Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

Mixed Marginal Copula Modeling

David Gunawan, Mohamad A. Khaled & Robert Kohn

To cite this article: David Gunawan, Mohamad A. Khaled & Robert Kohn (2018): Mixed Marginal Copula Modeling, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2018.1469998

To link to this article: https://doi.org/10.1080/07350015.2018.1469998

Mixed Marginal Copula Modeling

David Gunawan[1], Mohamad A. Khaled[2] and Robert Kohn[1]

[1]University of New South Wales and ACEMS

[2]University of Queensland

April 3, 2018

Abstract

This article extends the literature on copulas with discrete or continuous marginals to the case where some of the marginals are a mixture of discrete and continuous components. We do so by carefully defining the likelihood as the density of the observations with respect to a mixed measure. The treatment is quite general, although we focus focus on mixtures of Gaussian and Archimedean copulas. The inference is Bayesian with the estimation carried out by Markov chain Monte Carlo. We illustrate the methodology and algorithms by applying them to estimate a multivariate income dynamics model.

KEYWORDS: Bayesian analysis; Markov chain Monte Carlo; Mixtures of copulas; Multivariate income dynamics.

1 Introduction

Copulas are a versatile and useful tool for modeling multivariate distributions. See, for example, Fan and Patton (2014), Patton (2009), Durante and Sempi (2015) and

Trivedi and Zimmer (2007). Modeling non-continuous marginal random variables is a challenging task due to computational problems, interpretation difficulties and various other pitfalls and paradoxes; see Smith and Khaled (2012), for example.The main source of the computational issues arises from the difficulty of directly evaluating the likelihood. For example, when modeling a vector of m discrete random variables, evaluating the likelihood at one point generally requires computing 2^m terms. The literature on modeling non-continuous random marginal problems has mostly focused on cases where all the marginals are discrete, and less extensively, on cases where some marginals are discrete and some are continuous. See, for example, Genest and Neslehová (2007), Smith and Khaled (2012), De Leon and Chough (2013), Panagiotelis et al. (2012) and Zilko and Kurowicka (2016). Furthermore, a lot of the literature has focused on approaches restricted to certain classes to copulas. For example, this is the case for Gaussian copulas (see, for example, Shen and Weissfeld, 2006; Hoff, 2007; Song et al., 2009; de Leon and Wu, 2011; He et al., 2012; Jiryaie et al., 2016) or pair-copula constructions (Stöber et al., 2015). Relatively little attention has been paid to the case where some variables are a mixture of discrete and continuous components. In contrast, our approach, presents methodology for an arbitrary copula and can be applied quite generally as long as it is possible to compute certain marginal and conditional copulas either in closed-form or numerically.

Our article extends the Bayesian methodology used for estimating continuous marginals to the case where each marginal can be a mixture of an absolutely continuous random variable and a discrete random variable. In particular, we are interested in applying the new methodology to copulas that are mixtures of Gaussian and Archimedean copulas. To illustrate the methodology and sampling algorithm we apply them to estimate a multivariate income dynamics model. We use the copula framework here to model the dependence structure of random variables that are mixtures of discrete and continuous components, and apply the model to empirical economic data. We note that there are many other real world economic applications that involve such mixtures of random variables as marginals, and Section 5 briefly discussed them.

Our proposed methodology extends that introduced in Pitt et al. (2006) and Smith and Khaled (2012). Smith and Khaled (2012) allow the joint modeling of distributions of random vectors such that each component can be either discrete or continuous. However, neither paper covers the case where some random variables can be a mixture of an absolutely continuous random variable and a discrete random variable. In a financial econometrics application, Brechmann et al. (2014) consider the case where the marginal distributions are mixtures of continuous distributions and point masses at zero. Our paper builds on, and generalizes, Brechmann et al. (2014) by deriving the likelihood equations in a much more general setting that allows for the margins to be arbitrarily classified into three groups: absolutely continuous, discrete and mixtures of absolutely continuous and discrete random variables. Furthermore, there is no restriction on the number or location of the point masses present in each margin. This can occur in many economic data, for instance in cases where earnings are top-coded and have individuals with zero earnings. Equally, our setting covers the case of dependent interval-censored data.

