Health Care Utilization and Self-Assessed Health: Specification of Bivariate Models Using Copulas[*]

José M. R. Murteira
Faculdade de Economia, Universidade de Coimbra, and CEMAPRE

Óscar D. Lourenço
Faculdade de Economia, Universidade de Coimbra, and CEIS-UC

This version: December, 2007

Preliminary draft; comments welcome.

Abstract

The discernment of relevant factors driving health care utilization constitutes one important research topic in Health Economics. This issue is frequently addressed through specification of regression models for health care use (*y* – often measured by number of doctor visits) including, among other covariates, a measure of self-assessed health (*sah*). However, the exogeneity of *sah* has been questioned, due to the possible presence of unobservables influencing *y* and *sah*, and because individuals' health assessments may depend on the quantity of medical care received.

This paper circumvents the potential endogeneity of *sah* and its associated consequences within conventional regression models (namely the need to find valid instruments) by adopting a full information approach, with specification of bivariate regression models for the discrete variables (*sah*,*y*). The approach is implemented with copula functions, which enable separate consideration of each variable margin and their dependence structure. Estimation of these models is through maximum likelihood, with cross-section data from the Portuguese National Health Survey of 1998/99. Results indicate that estimates of regression parameters do not vary much between different copula models. The dependence parameter estimate is negative across joint models, which suggests evidence of simultaneity of (*sah*,*y*) and casts doubt on the appropriateness of limited information approaches.

JEL classification code: I10, C16, C51.
Key Words: health care utilization; self-assessed health; endogeneity; discrete data; copulas.

[*] Support from Fundação para a Ciência e Tecnologia, Program FEDER/POCI 2010 is gratefully acknowledged. Address for correspondence: José Murteira, Faculdade de Economia, Universidade de Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal. E-mail: jmurt@fe.uc.pt.

1. Introduction

The area of health economics has witnessed a steady increase in research activity over the last decades. To some extent, this growing interest can be seen as a consequence of the volume and continuous rise of health care expenditures in most industrialized countries, Portugal included.[1] Not surprisingly, one important ongoing research topic in this area has been the discernment of the relevant factors driving health care use by individuals. In the Portuguese context, the estimation of income-elasticities of utilization, and the assessment of the effect of supplementary health insurance on individuals' health care use provide well-known examples of applied work in this area, also relevant from a health policy perspective (Barros, 1999, Barros, Machado, Galdeano, 2005).

This type of concern is frequently addressed through specification of univariate regression models for health care use (often represented by a count, y, measuring the number of doctor visits). The literature includes diversified examples of count data specifications, *e.g.*, Poisson and negative binomial models, hurdle, zero-inflated and finite mixture regression models, to cite only the most usual. Applications can be found, among others, in Bago d'Uva (2006), Deb and Trivedi (1997, 2002), Gerdtham and Sundberg (1998), Lourenço, Quintal, Ferreira and Barros (2007), Sarma and Simpson (2006), Vera-Hernandez (1999), and Winkelmann (2004).

Most of the regression models discussed in the literature include, among other covariates, an indicator of self-assessed health (*sah*), which is usually found to be a relevant regressor for the dependent variable of interest. Frequently – and expectably – these models are estimated by use of methods that rely on the assumption of regressors' exogeneity, *sah* included. Some authors, however, have cast doubt on the exogeneity of *sah* within such models, due to the accepted fact that individuals' health assessments are, to a significant extent, both subjective and determined by the quantity of medical care recently received (see, *e.g.*, Windmeijer and Santos Silva, 1997, Barros, 1999, and Van Ourti, 2004).

Frequently, the endogeneity issue is handled within a limited information approach, through specification of one or two moments of the conditional probability

[1] In Portugal, according to the OECD Health Data (2006), the total expenditures on health, as share of GDP, increased from 7.3% in 1994 to 10.1% in 2004.

function (p.f.) of y given (sah, x). As is well known (see, *e.g.* Cameron and Trivedi, 1998, ch. 11) such an approach calls for estimation strategies, namely nonlinear instrumental variables (NLIV) or generalized method of moments (GMM), usually requiring valid instruments. When no such variables are available, researchers often face two options: either to exclude *sah* from the regression model, or to adopt a non-robust method – usually nonlinear least squares (NLS) or conditional maximum likelihood (ML). Clearly, either choice involves a considerable risk of producing inconsistent estimates.

This paper circumvents the possible endogeneity of *sah* and its associated consequences by specifying the joint p.f. of (y, sah), conditional on a set of exogenous regressors (x). This full information approach can be implemented using copula functions (Sklar, 1959). One advantage of copulas is that they enable separate consideration of the marginal distribution for each dependent variable, as well as their dependence structure. This flexibility makes it possible for researchers to capture the dependence structure of the data without knowing the exact form of the joint p.f., while, at the same time, preserving desirable characteristics of the chosen marginals for the response variables.

Here, the foregoing idea is applied to cross-sectional data taken from the National Health Survey (NHS) of 1998/99. The main goal of the present application is to compare estimates of the impact of *sah* on health care use, obtained from different modelling approaches. The study compares these and other inference results from bivariate (copula-based and mixture) models for the p.f. of (y, sah), given x, as well as from conventional regression count models for the conditional p.f. of y given (sah, x).

The paper is organized as follows. Section two details the main problem and surveys alternative econometric methodologies to deal with it. Section three presents the specification of models for the joint conditional p.f. of (y, sah), suggesting its estimation through ML. This section also includes a very brief account of copula theory, setting the general framework for the proposed specifications. Section four describes the empirical application and comments on its results. Finally, section five concludes the paper.

2. The Problem

2.1 Endogeneity

The present paper addresses the possible endogeneity of *sah* variables in regression models for health care utilization, a concern that poses relevant research issues. Formally, endogeneity (of *sah*) is referred to here according to the following, well established, definition (see Cameron and Trivedi, 1998, ch. 11): let the joint conditional p.f. of $z \equiv (y, sah)$, given *x*, be denoted as $f(z \mid x;\theta)$. The usual factorization has

$$f(z \mid x;\theta) = g(y \mid sah, x;\theta_1) f_2(sah \mid x;\theta_2),$$

where $\theta \equiv (\theta_1, \theta_2)$ denotes a parameter vector. If the marginal p.f. of *sah* depends on θ_1, estimating the parameters θ_1 by conditioning *y* on *sah* does not yield consistent estimates. In this case, *sah* is said to be endogenous.

