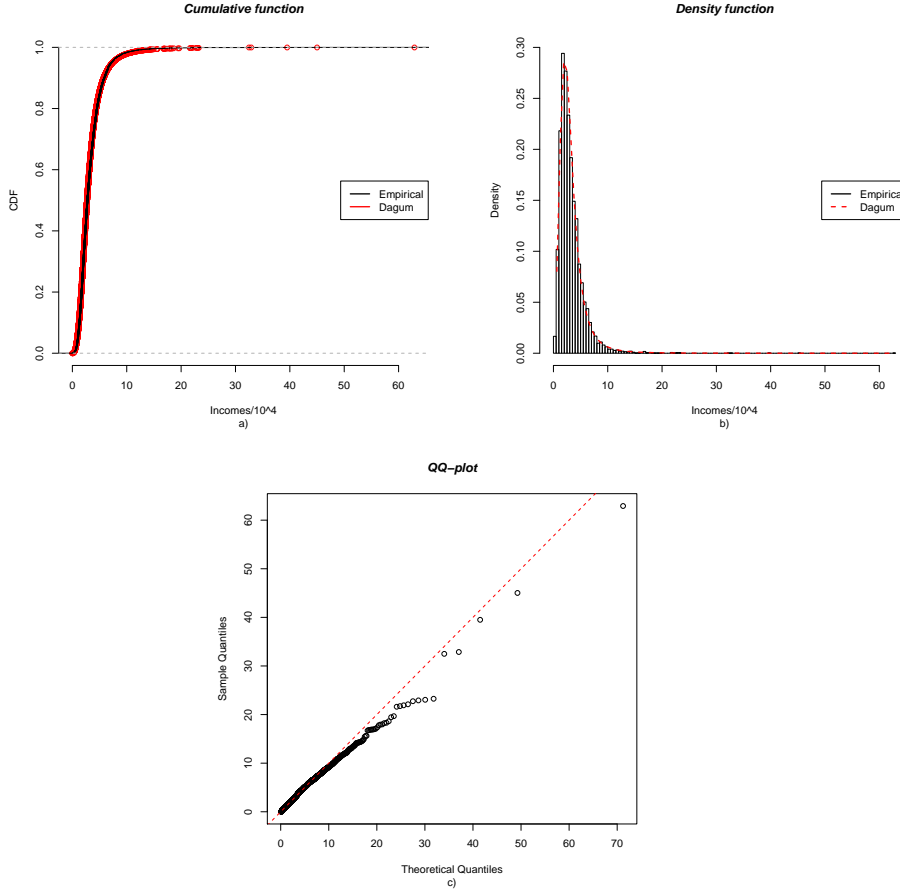


Fig. 1 Graphical goodness-of-fit on income data: (a) Empirical and fitted cumulative distribution functions; (b) Empirical and fitted density functions; (c) QQ-plot

We fit the Dagum distribution to the SHIW08 data. The maximum likelihood estimates (MLE) $\hat{\gamma}_c = (\hat{\beta}_c, \hat{\lambda}_c, \hat{\delta}_c)$ and $\hat{\gamma}_i = (\hat{\beta}_i, \hat{\lambda}_i, \hat{\delta}_i)$ are obtained by numerical methods since the solution of the maximum likelihood equations is not in closed form. An optimization procedure is implemented in the language R (2011).

Table 2 reports the MLE of the parameters of the two marginal distributions, the standard errors (SE), 95% asymptotic confidence intervals and the