The paper is organized as follows. Section 2 outlines the copula model and defines the likelihood as a density with respect to a mixed measure. Section 3 presents the simulation algorithms used for inference. Section 4 applies the methods and algorithms to model multivariate income dynamics. This section describes the data and presents the estimation results. Section 5 concludes. The paper has two appendices. Appendix A defines the difference notation which is useful for expressing the likelihood of our model in closed-form. Appendix B presents and proves the results required to define the likelihood as a density with respect to a mixed measure. The paper also has an online supplement whose sections are denoted as Sections S1, Eq. S1, etc. Section S1 describes the Gaussian and Archimedean copulas used in the

 ACCEPTED MANUSCRIPT

article, as well as the Markov chain Monte Carlo (MCMC) sampling scheme. Section S2 introduces a new three dimensional example to further illustrate the methods in the paper. Section S3 gives a proof of Lemma 3 which is discussed in Appendix B. Section S4 presents some additional empirical results.

2 Defining the Likelihood of a general copula

This section discusses the proposed model and shows how to write the likelihood of an i.i.d. sample from it. Each random vector is modeled using a marginal distribution-copula decomposition and each marginal is allowed to be a mixture of an absolutely continuous component and a discrete component. The MCMC sampling scheme in the next section is based on this definition of the likelihood.

Let $\boldsymbol{X} = (X_1, \ldots, X_m)$ be an $\mathbb{R}^m$-valued random vector. If, for example, X_j is categorical, then its support would be a finite subset of $\mathbb{R}$ and thus without loss of generality, we can work with $\mathbb{R}^m$. Let $\mathcal{M} = \{1, \ldots, m\}$ be the index set, and $2^{\mathcal{M}}$ its power-set (or the set of all of its subsets). Let the random variable X_j have cumulative distribution function (CDF) F_j for $j = 1, \ldots, m$. By the Lebesgue decomposition theorem (Shorack, 2000, Chapter 7, Theorem 1.1), and assuming there are no continuous singularities (see Durante and Sempi, 2015, for a detailed discussion), the distribution of each X_j can be written as a mixture of an absolutely continuous random variable and a discrete random variable. This means that F_j is allowed to have jumps at a countable number of points. In order to exploit this result, we would like to be able to decide at each point of $\mathbb{R}^m$, which indices have jumps in their corresponding CDFs.

We need a mapping $\mathcal{C} : \mathbb{R}^m \to 2^{\mathcal{M}}$ that, for each $\boldsymbol{x} \in \mathbb{R}^m$, picks out the subset of the indices of $\boldsymbol{x}$ where F_j is continuous at x_j for each $j \in \mathcal{C}(\boldsymbol{x})$.

$$\mathcal{C} : \mathbb{R}^m \longrightarrow 2^{\mathcal{M}} \quad \text{with} \quad \boldsymbol{x} \longrightarrow \mathcal{C}(\boldsymbol{x}).$$

Similarly, we define the set $\mathcal{D}(\boldsymbol{x}) = \mathcal{M} - \mathcal{C}(\boldsymbol{x})$ (the complement of $\mathcal{C}(\boldsymbol{x})$ in $\mathcal{M}$, that is the set of indices j for which F_j presents jumps at x_j). This means that for all $\boldsymbol{x} \in \mathbb{R}^m$, $\{\mathcal{C}(\boldsymbol{x}), \mathcal{D}(\boldsymbol{x})\}$ partitions the index set so that $\mathcal{C}(\boldsymbol{x}) \cap \mathcal{D}(\boldsymbol{x}) = \varnothing$ and $\mathcal{C}(\boldsymbol{x}) \cup \mathcal{D}(\boldsymbol{x}) = \mathcal{M}$.

As a first example, consider $\boldsymbol{X} = (X_1, X_2)$, where $X_1 \sim \mathcal{N}(0,1)$ and X_2 is a mixture of an exponential distribution with parameter λ and a point mass at 0 with probability p, i.e., $X_2 \sim p\delta_0 + (1-p)\mathcal{E}(\lambda))$. Then, $\mathcal{C}(x_1, 0) = \{1\}$ for all $x_1 \in \mathbb{R}$ and $\mathcal{C}(x_1, x_2) = \{1, 2\}$ for all $x_1 \in \mathbb{R}, x_2 > 0$. Similarly, $\mathcal{D}(x_1, 0) = \{2\}$ for all $x_1 \in \mathbb{R}$ and $\mathcal{D}(x_1, x_2) = \varnothing$.