Why can *sah* be endogenous? Two main arguments are usually invoked, that help explain the plausibility of this concern. The first reason is the possible existence of unobservables that condition individual self-assessments and, at the same time, influence the use of health care. Such factors as individual cultural background, personality characteristics or some dimensions of unmeasured health, like mental and social health (Jurges, 2007) are difficult to measure (hence, not included in the regression model) and likely to influence both the dependent variable and self-judgements. Take, for instance, the case of a hypochondriac individual (usually a characteristic not accounted for): by definition, such a person will tend to display negative feelings towards his/her own health, probably rating it worse than it actually is. At the same time, he/she may also present a clear predisposition to visit the doctor often. In this case, the assumption of independence between *sah* and unobservables influencing *y* beyond the effect of observed covariates does not hold.

The endogeneity of *sah* may also be due to simultaneity of this variable and *y*. It is noted that, under the data collection scheme, individuals evaluate their own health state after visiting the doctor. Expectably, in these visits they acquire objective information that allows them to revise, thus update, their views about their own health.[2] Therefore, it is reasonable to suppose that individuals' health assessments are, to some extent, determined by the quantity of medical care recently received, which gives rise to the simultaneity of *sah* and *y* in the classical demand equation.

2.2 Econometric Choices

Some authors have mentioned the possible endogeneity of *sah* in models for health care use – see Windmeijer and Santos Silva, 1997, Barros, 1999, and Van Ourti, 2004. Each of these papers adopts a different methodological course to meet the issue. While in Barros (1999) *sah* is simply excluded from the regression model, the opposite is proposed in Van Ourti (2004), with *sah* included in the set of regressors, alongside with remaining covariates. As previously mentioned, both approaches incur a serious risk of producing inconsistent estimates, due to the possible misspecification of the regression model for y given (sah, x). Windmeijer and Santos Silva (1997), in turn, do take into account the possible endogeneity of *sah*, resorting to GMM techniques to estimate a regression model for the number of visits to the doctor by individuals.

Addressing endogeneity within a limited information framework usually requires the availability of instrumental variables. For instance, Windmeijer and Santos Silva (1997) suggest using as instruments variables that influence health in the long run, *e.g.*, variables which reflect behavioural attitudes like smoking- and drinking-related variables. Valid instruments are also required for the Hausman test of endogeneity, comparing NLIV to NLS or quasi-ML estimates (see, *e.g.*, Grogger, 1990).

When no valid instruments are available, the above methods fail. Then, one alternative to the foregoing approaches is to adopt a full information strategy, specifying $f(z \mid x)$ and estimating the resulting model through likelihood-based methods. This goal can be achieved using a particular class of cumulative distribution functions (c.d.f.'s) known as copulas. Essentially, a copula function is a joint c.d.f. whose marginals are uniform. In formal terms, the model for the joint conditional c.d.f. of (y, sah) can be expressed as

$$F(y, sah \mid x) = C(F_1(y \mid x), F_2(sah \mid x) \mid x), \tag{1}$$

[2] This information is considered to be objective, because it is provided by the doctor, possibly based on diagnostic tests, like lab tests, x-rays, etc.

where C is the copula, and F, F_1 and F_2 denote, respectively, the joint and marginal c.d.f.'s.

The notion of copula has been well known for some time in statistics. It was introduced in the literature by Sklar (1959), although the main idea dates back to Hoeffding (1940). Its application to the study of economic problems is a recent but fast-growing field, namely in finance (see, *e.g.*, Bouyé, Durrleman, Nikeghbali, Riboulet and Roncalli, 2000). Lee (1983), in one early and seminal paper, was the first to use copulas in econometrics, introducing the "normal copula" as an alternative to Heckman's (1976) two-step procedure of modelling selectivity. General surveys on copulas can be found in Joe (1997), Nelsen (2006) and Trivedi and Zimmer (2005).

As in other areas, the use of copulas in health economics is recent but fast growing. Smith (2003) applies the copula approach to specify models for health care data that may suffer from selectivity bias. Zimmer and Trivedi (2006) use trivariate copulas to specify a regression joint model for three discrete response variables. These are, respectively, two counted measures of health care use by spouses, and a binary variable of insurance status. Dancer, Rammohan and Smith (2007) adopt a similar methodology to assess the degree of dependence between infant mortality and child nutrition. Quinn (2007) addresses the simultaneous determination of mortality risk, health and lifestyles with a reduced-form system of equations, using a copula to define the corresponding multivariate distribution. Other examples in the area of health economics and econometrics are mentioned in the excellent survey by Quinn (2007).

Alternatively, one can use a bivariate mixture model for the specification of $f(z \mid x)$. For instance, the joint p.f. of *sah* and *y* can be obtained upon mixing statistical independence, conditional on unobserved heterogeneity. Formally,

$$f(y, sah \mid x) = \int f_1(y \mid x, \varepsilon) f_2(sah \mid x, \varepsilon) h(\varepsilon \mid x) d\varepsilon\,, \qquad (2)$$

where f_1 and f_2 represent the marginal p.f.'s and ε denotes unobserved heterogeneity, with density *h*. Except for some particular cases, one disadvantage associated with this approach is that it generally leads to criterion functions without analytical expressions, which require simulation-based or numerical approximation methods of maximization. On the other hand, such a specification enables the control of rich heterogeneity structures.

Actually, a mixture joint model can be given a copula interpretation, with the copula function implicitly defined by $F(y, sah \mid x) = \sum_{i \le y} \sum_{j \le sah} f(i, j \mid x)$, and *f* as in (2). The next section presents the specification of a mixture model that is used in the present application and details its interpretation as a copula-based model.

3. Model Specification

This section presents models for the joint conditional p.f. of (*y*,*sah*), given a set of regressors. The section begins with a brief presentation of bivariate copulas, setting the general framework for the proposed copula-based models and subsequent empirical application.

3.1 Copulas

The main finding of copula theory is the fact that the joint c.d.f. of a set of real-valued random variables (r.v.'s) can be separated into its marginal c.d.f.'s and a copula, describing their dependence structure. More precisely, an *l*-variate copula (or *l*-copula) is defined as the c.d.f. of a random *l*-vector with uniform marginal c.d.f.'s. In the bivariate case, a 2-copula is a function $C:[0,1]^2 \mapsto [0,1]$ that satisfies the following properties:

i. For every $u \equiv (u_1, u_2) \in [0,1]^2$,

$C(u) = 0$, if at least one coordinate of *u* is zero;

$C(1, w) = C(w, 1) = w, w \in [0,1]$.

ii. $\forall (a_1, a_2), (b_1, b_2) \in [0,1]^2$, $a_j \le b_j, j = 1,2,$ $\Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} C(v) \ge 0$, where the two first-order differences of the function *C* are defined, respectively, as

$$\Delta_{a_1}^{b_1} C(v) \equiv C(b_1, v_2) - C(a_1, v_2), \;\; \Delta_{a_2}^{b_2} C(v) \equiv C(v_1, b_2) - C(v_1, a_2).$$

Expression $\Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} C(u)$ is naturally interpreted as $\Pr(a_1 \le u_1 \le b_1, a_2 \le u_2 \le b_2)$.

