HEDG Working Paper 07/25

Using copulas to estimate reduced-form systems of equations

Casey Quinn

October 2007

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

Using copulas to estimate reduced-form systems of equations

Casey Quinn
University of York, York YO10 5DD, England
Lehigh University, Bethlehem, PA 18015, USA

October 12, 2007

Abstract

This paper considers the simultaneous explanation of mortality risk, health and lifestyles, using a reduced-form system of equations in which the multivariate distribution is defined by the copula. By applying the theory of inference functions for margins the parameters of each lifestyle, health and mortality risk equation can be estimated separately to the parameters of association found in their joint distribution, simplifying analysis considerably. Copulas also enable estimation of skewed multivariate distributions for the latent variables in a multivariate model of discrete response variables.

Mortality, Self-assessed Health and 6 lifestyles were taken from Health and Lifestyle Survey of England. The results suggest that non-normal latent distributions are preferred for the margins of the multivariate distribution, and that the multivariate distribution is skewed. The copula method should therefore be considered over the multivariate probit.

JEL classification: C1, C3, C5, I3, I10

1 Introduction

In this paper, a reduced-form model of mortality risk, health and lifestyles (diet, exercise, smoking, sleeping, alcohol consumption and obesity) is considered, using a system of equations in which the multivariate distribution is defined by a copula. A copula approximation of the joint distribution can avoid the distributional assumptions implicit in other multivariate families such as the multivariate normal, Beta, etc., allowing potentially more robust estimation. Employing a method due to Lee (1983), McLeish and Small (1988) and Joe and Xu (1996) that uses inferencing for margins, the parameters of each lifestyle, health and mortality equation can be estimated separately to the parameters of association found in their joint distribution, simplifying analysis considerably.

Analysing lifestyles and health jointly stems from research on the correlation between socioeconomic status and health, as well as income inequalities and health inequalities (van Doorslaer *et al.* 1997; Wagstaff and van Doorslaer 2000; van Doorslaer and Koolman 2004). Lifestyles need to be considered in this context because they may contribute to both health status and mortality. Contoyannis and Jones (2004) and Balia and Jones (2007), for example, both show that the introduction of lifestyles into a model for health (and in the latter study, risk of mortality also) reduces the influence of socioeconomic characteristics, affecting their contribution to socioeconomic inequalities in health and mortality.

Moreover, lifestyles can be assumed to be endogenous: as well as determining health and mortality, lifestyles themselves can be determined by factors such as income and education. In the context of a structural-equations model this creates a pathway through which strictly exogenous variables, including income inequalities, may have both direct and indirect effects on health and mortality, so that some of the variation in health and mortality can be explained in part by these endogenous factors.

This paper is an extension of Contoyannis and Jones (2004) and Balia and Jones (2007), using the same data (the British Health and Lifestyle Survey, hereon HALS) and a similar underlying thesis: that individual lifestyle choices determine health outcomes, including health and mortality. These choices are influenced by socioeconomic characteristics. To some extent those socioeconomic characteristics have a direct effect on health outcomes also, controlling for lifestyle choices.

This paper is also a methodological extension of previous work. Contoyannis and Jones (2004) considered unobserved heterogeneity via a recursive system of equations for self-assessed health and some endogenous lifestyles. Balia and Jones (2007) use the HALS data also, including follow-up data on mortality and health-affecting lifestyles. They also use a recursive system, where endogenous lifestyles are used to explain self-assessed health and mortality. In both studies the multivariate probit model is used for estimation. Here I use copulas to define the joint distribution function, rather than a multivariate probit.

The results show that, at least in this instance, the assumptions underlying the multivariate probit and multivariate normality are robust to the non-normality uncovered: covariate estimates and the variance-covariance matrices are comparable across the multivariate probit and the copulas used. As well as providing efficient estimates with less computation that the multivariate probit, the copula is used to highlight the statistical significance of skewness in the multivariate distribution, and its relation to what could otherwise be recognised as tail dependence in a copula, or potentially not measured at all in a multivariate normal distribution. The copula provides more information during analysis than standard methods, which is useful for model selection and testing goodness of fit.

2 Mortality risk, health and lifestyles

The behavioural model is as in Balia and Jones (2007). Individuals are assumed to maximise all periods of lifetime utility simultaneously, wherein utility for each period is a function of health, lifestyles, exogenous covariates and some probability of survival. Three elements are to be estimated: health, lifestyles and mortality, the risk of which influences the utility and optimal levels of the other two. The outcomes mortality y_m, health y_h and lifestyles $y_1, .., y_6$ are indicated by dichotomous variables (including self-reported health). Making the assumption that these follow a linearly-determined latent scale, following Balia and Jones (2007) gives the reduced form

$$y^*_{im} = \beta'_m \mathbf{X}_{im} + \varepsilon_{im} \quad (1)$$

$$y^*_{ih} = \beta'_h \mathbf{X}_{ih} + \varepsilon_{ih} \quad (2)$$

$$y^*_{il} = \beta'_l \mathbf{X}_{il} + \varepsilon_{il} \quad (3)$$

for each individual i, where $l \in \{$obesity, diet, sleeping, exercise, smoking, alcohol consumption$\}$ such that

$$y_{im} = 1\left(y^*_{im} \geq 0\right) \quad (4)$$

$$y_{ih} = 1\left(y^*_{ih} \geq 0\right) \quad (5)$$

$$y_{il} = 1\left(y^*_{il} \geq 0\right) \quad (6)$$

The vectors $\mathbf{X}_{im}, \mathbf{X}_{ih}$ and $\mathbf{X}_{il}$ are individual-specific exogenous vectors explaining, respectively, mortality risk, health and lifestyle choices. Under a structural specification these would be distinct due to exclusion restrictions needed to satisfy simple rank and order conditions for identification, but in reduced-form this can be relaxed. In Balia and Jones (2007), for example, $\mathbf{X}_{im}$ contains only some of the variables in $\mathbf{X}_{ih}$ and $\mathbf{X}_{il}$. Here the vectors are the same.

Estimation in Balia and Jones (2007) is carried out using the method of Maximum Simulated Likelihood (MSL), assuming the error terms are correlated and the random components ε_l, ε_H and ε_M are jointly normally distributed. They use MSL because the standard multivariate probit is underlied by an 8-dimensional normal distribution, for which the standard method of Maximum Likelihood (ML) and (generalised) Method of Moments would require substantially more computation. ML, for example, requires, in this case, integration over 7 cumulative normal probabilities in order to find solutions. MSL on the other hand simulates the

likelihood so that approximations, rather than the likelihood itself, are maximised. Similarly the Method of Simulated Moments (or Scores) can be used in place of the more intensive method of moments (Gouriéorux and Monfort 1996).

The two issues taken with MSL and multivariate normality are, in the first place, the estimation itself, which can be cumbersome and not necessarily efficient compared to standard ML (Hajivassiliou 1997). Secondly, the (multivariate) normality assumption is not necessarily made according to the best description of the data-generating process, and the results may not be robust under non-normality or asymmetry. The method presented here will help identify if these are problems, while showing a more convenient procedure for estimation.

2.1 Considering multivariate (non-)normality

The motivations for moving away from the normality assumption in a multivariate framework are two-fold. The first is computation: although low orders of dimensionality rarely present problems for computing multivariate probits, maximising likelihoods across 8 dimensions is time-consuming and computationally intensive (Muthén 1979, 1984 discusses this in some detail). The second is robustness: although some authors have shown that departures from normality are not necessarily of great concern, they too become more problematic in higher dimensions (Keselman, *et al.* 2005; Prokhorov and Schmidt 2005). The robustness issues with standard t and F-tests under non-normality are also known (Mardia 1971; Ali and Sharma 1996; Curran, *et al.* 1996). In a structure-of-equations model, multivariate non-normality can also lead to erroneous rejection of some models or equations within the structure (see Klein 1998, for example).

The multivariate normal distribution is commonly selected for the convenience of its use; because the univariate normal distribution is robust under reasonable levels of non-normality, it also explains the margins of the joint distribution fairly well (Kowalski 1973). The normal distribution also tends to be more easily extended to higher dimensions: the density or characteristic function of the normal distribution can be used, or a linear combination of normally-distributed random variables (Fang, Kotz and Ng,1989). The preference for the multivariate normal then can exist even when the data appears not to be elliptically symmetric. While a multivariate distribution with one or more non-normally distributed margins is always non-normal however, a multivariate distribution with normally-distributed margins but a skewed or kurtotic correlation will be non-normal also.

Abandoning the multivariate probit/normality assumption has direct implications in terms of the econometric problem. The recursive model is one of conditionally-dependent random variables such that, in this case, endogenous health status is a function within a function: it is an explanatory variable for mortality

risk, while also being explained by endogenous lifestyle choices. At each level these are also explained by exogenous explanatory variables. This structure can only be maintained by assuming a symmetric distribution, such as the normal, and is subject to Borel's paradox otherwise (Kolmogorov 1950). In order to consider non-normal and/or skewed latent variables, the reduced-form must be used, rather than the structural, so that the health and mortality equations are not conditionally distributed according to the lifestyle variables.

One significant advantage offered by the multivariate normal distribution is its correlation: few families of distributions, including copulas, are so easily extended to multivariate distributions with generalisable correlation/dependence structures. The best approach among those discussed here is based upon the multivariate normal and t distributions. As copulas they allow a broad range of marginal distributions to be specified, while retaining the flexible multivariate dependence structure these distributions offer. Multivariate skewed elliptical distributions, which are presented later and used in the analysis, represent a very useful approach to combining the dual needs for reliable measures of multivariate dependence, as well as flexibility in the face of multivariate asymmetry.

3 The copula method

For univariate marginal distribution functions $F_1(x_1)$ and $F_2(x_2)$, a copula is a function that binds those margins precisely, to form the multivariate distribution function (Smith 2003). The copula parameterises the dependence between the margins, while the parameters of each marginal distribution function can be estimated separately. For the purposes of empirical analysis a copula is best described, as in Joe (1997), as a multivariate distribution function that separates each marginal distribution both from every other marginal distribution, and from the dependence between their associated random variables. Thus the two most important features of copulas: they exist as multivariate distribution functions which can feasibly contain any type and combination of marginal distributions; and each uniquely represents dependence. Depending on the functional form used, association of quite different types can either be assumed or tested, independently of the functional forms of the marginal distributions used.

By a theorem due to Sklar (1959) one can say that all multivariate distributions have a copula representation, in which each margin is invariant to transformations in every other margin, or independent of the choice of every other marginal distribution.[1] Consider two random variables X_1, X_2 with bivariate distribution function $H(x_1, x_2) = \Pr(X_1 \leq x_1, X_2 \leq x_2)$ and univariate marginal distributions $F_1(x_1)$ and $F_2(x_2)$ respectively. Then there exists a copula C that represents the joint distribution function in terms of the margins, such that

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{7}$$

for all real values of x_1, x_2 (or $(X_1, X_2) \in \mathbb{R}^2$). If F_1, F_2 are continuous, C is unique. Under discontinuity C is uniquely determined on its domain, the range of the margins $\text{Ran}F_1 \times \text{Ran}F_2$.[2] Moreover it can be seen using Sklar's theorem that, if C is a copula and F_1 and F_2 are distribution functions, then some function H as defined in Equation (7) is a joint distribution function (see Nelsen 2006 for this proof, as well as an explanation of quasi-inverses of non-strictly increasing margins, which can also be used to construct a copula). By taking the marginal distribution functions as explanators within which association is not contained, the copula separates the explanation of X_1 and X_2 from their association.

The most intuitive approach to constructing copulas is by using inverted distribution functions as arguments in known multivariate distribution functions - the so-called Inversion method of construction (Nelsen 2006). Thus, using inverses of the marginal distribution functions gives

$$C(F_1(x_1), F_2(x_2)) = H_{(1,2)}\left(H_1^{-1}(F_1(x_1)), H_2^{-1}(F_2(x_2))\right) \tag{8}$$

Note that $H_{(1,2)}$ can feasibly be any joint distribution; different formulations of $H_{(1,2)}$ will generate different forms of C, one of which will be the closest approximation to the true bivariate distribution H (the familiar bivariate normal distribution $\Phi_{(1,2)}(x_1, x_2)$, for example, is merely another such approximation). Using the uniformly-distributed $H_1^{-1}(F_1(x_1)), H_2^{-1}(F_2(x_2))$ allows, via inversion, the subsequent use of any type of distribution in the margin. Extending Equations (**??**) and (8) to higher dimensions, for $(X_1, ..., X_n) \in \mathbb{R}^n$ there exists the so-called n-copula

$$C(F_1(x_1), .., F_n(x_n)) = H_{(n)}\left(H_1^{-1}(F_1(x_1)), .., H_n^{-1}(F_n(x_n))\right) \tag{9}$$

and where, according to Sklar's theorem, there exists one n-copula such that

$$H(x_1, ..., x_n) = C(F_1(x_1), .., F_n(x_n)) \tag{10}$$

The so-called Gaussian copula, given by

$$C(F_1(x_1), F_2(x_2)) = \Phi_{(1,2)}\left(\Phi_1^{-1}(F_1(x_1)), \Phi_2^{-1}(F_2(x_2))\right) \tag{11}$$

is a well-known example of the inversion method, as is the t copula, which uses the multivariate student's t distribution. Nelsen (2006) discusses some more examples that use asymmetric distributions.

4 Inference functions and the Gaussian and t copulas

A two-step method due to Lee (1983), McLeish and Small (1988), Joe and Xu (1996), Xu (1996) and Joe (1997) is inferencing (IFM).[3] For some multivariate distribution function $H(X_1,..,X_n;\beta_1,..,\beta_n,\theta)$, consider the corresponding copula $C(F_1(X_1;\beta_1),..,F_n(X_n;\beta_n);\theta)$. The marginal parameter vectors $\beta_1,..,\beta_n$ can contain coefficients due to regression, and/or simple parameters for each distribution. The vector θ contains measures of assocation for the copula as a whole. The IFM method is a two-step procedure is as follows.

Step 1: Each marginal vector of coefficients $\beta_{i\in n}$ from marginal univariate distribution functions $F_1(X_1;\beta_1),...,F_n(X_n;\beta_n)$ is estimated first, and separately, to determine $\left\{\hat{\beta}_1,..,\hat{\beta}_n\right\}$ such that

$$\hat{\beta}_i = \arg\max_{\beta_i} \sum_{i=1}^{n} \ln f_i(x_i;\beta_i) \tag{12}$$

Step 2: The estimates $\hat{\beta}_i$ can be used to calculate the evaluated marginal distribution functions $\hat{F}_i(X_i;\beta_i) = F_i\left(X_i;\hat{\beta}_i\right)$. It is these, rather than $F_i(X_i;\beta_i)$, that are passed into the copula likelihood for estimation of θ.

