HEDG Health, Econometrics and Data Group
THE UNIVERSITY of York

HEDG Working Paper 06/05

Alternative methods for estimating systems of (health) equations

Casey Quinn

July 2006

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

Alternative methods for estimating systems of (health) equations*

Casey Quinn†

July 12, 2006

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. A copula approximation of the joint distribution allows one to avoid usually implicit distributional assumptions, allowing potentially more robust and efficient estimates to be retrieved. By applying the theory of inference functions 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.

The use of copulas also enables estimation of skewed multivariate distributions for the latent variables in a multivariate model of discrete response variables. This flexibility provides more precise estimates with more appropriate distributional assumptions, but presents explicit trade-offs during analysis. Information that can be retrieved concerning distributional assumptions, skewness and tail dependence require prioritisation such that different needs could generate a different 'best' model even for the same data.

*The author is grateful to Andrew Jones, Nigel Rice, Silvia Balia, Harry Joe and the Health, Econometrics and Data Group at the University of York. Financial support from the Centre for Health Economics is also gratefully acknowledged.

†Correspond to the author at the Centre for Health Economics, The University of York, York YO10 5DD, England, or email cq1@york.ac.uk.

Keywords: health, lifestyle, mortality multivariate models, copulas, inference functions.

JEL classification: **C1, C3, I1**

1 Introduction

In this paper the simultaneous explanation of mortality risk, health and lifestyles is considered, using a reduced-form 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 inference functions, 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 socio-economic status and health, as well as income inequalities and health inequalities (van Doorslaer, Wagstaff and Bleichrodt *et al* 1997, Wagstaff and van Doorslaer 2000, van Doorslaer and Koolman 2004). Lifestyles need to be considered in this context because they may determine health status and mortality. Contoyannis and Jones (2004) and Balia and Jones (2005), 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 socio-economic characteristics, altering the explanation of inequalities in health and mortality.

Moreoever, lifestyles can be supposed endogenous: as well as determining health and mortality, lifestyles can themselves be determined by factors such as income and education, for example. In the context of the 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 these studies, using the same data and a similar underlying thesis: that individual lifestyle choices determine health outcomes, including health and mortality. These choices are influenced by socio-economic characteristics; to some extent those socio-economic characteristics have a direct effect on health outcomes also, controlling for lifestyle choices. It is a methodological extension also: Contoyannis and Jones (2004) used the British Health and Lifestyle Survey (HALS), considering unobserved heterogeneity via a recursive system of equations for self-assessed health and some endogenous lifestyles. Balia and Jones (2005) 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.

Other empirical analyses have used various single-equation methods, including interaction effects, as well as instruments, to capture endogeneity using Two-Stage Least-Squares and the Generalised Method of Moments (Ruhm 2005, Aster 1969, Mullahy and Portney 1990 and Mullahy and Sindelar 1996 respectively, for example). The multivariate structure of equations provides more flexibity when approximating the true explanation of health because it not only considers endogeneity but gives it a structural representation in the model. This paper will use predominantly the same economic model as Contoyannis and Jones (2004) and Balia and Jones (2005), however the econometric problem is considered differently.

The results presented here 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 estimates of variance-covariance are comparable across the multivariate probit and the copulas used. As well as providing efficient estimates more simply than the multivariate probit however, the copula is used to highlight the statistical significance of skewness in the multivariate distribution, and its relation to what would otherwise be recognised as tail dependence. The copula provides much more information about the data and the accuracy of its analysis, as well as facilitating more choice for the

researcher about exactly what they want to analyse in a multivariate framework.

2 Econometric problem

The behavioural model is as in Balia and Jones (2005); that individuals are assumed to maximise simultaneously the utility function

$$\max_{O} \sum_{t=0}^{\infty} \beta_t \pi_t \times u\left(O_{lt}, H_t; X_U, \mu_U\right) \tag{1}$$

where t-th period utility is determined by the vector of l (in this case 6) lifestyles O_{lt}, health H_t and conditional upon exogenous variables X_U and unobserved μ_U, which influence individual preferences. Similarly β_t influences time preferences, while the probability of survival period-by-period is given by π_t. Thus 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 M, H and $O_1, .., O_l$ are indicated by dichotomous variables (including self-reported health). Making the assumption that these follow a linearly-determined latent scale, following Balia and Jones (2005) gives the reduced form

$$y^*_{im} = \beta'_m \mathbf{X}_{im} + \varepsilon_{im} \tag{2}$$

$$y^*_{ih} = \beta'_h \mathbf{X}_{ih} + \varepsilon_{ih} \tag{3}$$

$$y^*_{il} = \beta'_l \mathbf{X}_{il} + \varepsilon_{il} \tag{4}$$

such that

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

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

$$y_{il} = 1\left(y^*_{il} \geq 0\right) \tag{7}$$

The vectors $\mathbf{X}_{im}, \mathbf{X}_{ih}$ and $\mathbf{X}_{il}$ are individual-specific exogenous vectors explaining, respectively, mortality risk, health and lifestyle. Under a structural specification these would be distinct due to exclusion restrictions needed to satisfy simple order conditions for identification, but in reduced-form this can be relaxed. Here they are the same.

Estimation in Balia and Jones (2005) is done via the method of Maximum Simulated Likelihood (MSL), assuming the errors terms are correlated and the random components μ_l, μ_H and μ_M are jointly normally distributed (in latent form ε_l, ε_H and ε_M). MSL is used because the standard multivariate probit is underlied by an 8-dimensional normal distribution, for which the standard method of Maximum Likelihood and (generalised) Method of Moments would require substantially more computation. The method of maximum likelihood, 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 this approach is, in the first place, the estimation itself, which can be cumbersome and not necessarily efficient compared to standard methods of maximum likelihood (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. The method presented here will help identify if these are problems, while showing a more convenient procedure for esitmation.

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, Wilcox and Lix 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, West and Finch 1996). In a structure-

of-equations model, multivariate non-normality can also lead to erroneous rejection of some models within the structure (see Klein 1998, for example).

The multivariate normal distribution is commonly selected for the convenience of its use, and because the univariate normal distribution is robust under reasonable levels of non-normality, and so explains the margins of the joint distribution fairly well (Kowalski 1973). Its use is a result of the common practice of selecting a multivariate distribution according to identification of the margins. Since the normal distribution is among the most robust, it is preferred to others such as the multivariate Pareto, Burr or logistic, for example (Mardia 1962, Takahasi 1965, Satterthwaite and Hutchinson 1978; Cook and Johnson 1981 present a generalised model that nests each of these as special cases).

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 dominate even when the joint density of the data being analysed 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 joint relationship 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 three 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). The empirical relevance of this is that non-normal (quasi-) Maximum Likelihood estimates of conditional moments are valid only when the conditional mean is identically zero, or the assumed and true densities are both unimodal and conditionally symmetric

about zero (Newey and Steigerwald 1996, Verhoeven and McAleer 2003). When this is not the case the estimates are not necessarily consistent, and the location and scale of the conditional distribution will not be identified correctly. In order to consider non-normal and/or skewed latent variables then, 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 subsequently 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 flexibity in the face of multivariate asymmetry.

3 The copula method

Using Sklar's (1959) theorem, all multivariate distributions can be held to 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. For any multivariate distribution with given margins there exists a copula that binds those margins to form the joint distribution precisely (Smith 2003). A copula in practice is a dependence function, and each one represents a unique description of the relationship between its margins, while the distributions of its margins are assigned separately, and with no consideration given as to the form of the copula. For the purposes of analysis, a copula is a distribution function for uniformly-distributed random variables. Since univariate CDFs are uniformly-distributed, the marginal CDF of each dependent variable can be considered a monotonic transformation, and used as a random variable in the copula.

Consider two random variables X_1, X_2 with bivariate distribution function $H(x_1, x_2)$ and univariate marginal distributions $F_1(x_1)$ and $F_2(x_2)$ respectively. Then there exists a copula C such that

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{8}$$

for all real values of x_1, x_2 (or $(X_1, X_2) \in \mathbb{R}$). If F_1, F_2 are continuous, C is unique. Under discontinuity C is uniquely determined on its domain, the range of the margins $\text{Ran}F_1 \times \text{Ran}F_2$.[1] 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 (8) is a joint distribution function (see Nelsen 1999 for this proof). By taking the marginal distribution functions as dependent variables, which do not contain the dependence structure, the copula separates the explanation of X_1 and X_2 from their association, an important distinction between the copula and standard multivariate distributions.