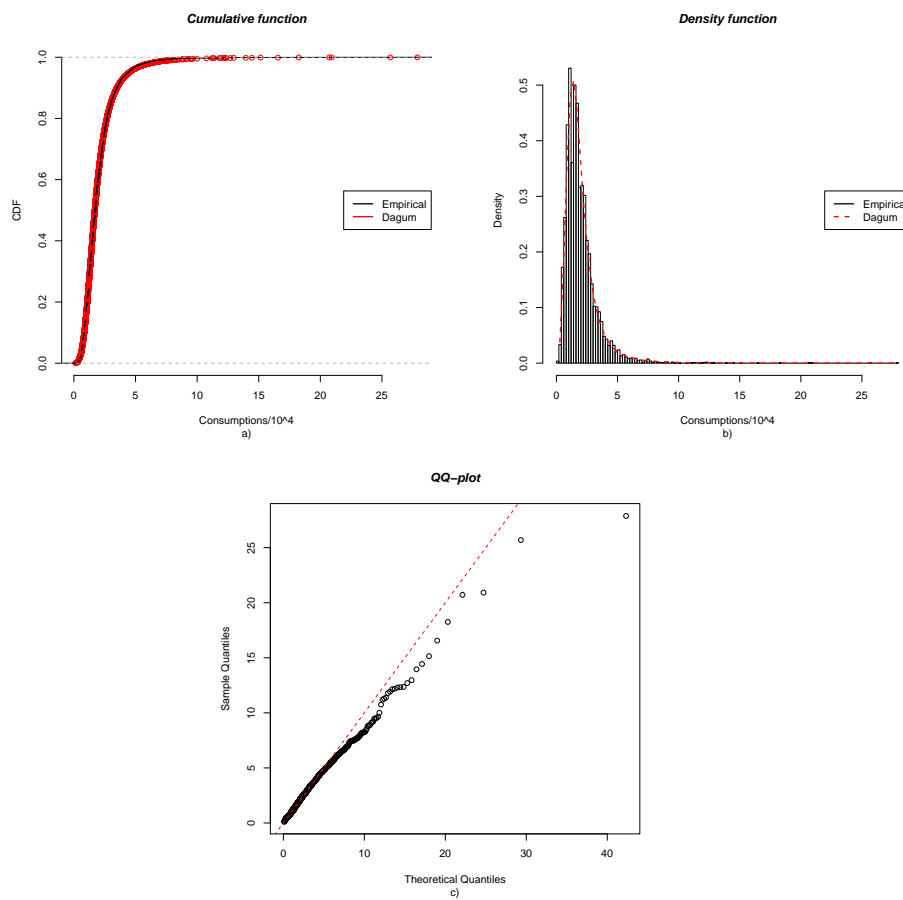


Fig. 2 Graphical goodness-of-fit on consumption data: (a) Empirical and fitted cumulative distribution functions; (b) Empirical and fitted density functions; (c) QQ-plot

maximum value of the log-likelihood functions (l). Moreover, the Anderson-Darling (AD) goodness-of-fit statistic values, in the last column, confirm that the Dagum model is suitable for these data.

We shall now check graphically the adequacy of the Dagum distribution to the analyzed data. The probability plots in Figures 1 and 2 show an excellent goodness-of-fit of the Dagum model. In fact, the fitted cumulative function well approximates the empirical curve in panels (a); the Dagum and the empirical density match appropriately in panels (b) and the points in the QQ-plots (c) lie extremely close to the 45° ray from the origin, except for a few extreme values. Thus, the overall fit of the Dagum distribution is satisfactory.

Table 2 MLE, SE, 95% confidence intervals, max log-likelihood, AD statistic test (p-value)

	γ	<i>MLE</i>	<i>SE</i>	<i>95% CI</i>	<i>l</i>	<i>AD test (p-value)</i>
Consumption	β_c	1.0837	0.0504	0.985–1.182	-11118.46	1.076 (<i>0.3197</i>)
	δ_c	2.9959	0.0543	2.889–3.102		
	λ_c	4.3327	0.4174	3.514–5.150		
Income	β_i	0.8582	0.0351	0.789–0.927	-15405.14	1.338 (<i>0.2204</i>)
	δ_i	2.9752	0.0541	2.869–3.081		
	λ_i	23.8284	2.7952	18.35–29.31		

4.2 Copula

A copula-based modeling of the joint distribution of income-consumption data with Dagum margins is now provided. The next practical step is the choice of an appropriate copula function to model the association of income-consumption data. The coverage of an appropriate dependence range, a graphical approach and a goodness-of-fit test are the three criteria that oriented our choice which we shall now discuss briefly.

