Journal of Data Science 13(2015), 771-794

On the classical estimation of bivariate copula-based Seemingly unrelated tobit models through the proposed inference function for augmented margins method

Francisco Louzada*, Paulo H. Ferreira

SME-ICMC, Universidade de Sao Paulo, Sao Carlos, SP, Brazil
PPGEs, Universidade Federal de Sao Carlos, Sao Carlos, SP, Brazil

Abstract: This paper extends the analysis of the bivariate Seemingly Unrelated (SUR) Tobit by modeling its nonlinear dependence structure through copula and assuming non-normal marginal error distributions. For model estimation, the use of copula methods enables the use of the (classical) Inference Function for Margins (IFM) method by Joe and Xu (1996), which is more computationally attractive (feasible) than the full maximum likelihood approach. However, our simulation study shows that the IFM method provides a biased estimate of the copula parameter in the presence of censored observations in both margins. In order to obtain an unbiased estimate of the copula association parameter, we propose/develop a modified version of the IFM method, which we refer to as Inference Function for Augmented Margins (IFAM). Since the usual asymptotic approach, that is the computation of the asymptotic covariance matrix of the parameter estimates, is troublesome, we propose the use of resampling procedures (bootstrap methods) to obtain confidence intervals for the copula-based SUR Tobit model parameters. The satisfactory results from the simulation and empirical studies indicate the adequate performance of our proposed model and methods. We illustrate our procedure using bivariate data on consumption of salad dressings and lettuce by U.S. individuals.

Keywords: bootstrap confidence interval, Clayton copula, data augmentation, logistic distribution

* Corresponding author. Tel.: +55 16 3373 6614. Email: louzada@icmc.usp.br.

1. Introduction

Tobit model refers to a class of regression models whose range of the dependent (or response) variable is somehow constrained. It was first proposed by Tobin (1958) to describe the relationship between a non-negative dependent variable y (the ratio of total durable goods expenditure to total disposable income, per household) and a vector of independent variables x (the age of the household head, and the ratio of liquid asset holdings to total disposable income). Tobin named his model the model of limited dependent variables; however, it and its several generalizations are popularly known among economists as Tobit models, a phrase coined by Goldberger (1964) because of similarities to probit models (the term Tobit aims to synthesize in one word the concept “Tobin’s probit”). Tobit models are also called censored or truncated regression models.

Particularly, censoring occurs when data on the dependent variable is lost (or limited). For example, antibody concentration values in Haitian 12-month-old infants vaccinated against measles are determined through neutralization antibody assays with the lower detection limit of 0.1 IU (Moulton and Halsey, 1995). Therefore, concentration values under or equal to 0.1 are reported as 0.1 (left-censoring).

The censoring problem also arises in situations with the presence of multiple dependent variables. For example, Chen and Zhou (2011) consider the joint problem of censoring and simultaneity when working with multivariate microeconomic data.

The next subsection describes a real data set that will be used to illustrate the approach proposed in this paper. It exhibits these two major characteristics: (left-) censoring and multiple (two) correlated dependent variables.

1.1 U.S. salad dressings and lettuce consumption data

The United States is the second largest lettuce-producing country after China. Total lettuce consumption, i.e. consumption of all lettuce varieties by Americans reached a high record of 34.5 pounds per capita in 2004 (Boriss and Brunke, 2005). Nevertheless, as discussed in Mintel’s ”Bagged Salad and Salad Dressings - U.S., July 2008”, salad dressings sales have declined since 2005. Among other reasons, it is due to the fact that health-oriented consumers who eat large amounts of lettuce and other vegetables are curtailing consumption of salad dressings perceived as high in fat, calories and sodium.

The empirical application of this paper aims at establishing some factors (age, region/location and income, among others) that influence the consumption of lettuce (including all plain, Boston and Romaine lettuce reported separately or as part of a mixed salad or sandwich) and salad dressings products (including mayonnaisetype salad dressing reported separately or as part of a sandwich, and pourable salad dressings reported separately or as part of a mixture such as a salad) by U.S. individuals. Our study is based on part of a data set extracted from the 1994-1996 Continuing Survey of Food Intakes by Individuals (CSFII) (USDA, 2000), hereafter salad dressings and lettuce data. In the CSFII, two nonconsecutive days of dietary data for individuals of all ages residing in the United States were collected through in-person interviews using 24-

hours recall. Each sample person reported the amount of each food item consumed. Where two days were reported there is also a third record containing daily averages. Socioeconomic and demographic data for the sample households and their members were also collected in the CSFII. The size of the extracted/selected sample here is n = 285 adults (we only consider one member per household) age 20 or older.

In Figure 1 the histograms, and in Figure 2 the scatter plot exhibit some features of the data and model we work on: both dependent variables (salad dressings and lettuce consumption) are limited (left-censored or lower-bounded by zero, since there are some individuals in the selected sample who did not consume salad dressings and/or lettuce during the two-day period), the assumption of normality of marginal errors, or equivalently, the assumption of left-censored normal distribution of the observed dependent variables is not a reasonable one to make (both distributions seem to have a right-tail heavier than the normal tail), and there is considerable positive association between salad dressings and lettuce consumption data (the Kendall tau rank correlation coefficient between them is 0.3914). All these features, along with the presence of covariates (age, region and income, among others), suggest that the relationship between the reported salad dressings and lettuce consumption could be modeled through a bivariate regression model with limited (left-censored at zero) dependent variables, and non-normal and heavy-tailed distribution (e.g., logistic distribution) for the marginal error terms.

1.2 Literature review

The multivariate Tobit models, which generalize the univariate Tobit models to systems of equations, is a class of models capable of accommodating the abovementioned issues. There are several generalizations available in the literature, each designed to uniquely capture characteristics of each particular application (see Lee (1993) for a survey). In this paper, we consider the Seemingly Unrelated Regression (SUR) Tobit model, which is a SUR-type model where all dependent variables are partially observed or censored (see, e.g., Zellner (1962), Greene (2003), and Zellner and Ando (2010) for more details on the SUR model, and Amemiya (1984) for a thorough review of various types of Tobit models).

Several estimation techniques have been proposed to implement the SUR To-

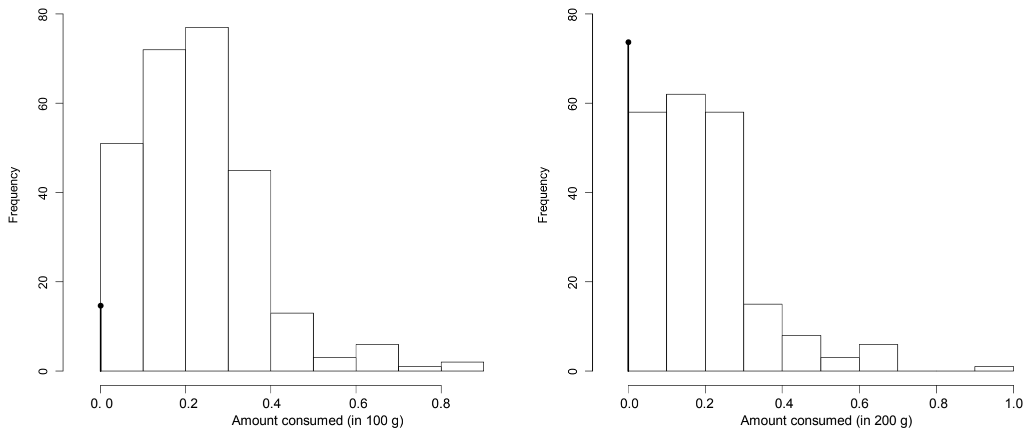


Figure 1: Distributions of the salad dressings (left panel) and lettuce (right panel) consumption. The vertical line at zero on x axis represents adults that did not consume salad dressings or lettuce during the survey period.