3.1 Multivariate FGM copulas

The FGM copula is the most commonly seen copula in exposition, since lower polynomials are more convenient for discussion (Smith 2003, Zimmer and Trivedi 2006). It is also derivative of another single-parameter family of copulas, the Frank.[2,3] The bivariate FGM copula for any u, v in $\mathbf{I} \in [0, 1]$ is C such that

$$C_\theta(u, v) = uv(1 + \theta(1 - u)(1 - v))\,|_{u=F_1(x_1), v=F_2(x_2)} \tag{9}$$

where $-1 \leq \theta \leq 1$, with positive and negative dependence for $\pm\theta$ respectively, and recalling that $F_1(x_1)$, $F_2(x_2)$ are (at least) monotonic.[4] In practical applications this copula has

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

[2] The FGM copula is a first-order Taylor approximation of the more flexible Frank copula. Its subsequent linearity in the margins has made it a popular exemplar (Smith 2003, Zimmer and Trivedi 2006).

[3] 'Single-parameter' refers to the parameterisation of association: single-parameter families use only one parameter of association. Joe's (1997) presentation of single-and multiple-parameter copulas is particularly useful in this regard.

[4] This is trivial: since $u = F_1(x_1), v = F_2(x_2)$ are univariate distribution functions they must be monotonic, at least (Nelsen 1999 contains an explanation of quasi-inverses of non-strictly increasing margins,

been shown to be a somewhat limited measure of dependence (Prieger 2002). Dependence $\theta \in [-1, 1]$ corresponds approximately to Spearman and Pearson's correlations $\rho \in \left[-\frac{1}{3}, \frac{1}{3}\right]$, such that $\rho = \frac{\theta}{\pi}$, and Kendall's $\tau \in \left[-\frac{2}{9}, \frac{2}{9}\right]$ such that $\tau = \frac{2\theta}{9}$ (de Matteis 2001. An appendix in Quinn 2000 also derives this condition). Mari and Kotz (2001) provide several extensions of the FGM copula, which can expand this range, but add more parameters.

The FGM is an example of a simple closed-form multivariate CDF, where Equation (9) is extended into n dimensions. Using the notation of Joe (1997), the multivariate FGM copula can be given as

$$C(x_1, .., x_8; \theta) = \prod_{i=1}^{8} u_i \left(1 + \sum_{1 \leq i < j}^{k=8} \theta_{ij} \left[1 - u_i\right] \left[1 - u_j\right]\right) \tag{10}$$

giving, like the 8-dimensional normal distribution, ${}^8C_2 = 28$ bivariate association parameters (since $\theta_{ij} = \theta_{ji} \; \forall \; i \neq j$).[5] Here $\theta_{ij} \in [-1, 1]$ as before, however more restrictions are introduced: θ_{ij} faces a limit also in sum, so that more margins means a narrower range of dependence for each non-zero θ_{ij}. Specifically

$$1 + \left| \sum_{1 \leq i < j < n}^{n-1} \theta_{ij} \right| \leq \theta_{1n} \leq 1 + \left| \theta_{12} - \sum_{2 \leq i < j \leq n}^{n} \theta_{ij} \right| \tag{11}$$

so that $\lim_{n \to \infty} \theta_{ij} = 0$. In fact this limit is much narrower: in practice much fewer than 28 unique values for θ_{ij} would be practicable. Thus, although the multivariate FGM offers a parameter for assocation in each bivariate margin, this is not typically feasible in practice.

which can also be used to construct a copula). This property is also necessary to ensure the measure of association, θ, 'obeys' the rules for measures of dependence.

[5] Nelsen (1999) and Mari and Kotz (2001), whose presentation draws on that of Nelsen (1999), provide a different form for the multivariate FGM, giving

$$C(x_1, ..., x_8; \theta) = \prod_{i=1}^{8} F_i(x_i) \left(1 + \sum_{k=2}^{8} \sum_{1 < j_1 < ... < j_k < n} \theta_{j_1, ... j_k} \left[1 - F_{j_1}(x_{1j})\right] ... \left[1 - F_{j_k}(x_{j_k})\right]\right)$$

which contains not ${}^nC_2$ but $2^n - n - 1$, or $\sum_{i=2}^{8} {}^8C_i$. In the current problem this would mean 247 different $\theta_{j_1, ... j_k}$terms, which is not considered practicable. Estimation issues aside, the limits on θ in multivariate FGM copulas would render them all null.

This limitation is commented upon specifically in Prieger's (2002) application of the FGM to the problem of sample selection, in a bivariate context.

3.2 Multivariate Archimedean copulas

Archimedean copulas are a particular class of copula that includes several popular families. These are copulas whose form, in n dimensions, is reduced to a single function, called a generator. This is a strictly decreasing, convex and continuous function $\varphi : [0,1] \rightarrow [0,\infty]$ in a set Ω of the same, where $\varphi(0) = \infty$, $\varphi(1) = 0$ and with inverse $\varphi^{-1} : [0,\infty] \rightarrow [0,1]$, $\varphi^{-1}(0) = 1$ and $\varphi^{-1}(\infty) = 0$.

For (u, v), an Archimedean copula is C such that

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{12}$$

An example is the Frank copula, given by (Frank 1979)

$$C(u,v;\theta) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right) \tag{13}$$

As an argument in only one margin, the generator φ can be used to extend Archimedean copulas into higher dimensions easily. For example, for u, v, w, z in $\mathbf{I}$,

$$C(u,v,w,z) = \varphi^{-1}(\varphi(u) + \varphi(v) + \varphi(w) + \varphi(z)) \tag{14}$$

All that is required to extend C is the addition of the generator function for a new margin. Note that φ_θ belongs to a single-parameter family of generators. Two-parameter generators also exist, but will not be used here. A pseudo-generation of a multiple-parameter copula will be achieved with mixtures below.

Multivariate Archimedean class copulas are a popular choice, however estimation in n-dimensions can be quite limited: for any $n > 2$-variate distribution to be a copula, the generator $\varphi_\theta^{-1} \in [0,\infty)$ must be completely monotonic. In Archimedean copulas that extend to negative dependence, φ_θ^{-1} fails to be monotonic when $\theta \in \tau < 0$ and $n > 2$: Archimedean copulas with $n > 2$ margins cannot contain negative dependence and still be a distribution.

In capturing positive multivariate dependence, Archimedean copulas are bound also by their parameterisation. Unlike the multivariate FGM, where θ_{jk} exists for each bivariate pair (u_j, u_k), φ_θ^{-1} is usually a function of a single parameter. Equation (14) shows that any bivariate pair will share a common association parameter. This can be seen for example in the Frank copula, whose trivariate form is given by

$$C(u,v,w;\theta) = -\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)\left(e^{-\theta w}-1\right)}{\left(e^{-\theta}-1\right)^{n-1}}\right) \tag{15}$$

3.2.1 The mixture-of-powers approach

An alternative for multivariate Archimedean class copulas is generation according to inverse Laplace transforms and mixtures of powers (Joe 1997, Zimmer and Trivedi 2006). This is a transform $\phi(s)$ of some univariate CDF $M(\alpha)$ such that

$$\phi(s) = \int_0^\infty e^{-s\alpha} M(\alpha) \tag{16}$$

for $s > 0$. Any arbitrary distribution function F will have a unique Laplace transform G, where

$$\begin{aligned} F(x) &= \int_0^\infty G^\alpha dM(\alpha) \\ &\equiv \phi(-\log G(x)) \end{aligned} \tag{17}$$

Zimmer and Trivedi (2006) present the parameter $\alpha > 0$ as a form of heterogeneity affecting the random variable X. Since copulas are distribution functions, like F, the bivariate case can be considered as

$$\begin{aligned} C(u,v;\theta) &= \int_0^\infty G_u^\alpha G_v^\alpha dM(\alpha) \\ &\equiv \phi(-\log G_u - \log G_v) \\ &\equiv \phi\left(\phi^{-1}(u) + \phi^{-1}(v)\right) \end{aligned} \tag{18}$$

where $G_u = \exp\left\{\phi^{-1}(u)\right\}, G_v = \exp\left\{\phi^{-1}(v)\right\}$. This can continue into any number of dimensions, using different Laplace transforms to overcome the singularity of the dependence

structure. Only $n-1$ distinct transforms exist across $\frac{n(n-1)}{2}$ bivariate margins in an n-copula, though, so that distinct bivariate margins nevertheless share a common association. A trivariate mixture using two distinct transforms $\phi(s) \neq \psi(s)$ will give

$$C(u,v,w;\theta) = \psi\left(\psi^{-1} \circ \phi\left(\phi^{-1}(u) + \phi^{-1}(v)\right) + \psi^{-1}(w)\right) \tag{19}$$

where $\psi^{-1} \circ \phi$ belongs to a class of infinitely differentiable increasing functions (Joe 1997).[6] Importantly, dependence is symmetric with respect to u and v, but not w now: this is an improvement upon, for example, Equation (14). This will produce 7 distinct measures of dependence for the 28 bivariate pairs of 8 distributions, but all positive. This is the Jouini and Clemen (1996) condition that $\theta > 0$ under Laplace transforms and multivariate Archimedean copulas. Correlations from the HALS data show 9 of these 28 pairs are negatively associated however, contra-indicating the use of Archimedean copulas.