Step 3: Using $\hat{F}_i(X_i;\beta_i)$, the copula likelihood $L\left(\hat{\beta}_1,..,\hat{\beta}_n,\theta\right)$ is maximised to find only $\hat{\theta}$ such that

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ln c\left(\hat{F}_1(x_1;\beta_1),..,\hat{F}_n(x_n;\beta_n);\theta\right) \tag{13}$$

for some copula C with density $c\left(\hat{F}_1(x_1;\beta_1),..,\hat{F}_n(x_n;\beta_n);\theta\right)$.

Ordinarily, ML solves $(\partial L/\partial\beta_1,..,\partial L/\partial\beta_n,\partial L/\partial\theta)=0$. Estimates from the method of IFM are such that $(\partial L_1/\partial\beta_1,..,\partial L_n/\partial\beta_n,\partial L/\partial\theta)=0$ for univariate log-likelihoods $L_1,..,L_n$ as well as the joint likelihood L. This holds under regularity conditions, and Joe (1997, 2005) shows that the IFM method is efficient relative to the method of maximum likelihood, particularly for discrete marginal distributions with few categories. It is less so for more categories, and for continuous marginal distributions with strong dependence, although standard errors for the parameters in this approach can corrected post-estimation using jackknife methods.

The method of IFM can be used to estimate the so-called Gaussian copula with a multivariate normal distribution, in this case given by

$$\begin{aligned} C(u_1,..,u_8) &= \Phi_8\left(\Phi_m^{-1}(F_m(y_{im}^*)),\Phi_h^{-1}(F_h(y_{ih}^*)),\Phi_{1_1}^{-1}\left(F_{l_1}\left(y_{il_1}^*\right)\right),\right. \\ &\quad \left.,..,\Phi_{1_6}^{-1}\left(F_{l_6}\left(y_{il_6}^*\right)\right)\right) \end{aligned} \tag{14}$$

In this approach the random variable has a different transformation. Where previously the copula used $F_m(y_{im}^*)$ instead of x_{im}, for example, these use - in the Gaussian case - $\Phi_m^{-1}(F_m(y_{im}^*))$. The original

combination $x'_{im}\hat{\beta}_m$ is used to estimate $F_m\left(\hat{y}^*_{im}\right)$, which in turn is transformed to $\Phi_m^{-1}\left(F_m\left(\hat{y}^*_{im}\right)\right)$, which is entered into the copula as a random variable. In fact it can be considered as a vector of pseudo-observations: $\Phi_m^{-1}\left(F_m\left(\hat{y}^*_{im}\right)\right)$ is a prediction of the erstwhile unobserved latent variable y^*_m from Equation (1).

Although the model in Equation (14) is a normal distribution function, the function of inverses results in tractability of the marginal distributions. It is subsequently much more straightforward than the multivariate probit, because exact Maximum Likelihood is available for the problem

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} \sum_{i=1}^{n} \ln \phi_8 \left(\Phi_m^{-1}\left(F_m\left(y^*_{im}\right)\right), \Phi_h^{-1}\left(F_h\left(y^*_{ih}\right)\right), \Phi_{1_1}^{-1}\left(F_{l_1}\left(y^*_{il_1}\right)\right),\right. \\ &\quad \left. .., \Phi_{1_6}^{-1}\left(F_{l_6}\left(y^*_{il_6}\right)\right); \theta\right) \end{aligned} \tag{15}$$

which is more easily implemented. For this reason the method of IFM is used: it is permitted with separated marginal distributions, and it is necessary due to the inversion, in order to get parameters with interpretable estimates. It is also a nice alternative to the multivariate probit irrespective of the issues discussed here, being much simpler to specify and estimate.

An alternative is the multivariate t-copula, which is narrower than the Gaussian and can capture tail dependence of extreme events (Embrechts, *et al.*. 2003; Demarta and McNeil 2004). In the Gaussian copula, as in the multivariate normal, such events become asymptotically independent. Moreover, uncorrelated events are not considered independent in the t-copula.

The composite, or pairwise, likelihood approach is another example of inference at higher orders than the univariate margins, wherein the joint likelihood is composed of valid bivariate likelihoods (Lindsay 1988; Kuk and Nott 2000; Andersen 2004; Bellio and Varin 2005; Liu and Zhao 1999 and Joe 2005 are examples), although with less efficiency than has been shown for the IFM. Hüsler and Reiss (1989) provide a similar approach: the dependence parameter for each margin can be estimated in each bivariate margin of the multivariate distribution. The process identified here as the IFM is also seen elsewhere, for instance in work by Arellano and Honoré (2000) and Arellano and Carrasco (2002) on panel data models with predetermined variables.

4.1 Considering skewness

This IFM approach does not restrict the margins of the multivariate distribution: any general form can be used (Nelsen 2006. Joe 2005 considers Pareto, Weibull and Gamma margins also). In this instance, for example, non-normal link functions can be considered alongside univariate probits for each margin.

Multivariate skewness can be also accommodated in the erstwhile symmetric normal and t distributions, via the multivariate skewed normal and/or multivariate skewed t distributions (Azzalini and Dalla Valle 1996, Azzalini and Capitanio 1999, 2003). The skewed normal distribution is generated by some random variable X whose PDF is of the form

$$f(x;\alpha) = 2\phi(x)\Phi(\alpha x) \tag{16}$$

where $\phi(.),\ \Phi(.)$ are the familiar standard normal density and distribution functions, respectively, and α is some scalar measuring skewness, such that the distribution of X is symmetric about the origin if $\alpha = 0$ (i.e. $X \sim N(0,1)$) and increasing in skewness with increases in $|\alpha|$. Then according to Azzalini and Dalla Valle (1996), X is skewed normal $X \sim SN(\alpha)$.[4] The multivariate skewed normal is given for some random vector $X_{[k\times 1]}$ where

$$f_k(x;\alpha) = 2\phi_k(x;\Omega)\Phi(\alpha' x) \tag{17}$$

for $x \in \mathbb{R}^k$, where $\alpha_{[k\times 1]}$ is a vector of skewness components and where X has correlation matrix Ω, and still assuming symmetry about 0. Then as above $X \sim SN_k(\Omega,\alpha)$, following the notation in Azzalini and Capitanio (1996). In general form, Azzalini and Capitanio (1996) show that, for the random vector X with distributional symmetry about 0, and some transformation $W(x)$ that is symmetric about 0 also (although μ_X could be used it is less simple), there exists some density function $f_k(x)$ such that

$$f_k(x) = 2f(x)F(W(x)) \tag{18}$$

where $f(.),\ F(.)$ are some k-dimensional density and distribution function, respectively. Any elliptical distribution can be accommodated in this manner, as can non-elliptical distributions.[5] Azzalini and Capitanio (2003) consider the multivariate skewed t distribution, using this generalisation, such that $X \sim St_{k,v}(\Omega,\alpha,v)$ with v degrees of freedom.

Although non-trivial, the multivariate skewed t distributional is implemented in the statistical package R, making estimation relatively straightforward: the R package, for example, provides estimates of the skewness for testing its statistical significance. . Vinod (2005) discusses the application of the skew densities and inference in Mathematica, also. Marrying the notation of Azzalini and Capitanio (2003) to the copula approach, consider Equation (14). The random vector $X = \left(\Phi_m^{-1}\left(F_m\left(y_{im}^*\right)\right), \Phi_h^{-1}\left(F_h\left(y_{ih}^*\right)\right), \Phi_{1_1}^{-1}\left(F_{l_1}\left(y_{il_1}^*\right)\right), .., \right.$ $\left.\Phi_{1_6}^{-1}\left(F_{l_6}\left(y_{il_6}^*\right)\right)\right)'$, for example, is estimated as $X \sim SN_k(\Omega,\alpha)$.

5 The HALS data

The Health and Lifestyle Survey of England (HALS) was a national survey of adults in private households, carried out (in the first wave) in 1984-5, during two home visits. The first of these was the survey interview; the second a visit by a nurse for physiological measurements and to test cognition. The survey has been followed up by 4 subsequent waves of information collection on the original interviewees: the principle follow-up in 1991-2 was used to capture change in characteristics, behaviours and beliefs over the 7 years, and included 5,352 interviewees. Subsequent follow-ups, the most recent of which took place in 2005, provided updated mortality data. Of the original 9,003 respondents 2,491 had died.[6]

This analysis follows Balia and Jones (2007), using information at the time of the first survey, coupled with the most recent follow-up data on mortality. The second survey is overlooked due in part to attrition, which can be problematic. In order to avoid confounding mortality with accident, injury or a genetic predisposition towards early death not related to lifestyle, only individuals 40 years of age and over at the time of the first survey are retained for analysis. At this age and over, initial states of health, education, income and so forth are considered to be stable, such that subsequent information is not required to explain mortality and health later in life.

5.1 Indicators of a healthy lifestyle

The lifestyle variables employed here are the same as those used in Balia and Jones (2007) and Contoyannis and Jones (2004), drawing on the analysis of Belloc and Breslow (1972) and Kenkel (1995). These are indicators for diet, weight, smoking and sleeping behaviour, alcohol consumption and exercise. All are dichotomous in this study. Diet is measured with an indicator for whether or not breakfast is eaten within one hour of waking (Kenkel 1995). Smoking is an indicator of whether or not the individual is currently smoking (any number of cigarettes per day). Exercise is measured by participation in one of 14 exercise categories in the fortnight preceding the survey. Alcohol consumption is a gender-specific measure of prudent alcohol consumption.[7] Sleep is measured as either optimal or not; optimal sleep shown by Belloc and Breslow (1972) to be between 7 and 9 hours per night. More or less is not considered separately, but together are suboptimal. Healthy weight is anything below obesity, as measured by a Body Mass Index (BMI) below 30 for males and 28.5 for females.[8]

5.2 Explanatory variables

Exogenous variables in the models are predominately dichotomous. They are given and described in Table 1.

Table 1: VARIABLE DEFINITIONS AND DESCRIPTIVE STATISTICS FOR THE HEALTH AND LIFESTYLE SURVEY OF ENGLAND

Variable	Definition	Mean/ Proportion = 1	Std.Dev.
Health status			
deceased	1 if deceased at June 2005	0.41	0.4911
sah	1 if self-assessed health is excellent or good (0 if fair or poor)	0.70	0.4572
Lifestyle			
non-smoker	1 if not currently smoking	0.70	0.4585
breakfast	1 if regularly eating a 'healthy' breakfast	0.71	0.4552
sleeping well	1 if sleeping between 7 and 9 hours	0.58	0.4932
prudent drinker	1 if consuming alcohol prudently	0.88	0.3251
non-obese	1 if under 'obese'	0.85	0.3538
exercising	1 if engaged in physical exercise	0.32	0.4677
Social Class			
sc1	1 if "professional/student", "managerial/intermediate"	0.32	0.4648
sc2	1 if "skilled", "armed service"	0.47	0.4990
sc3	1 if "partly skilled", "unskilled", "unclassified"	0.22	0.4128
Education			
degree	1 if University	0.13	0.3308
HVQ/A level	1 if Higher Vocational Qualifications or A level (or equivalent)	0.12	0.3305
CSE/O level	1 if CSE or O level (or equivalent)	0.09	0.2924
none	1 if no qualification	0.61	0.4882
other	1 if any other vocational or professional qualification	0.05	0.2130
Marital status			
married	1 if married	0.76	0.4268
widowed	1 if widowed	0.13	0.3339
divorced/separated	1 if divorced or separated	0.05	0.2280
single	1 if single	0.06	0.2312
Occupation			
full time	1 if employed full-time	0.36	0.4813
part time	1 if employed part-time	0.13	0.3384
shift/casual worker	1 if shift/casual worker	0.06	0.2327
unemployed	1 if unemployed	0.03	0.1716
absent (illness)	1 if absent from work due to illness/injury	0.03	0.1789
retired	1 if retired	0.34	0.4733
housekeeper	1 if housekeeper	0.10	0.3024

Variable	Definition	Mean/ Proportion = 1	Std.Dev.
Geography			
Scotland	1 if in Scotland	0.10	0.2954
Wales	1 if in Wales	0.06	0.2333
Northern England	1 if in the North of England	0.07	0.2468
North-western England	1 if in the North-west of England	0.13	0.3339
Yorkshire	1 if in Yorkshire	0.09	0.2807
West midlands	1 if in the West-midlands of England	0.08	0.2716
East midlands	1 if in the East-midlands of England	0.08	0.2660
Anglia	1 if in Anglia	0.04	0.1959
South-western England	1 if in the South-west of England	0.09	0.2839
South-eastern England	1 if in the South-east of England	0.19	0.3901
London	1 if in London	0.09	0.2924
Area			
Rural	1 if in Rural area	0.22	0.4132
Suburban	1 if in Suburban area	0.47	0.4993
Urban	1 if in Urban area	0.31	0.4627
Ethnicity			
European caucasian	1 if European caucasian	0.98	0.1436
Physical Characteristics			
gender (male)	1 if Male	0.46	0.4981
height	Height in inches	65.95	3.7032
age	Age in years	57.47	11.6733
age^2	Age^2/100	34.39	14.0761
Residential Characteristics			
owner	1 if owning own home	0.66	0.4746
household size	Number of people in the household	1.65	1.2723
smoking household	1 if anyone smokes in the household	0.35	0.4773
Parental Characteristics			
mother smoked	1 if only mother smoked/s	0.03	0.1731
father smoked	1 if only father smoked/s	0.60	0.4909
both smoked	1 if both smoked/s	0.25	0.4306
mother's drinking	Mother's drinking (0-4, non-to-heavy drinker)	0.91	0.9812
father's drinking	Father's drinking (0-4, non-to-heavy drinker)	1.89	1.2005

They are familiar considerations for explanators for health: variables representing social class, education, marital status, employment status, cultural background, geographical region and area type, residential tenure and physical, household and parental characteristics.

5.3 Some descriptive results

Some descriptive statistics for variables of interest are given in Table 1 also. After cleaning the data of missing values for variables of interest (including those lost to the official registry), and restricting analysis to people aged 40 years and over at the time of the first survey, 3,655 remain from the original 9,003.

The majority of respondents correspond to at least one healthy lifestyle, apart from exercising, of which only 32% partake. Around 41% now are deceased, while in the original HALS 70% considered themselves to be in good health. With an average age of 57, after censoring at 40, this is not necessarily surprising, particularly when considering the lifestyles.

For education the generation(s) under consideration become apparent, with around 61% of respondents offering no educational qualification. The proportions of full-time employed and retired, after dropping the younger-than-40, are also quite significant. As mentioned previously European Caucasians make up 98% of the sample. Also high is married respondents, 76%. Home ownership, another indicator of social class, is around 66%.[9]

Nine of the 28 pairwise correlations (not shown) are negative, effectively proscribing the use of the mixture copulas. Urban living has a negative correlation lifestyles and a (subsequent) positive correlation with mortality. A social gradient appears to exist across health, mortality and lifestyles. These however are the only two variable with relatively consistent correlation.