If *F* is a bivariate c.d.f. with margins F_1, F_2, then, there exists a 2-copula *C* such that, for any random vector $z \equiv (z_1, z_2) \in R^2$,

$$F(z_1, z_2) = C(F_1(z_1), F_2(z_2)).$$

If F_1, F_2 are continuous, then C is unique; otherwise C is uniquely determined on $RanF_1 \times RanF_2$ (*Ran G* denotes the range of the function *G*). Conversely, if *C* is a 2-copula and F_1, F_2 are c.d.f.'s, then the function *F* defined above is a bivariate c.d.f. with marginal c.d.f.'s F_1, F_2.

The above statement is the bivariate version of what is known as Sklar's theorem. It demonstrates the role of copulas as the link between multivariate distributions and their univariate margins. The result essentially follows from the probability integral transformation, under which, for a continuous random variable *w* with c.d.f. *F*, $F(w)$ is uniformly distributed over the range $(0,1)$. The theorem enables the construction of a joint c.d.f., once the marginal c.d.f.'s and copula are available.

The copula is not unique if any of the marginal c.d.f.'s exhibits discontinuities – as is the case for discrete r.v.'s (see Joe, 1997, p. 14, for details). Nevertheless, as Zimmer and Trivedi (2006, p. 64) point out, the non-uniqueness of copula in such cases is a theoretical issue that does not hinder its use in empirical applications. Finding a unique copula representation rests on full knowledge of the joint c.d.f.. Now, one of the reasons why researchers use copulas is precisely the fact that they ignore the true form of the joint c.d.f.. Thus, once the researcher decides which marginals to adopt, the issue, for him, is one of finding a copula that is able to reflect the dependence structure of the data while preserving desirable features of those marginals.

Given the purpose of the present paper, conditional c.d.f.'s and copulas must be considered. A bivariate conditional copula is a function $C:[0,1]^2 \mapsto [0,1]$, such that, conditional on some set (name it *H*), *C* corresponds to the above definition of copula. Sklar's theorem for conditional distributions leads to (see, *e.g.*, Patton, 2005)

$$F(z \mid H) = C(F_1(z_1 \mid H), F_2(z_2 \mid H) \mid H).$$

As previously mentioned, the copula describes the dependence structure of r.v.'s with a given joint c.d.f.. One trivial but important case is the bivariate product copula, $\Pi(u) \equiv u_1 u_2$, that results in case of independence. The close relationship between copulas and dependence is also reflected by the Fréchet-Hoeffding bounds inequality: for every copula *C* and every $u \in [0,1]^2$, it can be shown that (see, *e.g.*, Nelsen, 2006)

$$W_2(u) \equiv \max\{u_1 + u_2 - 1, 0\} \le C(u) \le \min\{u\} \equiv M_2(u).$$

Both bounds are themselves copulas in the bivariate case; the upper (lower) bound arises if and only if one r.v. is almost surely a strictly increasing (decreasing) transformation of the other. Between the extremes of independence and monotone functional dependence many forms of dependence can be considered, that are described by the properties of copulas. Besides the familiar notion of linear correlation, several dependence concepts and measures have been proposed in the literature (see Joe, 1997, for an extended survey). For present purposes it suffices to distinguish "positive" from "negative" bivariate dependence – with positive dependence expressing the idea that "large" (or "small") values of both r.v.'s tend to occur together, and negative dependence expressing the notion that "large" values of one r.v. tend to be associated with "small" values of the other.

In practice, marginal c.d.f.'s can be specified conditional on a set of regressors, leading to a conditional copula representation for the joint (conditional) c.d.f. of the dependent r.v.'s of interest. In addition, the copula can include one or more parameters intended to capture the dependence between the univariate margins – usually, in the bivariate case, a single dependence parameter is used.

Interpreting the dependence parameter of a copula in the discrete case is not as straightforward as for continuous r.v.'s. In the latter case, the dependence parameter is frequently converted into a concordance measure, such as Kendall's tau or Spearman's rho, both defined on the interval $[-1,1]$ and independent of the functional form of the margins. However, as shown by several authors (*e.g.* Marshall, 1996, Denuit and Lambert, 2005), this is not so with discrete data, for which these measures are no longer bounded on the above interval, and are sensitive to the choice of margins. Still, every copula defines a range of permissible values for its dependence parameter, thereby allowing for varying degrees of positive and/or negative dependence. Thus, a researcher should choose those families of copulas that best fit his intended application, being able to capture the dependence pattern in the available data.

3.2 Model Specification

This section presents several alternative specifications for the conditional c.d.f. $F(z \mid x)$, where $z \equiv (y, sah)$. Starting with copula-based models, the bivariate probabilistic model can be generally expressed as in (1),

$$F(z \mid x;\theta,\delta) = C(F_1(y \mid x_1;\theta_1), F_2(sah \mid x_2,\theta_2);\delta),$$

where $x \equiv (x_1', x_2')'$ represents the vector of conditioning variables (including intercept terms in both x_1 and x_2), $\theta \equiv (\theta_1', \theta_2')'$ denotes the vector of the margins' parameters, and δ represents a dependence parameter.

In the present application, *y* is a count variable with unbounded support. Following common practice (see Cameron and Trivedi, 1998), the function $F_1(y \mid x_1;\theta_1)$ is specified as the c.d.f. of a negative binomial p.f. with conditional mean $E(y \mid x_1) \equiv \mu_y = \exp(x_1'\beta_1)$ and variance $V(y \mid x_1) = \mu_y + \alpha\mu_y^2, \alpha > 0$. Formally, the marginal p.f. of *y* can be expressed as

$$f_1(y \mid x_1;\theta_1) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)\Gamma(y+1)}\left(\frac{\alpha}{\mu_y+\alpha}\right)^{\alpha}\left(\frac{\mu_y}{\mu_y+\alpha}\right)^{y}, \qquad (3)$$

with $\theta_1 \equiv (\beta_1', \alpha)'$. As is well known, this functional form allows for overdispersion in the data, with reference to the Poisson p.f. (which results for $\alpha = 0$), thereby providing considerable modelling flexibility.