One solution is to use the Laplace transform $\phi(s) = \max\left\{(1+\eta s)^{\frac{-1}{\eta}}, 0\right\}$, which does permit negative association.[7] Using this, and another other Laplace transform ψ, a multivariate copula can be constructed from Equation (19), such that each bivariate margin has the appropriate association, at least in terms of sign: the limit to the number of unique Laplace transforms that can be used still exists. There is, for example, only one known Laplace transform extending to negative dependence, such that there will always be fewer unique dependence parameters than bivariate correlations. In practical terms the procedure in Equations (16)-(18) need not be undertaken by the researcher. As well as families of copulas being widely available, so too are known Laplace transforms (Joe 1997).

[6] This is a condition assuring monotonicity of $\psi^{-1} \circ \phi$ mixtures, and hence the mixture-of-powers copula itself. Since known transforms are readily available in Joe's (1997) appendix, the requirement of infinite differentiability is not one the analyst will usually face.

[7] This is Joe's (1997) Laplace transform B, or Gamma-form LT, given by $\phi(s) = (1+\theta s)^{-\frac{1}{\theta}}$, where $\theta > 0$. The extension to negativity is, statistically, similar to that of the Clayton copula (Mari and Kotz 2001). That is, after extension the negative LTB is no longer strictly monotonic. This is also why Laplace-transformed multivariate copulas do not have mixture representations when extended to negative dependence.

3.2.2 The mixture of Max-ID approach

Consider instead mixtures of Max-Infinitely Divisible distributions, rather than standard Archimedean copulas (Joe and Hu 1996, Joe 1997). A multivariate distribution H is called Max-ID if H^γ is a CDF for all $\gamma > 0$ and for all n dimensions.[8] In fact the mixture-of-powers approach just discussed is a mixture of powers of a Max -or Min-ID multivariate distribution. This approach can be extended to negative dependence for some bivariate margins, although such extensions are less common or straightforward. If a copula is of the form in Equation (18), C can take the general form $C(u_1, .., u_n) = \phi(-\ln H(u_1, .., u_n))$, and C is a multivariate CDF if H is Max-ID and $-\ln\phi$ belongs to a class of infinitely differentiable increasing functions. This general form contains copulas of the form in Equation (19), however extensions to negative dependence (a sufficient condition for which is when $-\ln\phi$ is convex) do not have mixture representations. Moreover, such extensions tend to generate multivariate copulas whose margins all have Reverse Rule of Order 2 (RR_2), or negatively dependent: this is because the n-copula would be a mixture of Min-ID distributions, such that each bivariate margin is RR_2.[9] This is the case even with general dependence such as the FGM in Equation (10) that allows unique bivariate association. Consider the copula C such that

$$C(u_1, .., u_n) = \psi\left(-\sum_{i<j} \ln K_{ij}\left(e^{-p_i\psi^{-1}(u_i)}, e^{-p_j\psi^{-1}(u_j)}\right) + \sum_{i=1}^{n}(q_i + n - 2)\, p_i\psi^{-1}(u_i)\right) \tag{20}$$

where q_i is another Max-ID mixing parameter specific to each marginal CDF. Each K_{ij} in this expression is a bivariate margin; specifically a bivariate copula. Each K_{ij} then is Max-ID, giving $C(u_1, .., u_n)$ Positive Orthant Dependence (POD).[10] Using the survival

[8] A univariate CDF F is such that F^γ is also a CDF for all $\gamma > 0$, but this is not the case for multivariate distribution functions. In general the n-dimensional CDF H is such that H^γ is a CDF for all $\gamma > n-1$ (Joe 1997). Max-ID is therefore a stronger dependence condition - it is equivalent to Total Positivity of Order 2 where, for $x_1 < x_2$ and $y_1 < y_2$, F is TP_2 if $F(x_1, y_1)F(x_2, y_2) > F(x_1, y_2)F(x_2, y_1)$.

[9] RR_2, or Reverse Rule of Order 2, is essentially the negative-dependence equivalent of TP_2.

[10] Bivariate distributions are Positive Quadrant Dependent if higher values of one variable are correlated with higher values of the other, and vice versa (essentially $\tau > 0$). Positive Orthant Dependence is the

function in each case will instead give negative orthant dependence (Joe 1997, Belzunce and Semeraro 2004). A component-wise interpretation due to Joe (1997) is that the transform ψ is used to capture 'global' dependence (that is, a minimal level of pairwise dependence), while the specific copula K_{ij} captures the individual (in the context of the multivariate copula proper) pairwise dependence, and q_i contributes to bi/multivariate asymmetry.

From the general method of copulas, the model for mortality risk, health and lifestyles would require the use of the multivariate FGM and/or the mixture of Archimedean copulas in order to estimate a closed-form distribution. The FGM however is too limiting in the degree of dependence it can measure. The Mixture of Powers is too limiting in the number of dependence parameters it allows and the Mixture of Max-ID copulas allows only totally positive (or totally negative) dependence. Thus as the dimensionality of the multivariate distribution increases, these methods become less practical. In terms of likelihoods these distributions are also prone to some complexity when rendered as densities, making this approach less attractive also. In order to estimate the entire model another method can be employed, which uses inference about the joint distribution, taking advantage of the separation of marginal distributions in a copula from the joint distribution.

4 Inference functions and the Gaussian and t copulas

An alternative method due to Lee (1983), McLeish and Small (1988), Joe and Xu (1996), Xu (1996) and Joe (1997) is the method of Inference Functions for Margins (IFM).[11] For some multivariate distribution $H(X_1, .., X_n; \beta_1, .., \beta_n, \theta)$, 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-

multivariate equivalent. Note that these are weaker than (multivariate) Total Positivity of Order 2, which implies positive (orthant) quadrant dependence (Mari and Kotz 2001 is a good reference for these dependence concepts).

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

step procedure in which each marginal vector of coefficients $\beta_{i\in n}$ is estimated first, and separately, to determine $\left\{\hat{\beta}_1, .., \hat{\beta}_n\right\}$ such that

$$\hat{\beta}_i = \arg\max_{\beta_i} \sum_{i=1}^{n} \ln f_i\left(x_i; \beta_i\right) \tag{21}$$

and $L\left(\hat{\beta}_1, .., \hat{\beta}_n, \theta\right)$ is maximised to find only $\hat{\theta}$ such that

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ln c\left(F_1\left(x_1; \beta_1\right), .., F_n\left(x_n; \beta_n\right)\right) \tag{22}$$

for some copula C with density c.[12] Ordinarily, the method of Maximum Likelihood is to solve $(\partial L/\partial\beta_1, .., \partial L/\partial\beta_n, \partial L/\partial\theta) = 0$, such as would be expected in Equations (10) and (15), for example.

Estimates from the method of IFM then are such that $(\partial L_1/\partial\beta_1, .., \partial L_n/\partial\beta_n, \partial L/\partial\theta) = 0$. This holds under regularity conditions, and Joe (1997 and 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 jacknife methods.

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

$$C\left(u_1, .., u_8\right) = \Phi_8\left(\Phi_m^{-1}\left(F_m\left(y_{im}^*\right)\right), \Phi_h^{-1}\left(F_h\left(y_{ih}^*\right)\right), \Phi_{1_1}^{-1}\left(F_{l_1}\left(y_{il_1}^*\right)\right), .., \Phi_{1_6}^{-1}\left(F_{l_6}\left(y_{il_6}^*\right)\right)\right) \tag{23}$$

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

Figure 1 here

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

[12] Here $c = \frac{\partial C}{\partial F_1, .., \partial F_n}$ is the copula density (see deMatteis 2000 for his discussion on copula densities).

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 (2).

Although the model in Equation (23) is essentially a normal distribution, the function of inverses results in tractability of the marginal distributions also, as in a copula. It is subsequently much more straightforward than the multivariate probit, because exact Maximum Likelihood is available for the problem

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ln \phi_8 \left(\Phi_m^{-1}\left(F_m\left(y_{im}^*\right)\right), \Phi_h^{-1}\left(F_h\left(y_{ih}^*\right)\right), \Phi_{1_1}^{-1}\left(F_{l_1}\left(y_{il_1}^*\right)\right), .., \Phi_{1_6}^{-1}\left(F_{l_6}\left(y_{il_6}^*\right)\right); \theta\right) \tag{24}$$

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. This copula is a nice alternative the multivariate probit irrespective of the issues discussed here, being much simpler to specify and estimate.