6 Estimation

Estimation of the multivariate model for mortality risk, health and lifestyles is undertaken with the inferencing approach of the Gaussian and t copulas.

The log-likelihood for the problem is as in Equation (15). Unlike the multivariate probit, this considers the summed logs of multivariate normal densities, rather than distributions. Three functional forms for $F_{\cdot}$ can be considered, or three link functions for Equations (1)-(6). These are the probit,

$$\Pr\left(Y = 1|x;\beta\right) = \Phi\left(x'\beta\right) \tag{19}$$

the fatter-tailed logit,

$$\Pr\left(Y=1|x;\beta\right)=\exp\left(-\exp\left(x'\beta\right)\right) \tag{20}$$

and the complementary log-log

$$\Pr\left(Y=1|x;\beta\right)=1-\exp\left(\exp\left(-x'\beta\right)\right) \tag{21}$$

which is an asymmetric extension of the logit, useful in particular for fairly heavily right-skewed distributions of $x'\beta$, or in this case the inverted probabilities of mortality and exercising.

7 Results

Employing the method of IFM has specific implications in terms of the results. Being able to choose freely both the marginal distributions and the joint distributions, separately from one another, means more information must be considered overall, and considered separately. Some of this is information gained over and above standard methods of estimation; most of it will relate to goodness of fit and model selection.

7.1 Specifying and selecting marginal distributions

Following Joe (1997), the appropiate link function can be found useing information criteria such as the Akaike Information Criterion (AIC), given by $AIC=2k-2\ln(L)$ for log-likelihood L and k parameters, or Bayesian Information Criterion (BIC), given by $k\ln(n)-2\ln(L)$ and where n is the sample size.

A second approach recommended by Joe (1997) is analysis of the predictive ability of the models estimated. That is, some comparison of the predicted summaries from the models with the observed summaries of the data itself. Results from cell prediction and information criteria are contained in Table 2.

Table 2: PERCENTAGES OF CORRECTLY-PREDICTED OUTCOMES AND INFORMATION CRITERIA FROM PROBIT, LOGIT AND COMPLEMENTARY LOG-LOG MODELS (SHADED CELLS ARE THE OPTIMUM MODEL FOR EACH MARGIN ACCORDING TO EACH CRITERION)

	% Correct				AIC			BIC		
	Multivariate Probit	Probit	Logit	CLog-Log	Probit	Logit	CLog-Log	Probit	Logit	CLog-Log
Mortality	0.7839	0.7852	0.7874	0.7899	3304.612	3299.585	3314.392	3558.97	3553.943	3568.75
SAH	0.6416	0.6421	0.6430	0.6364	4126.101	4125.847	4125.206	4380.459	4380.205	4379.564
Breakfast	0.6492	0.6487	0.6506	0.6410	4106.177	4108.384	4101.67	4360.535	4362.742	4356.028
Not obese	0.6150	0.6164	0.6249	0.6049	2933.099	2930.946	2936.521	3187.457	3185.304	3190.879
Non-smoker	0.6722	0.6725	0.6780	0.6651	3993.578	3993.232	3998.705	4247.936	4247.59	4253.063
Sleeping well	0.5789	0.5765	0.5759	0.5735	4911.126	4911.547	4909.652	5165.484	5165.905	5164.009
Prudent drinker	0.6810	0.6802	0.6925	0.6635	2385.302	2382.104	2392.089	2639.659	2636.462	2646.447
Exerciser	0.6287	0.6276	0.6301	0.6328	4301.734	4301.714	4300.307	4556.092	4556.072	4554.665

.

The probit, logit and complementary log-log functions are equally complex, so the results are the same across the BIC and AIC tests, although the results from both are included.

The predictive accuracy of each function has also been included in Table 2, and offers different optimal specifications. Only the overall accuracy of predictions have been included: the predictions of 0 and 1 separately is not useful. Due to its structure, the complementary log-log consistently predicts more 0s accurately by virtue of predicting more of them. Moreover there is no preference *prima facie* for accuracy in one or the other outcome, so neither can be justified as a criterion for model selection (although this need not always be the case). The combination used here is according to the overall accuracy in cell predictions - a mixture of probit, logit and complementary log-log link functions for the margins.

There is no particular econometric justification for either of the probit or logit over the other. The complementary log-log is preferred for mortality and exercising, due to their much higher rates of failure (such that the probability that the indicator will be 0 is substantially higher), and due to the dichotomising of SAH one would reasonably expect a latent distribution of health with fatter tails, for which the logit is better-suited, but for the remaining lifestyles no similar information is available. Kolmogorov-Smirnoff tests can be used to compare the distributions of the predicted probabilities to determine whether there is any statistical relevance to the choice made.[10] Considering the dimensions *in toto*, Table 3 from the corresponding analysis of the joint distribution shows that, for each of the candidate copulas, the mixture of probit, logit and complementary log-log models is optimal.[11]

7.2 Comparing the t and Gaussian copulas, skewness and symmetry

7.2.1 Skewness

Degrees of freedom in the multivariate t copula (or distribution) can be fixed or estimated freely, as in this case. Maximum-Likelihood estimates of the degrees of freedom from the skewed t are around $df = 27.5$ whereas, for the symmetric t, $df = 12.9$. This follows on from the previous section: because the skewness affects the distribution principally around the mean, central tendency is estimated more precisely, so that less tail dependence is observed. The means of the inverse probabilities from each model (including the 'observed' data) are in Table 4.[12] They are generally similar; the results from the skewed and symmetric t models are slightly different.

In this instance the skewed t is tending fairly Gaussian, based on the degrees of freedom, which raises the question of whether skewness or tail dependence is more important. Considering extreme events, for example, would favour tail dependence over skewness. From a purely statistical standpoint, the choice is dependent upon the skewness estimates for the multivariate normal and t models. These are in Table 5.

Table 3: INFORMATION CRITERIA FROM THE JOINT (COPULA) DISTRIBUTIONS (SHADED ROWS CONTAIN THE PREFERRED MODEL, ACCORDING TO MINIMUM INFORMATION CRITERIA).

		Log-likelihood	AIC	BIC	ECVI
	MV Probit	-14548.75	29809.50	32018.07	8.16
Skewed	Gaussian copulas (IC)				
	Probit margins	-10204.68	21137.36	23395.56	5.78
	Probit/Clog-Log margins	-10387.55	21503.10	23761.30	5.88
	Logit/Clog-Log margins	-10137.91	21003.82	23262.02	5.75
	Gaussian copulas (% Correct)				
	Probit/Clog-Log margins	-10484.83	21697.66	23955.86	5.94
	Mixed margins	-9836.056	20400.11	22658.31	5.58
	t copulas (IC)				
	Probit margins	-10148.79	21027.58	23291.99	5.75
	Probit/Clog-Log margins	-10349.44	21428.88	23693.29	5.86
	Logit/Clog-Log margins	-10110.19	20950.38	23214.79	5.73
	t copulas (% Correct)				
	Probit/Clog-Log margins	-10215.59	21161.18	23425.59	5.79
	Mixed margins	-9836.055	20402.11	22666.52	5.58
Symmetric	Gaussian copulas (IC)				
	Probit margins	-10993.64	22699.28	24907.85	6.21
	Probit/Clog-Log margins	-11111.91	22935.82	25144.39	6.28
	Logit/Clog-Log margins	-10865.48	22442.96	24651.53	6.14
	Gaussian copulas (% Correct)				
	Probit/Clog-Log margins	-11005.55	22723.10	24931.67	6.22
	Mixed margins	-10721.94	22155.88	24364.45	6.06
	t copulas (IC)				
	Probit margins	-10706.44	22126.88	24341.65	6.05
	Probit/Clog-Log margins	-10881.76	22477.52	24692.29	6.15
	Logit/Clog-Log margins	-10657.82	22029.64	24244.41	6.03
	t copulas (% Correct)				
	Probit/Clog-Log margins	-10695.08	22104.16	24318.93	6.05
	Mixed margins	-10394.53	21503.06	23717.83	5.88

Table 4: PREDICTED MEAN (STANDARD DEVIATION) IN EACH DIMENSION OF THE MORTALITY RISK, HEALTH AND LIFESTYLE MODELS

	Skewed Normal	Symmetric Normal	Symmetric *t*	Skewed *t*	Observed
Mortality	-0.2739	-0.2731	-0.3453	-0.3051	-0.2850
	(1.0816)	(1.1263)	(1.0840)	(1.0100)	(1.0957)
SAH	0.5729	0.5666	0.6060	0.5868	0.5686
	(0.4650)	(0.4580)	(0.3833)	(0.4000)	(0.4653)
Breakfast	0.6094	0.5908	0.6282	0.5984	0.5959
	(0.4376)	(0.4687)	(0.4348)	(0.4217)	(0.4547)
Not obese	1.1389	1.1117	1.1117	1.1298	1.1227
	(0.3726)	(0.3621)	(0.3542)	(0.3658)	(0.3731)
Non-smoker	0.5914	0.5865	0.6223	0.6143	0.5960
	(0.5424)	(0.5528)	(0.5190)	(0.5240)	(0.5518)
Sleeping well	0.2171	0.2060	0.2266	0.2210	0.2148
	(0.2417)	(0.2510)	(0.2308)	(0.2344)	(0.2493)
Prudent drinker	1.3511	1.3491	1.3739	1.3566	1.3459
	(0.5641)	(0.5455)	(0.5266)	(0.5521)	(0.5502)
Exerciser	-0.4934	-0.5146	-0.4918	-0.4826	-0.5028
	(0.4554)	(0.4508)	(0.4020)	(0.4418)	(0.4461)

Table 5: ESTIMATED SHAPE PARAMETERS (STANDARD ERRORS) FROM R, FOR EACH DIMENSION OF THE MORTALITY RISK, HEALTH AND LIFESTYLE MODELS

	Skewed Normal		Skewed *t*	
	Shape/Skewness	Covariate/Shift	Shape/Skewness	Covariate/Shift
Mortality	13.8945	-1.6646	-2.7993	-0.3779
	(1.3244)	(0.0190)	(0.1755)	(0.0328)
SAH	1.0472	0.7137	-5.2300	0.9210
	(0.1696)	(0.0150)	(0.2363)	(0.0110)
Breakfast	-1.9399	0.3277	0.1935	0.7475
	(0.2085)	(0.0130)	(0.1094)	(0.0140)
Not obese	0.4219	1.0612	2.3438	1.1213
	(0.1300)	(0.0118)	(0.1265)	(0.0117)
Non-smoker	1.5039	0.3519	0.7909	0.7329
	(0.1889)	(0.0161)	(0.1076)	(0.0164)
Sleeping well	0.1234	0.3635	-0.4122	0.2965
	(0.0990)	(0.0071)	(0.0766)	(0.0071)
Prudent drinker	0.7700	1.1837	0.3279	1.3383
	(0.1286)	(0.0165)	(0.0922)	(0.0168)
Exerciser	0.9167	-0.0791	-0.9720	-0.3588
	(0.2489)	(0.0109)	(0.1429)	(0.0129)

Skewness parameters, as provided by R, suggest skewness with statistical significance according to the statistical significance of the paramaters. In this case only Sleeping Well in the Skewed Normal model is not statistically significant, suggesting robust evidence of skewness. Covariates, or shift parameters, are the degree to which each dimension in each model was shifted for the analysis of the multivariate skew distributions, according to Equations (16)-(18)

The parameter corresponding to mortality risk in the skewed normal distribution is quite large, relative to the t distribution, as well as having a different sign. Sign differences occur in other margins as well. For the multivariate Gaussian copula, the skewed normal distribution is preferred to the symmetric normal by virtue of the statistical significance of the skewness vector: the trade-off from the t-copula does not exist for the Gaussian, which ignores tail dependence.

For mortality risk, SAH, eating breakfast and exercising, skewness is counter-directional in the normal and t distributions. In all cases, except eating breakfast in the skewed t and sleeping well in the skewed normal, skewness is statistically significant at 5% (and at 10% for eating breakfast in the skewed t). Accordingly the distribution is considered to be a skewed t, although with reasonably high degrees of freedom.

Correlation matrices for the Normal and t models, as well as the set of marginal inverses passed into R for analysis via IFM, are provided in Tables 6 to 8. Emboldened values represent economic significance (correlation greater than 5%, in either direction).

Table 6: CORRELATIONS BETWEEN MORTALITY RISK, HEALTH AND LIFESTYLES FOR SYMMETRIC AND SKEWED NORMAL DISTRIBUTIONS

Symmetric Normal distribution								
	Mortality	SAH	Breakfast	Not obese	Non-smoker	Sleeping well	Prudent drinker	Exerciser
Mortality	**1**							
SAH	**-0.4076**	**1**						
Breakfast	**0.4567**	**0.2373**	**1**					
Not obese	0.0039	**0.5376**	**0.1522**	**1**				
Non-smoker	**0.1667**	**0.4443**	**0.7358**	**0.2385**	**1**			
Sleeping well	**-0.4915**	**0.4533**	0.0365	**0.2467**	**0.0715**	**1**		
Prudent drinker	**0.1898**	**-0.1877**	**0.4238**	**-0.4921**	**0.4579**	**-0.0693**	**1**	
Exerciser	**-0.8066**	**0.5941**	**-0.1535**	**0.2966**	-0.0125	**0.5127**	**-0.3186**	**1**

Skewed Normal distribution								
	Mortality	SAH	Breakfast	Not obese	Non-smoker	Sleeping well	Prudent drinker	Exerciser
Mortality	**1**							
SAH	**-0.4511**	**1**						
Breakfast	**0.3890**	**0.2308**	**1**					
Not obese	-0.0220	**0.4138**	**0.1149**	**1**				
Non-smoker	**0.1648**	**0.3879**	**0.7429**	**0.1698**	**1**			
Sleeping well	**-0.5404**	**0.4806**	**0.0782**	**0.2123**	**0.0988**	**1**		
Prudent drinker	**0.1822**	**-0.1180**	**0.4288**	**-0.5201**	**0.5056**	-0.0353	**1**	
Exerciser	**-0.8117**	**0.6221**	**-0.1452**	**0.3045**	-0.0538	**0.5348**	**-0.3314**	**1**

Differences in the matrices in Tables 6 to 8 can be observed, although the values for each pair are reasonably close, across the distributions. Overall the correlations are significant.