As a second example, let $\boldsymbol{X} = (X_1, X_2)$, where X_1 is Bernoulli and $X_2 \sim \mathcal{N}(0,1)$. Then $\mathcal{C}(\boldsymbol{x}) = \{2\}$ for all $\boldsymbol{x} \in \{0,1\} \times \mathbb{R}$. Similarly $\mathcal{D}(\boldsymbol{x}) = \{1\}$ for all $\boldsymbol{x}$.

Let $\boldsymbol{U} = (U_1, \ldots, U_m)$ be a vector of uniform random variables whose distribution is given by some copula C. We assume that F_j^{-1} is the quantile function corresponding to F_j; since F_j is not invertible when X_j is not absolutely continuous, this corresponds to picking one possible generalized inverse function.

The variables $\boldsymbol{U}$ are selected to satisfy the following criteria. If, at coordinate x_j, $j \in \mathcal{C}(\boldsymbol{x})$, then $u_j = F_j(x_j)$, resulting in a deterministic one-to-one relationship when conditioning on either U_j or X_j. Otherwise, $j \in \mathcal{D}(\boldsymbol{x})$, and we require $x_j = F_j^{-1}(u_j)$, resulting in an infinity of U_j corresponding to one X_j and spanning the interval $(F_j(X_j^-), F_j(X_j))$. This interval corresponds to gaps in the range of F_j. If $\mathcal{C}(\boldsymbol{x}) = \mathcal{M}$ for every $\boldsymbol{x}$, then C will be the copula of $\boldsymbol{X}$. Otherwise, the copula structure will still create dependence between the non-continuous marginal variables but will not be unique in general. Mathematically, the above description leads to the joint density

$$f(\boldsymbol{x}, \boldsymbol{u}) := c(\boldsymbol{u}) \prod_{j \in \mathcal{C}(\boldsymbol{x})} \mathcal{I}(u_j = F_j(x_j)) \prod_{j' \in \mathcal{D}(\boldsymbol{x})} \mathcal{I}(F_{j'}(x_{j'}^-) \leqslant u_{j'} < F_{j'}(x_{j'})), \tag{1}$$

where c is the density corresponding to C and $\mathcal{I}$ is an indicator variable. See Lemma 4, part (i), of Appendix B for a derivation of Eq. (1) and the corresponding measure. Notice that in Eq. (1), products over the indices j and j' correspond to different partitions for each $\boldsymbol{x}$.

With a small abuse of notation, we call $\boldsymbol{U}$ the vector of latent variables, even though U_j is a deterministic function of X_j if F_j is invertible.

To derive the likelihood function, that is the marginal density of $\boldsymbol{X}$, from the joint density $f(\boldsymbol{x}, \boldsymbol{u})$, we introduce some notation. Let $\boldsymbol{a}, \boldsymbol{b}$ be two vectors in $\mathbb{R}^k$ such that $\boldsymbol{a} \leqslant \boldsymbol{b}$ componentwise and let g be an arbitrary function from $\mathbb{R}^k$ into $\mathbb{R}$. We denote by $\triangle_{\boldsymbol{a}}^{\boldsymbol{b}} g(.)$ the sum of 2^k terms that are obtained by repeatedly subtracting $g(., a_j, .)$ from $g(., b_j, .)$ for each $j = 1, \ldots, k$. Appendix A contains more details on using this notation.

For each $\boldsymbol{x} \in \mathbb{R}^m$, denote by $\boldsymbol{b} = (F_1(x_1), \ldots, F_m(x_m))$ the vector of upper bounds and similarly denote by $\boldsymbol{a} = (F_1(x_1^-), \ldots, F_m(x_m^-))$ the vector of lower bounds. For each $j \in \mathcal{C}(\boldsymbol{x})$, $\boldsymbol{b}(j) = \boldsymbol{a}(j)$, otherwise we have the strict inequality $\boldsymbol{b}(j) > \boldsymbol{a}(j)$. Denote the partitions of $\boldsymbol{a}$ and $\boldsymbol{b}$ by $\boldsymbol{a}_{\mathcal{C}(\boldsymbol{x})}$, $\boldsymbol{a}_{\mathcal{D}(\boldsymbol{x})}$, $\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}$ and $\boldsymbol{b}_{\mathcal{D}(\boldsymbol{x})}$. For some sets $A, B \subset \mathcal{M}$, denote by c_A and $c_{A|B}$, the marginal copula density over the indices of A, the conditional copula density where the variables in A are conditioned on the variables with index set B. It is possible to do the same for the copulas C_A and $C_{A|B}$.