Comparing copulas, the Gaussian copula provides generalised dependence, unlike the multivariate mixture approaches, and the method of IFM has shown asymptotic efficiency relative to the multivariate probit in other studies (Joe 1997, 2005, Joe and Xu 1996). An alternative is the so-called multivariate t-copula, which is narrower than the Gaussian and can capture tail dependence of extreme events (Embrechts, Lindskog and McNeil 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, Zhao and Joe 2005 are examples), although with less efficiency than has been shown for the IFM. Hustler 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 approach does not restrict the IFM to multivariate symmetry: the general form $C(x_1, x_2) = H_{(1,2)}\left(H_1^{-1}(F_1(x_1)), H_2^{-1}(F_2(x_2))\right)$ for some distribution F with inverse H^{-1} can be used, generating so-called Inversion Copulas (Nelsen 1999. 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 accomodated, 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

$$f(x;\alpha) = 2\phi(x)\Phi(\alpha x) \tag{25}$$

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 at $\alpha = 0$ (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 $X \sim SN(\alpha)$.[13] The multivariate skewed normal is given for some random vector $X_{[k\times 1]}$ where

$$f_k(x;\alpha) = 2\phi(x;\Omega)\Phi(\alpha' x) \tag{26}$$

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 $X \sim SN_k(\Omega, \alpha)$. In general form, Azzalini and Capitanio 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

$$f_k(x) = 2f(x)F(W(x)) \tag{27}$$

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

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.[14] 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. Marrying this notation to the copula approach, consider Equation (23). The random vector $X = \left(\Phi_m^{-1}\left(F_m\left(y_{im}^*\right)\right), \Phi_h^{-1}\left(F_h\left(y_{ih}^*\right)\right), \Phi_{1_1}^{-1}\right.$ 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 in 2005, provided updated mortality data. Of the original 9,003 respondents 2,491 had died.[15] This analysis follows Balia and Jones (2005), using information at the time of the first survey, coupled with the most recent mortality data. 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

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

[15] 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.

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 (2005) 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.[16] 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.[17]

[16] '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 (2005) discuss.

[17] 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.

5.2 Explanatory variables

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

Table 1 here

As the table shows 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, we are left with 3,655 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.

Social class is fairly normally distributed, and gender is only slightly in favour of females. 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%.[18]

[18] 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

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

The presence of negative correlation precludes the reliable use of the mixture of Max-ID copulas, even if survival functions were used in some instances, and managing adequate representation of dependence is difficult with the Mixture of Powers copulas. After differentiation the Mixture-of-Powers density is too cumbersome to be estimated, unless the number of mixtures is reduced too far for the distribution to be informative: in fact, to be practicable the Mixture of Powers is not much better at capturing dependence than the multivariate FGM. Although the copulas can be more informative for sub-sected dimensions of the problem, estimating all of mortality risk, health and lifestyles is feasible only with the inferencing approach of the Gaussian and t copulas.

The log-likelihood for the problem is as in Equation (24). 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 (2)-(7). These are the probit,

$$\Pr\left(Y=1|x;\beta\right)=\Phi\left(x'\beta\right) \tag{28}$$

the fatter-tailed logit,

$$\Pr\left(Y=1|x;\beta\right)=\exp\left(-\exp\left(x'\beta\right)\right) \tag{29}$$

and the complementary log-log

$$\Pr\left(Y=1|x;\beta\right)=1-\exp\left(\exp\left(-x'\beta\right)\right) \tag{30}$$

with the dependent variables. Comparison however gives no indication that previous results were affected significantly by the higher value.

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 and discussion

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

The appropriate link function for each margin is selected according to varying criteria which, along with cell predictions, are contained in Table 2, below.

Table 2 here

The probit, logit and complementary log-log functions are equally complex, so the results are the same across the tests. The BIC will not identify differences without parsimony gains in one specification over another, and the Expected Cross-Validation Index (ECVI) will consistently give the same recommendation as the AIC in this case: while it is useful to minimise the underestimation of fit relative to the AIC, it is essentially the AIC divided by the sample size.

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 predisposition towards 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 imperative attached to either the probit or logit. 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 can be used to compare the distributions of the predicted probabilities to determine whether there is any statistical relevance to the choice made.[19] Considering the dimensions *in toto*, Table 3 (below) 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.[20]

Table 3 here

The differences between the skewed and symmetric distributions are worth consideration, particularly for the multivariate skewed t distribution. This is because the symmetric multivariate t distribution may have greater consideration of tail dependence than the the multivariate skewed t, which will affect the distribution at its centre, but one or the other will be of more importance to the researcher, who may have to choose between them.

[19] 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.

[20] 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.

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 lower tail dependence is observed. The means of the inverse probabilities from each model (including the 'observed' data) are in Table 4, below.[21] They are generally similar except for mortality risk and exercise, for which quite disparate results can be seen, particularly with the skewed normal distribution.

Table 4 here

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, below.

Table 5 here

The parameter corresponding to mortality risk in the skewed normal distribution is quite large, relative to the t, 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

[21] 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 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 (23).

vector: the trade-off from the t-copula does not exist for the Gaussian, which ignores tail dependence anyway. 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.

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 a marked improvement upon the symmetric. Between symmetric distributions the multivariate t copula is more noticably better than the Gaussian 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 tests. The results are in Table 6, below.

Table 6 here

Of most use is the left-hand column, which gives each replicated distribution's distance from the one upon which it was estimated. The implications are not as straightforward as

for other measures of fit, in part because 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 we would find the information on replication more useful, because our confidence in predictions may change accordingly.

Procedures for comparing the fit graphically can be found in de Matteis (2001). This can be done using Quantile-Quantile plots of non-parametric distribution functions, as well as analytical methods described in Frees and Valdez (1998), although these will not always be practicable, as in this case. Quinn (2006) also contains QQ-plots and empirical distance as criteria for copula selection in bivarariate models.

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 supressed 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 jackkifing procedure would be used.[22] 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

[22] 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).

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 2-9 (below) show the marginal effects and t-statistics for the covariates in the margins, for all of the feasible methods of analysis. The actual model used in each dimension is indicated, as is statistical significance at 5 and 10%. 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.

Figures 2-9 here

One noticeable result is the positive, statistically significant impact of being European caucasian on both reducing 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 (2005) 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, only in which model does age and age squared show markedly different results to the other models. In this case mortality risk is increasing with age but at a decreasing rate, which the other models do not predict. The effect of age on health is consistent across all models. So too 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 occassionally repeated in other equations, but not as consistently across all parameters as in this equation.

Balia and Jones (2005) 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 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 (2005) 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 (2005) 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 shows interesting results that 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). While the prevalent image of single urban living might include more time spent in a gym, for example, the sample here is older. Non-urban areas may also afford for space for outdoor sports, 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 the covariate in question would have been otherwise excluded from or included in the analysis. The differences that do exist however, together with those in the marginal effects between different models, suggest that the considerations made here are not without merit.

This can also be seen in the variance-covariances of the outcomes, given in Table 7, below.[23]

Table 7 here

These are of interest *vis a vis* the point taken previously from Klein (1998) that, under non-normality (or even asymmetric normality) a structure-of-equations model may recommend erroneous rejection of one or more equations in the structure, due to non-robust

[23] Correlations are used rather than the variance-covariance matrix so that comparisons can be made on the same scale (which for copulas will vary, unlike the multivariate probit). We would already expect, for example, some elements of the variance-covariance matrices to be scaled differently based on the estimates in Figure 1.

distributional assumptions. As such the matrices in Table 7 would indicate such a problem with a structural, normally-symmetric system of equations. This would be evident in, for example, statistically less significant correlation between mortality risk and/or SAH and the endogenous lifestyles, relative to the symmetric normal estimates. In fact there is no such indication.

8 Conclusion

The methods and results presented here lead to two conclusions: first, that more flexible approaches to estimating multivariate data, even multivariate dichotomous data, are a worthwhile enterprise.

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 was alone 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.

The copula approach allows such flexibility where traditional models do not, particularly when analysing jointly-distribution 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 idiosyncracies of the data such as skewness and tail dependence, while also being simpler to implement and estimate.

The analysis here also showed that this approach generates more information about the models and the results, which can be used to select better-fitting marginal and joint distributions. In fact several trade-offs were identified during estimation, between information of different types on different behaviour of the data. Which of these types of information is preferred depends upon the focus of the analysis, but estimation with copulas can be responsive to this need where standard multivariate analysis might not.

9 References

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 AM (2005) Mortality, lifestyle and socio-economic status, Working Paper 05/02, Health, Econometrics and Data Group, University of York, York.

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, 41(4): 961-974.

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

Cook RD, Johnson ME (1981) A family of distributions for modelling non-elliptically symmetric and multivariate data, Journal of the Royal Statistical Society B, 43(2): 210-218.

Deurenberg P, Yap M, van Staveren WA (1998) Body mass index and percent body fat: a meta analysis among different ethnic groups, International Journal of Obesity Related Metabolic Disorders, 22(12): 1164-1171.

van Doorslaer E, Koolman X (2004) Explaining the differences in income-related health inequalities across European countries, Health Economics, 13(7): 609-628.

van Doorslaer E, Wagstaff A, Bleichrodt H, Calonge S, Gerdtham Ulf-G, Gerfin M, Geurts J, Gross L, Häkkinen U, R Leu, 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, 16(1): 93-112.