The second dependent r.v., *sah*, is a rank variable ranging from 1 to 5. Again following established literature, its marginal p.f. is specified as ordered probit, conditional on x_2 (see *e.g.*, Maddala, 1983). Under this specification,

$$\Pr(sah = j \mid x_2;\theta_2) = \Phi(\lambda_{j+1} - x_2'\beta_2) - \Phi(\lambda_j - x_2'\beta_2),\ j = 1,\ldots,5, \qquad (4)$$

with $\Phi(\cdot)$ denoting the standard normal c.d.f., $\theta_2 \equiv (\beta_2', \lambda')'$, $\lambda \equiv (\lambda_2,\ldots,\lambda_5)'$, $\lambda_1 = -\infty$ and $\lambda_6 = \infty$. From this it follows

$$\begin{aligned}&F_2(j \mid x_2;\theta_2) =\\ &\Pr(sah \le j \mid x_2;\theta_2) = \textstyle\sum_{k=1}^{j}\Pr(sah = k \mid x_2;\theta_2) =\\ &\Phi(\lambda_{j+1} - x_2'\beta_2),\ j = 1,\ldots,5.\end{aligned}$$

As usual, identification requires a normalization, such as 0, for the intercept term in β_2 or one of the λ's.

The next step towards full specification of the c.d.f. of *z* consists on the choice of copula. In the present context, *y* and *sah* may well tend to move in opposite directions, thereby producing negative dependence in the data. This suggests the convenience of choosing a copula that allows for both positive and negative dependence. Among (few) others, two possible choices are the Frank copula (Frank, 1979) and the Farlie-Gumbel-Morgenstern (FGM) copula, first proposed by Morgenstern (1956). The formal expressions for these copulas can be written, respectively, as

Frank Copula

$$C(u_1,u_2;\delta)=\begin{cases}\dfrac{-1}{\delta}\log\left(1+\dfrac{(\exp(-\delta u_1)-1)(\exp(-\delta u_2)-1)}{\exp(-\delta)-1}\right), & \delta\neq 0,\\ u_1u_2, & \delta=0,\end{cases} \tag{5}$$

FGM copula

$$C(u_1,u_2;\delta)=u_1u_2(1+\delta(1-u_1)(1-u_2)),\quad |\delta|\leq 1. \tag{6}$$

Both functions nest the independence copula, which results for $\delta=0$. Positive and negative dependence occur with, respectively, $\delta>0$ and $\delta<0$. The Frank copula attains the Fréchet-Hoeffding upper and lower bounds, under, respectively, $\delta\to\infty$ and $\delta\to-\infty$. Despite its simplicity, the FGM copula is more restrictive, in that the dependence parameter is bounded on $[-1,1]$ and does not lead to either Fréchet-Hoeffding bound.

Let $(u_1,u_2)=(F_1(y\mid x_1),F_2(sah\mid x_2))$. Then, $F(z|x)$ immediately results by plugging $F_1(y\mid x_1)$ and $F_2(sah\mid x_2)$ into (5) or (6).

The joint conditional p.f. of (y,sah) can also be obtained as a bivariate mixture model. Conditional on x and unobserved heterogeneity, $\varepsilon\equiv(\varepsilon_1,\varepsilon_2)$, (y,sah) are assumed independent, with the same conditional margins as before: $y\mid(x_1,\varepsilon_1)$ is distributed as in (3), but with $E(y\mid x_1,\varepsilon_1)=\exp(x_1'\beta_1+\varepsilon_1)$, and $\Pr(sah=j\mid x_2,\varepsilon_2;\theta_2)$ $=\Phi(\lambda_{j+1}-x_2'\beta_2-\varepsilon_2)-\Phi(\lambda_j-x_2'\beta_2-\varepsilon_2)$, $j=1,\ldots,5$. Then, with $(\varepsilon_1,\varepsilon_2)$ assumed bivariate normal, independent of the regressors, with null mean vector, common variance, σ^2, and correlation coefficient δ, the model results as

$$\begin{gathered}f(y,sah\mid x;\theta,\sigma^2,\delta)=\\ \int f_1(y\mid x_1,\varepsilon_1;\theta_1)f_2(sah\mid x_2,\varepsilon_2;\theta_2)\phi_{\sigma^2,\delta}(\varepsilon_1,\varepsilon_2)d\varepsilon_1 d\varepsilon_2,\end{gathered} \tag{7}$$

where $\phi_{\sigma^2,\delta}$ denotes the bivariate normal density with parameters (σ^2,δ).

This formulation is naturally equivalent to a model with random intercepts in f_1 and f_2. The assumption of Gaussian heterogeneity is common in the literature (see, *e.g.*, Train, 2003). Although estimation is computationally demanding, requiring simulation-based methods or numerical approximations, the specification leads to easily interpretable parameters, namely the dependence parameter, δ. Within this framework, independence can easily be checked with the usual statistical tests. The assumptions of common variance and of independence from regressors do not seem unreasonable in the present context and add to computational convenience; other schemes can be considered, such as random coefficients (other than the intercepts), varying dispersion parameters and/or dependence with respect to regressors. However, the usefulness of such sophistications in the present context is questionable, namely in view of the added estimation difficulty they are bound to represent. In any case, it is noted that two correlated heterogeneity terms are allowed for, instead of a shared term in f_1 and f_2. In the present context, these terms can naturally be seen as correlated unobserved heterogeneity influencing both *y* and *sah*. The assumption is also useful because it enables the discernment of negative from positive dependence in the data (through the sign of δ), not just whether or not there is dependence (as the case would be with just one term).

As previously mentioned, the mixture model can be given a copula interpretation. In this case, the function *C* in (1) is defined as

$$F(y, sah \mid x) =$$
$$\sum_{i=0}^{y}\sum_{j=1}^{sah} f(i, j \mid x) = \int F_1(y \mid x_1, \varepsilon_1) F_2(sah \mid x_2, \varepsilon_2) \phi_{\sigma^2,\delta}(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 =$$
$$\int \Pi(F_1(y \mid x_1, \varepsilon_1), F_2(y \mid x_2, \varepsilon_2)) \phi_{\sigma^2,\delta}(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \,,$$

where F_k, $k = 1,2$, now denote the marginal c.d.f.'s given (x_k, ε_k), and Π denotes the (conditional) independence copula.