If $(\boldsymbol{X}, \boldsymbol{U})$ has the joint density given by Eq. (1), then the marginal density of $\boldsymbol{X}$ is

$$f(\boldsymbol{x}) = c_{\mathcal{C}(\boldsymbol{x})}(\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}) \prod_{j \in \mathcal{C}(\boldsymbol{x})} f_j(x_j) \, \triangle_{\boldsymbol{a}_{\mathcal{D}(\boldsymbol{x})}}^{\boldsymbol{b}_{\mathcal{D}(\boldsymbol{x})}} C_{\mathcal{D}(\boldsymbol{x})|\mathcal{C}(\boldsymbol{x})}(\cdot|\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}), \tag{2}$$

which corresponds to writing the formula for the density of $\boldsymbol{X}$ as the product of the (marginal) density of continuous components at $\boldsymbol{x}$

$$f(\boldsymbol{x}_{\mathcal{C}(\boldsymbol{x})}) = c_{\mathcal{C}(\boldsymbol{x})}(\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}) \prod_{j \in \mathcal{C}(\boldsymbol{x})} f_j(x_j),$$

and the (conditional) density of the non-continuous components conditional on the continuous ones

$$f(\boldsymbol{x}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{x}_{\mathcal{C}(\boldsymbol{x})}) = \triangle_{\boldsymbol{a}_{\mathcal{D}(\boldsymbol{x})}}^{\boldsymbol{b}_{\mathcal{D}(\boldsymbol{x})}} C_{\mathcal{D}(\boldsymbol{x})|\mathcal{C}(\boldsymbol{x})}(\cdot|\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}).$$

See Lemma 4, part (ii), of Appendix B for a derivation of Eq. (2) and the corresponding measure.

We now give a bivariate example to illustrate how the formulas can be used. This example is continued in later sections. See also Section S2 for a trivariate illustrative example.

Example 1 (running illustrative example)**.** *Let X_1 have a density that is a mixture of a point mass at zero and a normal distribution $f_1(x_1) \sim \pi\delta_{x_1}(0) + (1-\pi)\phi(x_1)$, where $\phi(\cdot)$ is the density of a standard normal. This implies that the cumulative distribution function of X_1 is*

$$F_1(x_1) = (1-\pi)\Phi(x_1) + \pi\mathcal{I}(x_1 \geqslant 0),$$

and thus there is a discontinuity in F_1 at the point 0. Let X_2 be a binary random variable with $\Pr\{X_2 = 0\} = \gamma$.

Let $C(\cdot)$ and $c(\cdot)$ be the Clayton copula and its density, with parameter $\theta = 1$, so that

$$C(u_1, u_2) = \left(\frac{1}{u_1} + \frac{1}{u_2} - 1\right)^{-1}, \quad c(u_1, u_2) = \frac{2}{u_1^2 u_2^2}\left(\frac{1}{u_1} + \frac{1}{u_2} - 1\right)^{-3}.$$

The conditional copula is given by

$$C_{2|1}(u_2|u_1) = \frac{1}{u_1^2}\left(\frac{1}{u_1} + \frac{1}{u_2} - 1\right)^{-2},$$

which has the conditional quantile function $C^{-1}(\tau|u_1) = \frac{\sqrt{\tau}u_1}{1+\sqrt{\tau}(u_1-1)}$ and the conditional density $c_{2|1}(u_2|u_1) = c(u_1, u_2)$ (because the marginal of u_1 is uniform).

The following details are necessary to construct the example.

$\mathcal{C}(\boldsymbol{x}) = \{2\}$ for $x_1 \neq 0$, for all x_2 and $\mathcal{C}(\boldsymbol{x}) = \{1, 2\}$ for $x_1 = 0$, for all x_2

Joint density of $\boldsymbol{x}$ and $\boldsymbol{u}$ (Eq. (1))

There are two cases. Case 1: $x_1 \neq 0$

$f(x_1, x_2, u_1, u_2) = c(u_1, u_2)\mathcal{I}(u_1 = F_1(x_1))\mathcal{I}(F_2(x_2-) \leqslant u_2 < F_2(x_2))$