7.2.2 Fitting a model vs. replicating data

Goodness of fit was considered previously with respect to the margins, using information criteria. Looking again at Table 5, the skewed joint distributions are an improvement upon the symmetric distributions. Between symmetric distributions, the multivariate t copula is noticeably better than the Gaussian moreso than between the skewed distribution, in which there is not much improvement due to the use of the t. This reflects previous comments concerning the Gaussian-tending degrees of freedom observed in the multivariate skewed t, relative to the symmetric. At $df = 27.5$, the t and Gaussian copulas are not as distinct, compared to their symmetric counterparts. Overall the skewed multivariate t-copula is still preferred, according to information criteria.

Goodness of fit can be considered also within the context of replication. The central question asked is, how close is the copula's approximation of the data-generating process to the process itself? This is a different question to the one answered with information criteria and the log-likelihood. By simulating dependent multivariate data using estimated means, covariances and skew (and degrees of freedom for the multivariate t) an appreciation is gained of the difference between the observed data and its behaviour according to each copula. This has been measured using relative distances between the distributions and Kolmogorov-Smirnoff

Table 7: CORRELATIONS BETWEEN MORTALITY RISK, HEALTH AND LIFESTYLES FOR SYMMETRIC AND SKEWED t DISTRIBUTIONS

Symmetric *t* distribution								
	Mortality	SAH	Breakfast	Not obese	Non-smoker	Sleeping well	Prudent drinker	Exerciser
Mortality	**1**							
SAH	**-0.4099**	**1**						
Breakfast	**0.4506**	**0.2726**	**1**					
Not obese	0.0137	**0.3997**	**0.1461**	**1**				
Non-smoker	**0.2360**	**0.4120**	**0.7695**	**0.2018**	**1**			
Sleeping well	**-0.5503**	**0.4833**	-0.0110	**0.1848**	**0.0284**	**1**		
Prudent drinker	**0.1930**	**-0.0945**	**0.4329**	**-0.4870**	**0.4901**	-0.0571	**1**	
Exerciser	**-0.8139**	**0.6060**	**-0.1820**	**0.2563**	**-0.0987**	**0.5562**	**-0.3217**	**1**

Skewed *t* distribution								
	Mortality	SAH	Breakfast	Not obese	Non-smoker	Sleeping well	Prudent drinker	Exerciser
Mortality	**1**							
SAH	**-0.4247**	**1**						
Breakfast	**0.4281**	**0.2801**	**1**					
Not obese	0.0004	**0.4973**	**0.1840**	**1**				
Non-smoker	**0.1683**	**0.4287**	**0.7431**	**0.2248**	**1**			
Sleeping well	**-0.5201**	**0.4595**	0.0193	**0.2149**	0.0533	**1**		
Prudent drinker	**0.2180**	**-0.2013**	**0.3989**	**-0.4901**	**0.4495**	**-0.0968**	**1**	
Exerciser	**-0.8145**	**0.6204**	**-0.1685**	**0.3158**	-0.0404	**0.5279**	**-0.3686**	**1**

Table 8: CORRELATIONS BETWEEN MORTALITY RISK, HEALTH AND LIFESTYLES FOR THE JOINT DISTRIBUTION OF INVERTED MARGINAL CDFS

Observed data								
	Mortality	SAH	Breakfast	Not obese	Non-smoker	Sleeping well	Prudent drinker	Exerciser
Mortality	**1**							
SAH	**-0.3998**	**1**						
Breakfast	**0.4034**	**0.3042**	**1**					
Not obese	0.0250	**0.4407**	**0.1760**	**1**				
Non-smoker	**0.1859**	**0.4328**	**0.7514**	**0.2379**	**1**			
Sleeping well	**-0.5175**	**0.4960**	**0.0754**	**0.2065**	**0.0868**	**1**		
Prudent drinker	**0.1828**	**-0.1405**	**0.4127**	**-0.4740**	**0.4787**	**-0.0668**	**1**	
Exerciser	**-0.8008**	**0.6135**	**-0.1200**	**0.2832**	-0.0425	**0.5432**	**-0.3273**	**1**

Table 9: P-VALUES FROM KOLMOGOROV-SMIRNOFF TESTS FOR DIFFERENCES BETWEEN DISTRIBUTIONS ($p < 0.05$ REPRESENTS A STATISTICALLY SIGNIFICANT DIFFERENCE AT 5 PERCENT LEVEL OF SIGNIFICANCE)

	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Mortality	0.001	0.094	0.002	0
Sah	0.003	0.01	0.089	0.682
Breakfast	0.291	0.125	0.005	0.039
Not obese	0.024	0.201	0.004	0.263
Non-smoker	0.002	0.017	0	0.037
Sleeping well	0.205	0.307	0.037	0.065
Prudent drinker	0.124	0.121	0.021	0.277
Exerciser	0.26	0.002	0.001	0.065

tests.[13] The results are in Table 6.

Table 6 illustrates goodness of replication, not of fit. It also raises another potential trade-off between poorer fit and better approximation of the data-generating process. In part this is because, certainly in this instance, the results are equivocal, relative to the information criteria. More importantly though it will depend upon the analysis. For an explanatory model, more precise estimates of coefficients would be preferable. For a predictive model the information on replication would be more useful.

7.2.3 Inference and the variance-covariance of the estimates

There is an efficiency loss from using inference functions for the margins of a multivariate distribution. This is due to the partitioning of the variance-covariance matrix since, under inference, $\frac{\partial^2 l}{\partial \beta_u \partial \beta_v} = 0$ for the parameters (or vector of parameters, but vector notation is suppressed here for convenience) β_u and β_v from two separated margins $u \neq v$. Similarly for dependence θ_{uv} between any two margins u and v, the cross-partial derivatives $\frac{\partial^2 l}{\partial \theta \partial \beta_u}$ and $\frac{\partial^2 l}{\partial \theta \partial \beta_v}$ are practically inaccessible when using elliptical copulas based upon inversion. If functions of these estimates are required (one may wish, for example, to gauge the association between one or more regressors in two margins and the dependence between their linear predictions), or Fisher Information on marginal parameters within the joint distribution, a jackknifing procedure would be used.[14] For other copulas such as the Archimedean class, jackknifing may be preferable anyway, relative to finding a matrix of analytical solutions.

The particular advantage of the jackknife approach is that far less needs to be coded for analysis - only

the marginal likelihoods in the first step of the IFM and the joint in the second. Asymptotically consistent estimates will then, under jackknifing, provide asymptotically efficient estimates for the variance-covariance matrix of the regressors (Joe 2005).

7.3 Comparing size and significance

Figures 1-8 show the marginal effects and t-statistics for the covariates in the margins, for all of the models estimated, separated by equation. The marginal distribution preferred, according to information criteria, is indicated. The reference individual is female, single, not European Caucasian, living in London (Urban), degree-qualified and in the first social class, and employed full-time.

One noticeable result is the positive, statistically significant impact of being European caucasian on both the risk of mortality and being in good health. The impact is more significant on health, but is significant in both equations nonetheless. Being male generates an increased risk of mortality but has a small and insignificant - though also positive - effect on health. Being married and owning a house has a marked effect on reducing mortality risk. Controlling for social class (though not income directly), home ownership improves health, also. Marriage does not have an effect. Balia and Jones (2007) in fact excluded marital status from the mortality equation in their reduced form model, however it does appear to be highly significant in the lifestyles they retained. Being married may therefore be having a substantial indirect effect on mortality risk. Household size was also supposed to affect lifestyles, rather than health or mortality risk directly, and this appears to be the case. Household size has an important positive and negative affect only on prudent drinking and exercise respectively. Since the sample is restricted to individuals aged 40 years and over it is reasonable to take this as representative of behaviour with large families, particularly children.

With specific regard to model selection, the marginal effects and t-statistics due to age and its quadratic in the equation for mortality risk are particularly interesting. Selection criteria favoured the use of the complementary log-log model for mortality risk, in which age and age^2 show different results to the other models. Mortality risk is increasing with age but at a decreasing rate, only for the complementary log-log model. This would show up as a difference between a copula and the multivariate probit. The effect of age on health is consistent across all models as are the estimates for the remaining explanatory variables. Moreover, in the mortality equation alone it also appears the complementary log-log returns estimates nearest to the multivariate probit. This is a pattern occasionally repeated in other equations, but not as consistently across all parameters as in this equation.

Balia and Jones (2007) excluded parental smoking from the health equation in their structural form, however it has a consistently significant negative effect in the reduced form, particularly in the case of both parents smoking. Their economic significance is, apart from illness-related absence from work and not being