3.3 Estimation

Maximum likelihood (ML) estimation of the above models requires the joint p.f. of (*y*,*sah*), given *x*, $f(z \mid x)$. Under copula-based models for continuous response variables this is obtained as the second-order derivative of the copula, that is (conditioning on *x* is omitted),

$$f(z_1, z_2) = \frac{\partial^2 F(z_1, z_2)}{\partial z_1 \partial z_2} = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} f_1(z_1) f_2(z_2),$$

where $(u_1, u_2) \equiv (F_1(z_1), F_2(z_2))$. In the present case, involving discrete r.v.'s, $f(z \mid x)$ is formed by taking differences. Formally,

$$f(z_1,z_2)=$$
$$F(z_1,z_2)-F(z_1-1,z_2)-F(z_1,z_2-1)+F(z_1-1,z_2-1)=$$
$$C(F_1(z_1),F_2(z_2))-C(F_1(z_1-1),F_2(z_2))-C(F_1(z_1),F_2(z_2-1))+C(F_1(z_1-1),F_2(z_2-1)).$$

Then, upon the choice of copula, the individual contribution to the log-likelihood is formed by taking the logarithm of this last expression. After simultaneous ML estimation of all the parameters, variances of the estimates are obtained through the robust sandwich formula. It is noted that, defined as above, both the Frank and FGM copulas are differentiable to order two at any particular value of δ, so independence can be assessed with the usual likelihood-based tests.

Estimation of model (7) requires either maximum simulated likelihood (MSL) or numerical approximation. The former is used here, with (7) being approximated by direct Monte Carlo (MC) integration, that is,

$$f(y,sah\mid x;\theta,\sigma^2,\delta)\approx\frac{1}{S}\sum_{s=1}^{S}f_1(y\mid x_1,\varepsilon_1^s;\theta_1)f_2(sah\mid x_2,\varepsilon_2^s;\theta_2),\tag{8}$$

where $(\varepsilon_1^s,\varepsilon_2^s)$, $s=1,\ldots,S$ denote random draws from the bivariate normal, $\phi_{\sigma^2,\delta}$, and *S* is the number of draws. Gouriéroux and Monfort (1991) show that, under regularity assumptions, the MSL estimator has the same asymptotic distribution as the ordinary ML estimator, provided that $\sqrt{n}/S\rightarrow 0$ as $n,S\rightarrow\infty$ (*n* denotes sample size). The number of draws used in the present application is selected *ad hoc*, mostly for reasons of computational convenience, on the basis of rough comparisons between results for various *S* values.

4. Empirical Study

This section presents empirical results from the application of the described approach to cross-sectional data on (y,sah) and a set of regressors, *x*. The data are taken from the Portuguese National Health Survey (NHS) conducted in 1998/99.[3] The main goal of the study is to estimate, via this methodology, the impact of *sah* on *y*, comparing its inference results with those from the mainstream conditional count models for *y*, given (sah,x).

[3] Details about the survey can be consulted, among others, in Barros *et. al* (2005), and Ministério da Saúde - Instituto Nacional de Saúde (1999).

4.1 Data and Summary Statistics

Tables 1 and 2 present summary statistics for, respectively, *y* and *sah*, and the set of regressors, *x*. The sample contains 27,044 observations, obtained after deleting incomplete records on any of the variables used in the study. As previously mentioned, *y* denotes the number of visits to the doctor in the last three months before the survey interview. The variable *sah* is a rank variable, ranging from 1 ("very bad" self-assessed health) to 5 ("excellent" self-assessed health). The covariates are described in table 2, being grouped under four headings: socioeconomic variables, health status variables, a variable measuring the supply of medical care services, and a binary variable indicating health insurance status. As detailed in the table legend, the marginals of the joint model do not share all regressors $(x_1 \neq x_2)$.[4] The selection of covariates for each marginal follows well-established research developed by other authors (Grossman, 1972; Muurinen, 1982; Wagstaff, 1986; Barros, 2003). Besides economic and behavioural criteria, practical considerations, such as data availability and computational tractability, are also relevant for the choice of covariates.

Table 1 – Dependent Variables

y	Rel. Freq.	*sah*	Rel. Freq.
0	.413		
1	.247	1 (very bad)	.045
2	.138	2 (bad)	.174
3	.104	3 (fair)	.379
4	.038	4 (good)	.363
5	.022	5 (excellent)	.039
6	.018		
7	.005		
8	.004		
9	.001		
10	.004		
11	.0004		
12	.003		
> 12	.003		
	sample size	27,044	
average	1.42	average	3.18
variance	4.33	variance	.84

[4] In any case, all *x* regressors are included in the regression models for *y* given (*sah*,*x*), to ensure that results from the two approaches are comparable.

Table 2
Regressors – Definition and Summary Statistics

Variable name	Variable Definition	average	st.dev.	min.	max.
	Socioeconomic				
age	Age, in years, divided by 10	4.26	2.49	0	9.50
agesq (2)	*age* squared	24.29	20.97	0	90.25
apr	= 1 if the individual lives in a rural area	.18	.39	0	1
educ	Years of schooling. If child, years of schooling of the most educated adult in the household.	5.68	4.27	0	24
fem	= 1 if the individual is female	.60	.49	0	1
inc	Monthly real income (unit: 100 euros)	3.63	2.77	.23	24.94
married (1)	= 1 if the individual is married	.54	.50	0	1
notw (1)	= 1 if the individual did not work in the two weeks prior to taking the survey	.67	.47	0	1
ret	= 1 if the individual is retired	.23	.42	0	1
	Health Status				
lim	= 1 if the individual has some physical handicap that prevents him from executing certain physical daily activities	.04	.19	0	1
nchrd	Number of chronic conditions reported	.96	1.03	0	6
nodent (2)	= 1 if the individual has no dental hygiene habits	.06	.23	0	1
noph (2)	= 1 if the individual's daily activities require no physical activity	.58	.49	0	1
smoke (2)	= 1 if the individual smokes on a daily basis	.11	.31	0	1
srill	= 1 if the individual reports being ill in the previous two weeks	.37	.48	0	1
stress	= 1 if the individual took sleeping pills in the last two weeks	.12	.33	0	1
	Supply Side				
ph1000 (1)	Total number of licensed physicians per 1000 inhabitants	2.75	2.22	.58	9.15
	Insurance Status				
nhs	= 1 if the individual is covered only through the NHS	.84	.36	0	1

(1) Regressor in $f_1(y|x_1)$ but not in $f_2(sah|x_2)$. (2) Regressor in $f_2(sah|x_2)$ but not in $f_1(y|x_1)$.

4.2 Estimation Results

Estimation results are presented in tables 3 and 4. The first table contains estimates for the parameters in $g(y \mid sah, x)$, whereas the estimation results from joint models for $f(y, sah \mid x)$ are included in table 4. All computations were performed using TSP 5.0 (Hall and Cummins, 2005).

In what concerns $g(y \mid sah, x)$ estimation, a noticeable result, often found in the literature, is the clear rejection of the Poisson model in favour of the NB2 model ($\hat{\alpha} = .622$, statistically significant). This result can be taken as indication of overdispersion in the data, with reference to the more restrictive Poisson specification. Thus, even in the case of a correct $E(y \mid sah, x)$ under both models, NB2's estimates are preferable to those of the Poisson.