The copula functions depend on one or more parameters, say θ , called association parameters, which are related to the degree of dependence between margins. Common measures of the amount of association that exists, such as Kendall's τ , Spearman's ρ and medial correlation \mathfrak{M} among others, are usually expressed as function of the association parameters.

For instance, Kendall's τ , defined as measure of discordance of the two variables, can be written as $\tau_\theta = 4 \int \int_{[0,1]^2} C(F, G; \theta) c(F, G; \theta) dF dG - 1$, while

Spearman's ρ , which assesses how well the relationship between two variables can be described using a monotonic function, is written in terms of copula $\rho_\theta = 12 \int \int_{[0,1]^2} C(F, G; \theta) dF dG - 3$. In particular, unlike Pearson's correlation coefficient, these measures do not depend on parameters of the marginal distributions but on the association parameters only.

It is worth noting that τ and ρ values corresponding to the θ domain do not necessarily cover the whole interval $[-1, 1]$ and the range of dependence that can be really achieved varies for different copulas. The Farlie-Gumbel-Morgenstern family of copulas, for example, allows the association measures to vary in a limited sub-interval only, i.e. $\tau \in [-0.22, 0.22]$ and $\rho \in [-0.33, 0.33]$. Therefore, the choice of an appropriate family of copulas should depend on the range of dependence covered. So, since the empirical values of τ and ρ on SHIW08 income-consumption data are $\tau^E = 0.494$ and $\rho^E = 0.677$ showing a medium-high degree of positive association, a critical step involves choosing one copula from the competitive families whose association parameter lies within a range which allows τ, ρ to cover at least that empirical value, or more generally, the positive dependence domain $(0, 1]$. We restrict our search to a copula which possesses this property.

Moreover, a graphical tool provides us with a rough copula selection strategy: the comparison between the plot of the pairs of fitted margins (\hat{F}_j, \hat{G}_j) , $j = 1, \dots, n$, where $\hat{F}_j = F(x_j; \hat{\gamma}_c)$, $\hat{G}_j = G(y_j; \hat{\gamma}_i)$, and the scatterplot of simulated data⁶ (U_j, V_j) , $j = 1, \dots, n$, from different copulas may orient the choice towards certain copula families whose scatterplot looks like that of the fitted marginal values on the unit square, whereas it excludes other copulas whose representations do not conform with the data. A sampling of scatterplots for some members of the Archimedean class of copulas, found in Nelsen (2006, pp. 120-121), guided our choice.

The third but key criterion which leads to the choice of a copula consists in performing a goodness-of-fit test. When testing the hypothesis that the dependence structure of a bivariate distribution is well represented by a specific copula family, a natural goodness-of-fit test involves quantifying a distance between the empirical copula C_n , which is a non-parametric estimator of the true unknown copula, and the estimated copula $C(\hat{\theta})$. Among the tests based on the empirical copula, we considered the statistic⁷ S_n which performed well in Genest et al. (2009). Large values of this statistic lead to the rejection of the hypothesized copula family. Approximated p-values are obtained by a parametric bootstrap-based procedure even if it is computationally demanding.

As candidate families of copulas we considered the widely applied class of the one-parameter Archimedean copulas. Its popularity is due to the fact that these copulas are easily constructed, possess nice properties, are mathematically tractable and many packages provide implemented procedures to fit them to data.

The Frank copula family is chosen on the above mentioned criteria. The alternative families we considered among the most common Archimedean copulas do not satisfy all the three criteria except the copula belonging to the Frank family. Some of the selected copulas (e.g. Ali-Mikhail-Haq and Farlie-Gumbel-Morgenstern) are discarded as they do not cover an appropriate range for the dependence measures, for other copulas (e.g. Clayton) the proposed scatterplot remarkably differs from the plot of fitted margins, and the hypothesis that the dependence structure between the margins is well-represented by a specific parametric family of copulas is accepted only for the Frank family.