Figure 1: PARAMETER ESTIMATES AND t-STATISTICS FROM THE MORTALITY RISK EQUATION, ALL MODELS

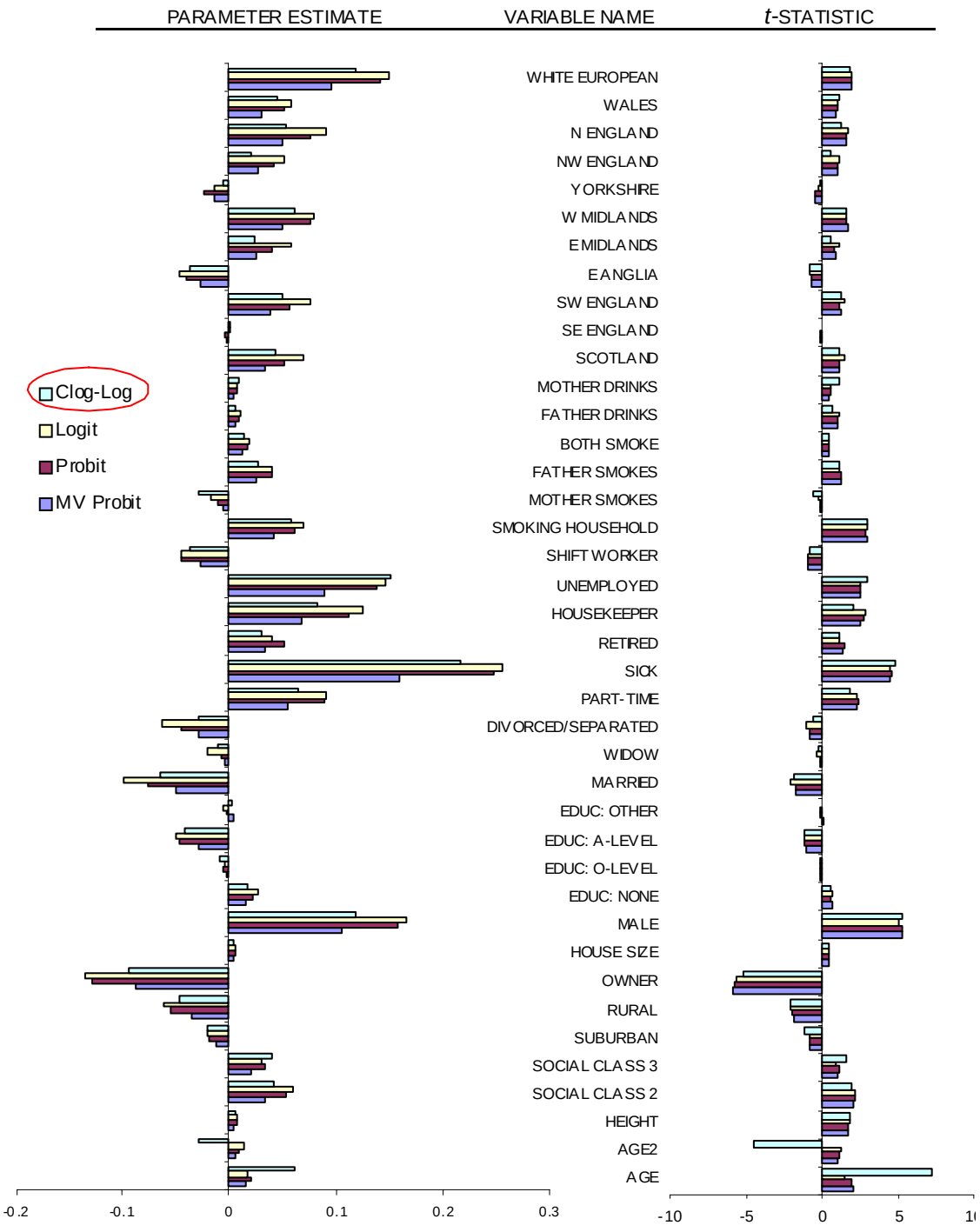


Figure 2: PARAMETER ESTIMATES AND t-STATISTICS FROM THE SELF-ASSESSED HEALTH EQUATION, ALL MODELS

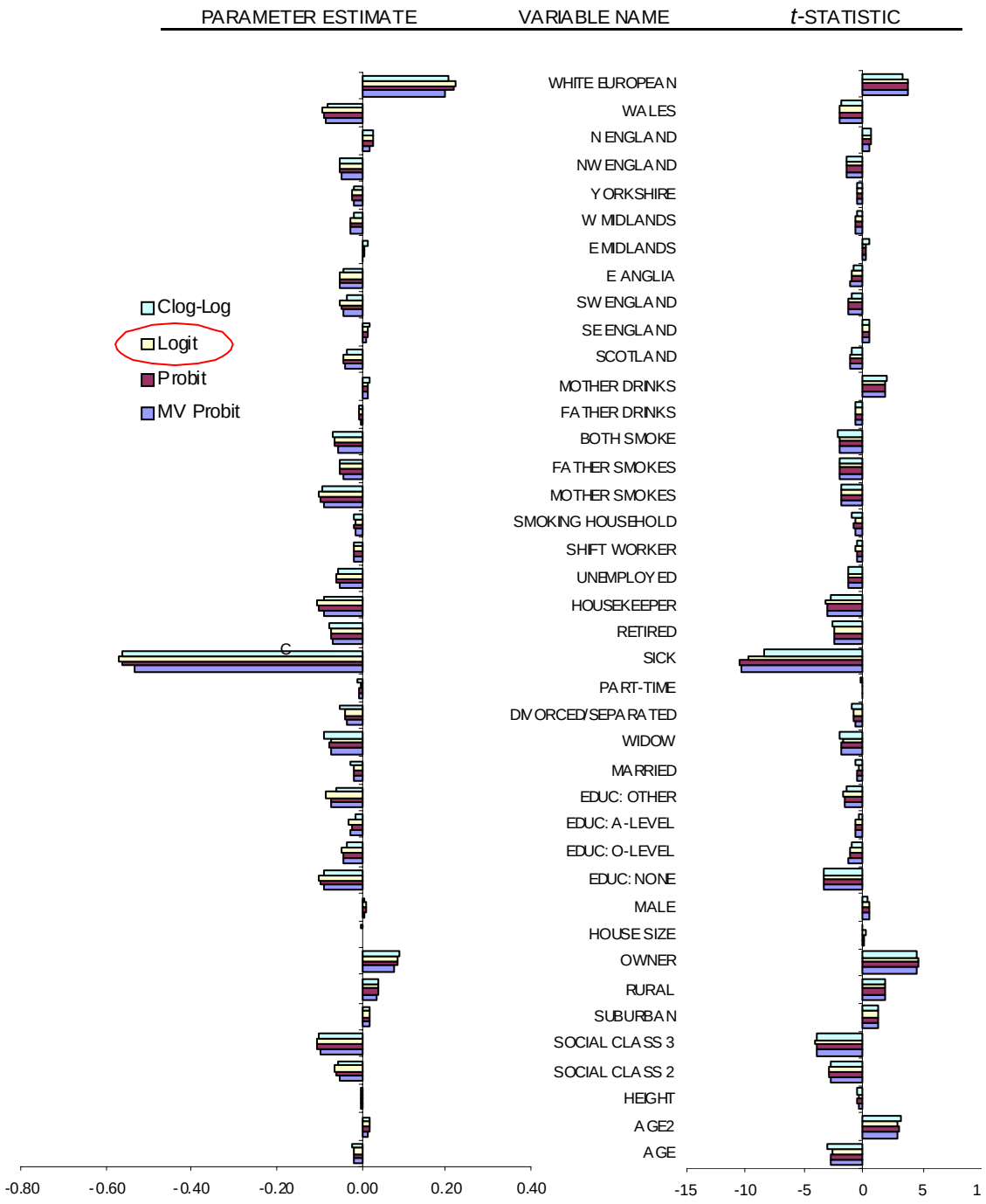


Figure 3: PARAMETER ESTIMATES AND t-STATISTICS FROM THE BREAKFAST EQUATION, ALL MODELS

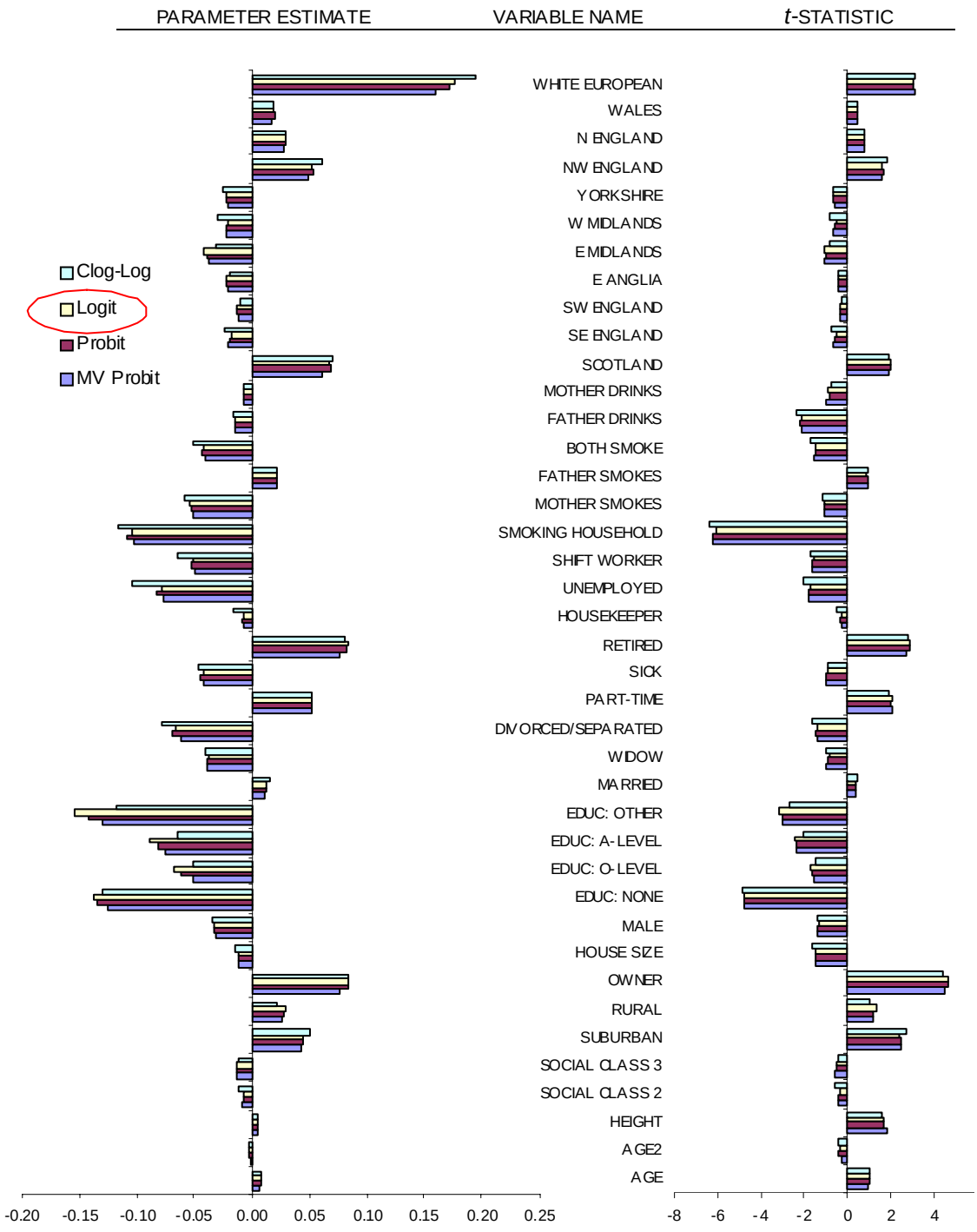


Figure 4: PARAMETER ESTIMATES AND t-STATISTICS FROM THE NOT OBESE EQUATION, ALL MODELS

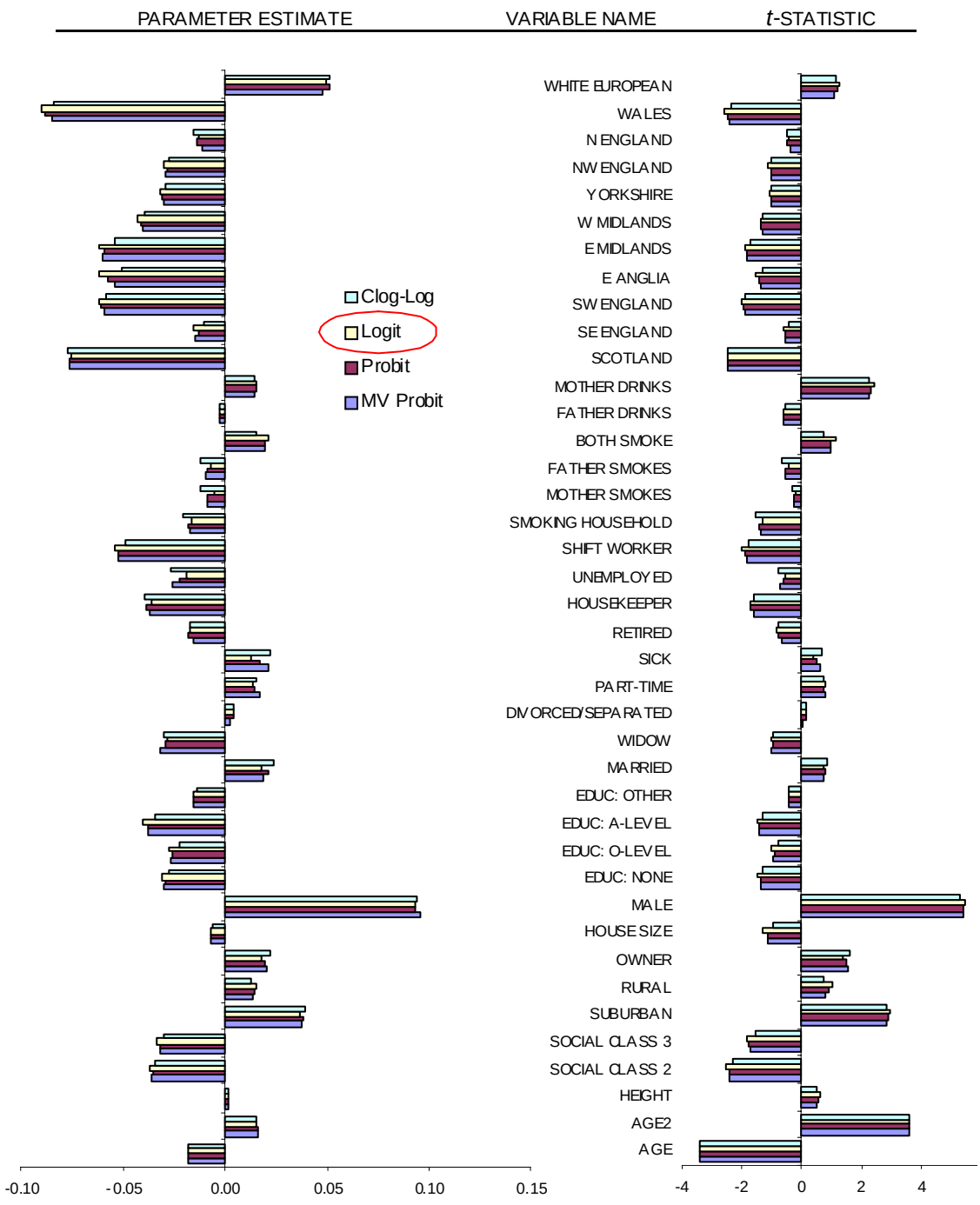


Figure 5: PARAMETER ESTIMATES AND t-STATISTICS FROM THE NOT SMOKING EQUATION, ALL MODELS

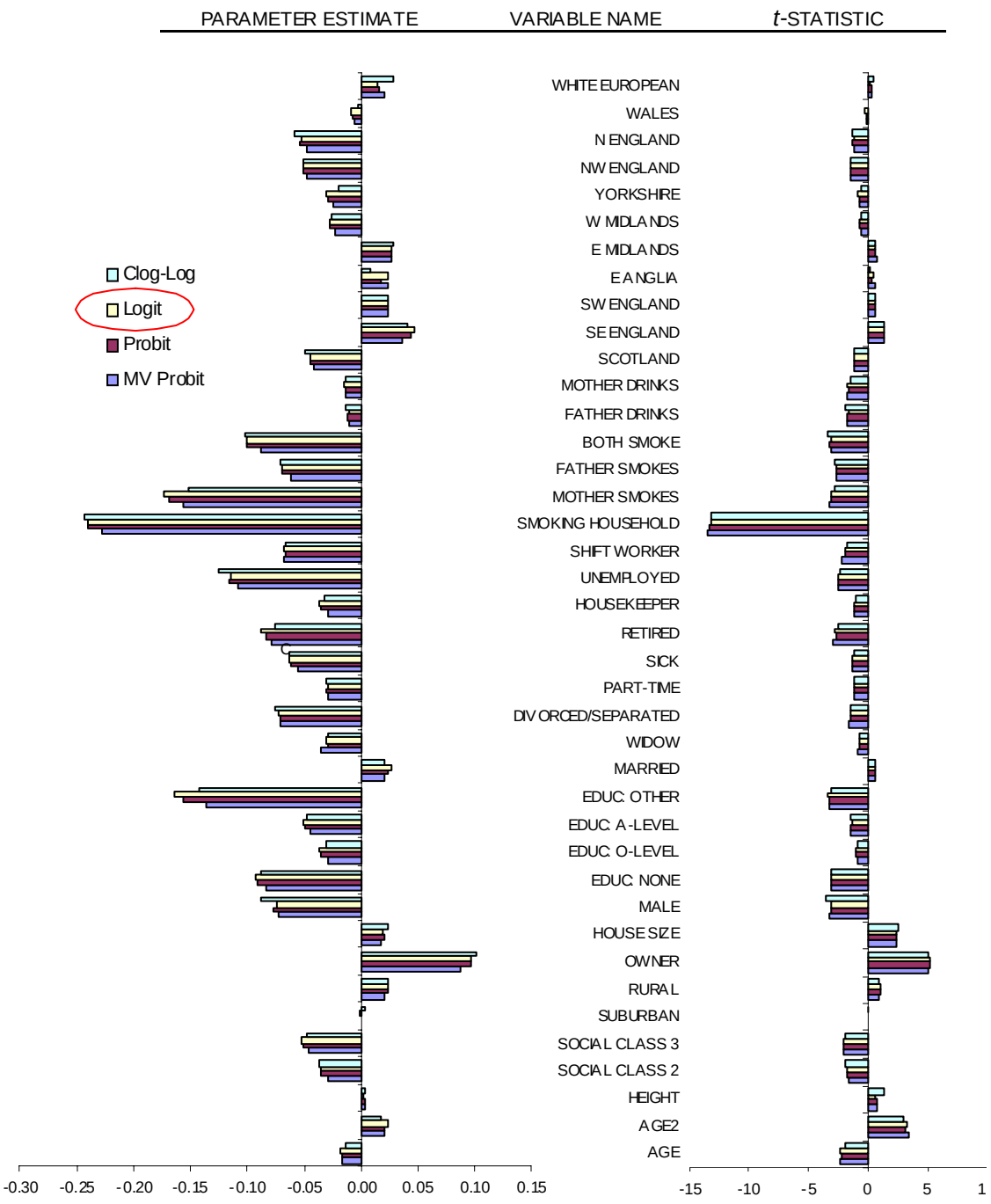


Figure 6: PARAMETER ESTIMATES AND t-STATISTICS FROM THE SLEEPING WELL EQUATION, ALL MODELS

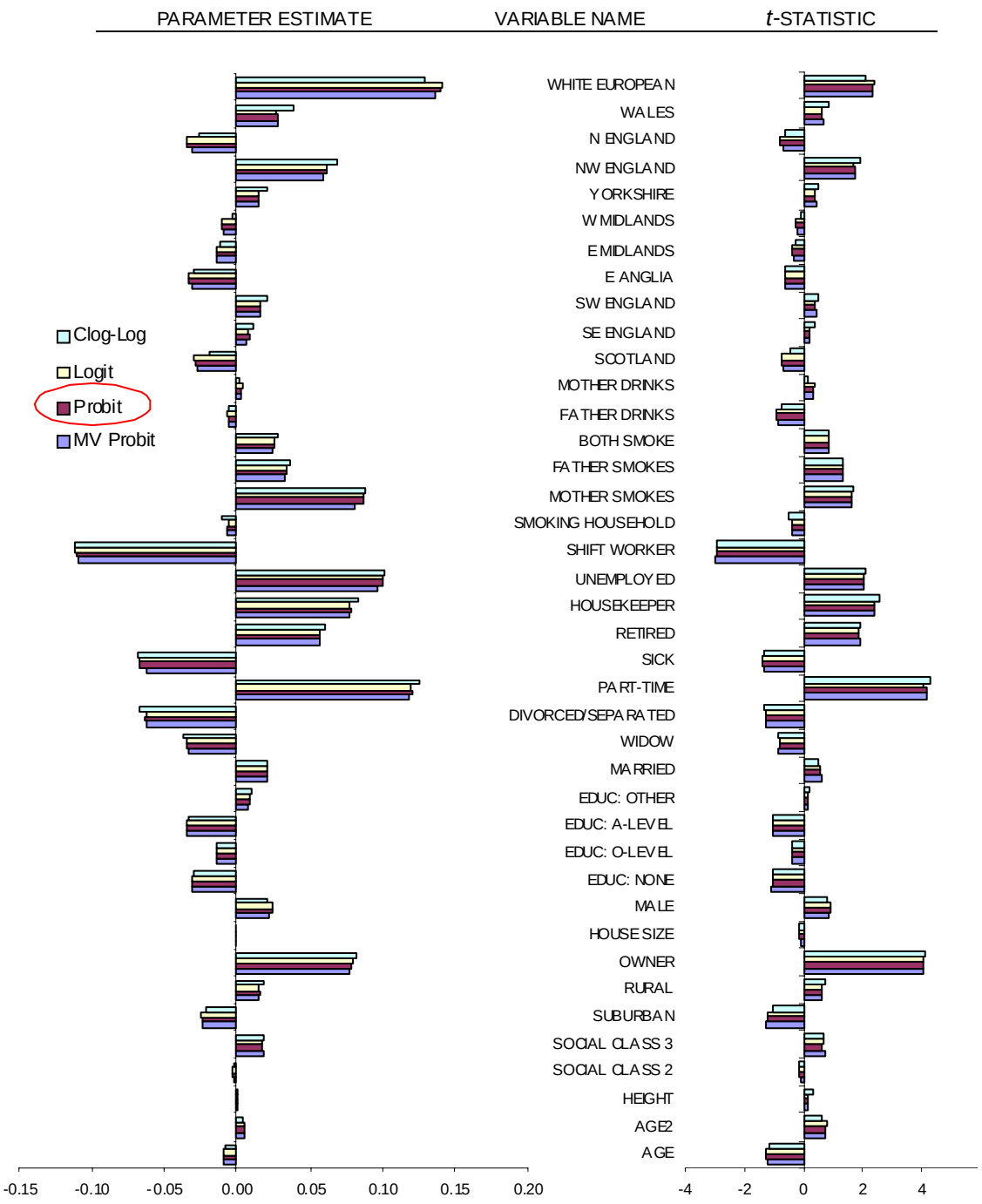


Figure 7: PARAMETER ESTIMATES AND t-STATISTICS FROM THE PRUDENT DRINKING EQUATION, ALL MODELS

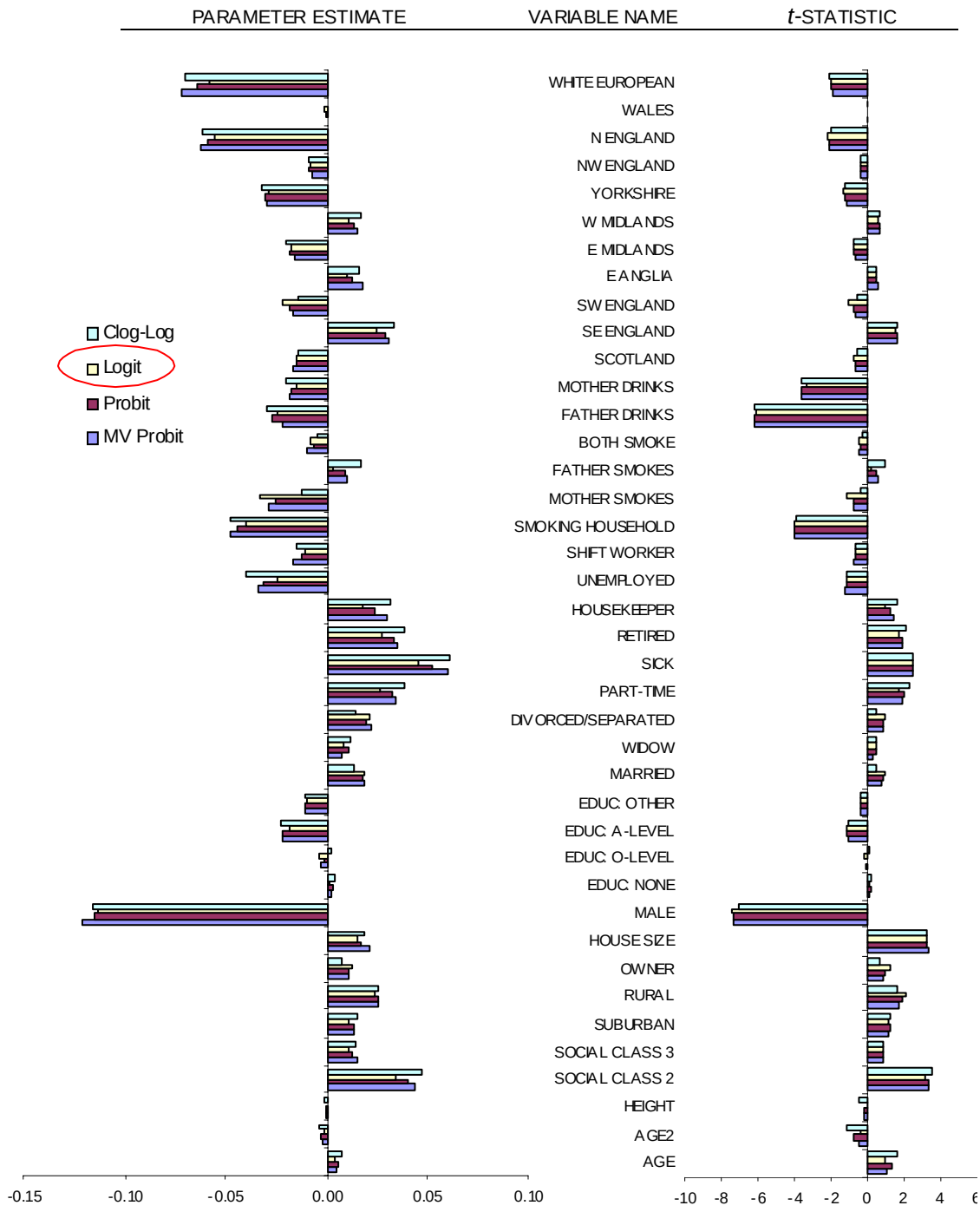
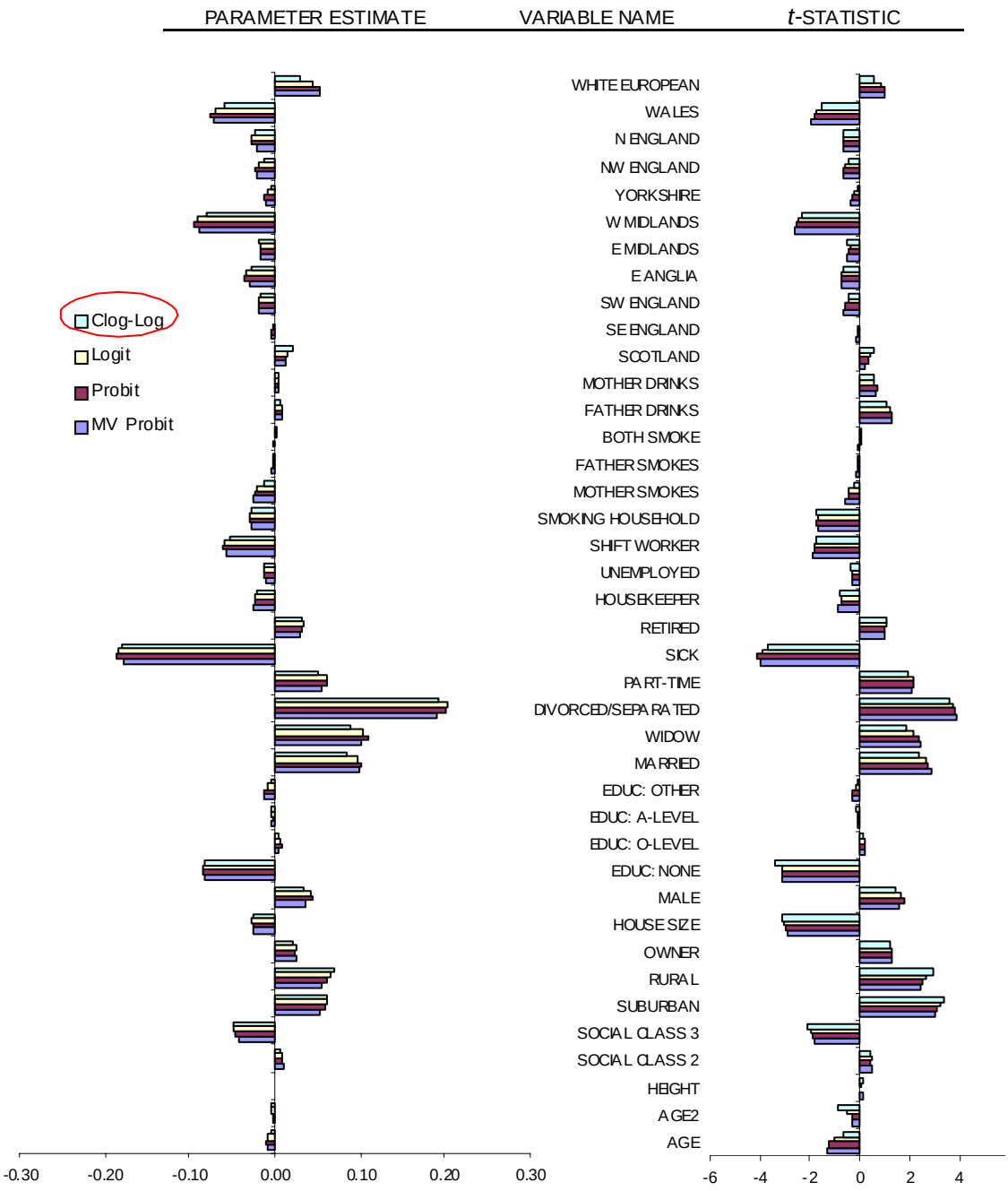


Figure 8: PARAMETER ESTIMATES AND t-STATISTICS FROM THE EXERCISING EQUATION, ALL MODELS



European caucasian, comparable to the other explanators of good or poor health, and only in the smoking equation is parental smoking elsewhere significant. This doesn't suggest Balia and Jones (2007) restricted parental smoking erroneously, though. The two effects could be reconciled by testing for a direct effect on health as well as the indirect effect via the propensity to smoke.

As stated above, home ownership has a significant role in determining eating breakfast, not smoking and sleeping well, with a reasonably-sized effect. Among the other statistically significant results are the large effect of being male on the likelihood of not being obese, the large effect of not being European caucasian on drinking imprudently (here being male has a larger effect still), and the non-smoking equation, in which the reference female seems the least likely of all to smoke. The strength of the effect of unemployment on the propensity to smoke stands out, also. Suburban living seems to provide lower chances of being obese, yet living almost anywhere else in the UK besides London increases those chances, which corresponds reasonably well with exercise, also.

Balia and Jones (2007) considered only not smoking, eating breakfast and sleeping well as endogenous, and non-obesity, prudent drinking and exercising as exogenous. The covariate explanation in these equations is significant overall though, as are the effects of covariates in each. This suggests some explanatory power may still be contained in these equations, for health and the risk of mortality. Exercise due to non-urban living and not being single might reflect the time of the survey and the age of the individuals (recall that the sample is restricted to individuals 40 years of age and over). Non-urban areas may also offer more space for outdoor sports and other activities, relative to cities.

In a few instances the non-normal marginal distribution has altered the significance of covariates, across the models. These are not considered to be drastic, however: something that was barely statistically significant may now barely be insignificant at a 5 or 10% level, not including age in the equation for mortality risk. The differences between the t-statistics are not large enough that a given covariate would have been either excluded from or included in the analysis. The differences that do exist however do suggest that copulas are a worthwhile comparator to an ordinary multivariate probit.