Expectably, the estimated coefficients of x_1 are quite different in models NB2 for $g(y \mid sah, x)$ and for $f_1(y \mid x_1)$ within joint models. Actually they are not even comparable, as they do not refer to the same quantity: in the former case each coefficient estimates the relative change of the conditional mean of *y*, given (sah, x), whereas, in the latter, each estimate refers to the relative change of the conditional mean of *y* marginal to *sah*.

With regard to full information approaches, it is noted that estimates from Frank and FGM models are almost identical – except for the dependence parameter (because of functional form differences between both copulas). The close resemblance to the results from the mixture model is also noticeable – with the exception of the dependence parameter and the overdispersion parameter, α (possibly due to the presence of ε_1, capturing part of the effects of unobserved heterogeneity). This similarity may be a consequence of the flexibility of copula functions, which are able to discern dependence from the marginals. Meanwhile, the (small) differences to the estimates from the mixture model may be due to the fact that the latter are not ML estimates, being obtained from maximization of an approximate the log-likelihood function. The MSL estimates for the mixture model are obtained using $S = 100$ draws of pseudo-random vectors from the bivariate normal. This number of draws is selected for computational convenience and from rough comparisons with results for larger *S* (*e.g.*, $S = 250$ leads to virtually the same estimates and standard errors). Results might be closer to those from Frank and FGM models with a significantly larger *S* but, no doubt, this would increase the computational burden. Here, instead of direct MC sampling, it may prove more efficient to use so-called "quasi-MC" methods (*e.g.*, Halton sequences) to approximate the integrals in the likelihood and estimate the model.[5]

[5] According to previous studies (*e.g.*, Bhat, 2001, Train, 2003), Halton draws can be computationally much more efficient than direct MC sampling.

With regard to regressors' coefficients in table 4, some estimates point to a varying degree of relevance of the corresponding covariates in the two margins: the variable *inc* (income) is irrelevant to explain health care use but it seems highly relevant in $f_2(sah \mid x_2)$; the opposite occurs with respect to *apr* (residence in a rural area), which is relevant in $f_1(y \mid x_1)$ and irrelevant in $f_2(sah \mid x_2)$. On the other hand, *nhs* (membership exclusively in the statutory public system) shows little relevance in both $f_1(y \mid x_1)$ and $f_2(sah \mid x_2)$.

Overall, estimation results for the joint models are in line with the usual findings in the literature. In general, worse-off individuals in terms of health status seek medical care more often (see the sign of covariates *lim* (+), *nchrd* (+), *srill* (+) and *stress* (+)) than those in better health. Nevertheless, higher income levels or education degrees are both linked to an increase in demand for health care.

Table 3 – Estimation Results – $g(y \mid sah, x)$

	Poisson		NB2	
Variable	coefficient	st. error	coefficient	st. error
intercept	.859	.077	.872	.065
sah	-.339	.014	-.352	.011
age	-.103	.022	-.147	.017
agesq	.007	.002	.012	.002
apr	-.040*	.022	-.055	.020
educ	.020	.003	.024	.003
fem	-.009**	.020	.022**	.016
inc	.012	.003	.014	.003
lim	.087	.043	.065*	.036
married	.100	.023	.121	.019
nchrd	.128	.009	.151	.008
nhs	-.050*	.027	-.042*	.023
nodent	-.094	.038	-.110	.033
noph	.085	.023	.089	.019
notw	.120	.025	.100	.021
ph1000	.021	.004	.022	.003
ret	.066	.026	.075	.025
smoke	-.125	.034	-.110	.027
srill	.479	.018	.488	.016
stress	.275	.023	.292	.021
α	-	-	.619	.013
Log-likelihood	-45198.4		-41440.8	
SBIC	45300.5		41547.9	

* Not significant at the .05 level. ** Not significant at the .10 level.

Table 4

Estimation Results – $f(y, sah \mid x)$

Model	Frank		FGM		Mixture	
Variable	coefficient	st. error	coefficient	st. error	coefficient	st. error
$f_1(y \mid x_1)$						
intercept	-.576	.039	-.573	.039	-.864	.045
age	-.009	.004	-.008*	.004	-.011	.005
fem	.060	.014	.059	.014	.071	.016
married	.084	.015	.084	.015	.100	.017
educ	.013	.002	.014	.002	.013	.002
ret	.146	.022	.140	.022	.157	.024
ph1000	.022	.003	.022	.003	.022	.004
inc	.003**	.003	.003**	.003	.005**	.004
notw	.139	.017	.146	.017	.166	.019
apr	-.048	.018	-.047	.018	-.057	.022
srill	.646	.014	.646	.014	.654	.016
lim	.233	.033	.232	.032	.196	.037
nchrd	.217	.008	.216	.008	.232	.008
stress	.375	.020	.373	.021	.398	.022
nhs	-.021**	.020	-.031**	.021	-.036**	.023
α	.684	.011	.683	.011	.158	.013
$f_2(sah \mid x_2)$						
intercept	3.755	.041	3.753	.041	4.060	.052
age	-.254	.013	-.254	.013	-.263	.014
agesq	.014	.002	.014	.002	.013	.002
fem	-.107	.015	-.107	.015	-.121	.016
educ	.045	.002	.046	.002	.046	.003
ret	-.246	.022	-.245	.022	-.266	.024
inc	.043	.003	.042	.003	.055	.003
apr	.004**	.018	.004**	.018	.013**	.020
noph	-.093	.018	-.092	.018	-.112	.019
nchrd	-.326	.008	-.327	.008	-.365	.009
stress	-.354	.021	-.355	.022	-.395	.024
nodent	-.147	.028	-.149	.028	-.147	.032
srill	-.702	.015	-.704	.015	-.769	.017
lim	-.811	.036	-.812	.036	-.866	.042
smoke	.042*	.023	.042*	.023	.055	.024
nhs	-.056	.022	-.051	.022	-.039*	.021
λ_3	1.351	.017	1.355	.017	1.461	.021
λ_4	2.998	.020	3.003	.020	3.246	.030
λ_5	5.064	.025	5.071	.025	5.498	.044
σ^2	-	-	-	-	.682	.025
δ	-1.498	.048	-.697	.022	-.693	.022
Log-likelihood	-67666.2		-67684.9		-67571.9	
SBIC	67701.9		67705.3		67597.4	

* Not significant at the .05 level. ** Not significant at the .10 level.

The dependence parameter estimate is both negative and significant across the three joint models. As expected, all three models point to a negative dependence between *y* and *sah*, after conditioning on observed factors. The precision of the estimates suggests evidence of simultaneity of both variables, which, as previously mentioned, can cause endogeneity of *sah* within limited information approaches.