The Frank family allows a range of both negative and positive dependence, shows a good fitting to the data ($S_n = 0.019, p\text{-value} = 0.2$) and is also in accordance with the graphical check. In this respect, as can be seen in Figure 3, the scatterplot of data simulated from a Frank copula with $\hat{\theta} = 5.489$ (i.e. the estimated value for θ as shown below) and that of the pairs of Dagum margins seem very similar confirming that the Frank copula can give an adequate description of the examined data. Other commonly used Archimedean copu-

⁶ Joe, 1997, page 146 illustrates a method for simulating random pairs from a copula C : if U, Q are independent random $U(0, 1)$ variables, then $(U, V) = (U, C_{2|1}^{-1}(Q|U))$ has distribution C , where the conditional distribution is $C_{2|1}(v|u) = \frac{dC(u,v)}{du}$.

⁷ The reader interested in details on S_n and the parametric bootstrap procedure to calculate p-values refers to the cited paper.

las, such as Clayton and Gumbel, exhibit greater dependence in the positive tail than in the negative, whereas the Frank copula better fits the symmetric dependence structure in the income-consumption data (see scatterplots in Nelsen 2006).

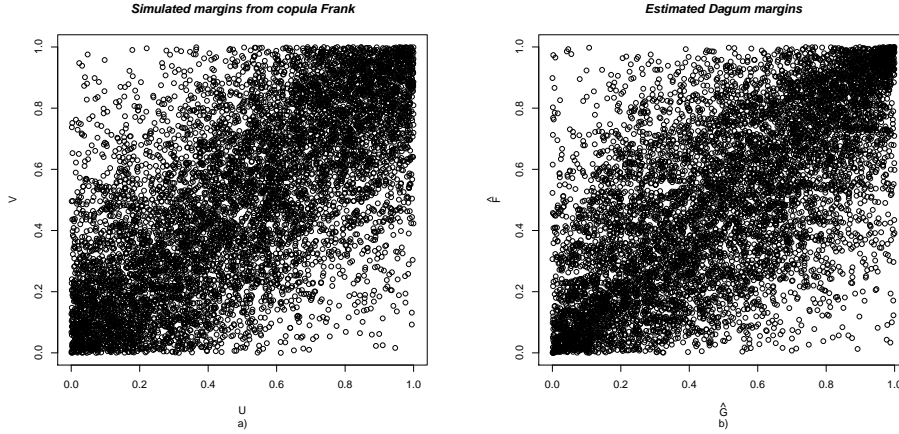


Fig. 3 Scatterplots: (a) U_j, V_j are simulated from a Frank copula with $\hat{\theta} = 5.489$; (b) $\hat{F}_j = F(x_j; \hat{\gamma}_c)$, $\hat{G}_j = G(y_j; \hat{\gamma}_i)$ are the estimated Dagum cumulative functions on income-consumption data

Next step is the estimation of the association parameter.

A basic notation is needed: the functional form of the Frank copula is

$$C(F, G; \theta) = -\theta^{-1} \log \left\{ 1 - \frac{(1 - e^{-\theta F})(1 - e^{-\theta G})}{1 - e^{-\theta}} \right\}$$

and the copula density is given by

$$c(F, G; \theta) = \frac{\theta(1 - e^{-\theta})e^{-\theta(F+G)}}{[1 - e^{-\theta} - (1 - e^{-\theta F})(1 - e^{-\theta G})]^2} \quad (4)$$

where $\theta \in \mathbb{R} \setminus \{0\}$.