Case 2: $x_1 = 0$

$f(x_1, x_2, u_1, u_2) = c(u_1, u_2)\mathcal{I}(F_1(0-) \leqslant u_1 < F_1(0))\mathcal{I}(F_2(x_2-) \leqslant u_2 < F_2(x_2))$

Likelihood at one point (Eq. (2))

If $x_1 \neq 0$, *then*

$$\begin{aligned} f(x_1, x_2) &= f(x_1) \triangle_{F_2(x_2-)}^{F_2(x_2)} C_{2|1}(\cdot|F(x_1)) \\ &= f_1(x_1)\{C_{2|1}(F_2(x_2)|F_1(x_1)) - C_{2|1}(F_2(x_2-)|F_1(x_1))\} \end{aligned}$$

because $c(u_1) = 1$ *as the one-dimensional margins of a copula are all uniform. If* $x_1 = 0$, *then*

$$\begin{aligned} f(0, x_2) &= \triangle_{F_1(0-)}^{F_1(0)} \triangle_{F_2(x_2-)}^{F_2(x_2)} C(\cdot) \\ &= \triangle_{F_1(0-)}^{F_1(0)} \{C(\cdot, F_2(x_2)) - C(\cdot, F_2(x_2-))\} \\ &= C(F_1(0), F_2(x_2)) - C(F_1(0), F_2(x_2-)) - C(F_1(0-), F_2(x_2)) + C(F_1(0-), F_2(x_2-)). \end{aligned}$$

The difficult part of implementing a simulation algorithm based on Eq. (1) and Eq. (2) is that the size of the vectors $\boldsymbol{x}_{\mathcal{C}(\boldsymbol{x})}$ and $\boldsymbol{x}_{\mathcal{D}(\boldsymbol{x})}$ changes with $\boldsymbol{x}$. A secondary difficulty is that the second term is a sum of $2^{|\mathcal{D}(\boldsymbol{x})|}$ terms for each $\boldsymbol{x}$, where $|\mathcal{D}(\boldsymbol{x})|$ is the cardinality of the set $\mathcal{D}(\boldsymbol{x})$.

3 Estimation and Algorithms

3.1 Conditional distribution of the latent variables

In any simulation scheme (such as MCMC or simulated EM) where the latent variables $\boldsymbol{U}$ are used to carry out inference, it is necessary to know the distribution of $\boldsymbol{U}|\boldsymbol{X}$. This distribution is singular due to the deterministic relationship over $\mathcal{C}(\boldsymbol{x})$ for each $\boldsymbol{x} \in \mathbb{R}^m$. For this reason, it is useful to work only with $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{X}$. A second issue is the need to work with different sized vectors $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}$ for each $\boldsymbol{x}$ in our sample (say $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$), so we will be working with n distributions over different spaces. Recursively using Bayes formula and similar integration arguments to the ones described during the derivation of the $\boldsymbol{X}$ density, we obtain the density for $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{X}$ as

$$f(\boldsymbol{u}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{x}) = \frac{c_{\mathcal{D}(\boldsymbol{x})|\mathcal{C}(\boldsymbol{x})}(\boldsymbol{u}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}) \prod_{j\in\mathcal{D}(\boldsymbol{x})} \mathcal{I}(a_j \leqslant u_j < b_j)}{\triangle_{\boldsymbol{a}_{\mathcal{D}(\boldsymbol{x})}}^{\boldsymbol{b}_{\mathcal{D}(\boldsymbol{x})}} C_{\mathcal{D}(\boldsymbol{x})|\mathcal{C}(\boldsymbol{x})}(\cdot|\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})}, \tag{3}$$

where the denominator is a constant of integration. As seen from the above conditional density, one of the complexities that arise is that the distribution $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{X} = \boldsymbol{x}$ depends on the whole vector $\boldsymbol{x}$ and not just on $\boldsymbol{x}_{\mathcal{D}(\boldsymbol{x})}$. See Lemma 4, part (iii), of Appendix B for a derivation of Eq. (3) and the corresponding measure.

We can now proceed in two ways. We can either draw each U_j in $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}$ separately conditionally on everything else. This is reminiscent of a single move Gibbs sampler. Alternatively, it turns out that in spite of the difficulties, the above distribution can also be sampled recursively without having to compute any of the above normalizing constants. By writing $\mathcal{D}(\boldsymbol{x})$ as $\{j_1, \ldots, j_{|\mathcal{D}(\boldsymbol{x})|}\}$, we can use the following scheme.