Negative dependence between *y* and *sah* also seems visible in the estimates for the average effects, included in table 5. The figures reported in the table refer to $\hat{E}(y \mid sah, \bar{x})$, for the five different values of *sah* and the sample averages of *x*. The values in the table are computed, respectively, as $\exp\left((sah, \bar{x}')\hat{\beta}\right)$ under the Poisson and NB2 models, and as $\sum_{y=0}^{30} y\hat{f}(y, sah \mid \bar{x}) \Big/ \sum_{y=0}^{30} \hat{f}(y, sah \mid \bar{x})$ under the joint models, where $\hat{f}$ denotes evaluation of *f* at the estimated parameters (for *y* > 30 both summands in the fraction are negligible). Under the mixture model, $f(y, sah \mid \bar{x})$ is approximated by (8). Unreported results indicate that the table estimates are almost identical to estimates obtained for *nhs* = 0 and *nhs* = 1. This suggests that, for an "average" individual, being covered only by the NHS or not, has no bearing on the impact of his own *sah* on health care utilization. The various specifications produce somewhat different estimates of the average effects, in spite of the negative impact of *sah*, consistently obtained across models. Under full information models these differences are somewhat noteworthy, in view of the aforementioned similarity of estimation results these models produce.

Table 5
Average Effects – $E(y \mid sah, \bar{x})$

	Model				
sah	Poisson	NB2	Frank	FGM	Mixture
1	2.538	2.589	2.835	3.368	2.255
2	1.809	1.820	2.758	3.238	1.606
3	1.290	1.279	2.349	2.485	1.191
4	.919	.899	1.891	1.494	.928
5	.655	.632	1.749	1.124	.680

Finally, it is interesting to see how the estimated models fit the data. To give an idea of the goodness of fit of the models, table 6 gives the true and fitted frequencies of the number of visits to the doctor. The fitted frequencies distribution is obtained as the average over observations of the predicted probabilities fitted for each count. Formally, $n^{-1}\sum_{i=1}^{n} f_1(y \mid sah_i, x_i)$, for y = 0, 1, 2, …; under models of the joint p.f. of (y, sah) this is computed as $n^{-1}\sum_{i=1}^{n} (f(y, sah_i \mid x_i)/f_2(sah_i \mid x_{2i}))$. Both the joint models and the NB2 model fit the data relatively well, being particularly good at predicting the number of individuals with few visits (up to 2). For y = 3 these four models under-predict the actual frequency, while the reverse occurs for $y > 3$.

Table 6 – Actual and Fitted Frequencies

		Model				
Visits	Actual	Poisson	NB2	Frank	FGM	Mixture
0	.413	.313	.414	.413	.413	.412
1	.247	.311	.256	.254	.254	.263
2	.138	.189	.140	.140	.140	.139
3	.104	.097	.076	.077	.078	.073
4	.038	.047	.043	.044	.044	.041
> 4	.060	.043	.071	.072	.071	.072

The statistical significance of the differences between actual and fitted frequencies can be assessed using a test for the joint moment conditions

$$\begin{cases} E(d_j(y) - \Pr(y = j \mid sah, x)) = 0, \ j = 0, ..., 4, \\ E(d_5(y) - \Pr(y > 4 \mid sah, x)) = 0, \end{cases}$$

with the binary variables d_j defined as $d_j(y) = 1$, if $y = j$, j = 0, …, 4, and $d_5(y) = 1$, if $y > 4$.[6] In order to try and reduce the effects of a large sample size on the outcome of the test, it is carried out with a sub-sample of about 25% of the initial size (6,436 observations). For each specification, the results of the test, asymptotically distributed as a chi-squared distribution with 5 degrees of freedom, are the following:

[6] For details on how to implement this type of tests and a simulation on their performance see Cameron and Trivedi (1998).

Poisson	NB2	Frank	FGM	Mixture
589.75	144.62	121.53	120.32	196.41

The null hypothesis is clearly rejected across all specifications, in spite of the reduced sample size. Nevertheless, the outcomes of the test suggest an ordering of the models, with the Poisson displaying a worse result than the remaining models, NB2 included. In accordance with the results of table 6, the NB2 model competes well with joint models, performing even better than the mixture model.

5. Conclusion

The study of the relevant factors influencing health care utilization constitutes one of the main research interests in health economics. In this context, the measurement of the impact of self-assessed health on the demand for health care stands as an important issue that requires careful methodological approaches. In particular, the possible endogeneity of *sah* within regression models for health care utilization should be met with GMM-type methods, usually requiring available instruments.

Alternatively, this issue can be circumvented by specifying the joint p.f. of (y, sah), conditional on a set of exogenous regressors (x). This full information approach is implemented here with copula functions, which enable separate consideration of the marginal distribution for each dependent variable, as well as their dependence structure.

The results of the paper indicate that copula-based models, though fully parametric, appear considerably flexible, seemingly able to capture the dependence structure in the data. The precision with which the dependence parameter is estimated across the joint models used here reinforces the suspicion of simultaneity of (y,sah), casting founded doubts on the appropriateness of conventional (NLS or conditional ML) methods. Thus, one conclusion of the foregoing results is that, when estimating the impact of *sah*-type covariates on health care use, then, either GMM or full information ML seems more trustworthy than conventional, limited information approaches. In any case, as the paper results also indicate, ML-based results for the NB2 model compete fairly well with the former methods. This finding, not unusual in the literature on count data, is a likely consequence of the presence of the overdispersion parameter in the NB2 model, greatly adding to its flexibility and potential goodness-of-fit, namely when compared to the more restrictive Poisson count model. If a researcher prefers to move along conventional lines, then a negative binomial specification seems a wise enough choice, clearly preferable to the Poisson. Nevertheless, dependence should be checked, as it may be an indication of causes for the endogeneity of *sah*. One way to do this, which avoids estimation of the joint model, is to compute a score test for independence upon estimation of the marginal models for y and *sah*.

The present study suggests some ideas for future applied research. One of these consists on the extension of copula models to the trivariate case, in order to try to deal with the possibility of sample selection within the NHS dataset, apart from the present endogeneity issue. In this sense, the present study can constitute a first step in that direction, which, in itself a complex issue, may well benefit from some of the ideas and methods set forth in the present work.

References

Bago D'Uva, T. (2006), “Latent Class Models for Utilization of Health Care”, *Health Economics,* 15**,** 329-343.

Barros, P. P. (1999), “Os sistemas privados de saúde e a reforma do sistema de saúde”, in Associação Nacional de Sistemas de Saúde, ed., *O Papel dos Sistemas Privados de Saúde num Sistema em Mudança*, Lisboa, ANSS.