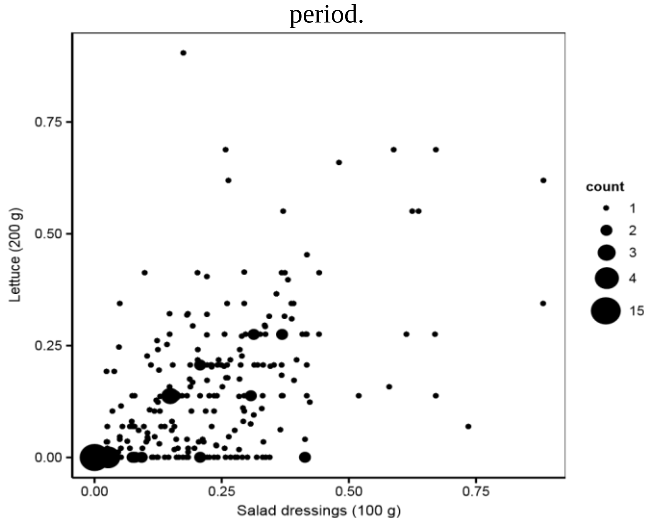


Figure 2: Scatter plot of salad dressings against lettuce consumption. The bold ball sizes are related to the number of pair of data with the same dependent variable values.

bit model. See, e.g., Wales and Woodland (1983), Brown and Lankford (1992), and Kamakura and Wedel (2001) for the maximum likelihood (ML) estimation; Huang et al. (1987) for the expectation-maximization; Meng and Rubin (1996) for the expectation-conditional maximization (ECM); and Huang (1999) for the Monte Carlo ECM (MCECM). Moreover, Huang (2001), Baranchuk and Chib (2008), and Taylor and Phaneuf (2009) implement the SUR Tobit model through Bayesian approach using Gibbs samplers, while Chen and Zhou (2011) estimate the model parameters in the semiparametric context. However, all these estimation methods are cumbersome (i.e. computationally demanding and difficult to implement), especially for high dimensions. Trivedi and Zimmer (2005) suggest this as a reason why the SUR Tobit model is not well applied. These methods also assume normal marginal error distributions, which may be inappropriate in many real applications (like the salad dressings and lettuce consumption dataset described in Section 1.1). In addition, modeling the dependence structure of the SUR Tobit model through the multivariate normal distribution is restricted to the linear relationship among marginal distributions through the correlation coefficients.

In order to relax the assumptions on the same normally-distributed margins and their linear dependence structure, we can employ copulas to analyze the SUR Tobit model (Wichitaksorn et al., 2012). According to Sklar (1959)'s theorem, copulas are used to model the nonlinear dependence structure of the margins that can follow any arbitrary distributions. For further details on copulas, see, e.g., Joe (1997), McNeil et al. (2005, Chapter 5), and Nelsen (2006). The copulas have been successfully applied in many financial and economic applications with continuous and discrete margins (Pitt et al., 2006; Smith and Khaled, 2012; Panagiotelis et al., 2012). Nevertheless, the case of censored (or semi-continuous) margins has not been widely studied and applied, as pointed out by Wichitaksorn et al. (2012). Moreover, the tail coefficients, especially the lower tail coefficient from some copulas can reveal the dependence at the tails where some data are censored. Trivedi and Zimmer (2005) implement the bivariate SUR Tobit model through a few copulas (Clayton, Frank, Gaussian and Farlie-Gumbel-Morgenstern) to model the U.S. out-of-pocket and nonout-of-pocket medical expenses data, finding that the two-stage ML/Inference Function for Margins (IFM) estimation results are unstable. This is not surprising considering the previous findings about the inconsistency of ML estimators of the parameters of the Tobit model with non-normal errors (Cameron and Trivedi, 2005). Yen and Lin (2008) estimate the copula-based censored equation system (a system of meat products - beef, pork, poultry and fish - consumed by U.S. individuals) via the quasi-ML estimation method, yet considering the Frank copula with generalized log-Burr margins (the generalized log-Burr distribution nests the logistic distribution, which is kin to the normal distribution) exclusively. Finally, Wichitaksorn et al. (2012) apply and combine the data augmentation techniques by Geweke (1991), Chib (1992), Chib and Greenberg (1998), Pitt et al. (2006), and Smith and Khaled (2012) to simulate the unobserved marginal dependent variables and proceed with the bivariate copula-based SUR Tobit models implementation through Bayesian Markov Chain Monte Carlo methods as in other copula models with continuous margins. In their work, the relationship between the self-reported out-of-pocket and non-out-ofpocket medical expenses of elderly Americans, as well as the relationship between the wage earnings income of household head and members living in the rural households in Thailand, are described by bivariate SUR Tobit models with Student-t margins through four different copulas (Gaussian, Student-t, Frank and Clayton).

1.3 Objectives and overview

In this paper, inspired by the (Bayesian) work of Wichitaksorn et al. (2012), we propose/develop a modified version of the (classical) IFM method by Joe and Xu (1996), hereafter Inference Function for Augmented Margins (IFAM) method, to implement the bivariate SUR Tobit model with arbitrary margins through copulas. For the present study, we consider only the Clayton copula, as well as logistic margins (i.e. logistic distribution for the marginal error terms). These choices were directed mainly by the data set features detected in Section 1.1, which indicate that the relationship between the reported salad dressings and lettuce consumption, in the presence of covariates (age, region and income, among others), could be modeled through the bivariate SUR Tobit model with logistic margins based on the Clayton copula.

Note from Figure 2 that there is a high number of zero observations of the lettuce variable, which corresponds to zero observations of the salad dressings variable (15 pairs of zero observations). This seems to indicate the strongest relationship between the two dependent variables/margins in their lower regions (i.e. for low or no consumption of salad dressings and lettuce), where data are most concentrated. Therefore, the use of the Clayton copula is justified in order to accommodate the possible existence of lower tail dependence, as well as positive nonlinear dependence structure.

In short, the IFAM method proposed here employs a (frequentist) data augmentation technique at the second stage of the IFM method (the IFM method provides a biased estimate of the Clayton copula parameter, as will be seen in Section 4.2) to generate the censored observations/margins and thus obtain a better (unbiased) estimate of the copula association parameter. This modification also aims to satisfy the Sklar (1959)'s theorem, which states that marginal distributions should be continuous to ensure the uniqueness of the resulting copula. Since the usual asymptotic approximation, that is the computation of the asymptotic covariance matrix of the parameter estimates, is troublesome, we consider resampling procedures (a parametric resampling plan) to obtain confidence intervals for the model parameters. More specifically, we apply the standard normal and percentile methods by Efron and Tibshirani (1993) to build bootstrap confidence intervals.

This paper is organized as follows. In Section 2 we present the bivariate copulabased SUR Tobit model, describing in particular the bivariate Clayton copula-based SUR Tobit model with logistic margins. Section 3 shows the model implementation through the proposed IFAM method (Section 3.1), and the confidence intervals construction using bootstrap approach (Section 3.2). Section 4 presents the simulation study. Section 5 provides the empirical application. Some final remarks and a few indications for further studies in Section 6 conclude the paper.

2. Bivariate copula-based SUR Tobit model

The bivariate SUR Tobit model is expressed as

$$y^*_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + \sigma_j\epsilon_{ij},$$

$$y_{ij} = \begin{cases} y^*_{ij} & \text{if } y^*_{ij} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1,\dots,n$ and $j = 1,2$, where n is the number of observations, y^*_{ij} is the latent (i.e. unobserved) dependent variable of margin j, y_{ij} is the observed dependent variable of margin j (which is defined to be equal to the latent dependent variable y_{ij^*} whenever y_{ij^*} is above zero and zero otherwise), x_{ij} is the $k \times 1$ vector of covariates, β_j is the $k \times 1$ vector of regression coefficients, σ_j is the scale parameter of margin j and $_{ij}$ is the margin j's error that follows some standard distribution (if $\epsilon_{ij} \sim N(0,1)$), then we have marginal standard Tobit models or Type I Tobit models; see Amemiya(1984)).

Given the error $_{ij}$ follows a standard logistic distribution, i.e. $\epsilon_{ij} \sim Logistic(0,1)$, the density function of y_{ij} is

$$f_j\left(y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\beta}_j,\sigma_j\right) = \begin{cases} 1 - G\left(\frac{\boldsymbol{x}'_{ij}\boldsymbol{\beta}_j}{\sigma_j}\right) & \text{if } y_{ij} = 0, \\ \frac{1}{\sigma_j} g\left(\frac{y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j}{\sigma_j}\right) & \text{if } y_{ij} > 0, \end{cases} \tag{1}$$

where $g(z) = e^z/(1 + e^z)^2$ and $G(z) = 1/(1 + e^{-z})$ are the *Logistic*(0,1) probability density function (p.d.f.) and cumulative distribution function (c.d.f.), respectively. The corresponding distribution function of y_{ij} is denoted by $F_j(y_{ij}|x_{ij},\beta_j,\sigma_j)$ and is obtained by replacing $g(.)$ with $G(.)$ and removing $1/\sigma_j$ from the second part of (1) (where $y_{ij} > 0$).