8 Discussion

The methods and results presented here lead to two conclusions: first, that more flexible approaches to estimating multivariate data, even multivariate dichotomous data, can provide more information on a given data-generating process.

Specific to the elliptical copulas, the results show an improvement in estimation due to approximating a joint distributions using the multivariate t distribution rather than the multivariate normal. There is

also an improvement above that when considering multivariate skewed distributions, rather than the more commonly-considered symmetric ones. The difference between skewness and symmetry alone is enough to alter what would otherwise have been thought about tail dependence in the joint distribution, which has significant implications for analysis concerning itself with extreme events.

9 Conclusion

This paper used a copula to represent a reduced-form system-of-equations model of mortality risk, health and lifestyles (diet, exercise, smoking, sleeping, alcohol consumption and obesity) is considered. Employing a method due to Lee (1983), McLeish and Small (1988) and Joe and Xu (1996) that uses inferencing for margins, the parameters of each lifestyle, health and mortality equation were estimated separately to the parameters of association that defined the joint distribution function.

The results showed that the assumptions underlying the multivariate probit are robust to the non-normality uncovered: covariate estimates and the variance-covariance matrices were comparable across the multivariate probit and the copulas used. As well as providing efficient estimates with less computation that the multivariate probit however, the copula enables estimation of multivariate skewed distributon functions for the system. In this instance, such skewness was found to be statistically significant, suggesting improved fit from the use of copulas.

The copula approach allows such flexibility where traditional models do not, particularly when analysing jointly-distributed discrete random variables, which is more cumbersome than with continuous random variables. The copula model for mortality risk, health and lifestyle was both able to capture idiosyncrasies of the data such as skewness and tail dependence, while also being simpler to implement and estimate.

Acknowledgements

This paper is a revised version of an earlier HEDG Working Paper (Quinn 2006). The advice of Murray Smith, Andrew Jones, Nigel Rice, Silvia Balia, Harry Joe, Stefano Conti and the Health, Econometrics and Data Group at the University of York. Financial support from the Centre for Health Economics is also acknowledged gratefully .

Notes

[1]As with the discussion of conditional distributions, transform invariance in this case is, for two random variables (X_1, X_2), invariance of $F_1(X_1)$ to the use of $F_2(X_2)$ or $F_3(X_2)$, if $F_2(.)$ and $F_3(.)$ are both almost surely increasing functions. Consequently, if the dependence between jointly uniformly-distributed (X_1, X_2) is defined by some copula $C(X_1, X_2)$, then $C(X_1, X_2) = C(X_1, F_2(X_2)) = C(X_1, F_3(X_2))$.

[2]This is not usually considered problematic since the region outside this is not usually of interest (Smith 2003).

[3]Lee (1983) does not refer to the method as IFM, though.

[4]Note that $X^2 \sim \chi^2$, irrespective of the value of α.

[5]Non-elliptical distributions will be valid for some α, conditional upon setting an appropriate transformation $W(.)$. Elliptical distributions will be valid for all α.