Embrechts P, Lindskog F, McNeil AJ (2003) Modelling dependence with copulas and

applications to risk management, in Handbook of Heavy Tailed Distributions in Finance, Rachev ST (ed.). Elsevier/North-Holland, Amserdam.

Frank MJ (1979) On the simultaneous associativity of f(x,y) and x + y - f(x, y), Aequationes Mathematica 19: 194–226.

Fang KT, Kotz S, Ng KW (1989) Symmetric Multivariate and Related Distributions. Chapman and Hall, New York.

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

Hajivassiliou VA (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.

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, 94(2): 401-419.

Joe H, Hu T (1996) Multivariate distributions from mixtures max-infinitely divisible distributions, Journal of Multivariate Analysis, 57(2): 240-265.

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

Jouini MN, Clemen RT (1996) Copula models for aggregating expert opinions, Operations Research, 44: 444-457.

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

Keselman HJ, Wilcox RR, Lix LM (2005). A generally robust approach to hypothesis testing in independent and correlated groups designs, Psychophysiology, 40(4): 586-96.

Kolmogorov AN (1950) Foundations of the Theory of Probability, Chelsea Publishing Company, New York.

Kowalski CJ (1973) Non-normal bivariate distributions with normal marginals, The American Statistician, 27(3): 103-106.

Lee L-F (1983) Generalized econometric models with selectivity, Econometrica, 51(2): 507-512.

Mardia KV (1962) Multivariate Pareto distributions, Annals of Mathematical Statistics, 33: 1008-1015.

Mari DD, Kotz S (2001) Correlation and dependence. Imperial College Press, London.

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

McLeish DL, Small CG (1988) The Theory and Applications of Statistical Inference Functions. Lecture Notes in Statistics 44. Springer, New York.

Mullahy J, Portney P (1990) Air pollution, cigarette smoking, and the production of respiratory health, Journal of Health Economics 9(2): 193-205.

Mullahy J, Sindelair J (1996) Employment, unemployment, and problem drinking, Journal of Health Economics, 15(4): 409-434.

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

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

Nelsen RB (1999) An introduction to copulas. Springer Verlag, New York.

Newey W, Steigerwald D (1997) Asymptotic bias for qasi-maximum likelihood estimators in conditional heteroskedasticity models, Econometrics, 65(3): 587-599.

Prieger JE (2000) A flexible parametric selection model for non-normal data with application to health care usage, Journal of Applied Econometrics 17(4): 367-392.

Prokhorov AB, 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.

Ruhm C (2005) Healthy living in hard times, Journal of Health Economics, 24(2): 341-363.

Satterthwaite SP, Hutchinson TP (1978) A generalisation of Gumbel's bivariate logistic distribution, Metrika, 25: 163-170.

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 MD (2003) Modelling sample selection using archimedean copulas, Econometrics Journal 6: 99-123.

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

Hajivassiliou, VA (1997) Some practical issues in maximum simulated likelihood, Working Paper, London School of Economics, London.

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.

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

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

Zimmer DM, Trivedi PK (2006) Using trivariate copulas to model sample selection and treatment effects: application to family health care demand, Journal of Business & Economic Statistics, 24(1): 63-76.

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

Table 1. Variable definitions and descriptive statistics.

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

Table 1(Cont). Variable definitions and descriptive statistics.

	% Correct				AIC			BIC		
	MV 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

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).

mortality		
	Probit	Logit
Probit		0.3160
CLog-Log	0.0030	0.3030

SAH		
	Probit	Logit
Probit		0.9940
CLog-Log	0.6490	0.2400

breakfast		
	Probit	Logit
Probit		0.9230
CLog-Log	0.3310	0.0770

not obese		
	Probit	Logit
Probit		0.5140
CLog-Log	0.3610	0.0150

non-smoker		
	Probit	Logit
Probit		0.8690
CLog-Log	0.1690	0.0280

sleeping well		
	Probit	Logit
Probit		1
CLog-Log	0.5710	0.4270

prudent drinker		
	Probit	Logit
Probit		0.0160
CLog-Log	0.0440	0.0000

exerciser		
	Probit	Logit
Probit		0.9970
CLog-Log	0.2520	0.6490

Table 3. *p*-values from Kolmogorov-Smirnoff tests for difference in distribution of linearly-predicted probabilities, Pr(y = 1) ($p < 0.05$ (shaded cells) represent statistically equivalent distributions).

	Symmetric Normal	Skewed Normal	Symmetric t	Skewed t	Observed
mortality	-0.2854	-1.6646	-0.3174	-0.3779	-0.2850
	(1.0955)	(1.7617)	(1.0461)	(1.0689)	(1.0957)
SAH	0.5686	0.7137	0.6117	0.9210	0.5686
	(0.4652)	(0.4873)	(0.3926)	(0.5396)	(0.4653)
breakfast	0.5957	0.3277	0.6209	0.7475	0.5959
	(0.4546)	(0.5278)	(0.4339)	(0.4630)	(0.4547)
not obese	1.1224	1.0611	1.1212	1.1213	1.1227
	(0.3730)	(0.3780)	(0.3633)	(0.3670)	(0.3731)
non-smoker	0.5958	0.3519	0.6202	0.7329	0.5960
	(0.5518)	(0.6033)	(0.5255)	(0.5507)	(0.5518)
sleeping well	0.2148	0.3635	0.2297	0.2965	0.2148
	(0.2504)	(0.2913)	(0.2352)	(0.2527)	(0.2493)
prudent drinker	1.3454	1.1837	1.3542	1.3383	1.3459
	(0.5502)	(0.5735)	(0.5309)	(0.5388)	(0.5502)
exerciser	-0.5025	-0.0791	-0.4842	-0.3588	-0.5028
	(0.4461)	(0.6152)	(0.4209)	(0.4520)	(0.4461)

Table 4. Predicted mean (standard deviation) in each dimension of the mortality risk, health and lifestyle models.

	Skewed Normal	Skewed t
mortality	13.8945	-2.7993
	(1.3244)	(0.1755)
SAH	1.0472	-5.2299
	(0.1696)	(0.2363)
breakfast	-1.9399	0.1935
	(0.2085)	(0.1094)
not obese	0.4219	2.3438
	(0.1300)	(0.1265)
non-smoker	1.5039	0.7909
	(0.1889)	(0.1076)
sleeping well	0.1234	-0.4122
	(0.0990)	(0.0766)
prudent drinker	0.7700	0.3279
	(0.1286)	(0.0922)
exerciser	0.9166	-0.9720
	(0.2489)	(0.1429)

Table 5. Estimated Skewness parameters (standard errors) for each dimension of the mortality risk, health and lifestyle models.

		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 6. Information criteria from the joint (copula) distributions
(shaded rows contain the 'best' model according to minimum information criterion).

Mortality	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.0010				
Skewed Normal	0.0940	0.0010			
Symmetric *t*	0.0020	0.1340	0.0010		
Skewed *t*	0.0000	0.0780	0.0220	0.2410	

SAH	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.0030				
Skewed Normal	0.0100	0.5010			
Symmetric *t*	0.0890	0.0010	0.0010		
Skewed *t*	0.6820	0.1210	0.0480	0.3410	

Breakfast	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.2910				
Skewed Normal	0.1250	0.4010			
Symmetric *t*	0.0050	0.0040	0.0060		
Skewed *t*	0.0390	0.5360	0.3700	0.1480	

Not obese	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.0240				
Skewed Normal	0.2010	0.1810			
Symmetric *t*	0.0040	0.7230	0.0620		
Skewed *t*	0.2630	0.5010	0.4010	0.1480	

Non-smoker	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.0020				
Skewed Normal	0.0170	0.9360			
Symmetric *t*	0.0000	0.0260	0.0110		
Skewed *t*	0.0370	0.3410	0.4330	0.2630	

Sleeping well	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.2050				
Skewed Normal	0.3070	0.3700			
Symmetric *t*	0.0370	0.0330	0.4010		
Skewed *t*	0.0650	0.0330	0.5730	0.2000	

Prudent drinking	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.1240				
Skewed Normal	0.1210	0.8880			
Symmetric *t*	0.0210	0.0970	0.1080		
Skewed *t*	0.2770	0.8600	0.8280	0.0870	

Exercise	Observed	Symmetric Normal	Skewed Normal	Symmetric *t*	Skewed *t*
Observed					
Symmetric Normal	0.2600				
Skewed Normal	0.0020	0.0040			
Symmetric *t*	0.0010	0.1340	0.0060		
Skewed *t*	0.0650	0.4010	0.0480	0.1640	

Table 7. *p*-values from Kolmogorov-Smirnoff tests for differences between distributions
(shaded cells <-> $p < 0.5$ represents a statistically significant difference at 5% level of significance).

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**

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**

Table 8. Correlation matrices from the joint distribution
(bold cells are significant at 5%).

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(Cont). Correlation matrices from the joint distribution
(bold cells are significant at 5%).