The dependence between ϵ_{i1} and ϵ_{i2} is usually modeled through a bivariate distribution, e.g. the classical bivariate logistic distribution as defined by Gumbel (1961) (this specification characterizes what we call here the basic bivariate SUR Tobit model with logistic margins). However, as commented before in Section 1.2, a restriction in applying a bivariate distribution to the bivariate SUR Tobit model is the linear relationship between marginal distributions through the correlation coefficient. To overcome this restriction, we can use a copula function to capture/model the possibly nonlinear dependence structure in the bivariate SUR Tobit model.

For the censored outcomes y_{i1} and y_{i2}, the bivariate copula-based SUR Tobit distribution is given by

$$F(yi1,yi2) = C(ui1,ui2|\theta),$$

where $u_{ij} = F_j(y_{ij}|x_{ij},\beta_j,\sigma_j)$, for $j = 1,2$, and θ is the parameter (or a parameter vector) of the copula, which is assumed to be scalar.

In this paper, we consider only the Clayton (1978) copula, also referred to as the Cook and Johnson (1981) copula, originally studied by Kimeldorf and Sampson (1975). It takes the form

$$C\left(u_{i1},u_{i2}|\theta\right) = \left(u_{i1}^{-\theta} + u_{i2}^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \tag{2}$$

with θ restricted to the region $(0,\infty)$. The dependence between the margins increases with the value of θ, with $\theta \to 0^+$ implying independence and $\theta \to \infty$ implying perfect positive dependence. Trivedi and Zimmer (2005) point out that the Clayton copula is widely used to study correlated risks because it exhibits strong left tail dependence and relatively weak right tail dependence. Indeed, when correlation between two events is stronger in the left tail of the joint distribution, Clayton is usually an appropriate modeling choice.

3. Inference

In this section, we discuss inference (point and interval estimation) for the parameters of the bivariate copula-based SUR Tobit model. Particularly, by considering/assuming the Clayton copula and logistic margins (which corresponds to the assumption that the SUR Tobit model's marginal errors are logistically distributed).

3.1 Estimation through the IFAM method

According to Trivedi and Zimmer (2005), the log-likelihood function for the bivariate copula-based SUR Tobit model can be written in the following form (that is the same form as in the case of continuous margins):

$$\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n} \log c\left(F_1\left(y_{i1}|\boldsymbol{x}_{i1}, \boldsymbol{\upsilon}_1\right), F_2\left(y_{i2}|\boldsymbol{x}_{i2}, \boldsymbol{\upsilon}_2\right)|\theta\right) + \sum_{i=1}^{n}\sum_{j=1}^{2} \log f_j\left(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\upsilon}_j\right), \tag{3}$$

where η = (υ1,υ2,θ) is the vector of model parameters, υj = (βj,σj) is the margin j's parameter vector, fj (yij|xij,υj) is the p.d.f. of yij, Fj (yij|xij,υj) is the c.d.f. of yij, and c(ui1,ui2|θ), with uij = Fj (yij|xij,υj), is the p.d.f. of the copula distribution.

For example, if we consider a bivariate SUR Tobit model with logistic margins based on the Clayton copula, then fj (yij|xij,υj) is given by (1) and c(ui1,ui2|θ) is calculated from (2) as

$$c\left(u_{i1}, u_{i2}|\theta\right) = \frac{\partial^2 C\left(u_{i1}, u_{i2}|\theta\right)}{\partial u_{i1}\partial u_{i2}} = (\theta+1)\left(u_{i1}u_{i2}\right)^{-(\theta+1)}\left(u_{i1}^{-\theta} + u_{i2}^{-\theta} - 1\right)^{-\left(\frac{2\theta+1}{\theta}\right)}.$$

For model estimation, the use of copula methods, along with the log-likelihood function form given by (3), enables the use of the (classical) two-stage ML/IFM method by Joe and Xu (1996), which estimates the marginal parameters υ_j at a first step through

$$\widehat{\boldsymbol{\upsilon}}_{j,\mathrm{IFM}} = \arg\max_{\boldsymbol{\upsilon}_j} \sum_{i=1}^{n} \log f_j\left(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\upsilon}_j\right), \tag{4}$$

for j = 1,2, and then estimates the association parameter θ given $\widehat{\boldsymbol{\upsilon}}_{j,\mathrm{IFM}}$ by

$$\widehat{\theta}_{\mathrm{IFM}} = \arg\max_{\theta} \sum_{i=1}^{n} \log c\big(F_1\left(y_{i1}|\boldsymbol{x}_{i1}, \widehat{\boldsymbol{\upsilon}}_{1,\mathrm{IFM}}\right), F_2\left(y_{i2}|\boldsymbol{x}_{i2}, \widehat{\boldsymbol{\upsilon}}_{2,\mathrm{IFM}}\right)|\theta\big). \tag{5}$$

Note that each maximization task (step) has a small number of parameters, which reduces the computational difficulty. However, the IFM method provides a biased estimate for the parameter θ in the presence of censored observations for both margins (as will be seen in Section 4.2). Since we are interested in the bivariate copula-based SUR Tobit model where both marginal distributions are censored/semi-continuous, we are dealing with the case that is not one-to-one relationship between the marginal distributions and the copula, i.e. there is more than one copula to join the marginal distributions. This constitutes a violation of the Sklar (1959)'s theorem. When it occurs, researchers often face with problems in the copula model fitting and validation.

In order to facilitate the implementation of copula models with semi-continuous margins, the semi-continuous marginal distributions could be augmented to achieve continuity (see Shih and Louis (1995) for an alternative way of implementing copula models with censored observations in the margins, but in a survival analysis framework). More specifically, we can employ a (frequentist) data augmentation technique to simulate the latent (unobserved) dependent variables in the censored margins, i.e. we generate the unobserved data with all properties, e.g., mean, variance and dependence structure that match with the observed ones, and obtain the continuous marginal distributions (Wichitaksorn et al., 2012). Thus, in order to obtain an unbiased estimate for the association parameter θ, we replace y_{ij} by the augmented data y_{ij}^{a}, or equivalently and more simply (thus, preferred by us), we can replace u_{ij} by the augmented uniform data u^{a}_{ij} at the second stage of the IFM method and proceed with the copula parameter estimation as usual for cases of continuous margins. This process (uniform data augmentation and copula parameter estimation) is then repeated until convergence occurs (IFAM method). The (frequentist) data augmentation technique we employ here is partially based on Algorithm A2 presented in Wichitaksorn et al. (2012).

In the remaining of this subsection, we discuss the proposed IFAM method when using the Clayton copula to describe the nonlinear dependence structure of the bivariate SUR Tobit model with logistic margins. But the proposed approach can be extended to other copula functions (and marginal distributions) by applying different sampling algorithms. For the cases where only one of the dependent variables/margins is censored (i.e. when $y_{i1} > 0$ and $y_{i2} = 0$, or $y_{i1} = 0$ and $y_{i2} > 0$), the uniform data augmentation is performed through the truncated conditional distribution of the Clayton copula. That is, if the inverse conditional distribution of the employed copula has a closed-form expression, which is the case of Clayton copula, we can generate random numbers from its truncated version by applying the method by Devroye (1986, p. 38-39); otherwise, numerical root-finding procedures are required. By observing the results in Oakes (2005), we see that the Clayton copula has a remarkable invariance property under truncation, such that the conditional distribution of u_{i1} and u_{i2} on a sub-region of a Clayton copula, with one corner at (0,0), can be written by means of a Clayton copula. That formulation enables a simple simulation scheme in the cases where both dependent variables/margins are censored (i.e. when $y_{i1} = y_{i2} = 0$). For copulas that do not have the invariance property, an iterative simulation scheme could be employed.

The implementation of the bivariate Clayton copula-based SUR Tobit model with logistic margins through the proposed IFAM method can be described as follows.

Stage 1. Estimate the marginal parameters using (4). Set $\hat{\boldsymbol{\upsilon}}_{j,\mathrm{IFAM}} = \hat{\boldsymbol{\upsilon}}_{j,\mathrm{IFM}} = \left(\hat{\boldsymbol{\beta}}_{j,\mathrm{IFM}}, \hat{\sigma}_{j,\mathrm{IFM}}\right)$, for $j = 1,2$ (in what follows, we drop the label subscript IFM from β̂j,IFM and σ̂j,IFM to simplify the notation).