[6]That is to say, 2,433 of the original interviewees still in the system are deceased. The original HALS was not intended for follow-up, so that not all interviewees were collected in for the second HALS. Only around 2% from the original HALS are lost from the most recent mortality update, due either to leaving the country or having otherwise been dropped from the official National Health Service registry. See Cox (1988, 1995) and Contoyannis and Jones (2004) for more information and discussion of the surveys.

[7]'Prudent' alcohol consumption is given as less than 21 units of alcohol per week for males and less than 14 units per week for females (Contoyannis and Jones 2004). This does not distinguish between moderate drinking and abstinence, despite evidence that moderate alcohol consumption can be beneficial, as both Contoyannis and Jones (2004), and Balia and Jones (2007) discuss.

[8]Evidence has shown some more dexterity is required when using BMI. Deurenberg, Yap and van Staveren (1998), for example, find that the obesity-rated BMI should be lower for some cultural backgrounds, such as South-East Asian. In our 1984-5 sample anglo Europeans constitute around 98% of the available cultural backgrounds, so any such differential is unlikely to be problematic. No indications were found of systematic variations in obesity according to other backgrounds.

[9]This differs from the earlier study by Contoyannis and Jones (2004) due to a previous coding error, corrected in this paper (in Contoyannis and Jones 2004, home ownership for the entire population was about 87% - in fact it is about 63%). It now, due largely to greater variation, has much stronger correlation with the dependent variables. Comparison however gives no indication that previous results were affected

significantly by the higher value.

[10]Results from these tests, not presented, indicate significant differences when the Complementary Log-Log function was used, but no difference between the Probit and Logit.

[11]This step assumes that each combination of margins (or inverse probabilites) is accurate, leaving only dependence to be captured by the joint distribution. Thus the 'best' copula is taken to be representing the correlation structure of the latent variables in each margins most accurately.

Although the Multivariate Probit is included, its likelihood is not directly comparable due to functional form: unlike it, the copula distributions are 8 dimensions of inverse probabilities estimated non-parametrically, although for the information criteria the full $k = 356$ was used. Thus the information criteria should be used to compare the copulas, but not to infer that they are better than the standard Multivariate Probit. The argument is that the information and efficiency gains in the joint distribution, and the fit in the margins, are the advantages due to the use of copulas.

[12]Some explanation of the 'observed' data is required. These are the inverse predicted probabilities used by the copula for the joint distribution, as shown in Figure 4.1. So the 'observed' mean is the mean of the inverted predicted probabilites of $y = 1$ in each dimension, which was passed into each copula according to Equation (14).

[13]A chi-squared test, using quintiles, is equally feasible in this instance.

[14]Still suppressing vector notation, the converse cross-partial derivative $\frac{\partial^2 l}{\partial \beta \partial \theta_v} = 0$ (from a proof in the appendix of Joe 2005, not reproduced here).

REFERENCES

Abdous, B., Ghoudi, K., Khoudraji, A., 1999. Nonparametric estimation of the limit dependence function of multivariate extremes, Extremes, 22: 243-265.

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, New York.

Ali, M. M., Mikhail, N. N., Haq, M. S. 1978. A class of bivariate distributions including the bivariate logistic, Journal of Multivariate Analysis. 8: 405-412.

Allison, R. A., Foster, J. E. 2004. Measuring health inequality using qualitative data, Journal of Health Economics. 23: 505-524.

Almeida, C., Mouchart 2003. A note on a copula approach to polychoric correlations, Discussion Paper 03-22, Institute de Statistique, Université Catholique de Louvain, Louvain-la-Neuve.

Andersen, E. W. 2004. Composite likelihood and two-stage estimation in family studies. Biostatistics. 5:15-30.

Arellano, M., Honore, Bo., 2001. Panel data models: some recent developments, in Heckman, J. J., Leamer, E. E., eds. Handbook of Econometrics. Elsevier.

Arellano, M., Carrasco, R., 2003. Binary choice panel data models with predetermined variables, Journal of Econometrics, 115(1): 125-157.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew-normal distribution, Journal of the Royal Statistical Society B, 61: 579-602.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, Journal of the Royal Statistical Society B, 65: 367-389.

Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution, Biometrika, 83: 715-726.

Balia, S., Jones, A. M., 2007. Mortality, lifestyle and socio-economic status, Journal of Health Economics, forthcoming.

Bellio, R. and Varin, C., 2005. A pairwise likelihood approach to generalized linear models with crossed random effects. Statistical Modelling, 5: 217–227.

Belloc, N., Breslow, L. 1972. Relationship of physical health status and health practices, Preventive Medicine 1: 409–421.

Belzunce, F., Semeraro, P., 2004. Preservation of some dependence concepts under mixtures and applications, Journal of Applied Probability, 414: 961-974

Billingsley, P., 1995. Probability and Measure, 3rd ed. Wiley, New York.

Bommier, A., Stecklov, G. 2002. Defining health inequality: why Rawls succeeds where social welfare theory fails, Journal of Health Economics. 21: 497-513.

Bosmans, K., Lauwers, L., Ooghe, E. 2006. A consistent multidimensional Pigou-Dalton transfer principle. Center for Economic Studies Discussion Paper. Katholieke Universiteit, Leuven.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T. 2000. Copulas for finance: a reading guide and some applications. Groupe de Recherche Operationnelle, Credit Lyonnais.

Box-Steffensmeier, J. M., Smith, R. M., 1997. Heterogeneity and individual party identification, presented at the annual meetings of the Midwest Political Science Association, Chicago, 1996.

Braveman, P., 2006. Health disparities and health equity: concepts and measurement. Annual Review of Public Health. 27: 167-94.

Briggs, A., 2005. Statistical methods for cost-effectiveness analysis alongside clinical trials, in: Jones, A. M., ed. Elgar Companion to Health Economics. Edward Elgar Publishing, Cheltenham.

Calabrese, E. J., Baldwin, L. A., 1998. A general classification of U - shaped dose - response relationships, Human and Experimental Toxicology, 17:353 - 364.

Cameron, A. C., Li, T., Trivedi, P. K., Zimmer, D. M., 2004. Modelling the differences in counted outcomes using bivariate copula models with an application to mismeasured counts. Econometrics Journal, 72: 566-584.

Cebrián, A. C., Denuit, M., Lambert, P., 2003. Analysis of bivariate tail dependence using extreme value copulas: an application to the SOA medical large claims database. Discussion Paper 03-02, Institute de Statistique, Université Catholique de Louvain, Louvain-la-Neuve.

Chateauneuf, A., Moyes, P. 2005. Lorenz non-consistent welfare and inequality measurement, Journal of Economic Inequality. 2(2): 1-87.

Chernick, M. R., 1999. Bootstrap Methods: A Practitioner's Guide. Wiley, New York.

Claxton, K., 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies, Journal of Health Economics, 18: 341–364.

Claxton, K., Posnett, J., 1996. An economic approach to clinical trial design and research priority-setting. Health Economics, 5: 513-524.

Clayton, D. G., Cuzick, J., 1985. Multivariate generalizations of the proportional hazards model, Journal of the Royal Statistical Society A, 1482: 82-117.

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. 65:141-151.

Contoyannis, P., Jones, A. M., 2004. Socioeconomic status, health and lifestyle, Journal of Health Economics. 23(5): 965-995.

Contoyannis, P., Wildman, J. 2006. Using relative distributions to investigate socioeconomic inequalities in the Body-Mass Index in England and Canada. Paper presented at the 5th IHEA World Congress. Barcelona, Spain.

Cook, R. D., Johnson, M. E., 1981. A family of distributions for modelling non-elliptically symmetric and multivariate data, Journal of the Royal Statistical Society B, 432: 210-218.

Cragg, J. G., 1971. Some statistical models for limited dependent variables with applications to the demand for durable goods, Econometrica, 39: 829-844.

Dall'Aglio, G., Kotz, S., Salinetti, G., eds. 1991. Advances in Probability Distributions With Given Marginals: Beyond the Copulas. Kluwer Academic Publishers, Dordrecht.

Dardanoni, V., Lambert, P. J. 2001. Horizontal inequity comparisons, Social Choice. and Welfare. 18: 799-816.

Darsow, W., Nguyen, B., Olsen, E., 1992. Copulas and Markov processes, Illinois Journal of Mathematics, 36: 600-642.

De Castro, S., Goncalves, F., 2002. False contagion and false convergence clubs in stochastic growth theory, Discussion Paper 237, Departamento de Economia, Universidade de Brasilia, Brasilia.

Deb, P., Trivedi, P. K., 1997. Demand for medical care by the elderly: a finite mixture approach, Journal of Applied Econometrics, 12: 313-336.

Deheuvels, P., 1979. La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance, Bulletin de la Classe des Sciences Académie Royale de Belgique, 65: 274–292.

Demarta, S., McNeil, A. J., 2005. The t-copula and related copulas, International Statistical Review 73: 111-129.

Deurenberg, P., Yap, M., van Staveren, W. A., 1998. Body mass index and percent body fat: a meta analysis among different ethnic groups, International Journal of Obesity Related Metabolic Disorders, 2212: 1164-1171.

Devroye, L., 1986. Nonuniform Random Variate Generation, Springer Verlag, New York.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., Stoddart, G. L., 2005. Methods for the economic evaluation of health care programmes. 3rd Edition. Oxford University Press, Oxford.

Durrleman, V., Nikeghbali, A., Roncalli, T., 2000. Which copula is the right one? Groupe de Recherche Operationnelle, Credit Lyonnais.

Easwaran, K., 2004. Two analyses of conditional probability in terms of unconditional, mimeo, University of California, Berkley.

Easwaran, K., 2005. What conditional probability must (almost) be, mimeo, University of California, Berkley.

El Gouch, A., van Keilegom, I., McKeague, I. W., 2005. Empirical likelihood confidence intervals for dependent duration data, Discussion Paper 05-21, Institute de Statistique, Université Catholique de Louvain, Louvain-la-Neuve.

Embrechts, P., Lindskog, F., McNeil, A. J., 2003. Modelling dependence with copulas and applications to risk management, in Handbook of Heavy-tailed Distributions in Finance, Rachev, S. T., ed. Elsevier/North-Holland, Amsterdam.

Fang, K. T., Kotz, S., Ng, K. W., 1989. Symmetric Multivariate and Related Distributions. Chapman and Hall, New York.

Fenwick, E., Claxton, K., Sculpher, M., 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves, Health Economics Letters, 10: 779–787.

Fenwick, E., O'Brien, B.J., Briggs, A., 2004. Cost-effectiveness acceptability curves - facts, fallacies and frequently asked questions, Health Economics, 13: 405–415.

Fermanian, J-D., Wegkamp, M., 2004. Time dependent copulas. Working Paper, Centre de Recherche en Économie et Statistique, Malakoff.

Fermanian, J-D., 2005. Goodness of fit tests for copulas, Journal of Multivariate Analysis, 951: 119–152.

Frank, M.J., 1979. On the simultaneous associativity of F(x,y). and x + y – F(x,y). Aequationes Mathematica, 19: 194–226.

Fredericks, G. A., Neslen, R. B., 20007. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables, Journal of Statistical Planning and Inference. 137: 2143-2150.

Frees, E. W., Valdez, E. A., 1998. Understanding relationships using copulas, North American Actuarial Journal 2: 1-25.

Galambos, J., 1987. The Asymptotic Theory of Extreme Order Statistics, 2nd ed. Kreiger Publishing Company, Malabar.

Garry, R., Fountain, J., Brown, J., Manca, A., Mason, S., Sculpher, M., Bridgman, S., Gray, J., Lilford, R., 2004. eVALuate hysterectomy trial: a multi-centre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy. Health Technology Assessment, 826.: 1–154.

Garry, R., Fountain, J., Mason, S., Napp, V., , Hawe, J., Clayton, R., Abbott, J., Phillips, G., Whittaker, M., Lilford, R., Bridgman, S., Brown, J., 2004b. The eVALuate study: two parallel randomised trials, one comparing laparoscopic with abdominal hysterectomy, the other comparing laparoscopic with vaginal hysterectomy. British Medical Journal, 328: 129.

Genest, C., MacKay, R. J., 1986. The joy of copulas: bivariate distributions with uniform marginals, The American Statistician, 40: 280-283.

Genest, C., Rivest, L-P. 1993. Statistical inference procedures for bivariate Archim edean copulas, Journal of the American Statistical Association. 88 (423): 1034-1043.

Genius, M., Strazzera, E., 2004. The copula approach to sample selection modelling: an application to the recreational value of forests, Nota di Lavoro 73.2004. Fondazione Eni Enrico Matei, Milan.

Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C., 1996. Cost-effectiveness in Health and Medicine. Oxford University Press, New York.

Gordy, M. B., 1998. A generalization of generalized beta distributions, Finance and Economics Discussion Series 1998-18, Board of Governors of the Federal Reserve System, Washington DC.

Gouriéroux, C., Monfort, A., 1996. Simulation Based Econometric Methods. Oxford University Press, Oxford.

Gravelle, H. 2003. Measuring income related inequality in health: standardisation and the partial concentration index. Health Economics. 12: 803-819.

Greene, W. H., 2002. Econometric Analysis, 5th ed. Prentice-hall International Inc. New York.

Hajivassiliou, V. A., 1999. Some practical issued in Maximum Simulated Likelihood, in Mariano R, Schuermann T, Weeks MJ eds. Simulation-based Inference in Econometrics: Methods and Applications. Cambridge University Press, Cambridge.

Hernández Quevedo, C., Jones, A. M., López Nicolás, Á., Rice, N. 2006. Socioeconomic inequalities in health: a comparative longitudinal analysis using the European Community Household Panel. Social Science and Medicine. 635: 1246-61.

Hoch, J.S., Briggs, A.H., Willan, A., 2002. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Economics, 11: 415–430.

Hüsler, J., Reiss, R-D., 1989. Maxima of normal random vectors: between independence and complete dependence, Statistics and Probability Letters, 7: 283-286.

Hutchinson, T. P., Lai, C. D., 1990. Continuous Bivariate Distributions, Emphasizing Applications. Rumsby Scientific Publishing, Adelaide.

International Society for Equity in Health. Working definitions 2001.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman and Hall, London.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models, Journal of Multivariate Analysis. 942: 401-419.

Joe, H., Hu, T., 1996. Multivariate distributions from mixtures max-infinitely divisible distributions, Journal of Multivariate Analysis, 572: 240-265.

Joe, H., Xu, J. J. 1996. The estimation method of inference functions for margins for multivariate models. Technical Report no. 166. Department of Statistics. University of British Columbia, Vancouver.

Johnson, N. L., Kotz, S., 1972. Distributions in Statistics: Continuous Multivariate Distributions. John Wiley & Sons, New York.