The estimation procedure employed is the IFM (Inference Functions for Margins), see e.g. Joe (1997), that provides estimates of the association parameter given the estimated marginal parameters. It consists of two steps: the parameters of the marginal distributions are estimated separately in the first step and then, given these, the procedure calculates the estimate of the association parameter of the copula function. Here, this means that the maximum likelihood estimates $\hat{\gamma}_c$ and $\hat{\gamma}_i$ of the Dagum distributions for consumption and income margins are provided in the first step. They are then plugged into the log-likelihood function

$$l(\theta) = \sum_{j=1}^n \log [c(F(x_j; \hat{\gamma}_c), G(y_j; \hat{\gamma}_i); \theta)]$$

which is maximized with respect to θ , $c(\cdot)$ being the Frank density copula (4) and $F(x_j; \hat{\gamma}_c), G(y_j; \hat{\gamma}_i)$ are the estimated Dagum cumulative functions (2). Joe (2005) showed that the traditional asymptotic properties of the maximum likelihood estimates still hold for the IFM estimates.

The starting value $\theta_0 = 5.62$ used in the maximization procedure is obtained by the inversion of Kendall's τ , a method-of-moment estimate for one-parameter copulas; the resulting MLE on income-consumption data is $\hat{\theta} = 5.489$, $SE = 0.084$, $l = 2395.362$, and the estimated Kendall's τ , $\tau(\hat{\theta}) = 0.486$, and Spearman's ρ , $\rho(\hat{\theta}) = 0.678$, well approximate the empirical values $\tau^E = 0.494$ and $\rho^E = 0.677$.

4.3 Estimating \mathfrak{F} with Dagum margins and Frank copula

We shall proceed to estimate the measure of Italian household fragility \mathfrak{F} under the assumption that income and consumption are non-identically Dagum distributed variables and their dependence is modeled using a Frank copula function.

Therefore, substituting the Dagum and Frank copula densities (3) and (4) in the expression (1), \mathfrak{F} turns out to be

$$\mathfrak{F} = a \int_0^{+\infty} \int_0^x \frac{e^{-\theta(F+G)} (1 + \lambda_c t^{-\delta_c})^{-\beta_c-1} (1 + \lambda_i y^{-\delta_i})^{-\beta_i-1}}{t^{\delta_c+1} y^{\delta_i+1} [1 - e^{-\theta} - (1 - e^{-\theta F})(1 - e^{-\theta G})]^2} dy dt \quad (5)$$

where $a = \theta(1 - e^{-\theta})\beta_i\lambda_i\delta_i\beta_c\lambda_c\delta_c$, and the marginal cumulative functions $F = F(x; \gamma_c), G = G(y; \gamma_i)$ defined in (2) are represented symbolically for the sake of simplicity.

Thus, \mathfrak{F} depends on both marginal and association parameters, $\mathfrak{F}(\gamma_c, \gamma_i, \theta)$. Then, from the invariance property of MLEs, computing $\hat{\mathfrak{F}}$ as function of all the estimated parameters $\mathfrak{F}(\hat{\gamma}_c, \hat{\gamma}_i, \hat{\theta})$ provides the MLE of \mathfrak{F} . We solve the integral (5) numerically by using the Mathematica software as it is unlikely that a solution in closed form can be found. This provided the estimated value $\hat{\mathfrak{F}} = 0.172$, whereas the empirical value of \mathfrak{F} evaluated for the surveyed households in 2008 is

$$\mathfrak{F}^E = \frac{\text{n. of households with (consumption - income) > 0}}{\text{n. of surveyed households}} = 0.156 \quad (6)$$

i.e. the yearly consumption of 15 households out of 100 is beyond their current income.

The estimated $\hat{\mathfrak{F}}$ accounting for the existing dependence in the data seems to well approximate the empirical value.

At this point it is helpful to find out whether it is worth taking into account the dependence between variables and, to do this, we calculate the value of \mathfrak{F} in case of the consumption and income variables are assumed independent and compare it with $\hat{\mathfrak{F}}$ and \mathfrak{F}^E .