Stage 2. Estimate the copula parameter using e.g., (5). Set $\hat{\theta}^{(1)}_{\mathrm{IFAM}} = \hat{\theta}_{\mathrm{IFM}}$ and then consider the algorithm below.

For $\omega = 1,2,...,$

For $i = 1,2,...,n,$

If $y_{i1} = y_{i2} = 0$ then draw (u^{a}_{i1}, u^{a}_{i2}) from $C\left(u^{a}_{i1}, u^{a}_{i2} | \hat{\theta}^{(\omega)}_{\mathrm{IFAM}}\right)$ truncated to the region $(0,b_{i1}) \times (0,b_{i2})$. This can be performed relatively easily with the following steps.

1. $C\left(p,q|\hat{\theta}^{(\omega)}_{\text{IFAM}}\right)=\left(p^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+q^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}-1\right)^{-1/\hat{\theta}^{(\omega)}_{\text{IFAM}}}$ Draw (p,q) from. See, e.g., Armstrong (2003) for the Clayton copula data generation.
2. Compute $b_{ij}=G\left(-\boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}}_j\,/\,\hat{\sigma}_j\right)=1-G\left(\boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}}_j\,/\,\hat{\sigma}_j\right)$, for j = 1,2.
3. Set $\left[\left(b_{i1}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+b_{i2}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}-1\right)p^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+1-b_{i2}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}\right]^{-1/\hat{\theta}^{(\omega)}_{\text{IFAM}}}$
4. Set $\left[\left(b_{i1}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+b_{i2}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}-1\right)q^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+1-b_{i1}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}\right]^{-1/\hat{\theta}^{(\omega)}_{\text{IFAM}}}$.

If $y_{i1}=0$ and $y_{i2}>0$ then draw u^{a}_{i1} from $C\left(u^{\text{a}}_{i1}|u_{i2},\hat{\theta}^{(\omega)}_{\text{IFAM}}\right)$ truncated to the interval $(0,b_{i1})$. This can be done according to the following steps.

1. Compute $u_{i2}=G\left(\left(y_{i2}-\boldsymbol{x}'_{i2}\hat{\boldsymbol{\beta}}_2\right)/\,\hat{\sigma}_2\right)$.
2. Compute $b_{i1}=G\left(-\boldsymbol{x}'_{i1}\hat{\boldsymbol{\beta}}_1\,/\,\hat{\sigma}_1\right)=1-G\left(\boldsymbol{x}'_{i1}\hat{\boldsymbol{\beta}}_1\,/\,\hat{\sigma}_1\right)$.
3. Draw t from *Uniform*(0,1).
4. Compute $$v_{i1}=t\left[\left(b_{i1}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+u_{i2}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}-1\right)^{-\left(\hat{\theta}^{(\omega)}_{\text{IFAM}}+1\right)/\hat{\theta}^{(\omega)}_{\text{IFAM}}}\right]u_{i2}^{-\left(\hat{\theta}^{(\omega)}_{\text{IFAM}}+1\right)}.$$
5. Set $$u^{\text{a}}_{i1}=\left[\left(v_{i1}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}/\left(\hat{\theta}^{(\omega)}_{\text{IFAM}}+1\right)}-1\right)u_{i2}^{-\hat{\theta}^{(\omega)}_{\text{IFAM}}}+1\right]^{-1/\hat{\theta}^{(\omega)}_{\text{IFAM}}}.$$

If $y_{i1}>0$ and $y_{i2}=0$ then draw u^{a}_{i2} from $C\left(u^{\text{a}}_{i2}|u_{i1},\hat{\theta}^{(\omega)}_{\text{IFAM}}\right)$ truncated to the interval $(0,b_{i2})$. This can be done by following the five steps of the previous case ($y_{i1}=0$ and $y_{i2}>0$) by switching subscripts 1 and 2.

If $y_{i1}>0$ and $y_{i2}>0$ then set $u^{\text{a}}_{i1}=u_{i1}=G\left(\left(y_{i1}-\boldsymbol{x}'_{i1}\hat{\boldsymbol{\beta}}_1\right)/\,\hat{\sigma}_1\right)$ and $u^{\text{a}}_{i2}=u_{i2}=G\left(\left(y_{i2}-\boldsymbol{x}'_{i2}\hat{\boldsymbol{\beta}}_2\right)/\,\hat{\sigma}_2\right)$.

Given the generated/augmented marginal uniform data u^{a}_{ij} [1], we estimate the association parameter θ by

$$\hat{\theta}^{(\omega+1)}_{\text{IFAM}}=\arg\max_{\theta}\sum_{i=1}^{n}\log c\left(u^{\text{a}}_{i1},u^{\text{a}}_{i2}|\theta\right).$$

The algorithm stops if a termination criterion is fulfilled, e.g. if $|\hat{\theta}^{(\omega+1)}_{\text{IFAM}}-\hat{\theta}^{(\omega)}_{\text{IFAM}}|<\xi$, where ξ is the tolerance parameter (e.g., $\xi=10^{-3}$).

[1] The generated/augmented marginal uniform data uaij should carry (ω) as a superscript $(\text{i.e. } u^{\text{a}(\omega)}_{ij})$, but we omit it so as not to clutter the notation.

3.2 Interval estimation

Joe and Xu (1996) suggest the use of the jackknife method for estimation of the standard errors of the multivariate model parameter estimates when using the IFM approach. It makes the analytic derivatives no longer required to compute the inverse Godambe information matrix, that is the asymptotic covariance matrix associated with the vector of parameter estimates. See Joe (1997, p. 301-302) for the form of this matrix. Nevertheless, we carried out a pilot simulation study whose results revealed that the jackknife is not valid to obtain standard errors of parameter estimates when using the IFAM approach, i.e. in the context of copula-based models with censored/semi-continuous margins (the jackknife method produces an overestimate of the standard error of the association parameter estimate). This implies that confidence intervals for the parameters of the bivariate copula-based SUR Tobit model cannot be constructed using this resampling technique. To surpass this problem, we propose the use of bootstrap methods to build confidence intervals.

Our bootstrap approach can be described as follows. Let η_h, $h = 1,...,k$, be a component of the parameter vector η of the bivariate copula-based SUR Tobit model (see Section 3.1). By using a resampling plan (e.g., a parametric resampling plan), we obtain the bootstrap estimates $\hat{\eta}^*_{h1}, \hat{\eta}^*_{h2}, ..., \hat{\eta}^*_{hB}$ of η_h through the IFAM method, with B being the number of bootstrap samples (Hinkley (1988) suggests that the minimum value of B will depend on the parameter being estimated, but that it will often be 100 or more). Thus, we can derive confidence intervals from the bootstrap distribution through the following two methods, for instance.

Percentile bootstrap (Efron and Tibshirani, 1993, p. 171). The $100(1-2\alpha)$% percentile confidence interval is defined by the $100(\alpha)$th and $100(1-\alpha)$th percentiles of the bootstrap distribution of $\hat{\eta}_h^*$:

$$\left[\hat{\eta}_h^{*(\alpha)}, \hat{\eta}_h^{*(1-\alpha)}\right].$$

For Carpenter and Bithell (2000), simplicity is an attractive characteristic of this method. Furthermore, no invalid parameter values can be included in the interval.

Standard normal interval (Efron and Tibshirani, 1993, p. 154). Since most statistics are asymptotically normally distributed, in large samples we can use the standard error estimate, $\widehat{se}_h$, along with the normal distribution, to yield a $100(1-2\alpha)$% conb

fidence interval for η_h based on the original estimate (i.e. from the original data) $\hat{\eta}_h$:

$$\left[\hat{\eta}_h - z^{(1-\alpha)}\widehat{se}_h, \hat{\eta}_h - z^{(\alpha)}\widehat{se}_h\right],$$

where $z^{(\alpha)}$ represents the $100(\alpha)$th percentile point of a standard normal distribution, and $\widehat{se}_h$ is the hth entry on the diagonal of the bootstrap-based covariance matrix b

estimate of the parameter vector estimate $\hat{\eta}$, which is given by

$$\widehat{\boldsymbol{\Sigma}}_{\text{boot}} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\boldsymbol{\eta}}_b^* - \overline{\hat{\boldsymbol{\eta}}}^*\right)\left(\hat{\boldsymbol{\eta}}_b^* - \overline{\hat{\boldsymbol{\eta}}}^*\right)', \qquad (6)$$

where $\hat{\eta}_b^*$, $b = 1,\ldots,B$, is the bootstrap estimate of η and

$$\overline{\hat{\boldsymbol{\eta}}}^* = \left(\frac{1}{B}\sum_{b=1}^{B}\hat{\eta}_{1b}^*, \frac{1}{B}\sum_{b=1}^{B}\hat{\eta}_{2b}^*, \ldots, \frac{1}{B}\sum_{b=1}^{B}\hat{\eta}_{kb}^*\right).$$

4. Simulation study

A simulation study was performed to investigate the behavior of the IFAM estimates and check the coverage probability of bootstrap confidence intervals (constructed using the two methods described in Section 3.2) for the parameters of the bivariate copula-based SUR Tobit model. Here, we considered some circumstances that might arise in the development of bivariate copula-based SUR Tobit models, involving the sample size, the censoring percentage (i.e. the percentage of zero observations) in the dependent variables/margins and their interdependence degree.