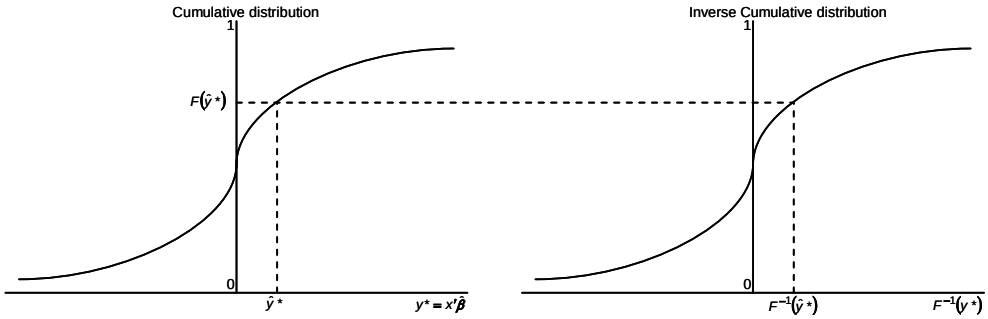

Figure 1. Relating inverse functions to random variables.

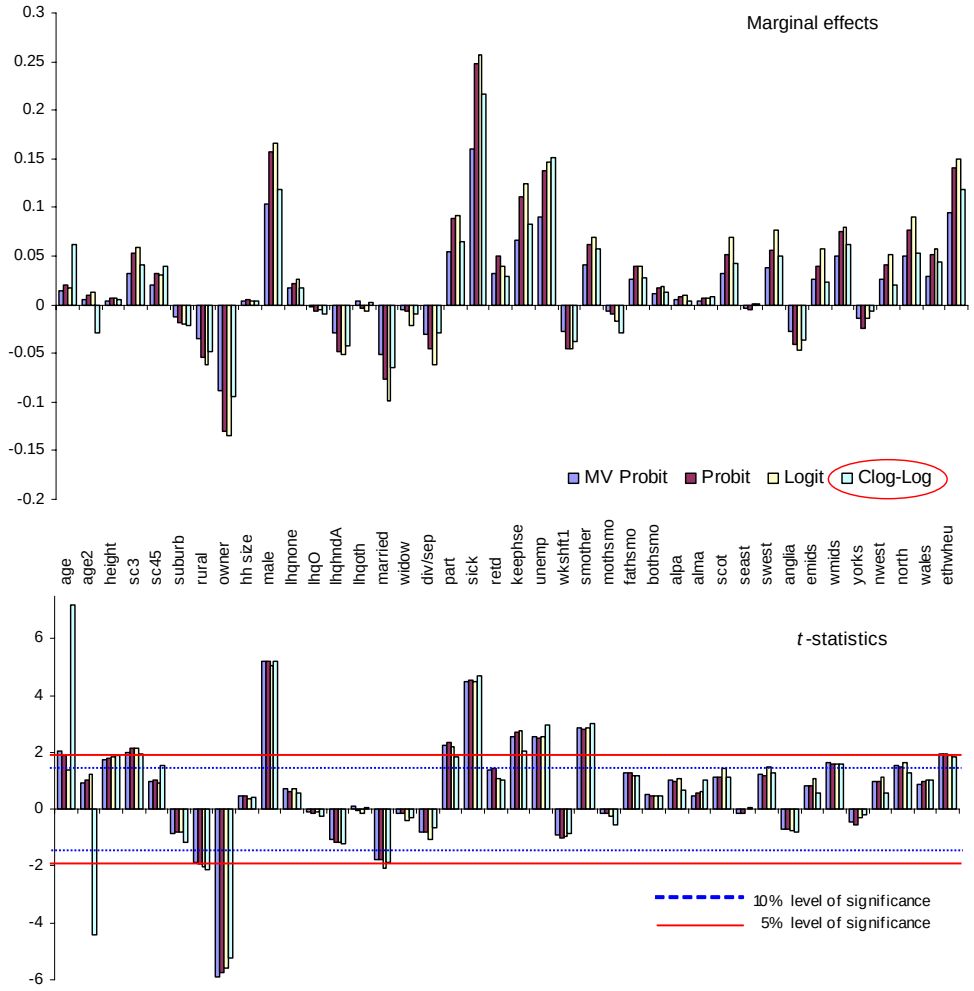

Figure 2. Marginal effects and *t*-statistics for Mortality Risk (all models).

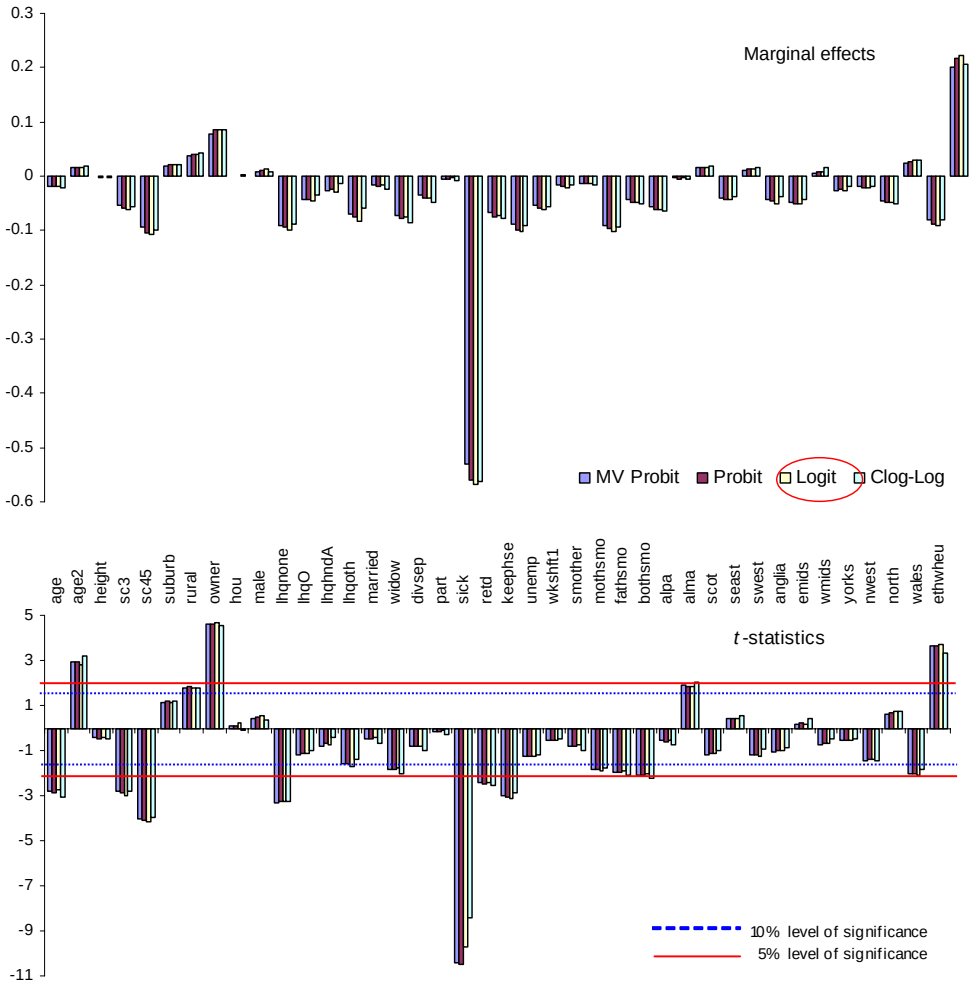

Figure 3. Marginal effects and *t*-statistics for Self-Assessed Health (all models).