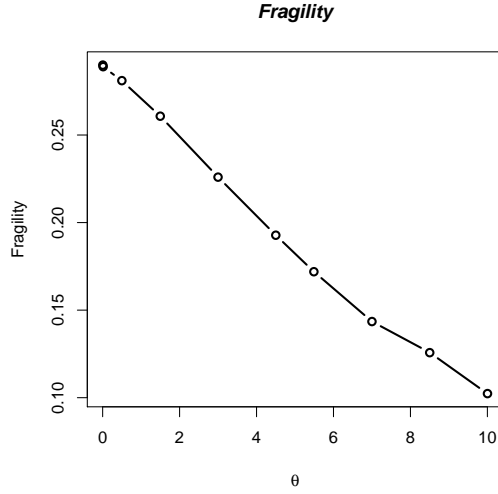


Fig. 4 \mathfrak{F} for $\theta = 0.001, 0.01, 0.5, 1.5, 3, 4.5, 5.489, 7, 8.5, 10$

For $\theta \rightarrow 0$ in the Frank copula, the marginal variables are independent. So, values of θ less than $\hat{\theta}$ and approaching to zero reveal departure from the actual presence of dependence in the data towards the independence assumption. In particular, decreasing θ leads to increasing values for \mathfrak{F} (e.g. \mathfrak{F} evaluated at $\theta = 0.001$ is 0.29) and Figure 4 shows this behavior for some values of the association parameter, for example $\theta \in (0, 10]$. Hence, neglecting the existing dependence between income and consumption actually overestimates household financial fragility.

Consideration of the dependence between margins enables us to evaluate \mathfrak{F} more accurately; in fact the estimated $\hat{\mathfrak{F}}$ is closer to the empirical \mathfrak{F}^E than whatever \mathfrak{F} under assumptions approaching to independence.

The change in the estimation of \mathfrak{F} due to the contribution of the dependence between X and Y is remarkable and leads to the obvious conclusion that this feature of the data cannot be omitted.

As final comment, a referee suggested stressing that in our assumption of financial fragile household there is no role for savings, but extraordinary expenses are typically financed by pre-existing household savings, especially in Italy. For this reason, we considered extraordinary the expenses for means of transport, furniture, furnishings, appliances and for maintenance of dwelling, then we defined another consumption variable by dropping out all the costs for them and estimated the financial fragility under this new specification. The results are robust and confirm the initial conclusion, in fact the fragility measure estimated taking the dependence between income and the new consumption into account is $\hat{\mathfrak{F}} = 0.1276$, while the empirical value is $\mathfrak{F}^E = 0.1143$ and $\mathfrak{F} = 0.2416$ under the independence assumption (a value of θ close to zero is considered).

5 Further developments

Since the independence between the household income and consumption variables is not the most appropriate assumption to work with, we have concentrated on estimating a measure of household fragility considering the relationship between the margins.

The proposed copula-based approach has been both promising and appealing, but more needs to be done. One clear extension of this work, according to the financial fragility literature, would be to finetune the estimation procedure to account economic and individual variables affecting household fragility. In fact, household consumption, in general, depends on individual (including wealth), institutional and market characteristics and these covariates should be considered. This detailed information on the surveyed households is collected in the SHIW08 dataset.

Another aspect of interest is the temporal trend of household fragility. As income and consumption of Italian households have been available since the 1970s from the Bank of Italy, we are stimulated to investigate the evolution of the measure of fragility over the years.

All these aspects are still open and in need of an in-depth study.

Acknowledgements The authors gratefully acknowledge Vincenzo Scoppa for his helpful comments pertaining to the economic aspects of the paper and also thank the anonymous referee for his/her valuable suggestions.