4.1 General specifications

In the simulation study, we applied the Clayton copula to model the nonlinear dependence structure of the bivariate SUR Tobit model. We set the true value for the association parameter θ at 0.67, 2 and 6, corresponding to a Kendall's tau association measure [2] of 0.25, 0.50 and 0.75, respectively. See, e.g., Armstrong (2003) for the Clayton copula data generation. For $i = 1,\ldots,n$, the model errors ϵ_{i1} and ϵ_{i2} were assumed to follow a logistic distribution with $\epsilon_{i1} \sim Logistic(0,\sigma_1)$ and $\epsilon_{i2} \sim Logistic(0,\sigma_2)$, where $\sigma_1 = 1$ and $\sigma_2 = 2$ are the scale parameters for margins 1 and 2, respectively. The covariates for margin 1, xi1 = (xi1,0,xi1,1) , were xi1,0 = 1 and xi1,1 was randomly simulated from a standard normal distribution; while the covariates for margin 2, xi2 = (xi2,0,xi2,1) , were generated as xi2,0 = 1 and xi2,1 was randomly simulated from N (1,2). To ensure a percentage of censoring (i.e. of zero observations) for both margins of approximately 5%, 15%, 25%, 35% and 50%, we assumed the following true values for $\boldsymbol{\beta}_1 = (\beta_{1,0},\beta_{1,1})'$ and $\boldsymbol{\beta}_2 = (\beta_{2,0},\beta_{2,1})'$: $\beta_1 = (3.3,1)$ and $\beta_2 = (5.8,1)$, $\beta_1 = (2.1,1)$ and $\beta_2 = (3.1,1)$, $\beta_1 = (1.3,1)$ and $\beta_2 = (1.7,1)$, $\beta_1 = (0.8,1)$ and $\beta_2 = (0.5,1)$, and $\beta_1 = (-0.05,1)$ and $\beta_2 = (-0.9,1)$, respectively. We generated 100 data sets each with n = 200, 800 and 2000. For each data set (original sample), we obtained 500 bootstrap samples through a parametric resampling plan (parametric bootstrap approach), i.e. we fitted a bivariate Clayton copula-based SUR Tobit model with logistic margins to each data set and then generated a set of 500 new data sets (the same size as the original data set/sample) from the estimated parametric model. The computing

[2] The Kendall's tau for Clayton copula is given by $\tau = \theta/(\theta + 2)$; see McNeil et al. (2005, p. 222).

language was written in R version 3.1.0 and ran on a virtual machine of the Cloud-USP at ICMC, with Intel Xeon processor E5500 series, 8 core (virtual CPUs), 32 GB RAM.

We assessed the performance of the proposed model and methods through the coverage probabilities of the nominally 90% standard normal and percentile bootstrap confidence intervals, the Bias and the Mean Squared Error (MSE), in which the Bias and the MSE of each parameter η_h, $h = 1,\dots,k$, are given by Bias $= M^{-1}\sum_{r=1}^{M}\left(\hat{\eta}_h^r - \eta_h\right)$ and MSE $= M^{-1}\sum_{r=1}^{M}\left(\hat{\eta}_h^r - \eta_h\right)^2$, respectively, where $M = 100$ is the number of replications (original data sets/samples) and $\hat{\eta}_h^r$ is the estimated value of η_h at the rth replication.

4.2 Simulation results

In this subsection, we present the main results obtained from the simulation study performed with samples (data sets) of different sizes, percentages of censoring in the margins and degrees of dependence between them, regarding the bivariate Clayton copula-based SUR Tobit model with logistic margins' parameters estimated using the IFAM approach. Since both IFAM and IFM methods provide the same marginal parameter estimates (the first stage of the proposed method is similar to the first stage of the usual one, as seen in Section 3.1), we focus here on the Clayton copula parameter estimate. We also exhibit the results related to the estimated coverage probabilities of the 90% confidence intervals for θ, obtained via bootstrap methods (standard normal and percentile intervals).

It took about four days, one and a half week and three weeks for the simulation for n = 200, 800 and 2000, respectively, to complete; which represent very long computation times.

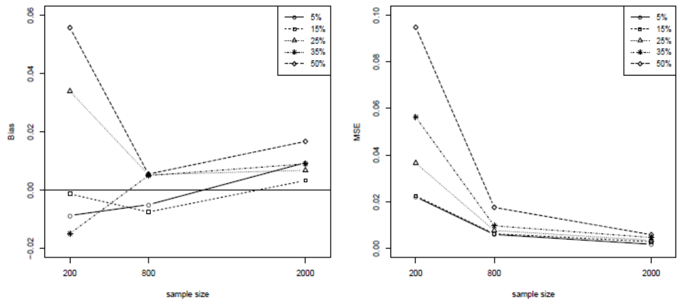
Figure 3 shows the Bias and MSE of the observed IFAM estimates of θ. From this figure, we observe that, regardless of the percentage of censoring in the margins and their interdependence degree, the Bias and MSE of the IFAM estimator of θ are relatively low and tend to zero for large n, i.e. the IFAM estimator is asymptotically unbiased (despite some random fluctuations of Bias) and consistent for the Clayton copula parameter.

In Figure 4, we see that the estimated coverage probabilities of the bootstrap confidence intervals for θ are sufficiently high and close to the nominal value of 0.90, except for a few cases in which n is small to moderate (n = 200 and 800) and the degree of dependence between the margins is high (θ = 6).

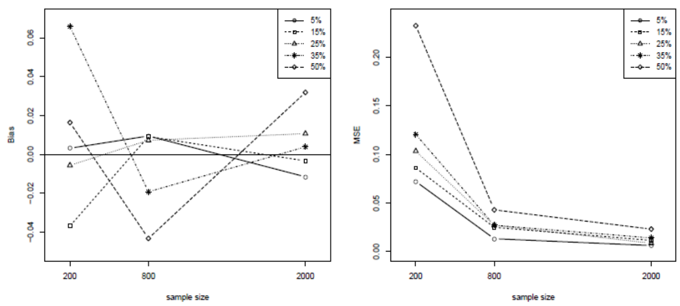
Finally, Figure 5 compares, via boxplots, the observed IFAM estimates of θ with its estimates obtained through the IFM method, for n = 2000. It can be seen that there is certain equivalence between the two estimation methods (with a slight advantage for the IFAM method over the IFM method in terms of bias) when the degree of dependence between the margins is moderate, that is θ = 2 (see Figure 5(b)). Note also that the IFM method overestimates θ for dependence at a lower level, that is θ = 0.67 (see Figure 5(a)), and underestimates θ for dependence at a higher level, that is θ = 6 (see Figure 5(c)). In these cases, the difference (distance) between the distributions of the IFM and IFAM estimates increases as the percentage of censoring in the margins increases.