- $U_{j_1}|\boldsymbol{X}$
- $U_{j_2}|U_{j_1}, \boldsymbol{X}$
- $\vdots$
- $U_{j_{|\mathcal{D}(\boldsymbol{x})|}}|U_{j_1}, \ldots, U_{j_{|\mathcal{D}(\boldsymbol{x})|-1}}, \boldsymbol{X}$

We now note that the order of the indices $j_1, ..., j_{|\mathcal{D}(\boldsymbol{x})|}$ is irrelevant for the sampling scheme. Although it might appear that the sampling procedure depends on the ordering of these indices, the acceptance or rejection of such samples also depends on the ordering and the next subsection shows that such a procedure will always result in a correct MCMC draw from the conditional distribution $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}|\boldsymbol{X}$.

The above sampling scheme requires knowing the marginal distribution of $\boldsymbol{U}_{\mathcal{J}}|\boldsymbol{X}$ for $\mathcal{J} \subset \mathcal{D}(\boldsymbol{x})$ and the conditional decomposition $U_j|\boldsymbol{U}_{\mathcal{K}}, \boldsymbol{X}$, where $(\{j\}, \mathcal{K})$ is a

 ACCEPTED MANUSCRIPT

partition of $\mathcal{J}$ (meaning $\{j\} = \mathcal{J}\backslash\mathcal{K}$, the complement of $\mathcal{K}$ in $\mathcal{J}$). This distribution can be derived as

$$\begin{aligned} f(\boldsymbol{u}_{\mathcal{J}}|\boldsymbol{x}) &= \frac{c(\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})\prod_{j\in\mathcal{C}(\boldsymbol{x})} f(x_j)}{f(\boldsymbol{x})} c(\boldsymbol{u}_{\mathcal{J}}|\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})}) \\ &\times \left[\triangle_{\boldsymbol{a}_{\mathcal{J}^c}}^{\boldsymbol{b}_{\mathcal{J}^c}} C_{\boldsymbol{U}_{\mathcal{J}^c}|\boldsymbol{U}_{\mathcal{J}},\boldsymbol{U}_{\mathcal{C}(\boldsymbol{x})}}(\cdot|\boldsymbol{u}_{\mathcal{J}},\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})\right] \prod_{j\in\mathcal{J}} \mathcal{I}(a_j \leqslant u_j < b_j) \end{aligned}$$

with $\mathcal{J}^c = \mathcal{D}(\boldsymbol{x})\backslash\mathcal{J}$ and

$$\begin{aligned} f(u_j|\boldsymbol{u}_{\mathcal{K}},\boldsymbol{x}) &= c(u_j|\boldsymbol{u}_{\mathcal{K}},\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})\mathcal{I}(a_j \leqslant u_j < b_j) \\ &\times \frac{\triangle_{\boldsymbol{a}_{\mathcal{J}^c}}^{\boldsymbol{b}_{\mathcal{J}^c}} C_{\boldsymbol{U}_{\mathcal{J}^c}|\boldsymbol{U}_{\mathcal{J}},\boldsymbol{U}_{\mathcal{C}}}(\cdot|\boldsymbol{u}_{\mathcal{J}},\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})}{\triangle_{\boldsymbol{a}_{\mathcal{K}^c}}^{\boldsymbol{b}_{\mathcal{K}^c}} C_{\boldsymbol{U}_{\mathcal{K}^c}|\boldsymbol{U}_{\mathcal{K}},\boldsymbol{U}_{\mathcal{C}}}(\cdot|\boldsymbol{u}_{\mathcal{K}},\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})}, \end{aligned}$$

where $\mathcal{K}^c = \mathcal{J}^c \cup \{j\}$.

We continue to illustrate how to apply the conditional formulas for the latent variables by considering Example 1.

Example 1 (continued)**.** *If $x_1 \neq 0$, then*

$$f(u_2|\boldsymbol{x}) = \frac{c_{2|1}(u_2|F_1(x_1))\mathcal{I}(F_2(x_2-) \leqslant u_2 < F_2(x_2))}{C_{2|1}(F_2(x_2)|F_1(x_1)) - C_{2|1}(F_2(x_2-)|F_1(x_1))};$$

u_1 is deterministically equal to $F_1(x_1)$, so we only need to sample u_2.