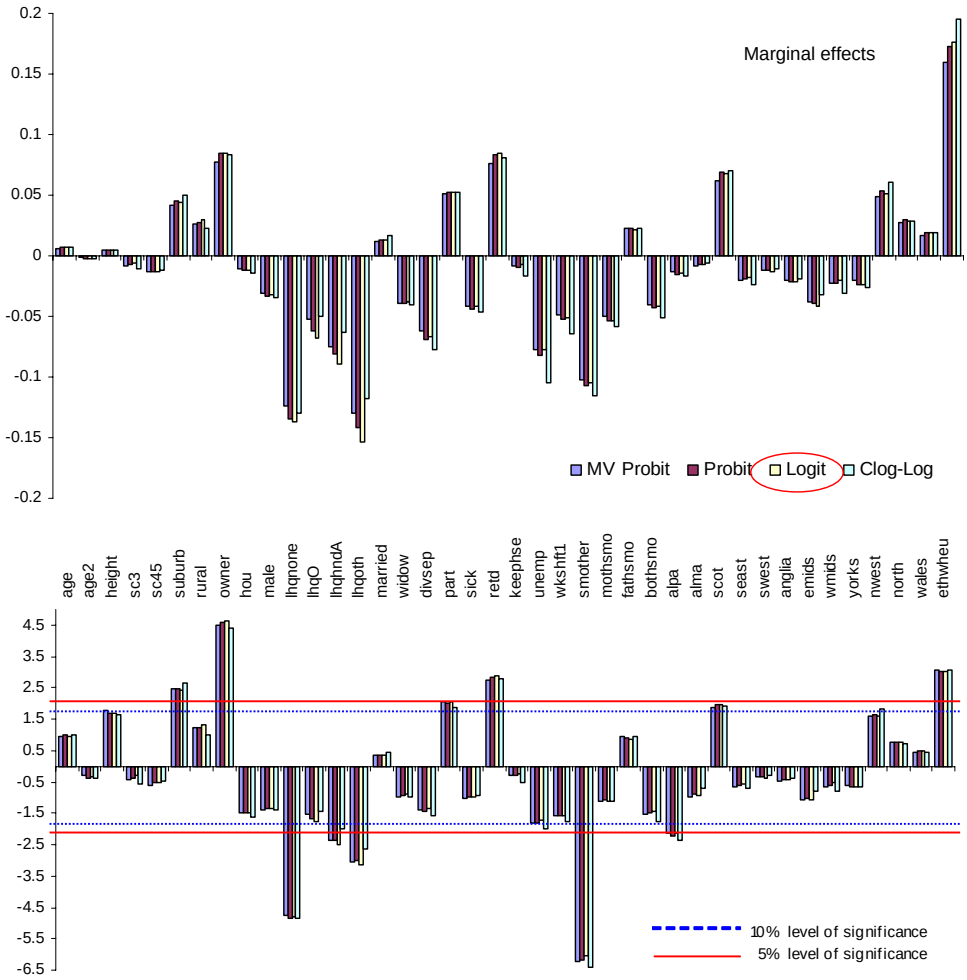

Figure 4. Marginal effects and *t*-statistics for eating Breakfast (all models).

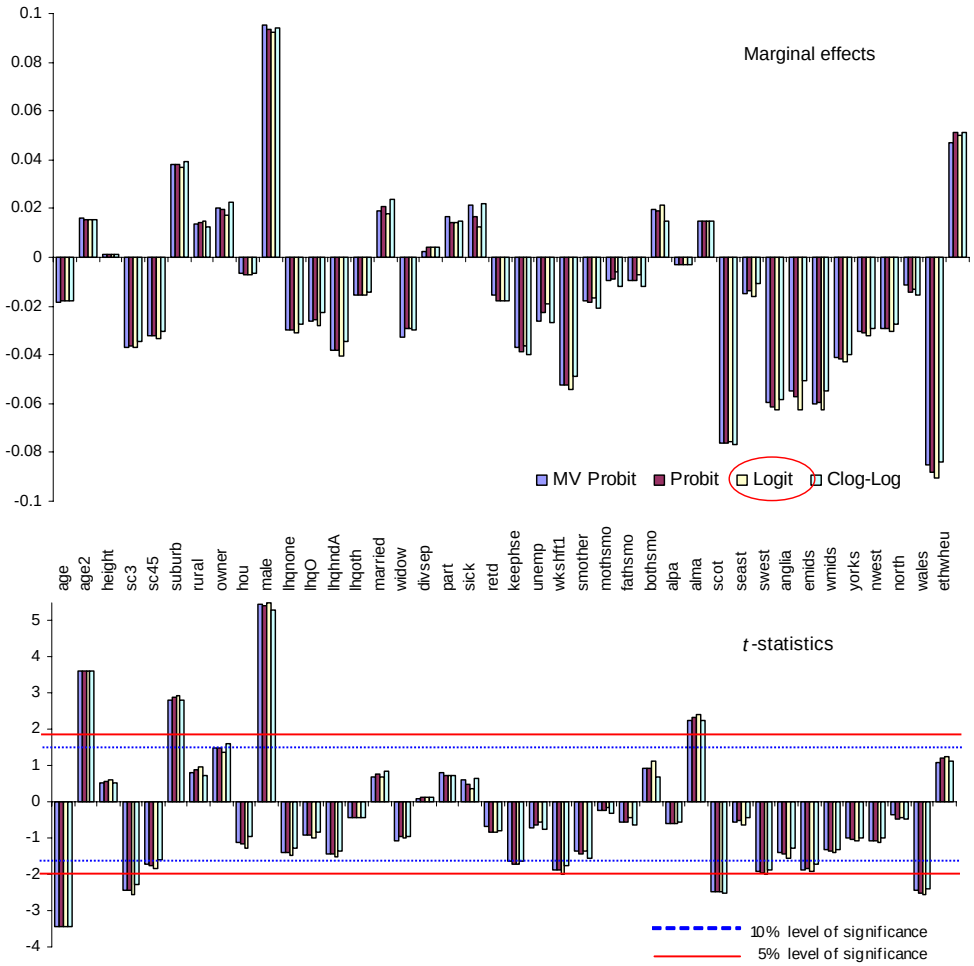

Figure 5. Marginal effects and *t*-statistics for Non-obesity (all models).

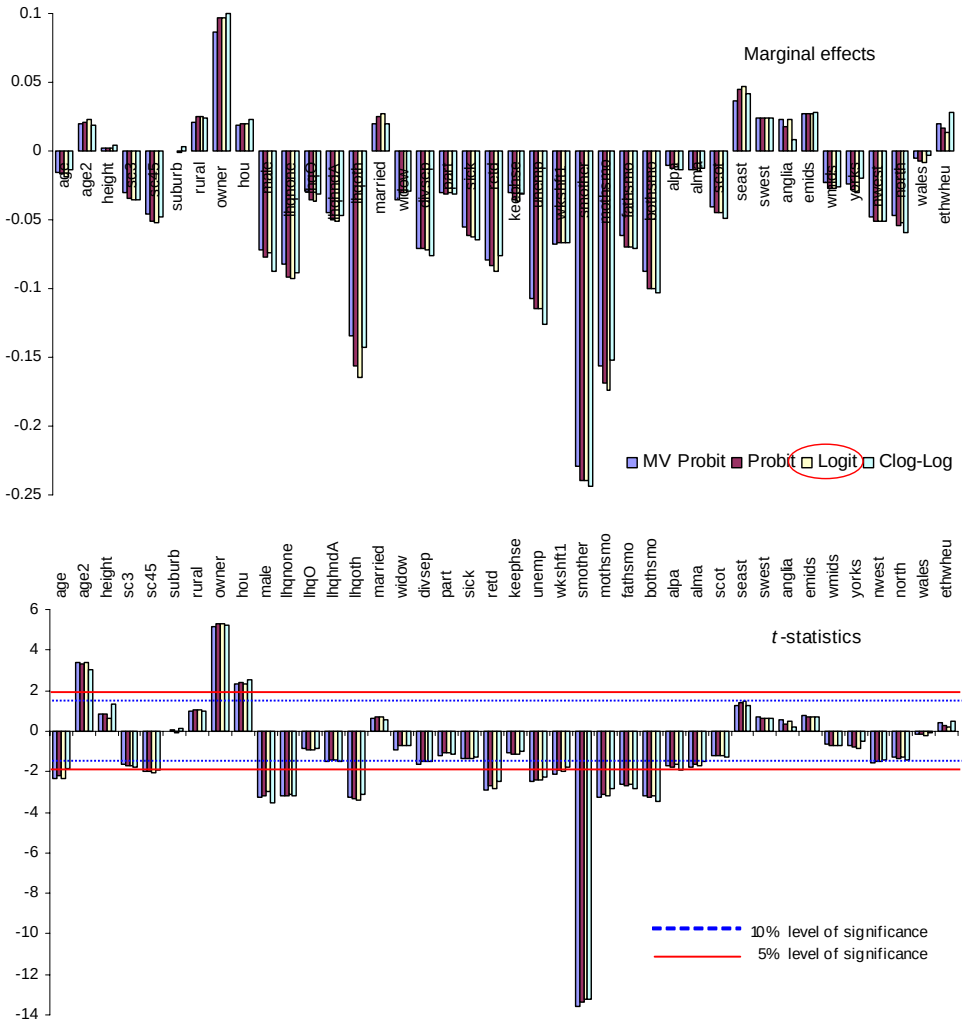

Figure 6. Marginal effects and *t*-statistics for Not Smoking (all models).