5. Application

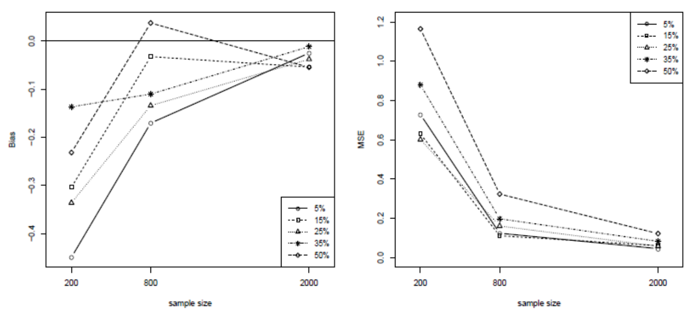
In this section, our methodology is illustrated on the consumption dataset discussed earlier in Section 1.1. In this application, the relationship between the re-



(a) $\theta = 0.67$

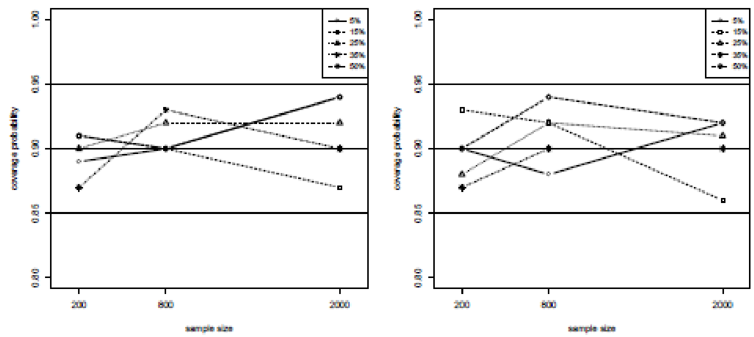


(b) $\theta = 2$

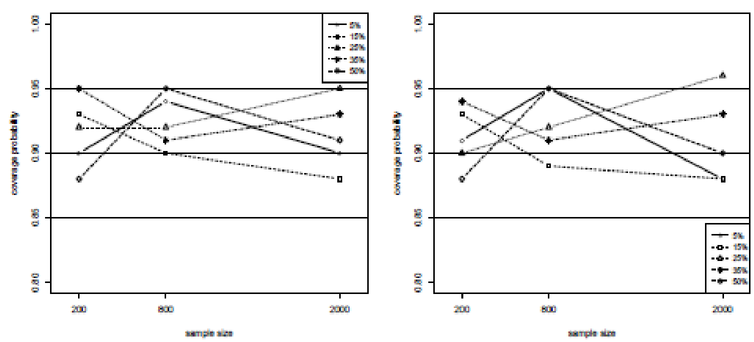


(c) $\theta = 6$

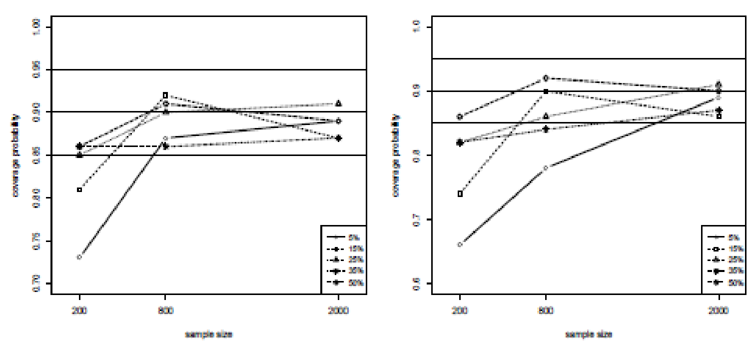
Figure 3: Bias and MSE of the IFAM estimate of the Clayton copula parameter versus sample size, percentage of censoring in the margins and degree of dependence between them.



(a) $\theta = 0.67$

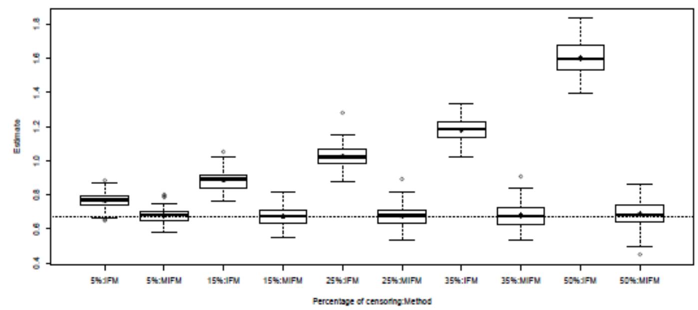


(b) $\theta = 2$

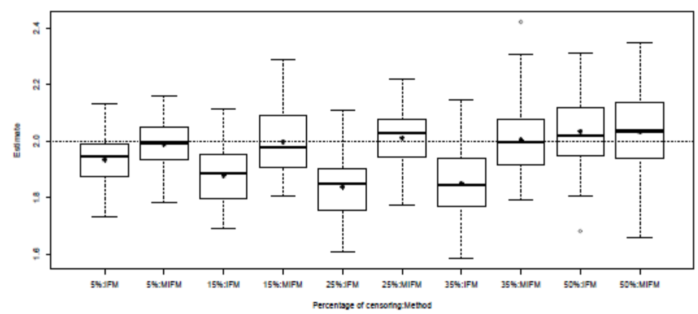


(c) $\theta = 6$

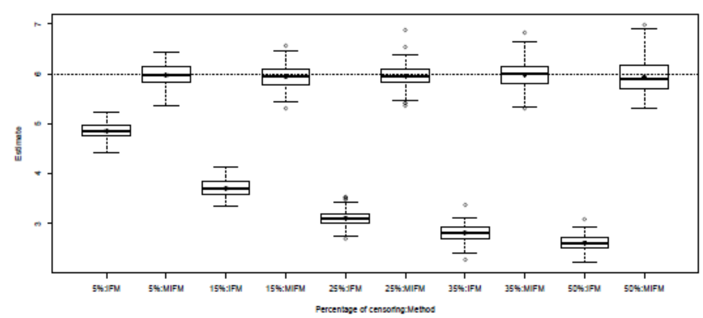
Figure 4: Coverage probabilities (CPs) of the 90% standard normal (panels on the left) and percentile (panels on the right) confidence intervals for the Clayton copula parameter versus sample size, percentage of censoring in the margins and degree of dependence between them. The horizontal line at CP = 0.90 and the two horizontal lines at CP = 0.85 and 0.95 correspond, respectively, to the lower and upper bounds of the 90% confidence interval of the CP = 0.90. Thus, if a confidence interval has exact coverage of 0.90, roughly 90% of the observed coverages should be between these lines.



(a) $\theta = 0.67$



(b) $\theta = 2$



(c) $\theta = 6$

Figure 5: Comparison between the IFM and IFAM estimates of the Clayton copula parameter (n = 2000). The averages of the parameter estimates are shown with a star symbol. The dotted horizontal line represents the true value of the Clayton copula parameter.

Table 1: Variable definitions and sample statistics (n = 285).

Table 1: Variable definitions and sample statistics ($n = 285$).

Variable	Definition	Mean	S.D.[a]
Dependent variables: amount consumed			
Salad dressings (in 100 g)	Quantity of salad dressings consumed	0.2159	0.1522
	Among the consuming ($n = 270$; 94.74%)	0.2279	0.1474
Lettuce (in 200 g)	Quantity of lettuce consumed	0.1403	0.1523
	Among the consuming ($n = 211$; 74.04%)	0.1896	0.1483
Continuous explanatory variable			
Income	Household income as the proportion of poverty threshold	2.3160	0.8404
Binary explanatory variables (yes = 1; no = 0)			
Age 20-30	Age is 20-30	0.1123	
Age 31-40	Age is 31-40	0.1684	
Age 41-50	Age is 41-50	0.1825	
Age 51-60	Age is 51-60	0.2175	
Age > 60	Age > 60 (reference)	0.3193	
Northeast	Resides in the Northeastern states	0.1719	
Midwest	Resides in the Midwestern states	0.2632	
South	Resides in the Southern states (reference)	0.3228	
West	Resides in the Western states	0.2421	

Source: Compiled from the CSFII, USDA, 1994-1996.
[a]S.D. = Standard Deviation.

ported salad dressings (amount consumed, in 100 grams) and lettuce (amount consumed, in 200 grams) consumption of 285 U.S. adults is modeled by the bivariate SUR Tobit model with logistic margins through the Clayton copula (see Sections 1.1 and 1.3 for the reasons for this modeling choice). We include age, location (region) and income as the covariates and use them for both margins. From the KolmogorovSmirnov goodness-of-fit tests of augmented marginal residuals [3], we obtain p-values equal to 0.3880 and 0.7944 for the salad dressings and lettuce models, respectively. Therefore, the logistic assumption for the marginal errors is valid. Table 1 provides the definitions and sample statistics for all considered variables, where we observe the proportions of consuming individuals in the data set to range from 94.74% for salad dressings, to 74.04% for lettuce. Among those consuming, an individual on average consumes 22.79 g of salad dressings and 37.92 g of lettuce per day.