References

1. Adimari G, Chiogna M (2006) Partially parametric interval estimation of $Pr(Y > X)$. *Comput Stat Data Anal* 51:1875–1891.
2. Azzalini A, Dal Cappello T, Kotz S (2003) Log-skew-normal and log-skew-t distributions as models for family income data. *J Income Distrib* 11:13–21.
3. Balakrishnan N, Lai CD (2009) *Continuous Bivariate Distributions* (2nd edn). Springer, New York.
4. Bandourian R, McDonald JB, Turley RS (2003) A comparison of parametric models of income distribution across countries and over time. *Estadistica* 55:135–152.
5. Brandolini A, Cannari L (1994) Methodological Appendix: The Bank of Italy's Survey of Household Income and Wealth. In: Ando A, Guiso L, Visco I eds. *Saving and the accumulation of wealth: Essays on Italian household and government saving behavior*, Cambridge University Press, Cambridge pp 369–386.
6. Brown S, Taylor K (2008) Household debt and financial assets: evidence from Germany, Great Britain and the USA. *J R Stat Soc, Ser A*, 171(3):615–643.
7. Christelis D, Jappelli T, Paccagnella O, Weber G (2009) Income, wealth and financial fragility in Europe. *Journal of European Social Policy*, 19:359–377.
8. Dagum C (1977) A new model of personal income distribution: specification and estimation. *Economie Appliquée* XXX 413–437.
9. Domma F, Giordano S (2009) Dependence in stress-strength models with FGM copula and Burr III margins. In: Ingrassia S, Rocci R (ed) *Book of Short Papers of the Meeting of the Classification and Data Analysis*, CLEUP, Padova pp 477–480.
10. Genest C, Remillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: a review and a power study. *Insur: Math Econ* 44:199–213.
11. Gupta RD, Gupta RC (1990) Estimation of $P(a'x > b'y)$ in the multivariate normal case. *Statistics* 1:91–97.

12. Gupta RC, Ghitany ME, Al-Mutairi DK (2012) Estimation of reliability from a bivariate log-normal data. Forthcoming in *J Stat Comput Simulation*.
13. Huang K, Mi J, Wang Z (2012) Inference about reliability parameter with gamma strength and stress. *J Stat Plan Inference* 142(4):848-854.
14. Jappelli T., Pagano M., Di Maggio M. (2008) Households' indebtedness and financial fragility. CSEF Working Papers, N. 208.
15. Jappelli T, Pistaferri L (2010) The consumption response to income changes. *Annu Rev Econ* 2:479-506.
16. Joe H (1997) *Multivariate Models and Dependence Concepts*, Monographs in Statistics and Probability 73. Chapman and Hall, New York.
17. Joe H (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *J Multivar Anal* 94:401-419.
18. Kleiber C (2008) A guide to the Dagum distribution. In: C. Duangkamon (ed) *Modeling Income Distributions and Lorenz Curves Series: Economic Studies in Inequality, Social Exclusion and Well-Being* 5, Springer, New York.
19. Kotz S, Lumelskii Y, Pensky M (2003) *The Stress-Strength Models and its Generalizations*. World Scientific Publishing, Singapore.
20. Krishnamoorthy K, Lin Y (2010) Confidence limits for stress-strength reliability involving Weibull models. *J Stat Plann Inference* 140:1754-1764.
21. Lusardi A, Schneider D, Tufano P (2011) Financially fragile households: evidence and implications. CeRP Working Papers, N. 116.
22. Nadarajah S (2005) Reliability for some bivariate beta distributions, *Mathematical Problems in Engineering*, 2005, 2, 101-111, DOI:10.1155/MPE.2005.101.
23. Nadarajah S, Kotz S (2005) Reliability for some exponential distributions, *Math Probl Eng*, 2006:1-14.
24. Nelsen RB (2006) *An Introduction to Copulas*. Springer-Verlag, New York.
25. R Development Core Team (2011) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
26. Rezaei S, Tahmasbi R, Mahmoodi M (2010) Estimation of $P(Y < X)$ for generalized Pareto distribution. *Journal of Statistical Planning and Inference*, 140, 480-494.
27. Sengupta S (2011) Unbiased estimation of $P(X > Y)$ for two-parameter exponential populations using order statistics. *Statistics* 45(2):179-188.