Johnson, N. L., Kotz, S., Balakrishnan, N., 1997. Discrete Multivariate Distributions. John Wiley & Sons, New York.

Jones, A. M., Rice, N. 2004. Using longitudinal data to investigate socio-economic inequality in health. In: Health Policy and Economics: Opportunities and Challenges. Smith PC, Ginnelly L, Sculpher M. eds.. Open University Press, Berkshire.

Jouini, M. N., Clemen, R. T., 1996. Copula models for aggregating expert opinions, Operations Research, 44: 444-457.

Kadane, J. B., Schervish, M. J., Seidenfeld, T., 1986. Statistical implications of finitely additive probability, in Bayesian Inference and Decision Techniques With Applications. Goel, P., Zellner, A., eds. 59–76. North-Holland, Amsterdam.

Kakwani, N., 1977. Measurement of tax progressivity: an international comparison. The Economics Journal. 87: 71-80.

Kakwani, N., 1980. Income Inequality and Poverty, World Bank, New York.

Kakwani, N., Wagstaff, A., van Doorslaer, E., 1997. Socioeconomic inequalities in health: measurement, computation, and statistical inference, Journal of Econometrics. 771: 87-103.

Kenkel, D., 1995. Should you eat breakfast? Estimates from health production functions, Health Economics 4: 15–29.

Keselman, H. J., Wilcox, R. R., Lix, L. M., 2005. A generally robust approach to hypothesis testing in independent and correlated groups designs, Psychophysiology, 404: 586-96.

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes, Annals of Statistics 255: 2084–2102.

Kolesárová, A., Mordelová, J., 2006. Quasi-copulas and copulas on a discrete scale. Soft Computing. 106:495-501.

Kolmogorov, A.N., 1950. Foundations of theory of Probability. Chelsea Publishing Company, New York.

Kotz, S., Balakrishnan, N., Johnson, N. L., 2000. Continuous Multivariate Distributions, 2nd edition. John Wiley & Sons, New York.

Kowalski, C. J., 1973. Non-normal bivariate distributions with normal marginals, The American Statistician, 273: 103-106.

Kuk, A.Y.C., Nott, D.J., 2000. A pairwise likelihood approach to analyzing correlated binary data.Statistics and Probability Letters. 47: 329–35.

Kulpa, T., 1999. On approximation of copulas, International Journal of Mathematics and mathematical Science, 222: 259-269.

Lambert, P., Vandenhende, F., 2002. A copula based model for multivariate non normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. Statistics in Medicine, 21: 3197–3217.

Leclerc, A., Lert, F., Fabien, C., 1990. Differential mortality: some comparisons between England and Wales, Finland and France, based on inequality measures, International Journal of Epidemiology, 19: 1001-1010.

Lee, L-F., 1983. Generalized econometric models with selectivity, Econometrica, 512: 507-512.

Lee, M-J., 1996. Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models. Springer, New York.

Lessard, C., 2007. Complexity and reflexivity: two important issues for economic evaluation in health care. Social Science and Medicine forthcoming.

Lin, D.Y., 2002. Regression analysis of incomplete medical cost data. Statistics in Medicine, 227.: 1181–200.

Lindsay, B.L., 1988. Composite likelihood methods, in Statistical Inference from Stochastic Processes. Prabhu, N.U. American Mathematical Society, Providence.

Lindskog, F., 2000. Modelling dependence with copulas and applications to risk management, Working Paper, RiskLab, ETH Zurich, Zurich.

Liu, G., Zhao, Z., 1999. Stochastic cost-effectiveness analysis: a simultaneous marginal-effect approach. Value in Health, December: 420-428.

Maas, C. J. M., Hox, J. J., 2004. The influence of violations of assumptions on multilevel parameter estimates and their standard errors, Computational Statistics and Data Analysis, 46: 427-440.

Mardia, K. V., 1962. Multivariate Pareto distributions, Annals of Mathematical Statistics, 33: 1008-1015.

Mari, D.D., Kotz, S., 2001. Correlation and dependence. Imperial College Press, London.

de Matteis, R., 2001. Fitting copulas to data. Diploma thesis. University of Zurich, Zurich.

McLeish, D. L., Small, C. G., 1988. theory and Applications of Statistical Inference Functions. Lecture Notes in Statistics 44. Springer, New York.

Melchiori, M., 2003. Which Archimedean copula is the right one? http://www.defaultrisk.com/pp_corr_68.htm.

Mullahy, J., Portney, P., 1990. Air pollution, cigarette smoking, and the production of respiratory health, Journal of Health Economics 92: 193-205.

Mullahy, J., Sindelair, J., 1996. Employment, unemployment, and problem drinking, Journal of Health Economics, 154: 409-434.

Mullahy, J., Manning, W., 1995. Statistical issues in cost-effectiveness analyses. In: Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies, Sloan, F., ed. Cambridge University Press, New York.

Muthen, B., 1979. A structural probit model with latent variables, Journal of the American Statistical Association, 74: 807-811.

Muthen, B., 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators, Psychometrika, 49: 115-132.

Nelsen, R. B., Quesada Molina, J. J., Rodríguez Lallena, J. A., Úbeda Flores, M., 2004. Best-possible bounds on sets of bivariate distribution functions, Journal Multivariate Analysis, 902: 348-358.

Nelsen, R. B., Úbeda Flores, M., 2004. A comparison of bounds on sets of joint distribution functions derived from various measures of association, Communications in Statistics - Theory & Methods, 3310: 2299-2305.

Nelsen, R. B. 1999. An Introduction To Copulas. Springer Verlag, New York.

Nelsen, R. B. 2006. An Introduction To Copulas. 2nd Ed. Springer Verlag, New York.

Newey, W., Steigerwald, D., 1997. Asymptotic bias for quasi maximum likelihood estimators in conditional heteroskedasticity models. Econometrica, 65: 587–599.

Nixon, R.M., Thomas, S.G., 2005. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. Health Economics, 1412:1217-29

Nixon, R.M., Wonderling, D., Grieve, R., 2005. How to estimate cost-effectiveness acceptability curves, confidence ellipses and incremental net benefits alongside randomised controlled trials, technical report of the Medical Research Council Biostaticis Unit, Cambridge.

O'Hagan, A., Stevens, J.W., 2002. The probability of cost-effectiveness. BMX Medical Research Methodology, 2: 5-10.

Owen, A., 2001. Empirical Likelihood. Chapman& Hall/CRC, New York.

Panas, E., 2005. Generalized beta distributions for stock market data: testing the U-shape pattern, Applied Economics, 37: 191-199.

Patton, A. J., 2003. Applications of Copula Theory in Financial Econometrics. PhD dissertation, University of California, San Diego.

Peracchi., F. 2002. The European Community Household Panel: a review. Empirical Economics. 27: 63-90.

Perkins, P., Lane, T., 2003. Monte-Carlo simulation in MATLAB using copulas. MATLAB News & Notes, November 2003.

Phelps, C. E., Mushlin, A., 1991. On the near equivalence of cost-effectiveness and cost-benefit analysis. International Journal of Technology Assessment in Health Care, 71: 12-21.

Pitt, M. D., Chan, D., Kohn, R., 2006. Efficient bayesian inference for Gaussian copula regression, Biometrika, 933: 537-554.

Pollard, D., 2002. A User's Guide to Measure Theoretic Probability. Cambridge University Press, Cambridge.

Preston, S. H., Haines, M. R., Pdmuk, E., 1981. Effects of industrialization and urbanization on mortality in developed countries, in Solicited Papers Vol 2, HJSSP 19th International Population Conference, Manila. IUSSP, Liege, 1981.

Prieger, J.E., 2000. A flexible parametric selection model for non-normal data with application to health care usage, Journal of Applied Econometrics 174: 367-392.

Prokhorov, A. B., Schmidt, P., 2006. Redundancy, Robustness, and Validity of Copulas in Likelihood Models, presented at the North American Winter Meeting of the Econometric Society, Boston, 2006.

Quinn, C., 2005. Generalisable regression methods for cost-effectiveness using copulas. Working Paper 05/13 of the Health, Econometrics and Data Group at the University of York, York.

Quinn, C., 2006. Alternative methods for estimating systems of health equations, Working Paper 06/05, Health, Econometrics and Data Group, University of York, York.

Quinn, C., 2007a. Estimating reduced-form systems of equations using copulas, mimeo.

Quinn, C., 2007b. Improved precision in cost-effectiveness analysis using copulas, mimeo.

Quinn, C., 2007c. Measuring income-related inequalities in health using copulas, mimeo.

Resnick, S. I., 1987. Extreme Values, Regular Variation, and Point Processes. Springer Verlag, New York.

Romano, C., 2002. Calibrating and simulating copula functions: an application to the Italian stock market, Working Paper 12/2002, Centro Interdipartimentale sul Diritto e l'Economia dei Mercati, Rome.

Roncalli, T., 2001. Modelling dependence in finance using copulas, Groupe de Recherche Operationnelle, Credit Lyonnais.

Rosa Dias, P., Jones, A. M. 2007. Giving equality of opportunity a fair innings. Health Economics. 16: 109-112.

Ruhm, C., 2005. Healthy living in hard times, Journal of Health Economics, 242: 341-363.

Satterthwaite, S. P., Hutchinson, T. P., 1978. A generalisation of Gumbel's bivariate logistic distribution, Metrika, 25: 163-170.

Savu, C., Trede, M. 2006. Hierarchical Archimedean copulas. Mimeo.

Schweizer, B., 1991. Thirty years of copulas, in Advances in Probability Distributions With Given Marginals: Beyond the Copulas. Dall'Aglio, G., Kotz, S., Salinetti, G., eds. Kluwer Academic Publishers, Dordrecht.

Sculpher, M., Manca, A., Abbott, J., Fountain, J., Mason, S., Garry, R., 2004. Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. British Medical Journal, 328:134.

Seger, J., 2004. Extreme-value copulas, Nekst, 131: 38–41.

Seidenfeld, T., Schervish, M., Kadane, J., 2001. Improper regular conditional distributions, The Annals of Probability 29: 1612-1624.

Sen, A., 1973. On Economic Inequality. Norton, New York.

Shao, J., Tu, D., 1995, The Jackknife and Bootstrap. Springer-Verlag,New York.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leur marges. Publications of the Institute of Statistics. University of Paris. 8: 229-231.

Smith, M. D., 2003. Modelling sample selection using archimedean copulas. Econometrics Journal, 6: 99–123.

Smith, M. D., 2005. Stochastic frontier models with dependent error components, Proceedings of the Econometric Society World Congress, London, 2005.

Smithson, M., Verkuilen, J., 2006. A better lemon-squeezer? Maximum-likelihood regression with Beta-distributed dependent variables. Psychological Methods, 111: 54-71.

Stinnet, A., Mullahy, J., 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making, 18: S68–S80.

Tajar, A., Denuit, M., Lambert, P., 2001. Copula-type representation for random couples with Bernoulli margins, Discussion paper 01-18, Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve.

Takahasi, K., 1965. Note on the multivariate Burr's distribution, Annals of the Institute of Statistical Mathematics, 17: 257-260.

Trivedi, P. K., Zimmer, D. M., 2006. Copula modeling: an introduction for practitioners, Foundations and Trends in Econometrics, 11: 1-110.

van Doorslaer, E., Jones A. M. 2003. Inequalities in self-reported health: validation of a new approach to measurement Journal of Health Economics. 221: 61-87.

van Doorslaer, E., Koolman, X., 2004. Explaining the differences in income-related health inequalities across European countries, Health Economics, 137: 609-628.

van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, Ulf-G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., Leu, R. E., O'Donnell, O., Propper, C., Puffer, F., Rodriguez, M., Sundberg, G., Winkelhake, O., 1997. Socioeconomic inequalities in health: some international comparisons, Journal of Health Economics, 161: 93-112.

van Hout, B.A., Al, M.J., Gordon, G.S., Rutten, F.F., 1994. Costs, effects and C/E-ratios alongside a clinical trial. Health Economics, 35: 309-319.

Vandenhende, F., Lambert, P., 2000. Modeling repeated ordered categorical data using copulas, Discussion Paper 00-25, Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve.

Vandenhende, F., Lambert, P. 2003. Improved rank-based dependence measures for categorical data. Statistics & Probability Letters. 63: 157-163.

Vandenhende, F., Lambert, P., 2004. Local dependence estimation using non-parametric archimedean copulas. Discussion Paper 04-02, Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve.

Vanness, D.J., Mullahy, J., 2005. Perspectives on mean-based evaluation of health care, Jones AM ed. Elgar Companion to Health Economics. Cheltenham, UK: Edward Elgar Publishing, forthcoming.

Vázquez-Polo, F.J., Negrín Hernández, M.A., González López-Valcárcel, B., 2004. Using covariates to reduce uncertainty in the economic evaluation of clinical trial data. Health Economics, 146: 545–557.

Vella, F., 1998. Estimating models with sample selection bias: a survey, Journal of Human Resources, 3: 127-144.

Verhoeven, P., McAleer, M., 2003. Fat tails and asymmetry in financial volatility models, Centre for International Research on the Japanese Economy Discussion Paper F-211, Centre for International Research on the Japanese Economy, The University of Tokyo, Tokyo.

Vinod, H. D., 2005. Skew densities and ensemble inference for financial economics, The Mathematica Journal. 9(4): 852-862.

Wagstaff, A., van Doorslaer, E., 2000. Income inequality and health: what does the literature tell us? Annual Review of Public Health, 21: 543-67.

Wagstaff, A., van Doorslaer, E., Paci, P. 1989. Equity in the finance and delivery of health care: some tentative cross-country comparisons. Oxford Review of Economic Policy. 51: 89-112.

Wagstaff, A., van Doorslaer, E., Watanabe, N. 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam, Journal of Econometrics. 112: 207-223.

Wagstaff, A., Paci, P., van Doorslaer, E. 1991.. On the measurement of inequalities in health. Social Science and Medicine. 335: 545-557.

Willan, A.R., Briggs, A.H., Hoch, J.S., 2004. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Economics, 13: 461–475.

Willan, A.R., Lin, D.Y., Manca, A., 2005. Regression methods for cost-effectiveness analysis with censored data. Statistics in Medicine, in press.

Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.

Xu, J. J., 1996. Statistical modelling and inference for multivariate and lingitudinal discrete response data. PhD thesis. Department of Statistics. University of British Columbia.

Zarnoch, S. J., English, D. B. K., & Kocis, S. M. 2004. A model for evaluating dispersed outdoor recreation use estimation. In: Proceedings of the Second International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas, June 16-20, Rovaniemi, Finland.

Zheng, B., 2006. Measuring health inequality and health opportunity. paper presented at the UNU-WIDER conference. United Nations University. Helsinki.

Zimmer, D. M., Trivedi, P. K., 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.