Table 2 shows the IFAM estimates for the marginal and Clayton copula parameters, as well as the 90% confidence intervals obtained through the standard normal and percentile bootstrap methods. The results reveal that individuals age 31-40 consume more salad dressings than those over 60 years of age. Su and Arab (2006) found similar effect of age on salad dressings consumption. Regional effect is also notable, with individuals from the Midwest and West consuming more lettuce than individuals residing in the South. The household income has a positive effect on the consumption of both salad dressings and lettuce. The IFAM estimate of the

[3] The augmented residuals are the differences between the augmented observed and predicted responses, (i.e. $e^{a}_{ij} = y^{a}_{ij} - \boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}}_j$, for i = 1,...,n and j = 1,2, where $y^{a}_{ij} = \boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}}_j + \hat{\sigma}_j G^{-1}\left(u^{a}_{ij}\right)$, with $e^{a}_{ij} = \hat{\sigma}_j G^{-1}\left(u^{a}_{ij}\right)$ G^{-1}(.) being the inverse function of the *Logistic*(0,1) c.d.f.; or simply,.

Clayton copula parameter $\hat{\theta}_{\text{IFAM}} = 1.2799$,obtained after 26 iterations and its 90% bootstrap-based confidence intervals show us that the relationship between salad dressings and lettuce consumption is positive (the estimated Kendall's tau is $\hat{\tau} = \hat{\theta}_{\text{IFAM}}/(\hat{\theta}_{\text{IFAM}} + 2) = 0.3902$, which is close to the value of the nonparametric association measure presented in Section 1.1) and significant at the 10% level (the lower limits of the 90% bootstrap-based confidence intervals for θ are greater than and far above zero), justifying joint estimation of the censored equations through the Clayton copula to improve statistical efficiency. Furthermore, the estimated coefficient of tail dependence for Clayton copula, $\hat{\lambda}_L = 0.5818$, obtained from $2^{-1/\hat{\theta}_{\text{IFAM}}}$ (see McNeil et al., 2005, p. 209), shows the positive dependence at the lower tail, i.e. for low or no consumption of salad dressings and lettuce.

For comparison purposes, we fit the basic bivariate SUR Tobit model with logistic margins, using the MCECM algorithm (Huang, 1999). The estimation results, obtained after 5 iterations (i.e. in much fewer iterations than required by the IFAM method, but the MCECM algorithm is much more time consuming), are presented in Table 3. The standard errors in Table 3 were derived from the bootstrap-based covariance matrix estimate given by (6), but now with η denoting the parameter vector of the basic bivariate SUR Tobit model with logistic margins (bootstrap standard errors). Note that the marginal parameter estimates obtained through the MCECM and IFAM methods are similar. However, unlike the fitted copula-based model, the fitted basic model exhibits a (significant) positive effect of the dummy variable age 31-40 on lettuce consumption, i.e. with individuals age 31-40 consuming more lettuce than those over 60 years of age. Tables 2 and 3 also present the log-likelihood values for the two fitted models. We can then compare the bivariate Clayton copulabased SUR Tobit model with logistic margins and basic bivariate SUR Tobit model with logistic margins by using some information criterion, e.g. the Akaike information criterion (AIC), which is defined by $-2\ell(\eta) + 2k$. The preferred model is the b

one with the smaller AIC value. The AIC criterion value for the bivariate Clayton copula-based SUR Tobit model with logistic margins is equal to −274.1849. While the AIC criterion value for the basic bivariate SUR Tobit model with logistic margins is −269.0044. Thus, the bivariate Clayton copula-based SUR Tobit model with logistic margins overcomes the basic bivariate SUR Tobit model with logistic margins in the considered criterion.

Table 2: Estimation results of bivariate Clayton copula-based SUR Tobit model with logistic margins for salad dressings and lettuce consumption in the U.S. in 1994-1996.

		90% Confidence Intervals	
Salad dressings	Estimate	Standard Normal	Percentile
Intercept	0.1164	[0.0676; 0.1653]	[0.0733; 0.1664]
Age 20-30	0.0431	[-0.0095; 0.0956]	[-0.0103; 0.0923]
Age 31-40	0.0759	[0.0294; 0.1224]	[0.0304; 0.1208]
Age 41-50	0.0328	[-0.0109; 0.0766]	[-0.0098; 0.0774]
Age 51-60	0.0090	[-0.0302; 0.0481]	[-0.0329; 0.0475]
Northeast	0.0085	[-0.0365; 0.0535]	[-0.0365; 0.0521]
Midwest	0.0066	[-0.0288; 0.0420]	[-0.0284; 0.0421]
West	0.0133	[-0.0243; 0.0510]	[-0.0249; 0.0522]
Income	0.0238	[0.0074; 0.0403]	[0.0074; 0.0409]
σ_1	0.0841	[0.0773; 0.0909]	[0.0758; 0.0894]
		90% Confidence Intervals	
Lettuce	Estimate	Standard Normal	Percentile
Intercept	-0.0328	[-0.0949; 0.0293]	[-0.0957; 0.0248]
Age 20-30	0.0442	[-0.0192; 0.1076]	[-0.0161; 0.1045]
Age 31-40	0.0442	[-0.0107; 0.0991]	[-0.0083; 0.1013]
Age 41-50	0.0287	[-0.0246; 0.0820]	[-0.0228; 0.0820]
Age 51-60	0.0340	[-0.0159; 0.0839]	[-0.0146; 0.0820]
Northeast	0.0036	[-0.0485; 0.0556]	[-0.0476; 0.0516]
Midwest	0.0458	[0.0021; 0.0896]	[0.0039; 0.0881]
West	0.0630	[0.0153; 0.1107]	[0.0161; 0.1077]
Income	0.0369	[0.0158; 0.0580]	[0.0155; 0.0572]
σ_2	0.1015	[0.0914; 0.1117]	[0.0902; 0.1097]
θ	1.2799	[0.9484; 1.6115]	[0.9835; 1.6304]
Log-likelihood	158.0924		

Table 3: Estimation results of basic bivariate SUR Tobit model with logistic margins for salad dressings and lettuce consumption in the U.S. in 1994-1996.

Salad dressings	Estimate	Standard Error
Intercept	0.1190 [a]	0.0346
Age 20-30	0.0455	0.0314
Age 31-40	0.0769 [a]	0.0247
Age 41-50	0.0389	0.0241
Age 51-60	0.0099	0.0249
Northeast	0.0038	0.0262
Midwest	0.0057	0.0209
West	0.0176	0.0230
Income	0.0231 [a]	0.0105
σ_1	0.0840 [a]	0.0036
Lettuce	Estimate	Standard Error
Intercept	-0.0294	0.0363
Age 20-30	0.0381	0.0390
Age 31-40	0.0491 [a]	0.0292
Age 41-50	0.0338	0.0286
Age 51-60	0.0374	0.0270
Northeast	0.0011	0.0299
Midwest	0.0578 [a]	0.0259
West	0.0652 [a]	0.0265
Income	0.0360 [a]	0.0127
σ_2	0.1004 [a]	0.0061
Log-likelihood	154.5022	

[a] Denotes significant at the 10% level.

6. Final remarks and further researches

In this paper, we extended the analysis of the bivariate SUR Tobit by modeling its nonlinear dependence structure through copula and assuming non-normal marginal error distributions. For the present study, we chose to work with Clayton copula and assumed logistic distribution for the marginal errors, mainly motivated by the real data at hand (consumption data). The ability to capture/model the tail dependence, especially the lower tail dependence where some data were censored, is an attractive feature of Clayton copula. Some advantages arose from this copula choice, regarding the development of the proposed IFAM method for obtaining the estimates of the bivariate copula-based SUR Tobit model parameters. First, Clayton copula is known to be preserved under bivariate truncation (Oakes, 2005). This invariance property enabled a simple simulation scheme in the cases where both dependent variables/margins were censored. Second, the existence of a closed-form expression for the inverse of the conditional Clayton copula distribution (see, e.g., Armstrong, 2003) enabled a simple simulation scheme when just a single dependent variable/margin was censored, by applying the method by Devroye (1986, p. 38-39).