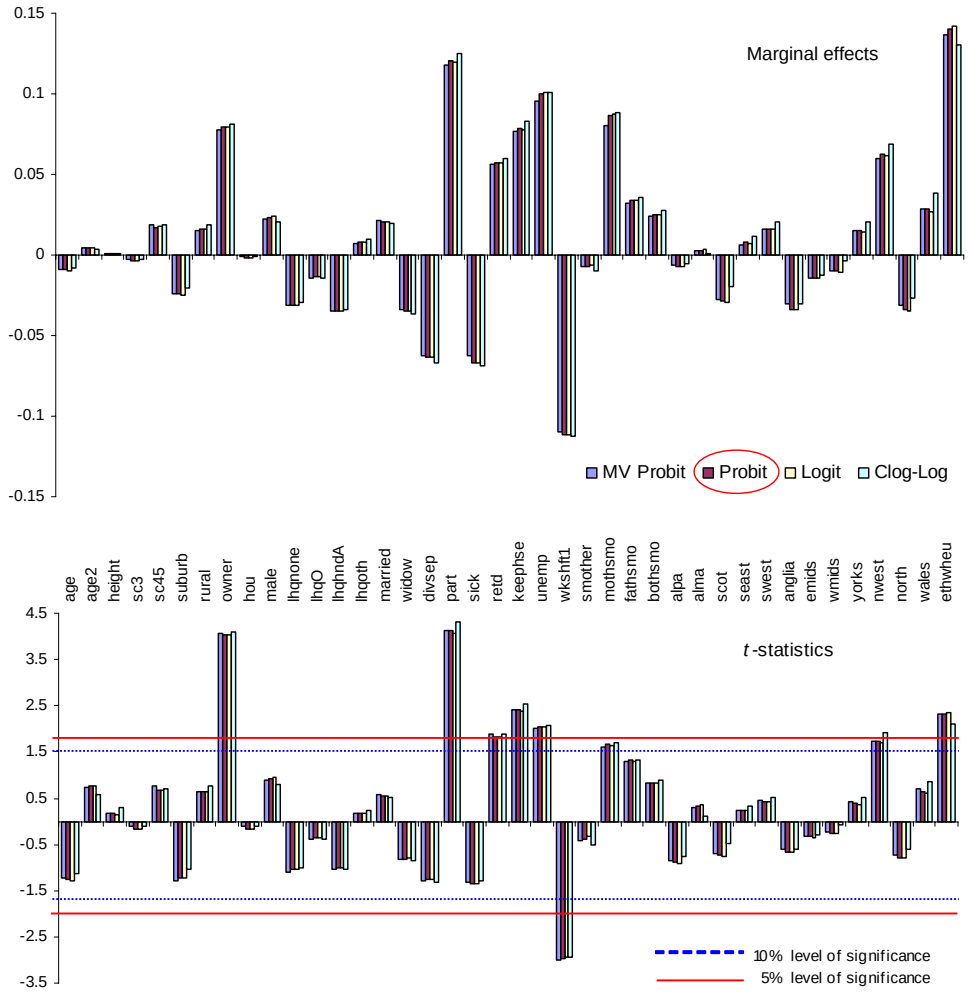

Figure 7. Marginal effects and *t*-statistics for Sleeping Well (all models).

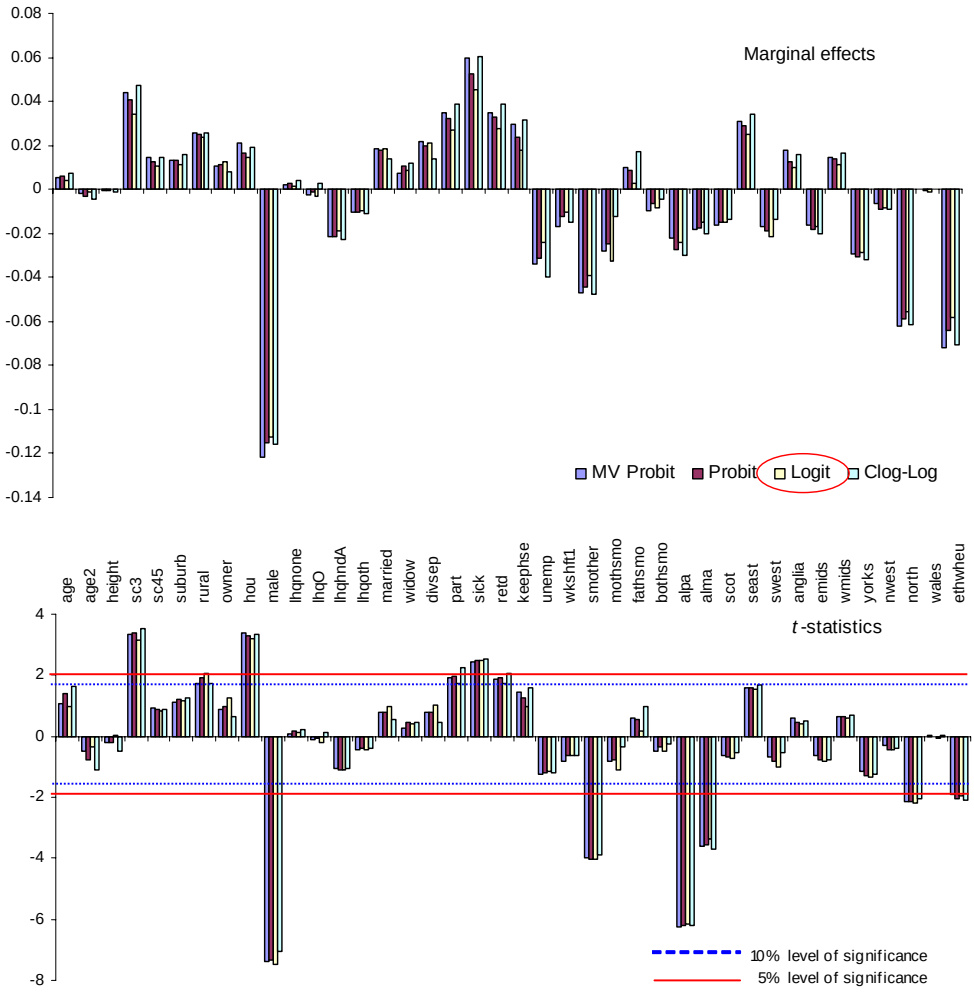

Figure 8. Marginal effects and *t*-statistics for Prudent Drinking (all models).

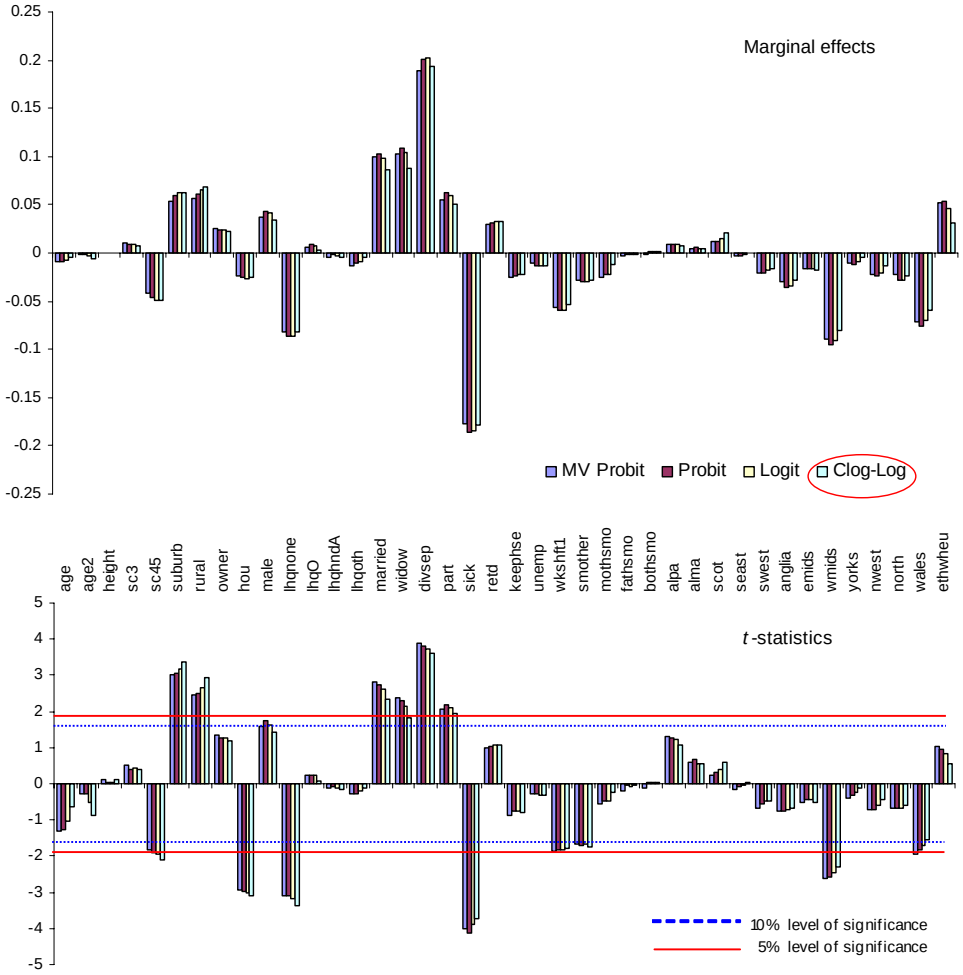

Figure 9. Marginal effects and *t*-statistics for Exercising (all models).