Barros, P. P. (2003), “Estilos de vida e estado de saúde: uma estimativa da função de produção de saúde”, *Revista Portuguesa de Saúde Publica,* 3.

Barros, P. P., M. P. Machado, A. Galdeano (2005), “Moral Hazard and the Demand For Health Services: A Matching Estimator Approach”, *Working paper 05/59*, Departamento de Economia, Universidade Carlos III de Madrid.

Bhat, C. R. (2001), “Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model”, *Transportation Research: Part B: Methodological*, 35, 677-693.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet and T. Roncalli (2000), “Copulas for Finance: A Reading Guide and Some Applications”, Unpublished Manuscript, London, Financial Econometrics Research Centre, City University Business School.

Cameron, A. C., P. K. Trivedi (1998), *Regression analysis of count data*, New York, Cambridge University Press.

Dancer, D., A. Rammohan, Smith (2007), “Infant Mortality and Child Nutrition in Bangladesh: Modelling Sample Selection Using Copulas”, 16th European Workshop on Econometrics and Health Economics, Bergen, Norway.

Deb, P., P. K. Trivedi (1997), “Demand for Medical Care by the Elderly: A Finite Mixture Approach”, *Journal of Applied Econometrics,* 12**,** 313-336.

Deb, P., P. K. Trivedi (2002), “The Structure of Demand for Health Care: Latent Class Versus Two-part Models”, *Journal of Health Economics,* 21**,** 601-625.

Denuit, M., P. Lambert (2005), “Constraints on concordance measures in bivariate discrete data”, J. Multivariate Analysis, 93, 40–57.

Frank, M. J. (1979), “On the Simultaneous Associativity of $F(x,y)$ and $x + y - F(x,y)$”, *Aequationes Mathematicae*, 19, 194-226.

Gerdtham, U. G., G. Sundberg (1998), “Equity in the Delivery of Health Care in Sweden”, *Scandinavian Journal of Social Medicine*.

Gouriéroux, C., A. Monfort (1991), “Simulation Based Econometrics in Models with Heterogeneity”, *Annales d’Economie et de Statistique*, 20, 69-107.

Grogger, J. (1990), “A Simple Test for Exogeneity in Probit, Logit and Poisson Regression Models”, *Economics Letters*, 33, 329-332.

Grossman, M. (1972). “Concept of Health Capital and Demand for Health”, Journal of Political Economy 80(2): 223-225.

Hall, B. H., C. Cummins (2005), TSP User's Guide, Version 5, TSP International, Palo Alto, Ca.

Heckman, J. (1976), “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models”, *Annals of Economics and Social Measurement*, 5, 475-492.

Hoeffding, W. (1940), "Masstabinvariante Korrelationstheorie", Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin, 5, Heft 3, 179-233. [Reprinted as "Scale-invariant Correlation Theory", in

N. I. Fisher, P. K. Sen (eds.), *The Collected Works of Wassily Hoeffding*, New York, Springer.]

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, New York, Chapman & Hall.

Jurges, H. (2007), "True Health vs Response Styles: Exploring Cross-country Differences in Self-reported Health", *Health Economics,* 16**,** 163-178.

Lee, L. (1983), "Generalized Econometric Models With Selectivity", *Econometrica*, 51, 507-512.

Lourenço, O. D., C. Quintal, P. Ferreira, P. P. Barros (2007), "A equidade na utilização de cuidados de saúde em Portugal: Uma avaliação baseada em modelos de contagem", *Notas Económicas,* 25**,** 6-26.

Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.

Marshall A. (1996), "Copulas, Marginals and Joint Distributions", in L. Ruschendorf, B. Schweizer, M. D. Taylor, eds., *Distributions With Fixed Margins and Related Topics*, Hayward, Ca., Institute of Mathematic Statistics, 213-222.

Ministério da Saúde – Instituto Nacional de Saúde (1999), *INS 1998/1999. Continente. Dados Gerais.*, INSA – Instituto Nacional de Saúde.

Morgenstern, D. (1956), "Einfache Beispiele Zweidimensionaler Verteilungen", *Mitteilingsblatt für Mathematische Statistik*, 8, 234-235.

Muurinen, J.-M. (1982), "Demand for health: A generalised Grossman model", Journal of Health Economics 1(1): 5-28.

Nelsen, R. B. (2006), *An Introduction to Copulas*, 2nd. ed., New York, Springer.

OECD Health Data 2006, Organization for the Economic Co-operation and Development, Paris.

Patton, A. J. (2005), "Estimation of Multivariate Models for Time Series of Possibly Different Lengths", *Journal of Applied Econometrics*.

Quinn, C. (2007), "Using Copulas to Estimate Reduced-form Systems of Equations", *HEDG Working Papers 07/25*, University of York.

Quinn, C. (2007), "The Health-economic Applications of Copulas: Methods in Applied Econometric Research", *HEDG Working Papers 07/22*, University of York.

Sarma, S., W. Simpson (2006), "A Microeconometric Analysis of Canadian Health Care Utilization", *Health Economics,* 15**,** 219-239.

Sklar, A. (1959), "Fonctions de Répartition à *n* Dimensions et Leurs Marges", *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.

Smith, M. (2003), "Modeling Selectivity Using Archimedean Copulas", *Econometrics Journal*, 6, 99-123.

Train, K. E. (2003), *Discrete Choice Methods with Simulation*, New York, Cambridge University Press.

Trivedi, P. K., D. M. Zimmer (2005), *Copula Modeling: An Introduction for Practitioners*, Boston, Now.

Van Ourti, T. (2004), "Measuring Horizontal Inequity in Belgian Health Care Using a Gaussian Random Effects Two-part Count Data Model", *Health Economics,* 13**,** 705-724.

Vera-Hernandez, A. M. (1999), "Duplicate coverage and demand for health care. The case of Catalonia", *Health Economics,* 8**,** 579-598.

Wagstaff, A. (1986). "The Demand for Health - Some New Empirical-Evidence", Journal of Health Economics 5(3): 195-233.

Windmeijer, F. A., J. M. C. Santos Silva (1997), “Endogeneity in count data models: An application to demand for health care”, *Journal of Applied Econometrics*, 12, 281-294.

Winkelmann, R. (2004), “Health Care Reform and the Number of Doctor Visits – An Econometric Analysis”, *Journal of Applied Econometrics,* 19**,** 455-472.

Zimmer, D. M., P. K. Trivedi (2006), “Using Trivariate Copulas to Model Sample Selection and Treatment Effects: Application to Family Health Care Demand”, *Journal of Business and Economic Statistics*, 24, 63-76.