In the performed simulation study, we assessed the performance of the proposed model (bivariate Clayton copula-based SUR Tobit model with logistic margins) and methods (IFAM method, standard normal and percentile bootstrap-based confidence intervals), obtaining satisfactory results (i.e. unbiased estimates of the Clayton copula parameter, high and near the nominal value coverage probabilities of the bootstrap-based confidence intervals) regardless of the censoring percentage in the margins and their interdependence degree.

We pointed the applicability of the model and methods to a real data set of consumption of salad dressings and lettuce by U.S. individuals. We also compared our results with the basic bivariate SUR Tobit model with logistic margins (which assumes a bivariate logistic distribution for the vector of error terms) and found that the gain for introducing the Clayton copula to model the nonlinear dependence structure of the SUR Tobit model, was substantial for this data set.

Although it is relatively rare to analyze the SUR Tobit model with over two dimensions, unless it is modeled in the longitudinal setting (see, e.g., Baranchuk and Chib (2008) for an example of the longitudinal Tobit model), our proposed model and methods can be straightforwardly applied to high dimensional SUR Tobit models.

We leave to further studies the issue of deriving other asymptotic properties (such as asymptotic normality) for the copula parameter estimate obtained through the IFAM method in our framework of SUR models with limited (partially observed or censored) dependent variables.

References

[1] Amemiya, T., 1984. Tobit Models: A survey. Journal of Econometrics 24 (1-2), 3–61.

[2] Armstrong, M., 2003. Copula catalogue. Part 1: Bivariate Archimedean copulas, unpublished paper available at: http://www.cerna.ensmp.fr/Documents/MACopulaCatalogue.pdf.

[3] Baranchuk, N., Chib, S., 2008. Assessing the role of option grants to CEOs: How important is heterogeneity? Journal of Empirical Finance 15 (2), 145–166.

[4] Boriss, H., Brunke, H., 2005. Commodity profile: Lettuce (pdf), university of California. Retrieved 2014-04-17.

[5] Brown, E., Lankford, H., 1992. Gifts of money and gifts of time: Estimating the effects of tax prices and available time. Journal of Public Economics 47 (3), 321– 341.

[6] Cameron, A. C., Trivedi, P. K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

[7] Carpenter, J., Bithell, J., 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Statistics in Medicine 19 (9), 1141–1164.

[8] Chen, S., Zhou, X., 2011. Semiparametric estimation of a bivariate Tobit model. Journal of Econometrics 165 (2), 266–274.

[9] Chib, S., 1992. Bayes inference in the Tobit censored regression model. Journal of Econometrics 51 (1-2), 79–99.

[10] Chib, S., Greenberg, E., 1998. Analysis of multivariate probit models. Biometrika 85 (2), 347–361.

[11] Clayton, D. G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65 (1), 141–152.

[12] Cook, R. D., Johnson, M. E., 1981. A family of distributions for modelling nonelliptically symmetric multivariate data. Journal of the Royal Statistical Society, Series B 43 (2), 210–218.

[13] Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer-Verlag, New York.

[14] Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

[15] Geweke, J., 1991. Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In: COMPUTING SCIENCE AND STATISTICS: PROCEEDINGS OF THE 23RD SYMPOSIUM ON THE INTERFACE. pp. 571–578, seattle, USA,

[16] April 22-24.

[17] Goldberger, A. S., 1964. Econometric Theory. Wiley, New York.

[18] Greene, W. H., 2003. Econometric Analysis, 5th Edition. Prentice Hall, New Jersey.

[19] Gumbel, E. J., 1961. Bivariate logistic distributions. Journal of the American Statistical Association 56 (294), 335–349.

[20] Hinkley, D. V., 1988. Bootstrap methods. Journal of the Royal Statistical Society, Series B 50 (3), 321–337.

[21] Huang, C. J., Sloan, F. A., Adamache, K. W., 1987. Estimation of seemingly unrelated Tobit regressions via the EM algorithm. Journal of Business and Economic Statistics 5 (3), 425–430.

[22] Huang, H. C., 1999. Estimation of the SUR Tobit model via the MCECM algorithm. Economics Letters 64 (1), 25–30.

[23] Huang, H. C., 2001. Bayesian analysis of the SUR Tobit model. Applied Economics Letters 8 (9), 617–622.

[24] Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

[25] Joe, H., Xu, J. J., 1996. The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

[26] Kamakura, W. A., Wedel, M., 2001. Exploratory Tobit factor analysis for multivariate censored data. Multivariate Behavioral Research 36 (1), 53–82.

[27] Kimeldorf, G., Sampson, A. R., 1975. Uniform representations of bivariate distributions. Communications in Statistics - Theory and Methods 4 (7), 617–627.

[28] Lee, L. F., 1993. Multivariate Tobit models in econometrics. Handbook of Statistics 11, 145–175, in: G.S. Maddala and C.R. Rao, eds., North-Holland.

[29] McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, New Jersey.

[30] Meng, X., Rubin, D. B., 1996. Efficient methods for estimating and testing seemingly unrelated regressions in the presence of latent variables and missing observations. Bayesian Analysis in Statistics and Econometrics, 215–227In: D.A. Berry, K.M. Chaloner, and J.K. Geweke, eds., New York: John Wiley & Sons.

[31] Moulton, L. H., Halsey, N. A., 1995. A mixture model with detection limits for regression analyses of antibody response to vaccine. Biometrics 51 (4), 1570–1578. Nelsen, R. B., 2006. An Introduction to Copulas, 2nd Edition. Springer, New York.

[32] Oakes, D., 2005. On the preservation of copula structure under truncation. The Canadian Journal of Statistics 33 (3), 465–468.

[33] Panagiotelis, A., Czado, C., Joe, M., 2012. Pair copula constructions for multivariate discrete data. Journal of the American Statistical Association 107 (499), 1063– 1072.

[34] Pitt, M., Chan, D., Kohn, R., 2006. Efficient Bayesian inference for Gaussian copula regression models. Biometrika 93 (3), 537–554.

[35] Shih, J. H., Louis, T. A., 1995. Inferences on the association parameter in copula models for bivariate survival data. Biometrics 51 (4), 1384–1399.

[36] Sklar, A., 1959. Fonctions de r´epartition a` n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Universit´e de Paris 8, 229–231.

[37] Smith, M. S., Khaled, M. A., 2012. Estimation of copula models with discrete margins via Bayesian data augmentation. Journal of the American Statistical Association 107 (497), 290–303.

[38] Su, L. J., Arab, L., 2006. Salad and raw vegetable consumption and nutritional status in the adult US population: results from the Third National Health and Nutrition Examination Survey. Journal of the American Dietetic Association 106 (9), 1394– 1404.

[39] Taylor, M. R., Phaneuf, D., 2009. Bayesian estimation of the impacts of food safety information on household demand for meat and poultry. In: PROCEEDINGS OF THE 2009 AAEA & ACCI JOINT ANNUAL MEETING. Milwaukee, WI, July

[40] 26-28.

[41] Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36.

[42] Trivedi, P. K., Zimmer, D. M., 2005. Copula modeling: An introduction for practitioners. Foundations and Trends in Econometrics 1 (1), 1–111.

[43] USDA, 2000. Continuing survey of food intakes by individuals 1994-1996. CD-ROM, agricultural Research Service, Washington, DC.

[44] Wales, T. J., Woodland, A. D., 1983. Estimation of consumer demand systems with binding non-negativity constraints. Journal of Econometrics 21 (3), 263–285.

[45] Wichitaksorn, N., Choy, S. T. B., Gerlach, R., 2012. Estimation of bivariate copulabased seemingly unrelated Tobit models, discipline of Business Analytics, University of Sydney Business School, NSW 2006, Australia. Eletronic copy available at: http://ssrn.com/abstract=2122388.

[46] Yen, S. T., Lin, B. H., 2008. Quasi-maximum likelihood estimation of a censored equation system with a copula approach: meat consumption by U.S. individuals. Agricultural Economics 39 (2), 207–217.

[47] Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57 (298), 348–368.

[48] Zellner, A., Ando, T., 2010. Bayesian and non-Bayesian analysis of the seemingly unrelated regression model with Student-t errors, and its application for forecasting.

[49] International Journal of Forecasting 26 (2), 413–434.

Received July 15, 2013; accepted September 10, 2015

Francisco Louzada*
SME-ICMC, Universidade de Sao Paulo, Sao Carlos, SP, Brazil
Paulo H. Ferreira
PPGEs, Universidade Federal de Sao Carlos, Sao Carlos, SP, Brazil