If $x_1 = 0$, then

$$f(u_1,u_2|\boldsymbol{x}) = \frac{c(u_1,u_2)\mathcal{I}(F_1(0-) \leqslant u_1 < F_1(0))\mathcal{I}(F_2(x_2-) \leqslant u_2 < F_2(x_2))}{C(F_1(0),F_2(x_2)) - C(F_1(0),F_2(x_2-)) - C(F_1(0-),F_2(x_2)) + C(F_1(0-),F_2(x_2-))}$$

3.2 Metropolis-Hastings sampling

It is clear from the formulas for $f(u_j|\boldsymbol{u}_{\mathcal{K}},\boldsymbol{x})$ that they are quite intricate. They correspond to a product of a simple term $c(u_j|\boldsymbol{u}_{\mathcal{K}},\boldsymbol{b}_{\mathcal{C}(\boldsymbol{x})})\mathcal{I}(a_j \leqslant u_j < b_j)$ (a truncated conditional copula density) and a complicated term that depends on ratios of normalizing constants for $f(\boldsymbol{u}_{\mathcal{J}}|\boldsymbol{x})$ and $f(\boldsymbol{u}_{\mathcal{K}}|\boldsymbol{x})$. One of the most useful aspects of the Metropolis-Hastings (MH) algorithm is that it does not require knowledge of normalizing constants. The trick here is that those normalizing constants are obtained recursively. Assume that we sample

- U_{j_1} from $c(u_{j_1})\mathcal{I}(a_{j_1} \leqslant u_{j_1} < b_{j_1})$
- U_{j_2} from $c(u_{j_2}|u_{j_1})\mathcal{I}(a_{j_2} \leqslant u_{j_2} < b_{j_2})$
- $\vdots$
- $U_{j_{|\mathcal{D}(\boldsymbol{x})|}}$ from $c(u_{j_{|\mathcal{D}(\boldsymbol{x})|}}|u_{j_1}, \ldots, u_{j_{|\mathcal{D}(\boldsymbol{x})|-1}})\mathcal{I}(a_{j_{|\mathcal{D}(\boldsymbol{x})|}} \leqslant u_{j_{|\mathcal{D}(\boldsymbol{x})|}} < b_{j_{|\mathcal{D}(\boldsymbol{x})|}})$;

i.e., if we use as proposal a truncated form of the copula marginal density over $\mathcal{D}(\boldsymbol{x})$, then computing the MH accept/reject ratio results in the computationally simple formula

$$\alpha(\boldsymbol{x}_i) = \prod_{k=1}^{|\mathcal{D}(\boldsymbol{x})|} \frac{C(F_{j_k}(x_{i,j_k})|u^N_{i,j_1}, \ldots, u^N_{i,j_{k-1}}, \boldsymbol{b}_{\mathcal{C}(\boldsymbol{x}_i),i}) - C(F_{j_k}(x^-_{i,j_k})|u^N_{i,j_1}, \ldots, u^N_{i,j_{k-1}}, \boldsymbol{b}_{\mathcal{C}(\boldsymbol{x}_i),i})}{C(F_{j_k}(x_{i,j_k})|u^O_{i,j_1}, \ldots, u^O_{i,j_{k-1}}, \boldsymbol{b}_{\mathcal{C}(\boldsymbol{x}_i),i}) - C(F_{j_k}(x^-_{i,j_k})|u^O_{i,j_1}, \ldots, u^O_{i,j_{k-1}}, \boldsymbol{b}_{\mathcal{C}(\boldsymbol{x}_i),i})},$$

where i represents the observation index. The complexity of this formula is much smaller than $2^{|\mathcal{D}(\boldsymbol{x})|}$.

We now illustrate the Metropolis-Hastings acceptance probabilities by again considering Example 1.

Example 1 (continued). *If $x_1 \neq 0$, then the ratio is $\alpha(x_2) = 1$ and if $x_1 = 0$ (first draw u^N_1 from a uniform on $(F_1(0^-), F_1(0))$ and compare to the previous draw u^O_1)*

$$\alpha(0, x_2) = \frac{C_{2|1}(F_2(x_2)|u^N_1) - C_{2|1}(F_2(x^-_2)|u^N_1)}{C_{2|1}(F_2(x_2)|u^O_1) - C_{2|1}(F_2(x^-_2)|u^O_1)}.$$

We note that here the ordering does not matter, as we could have computed the other ratio (if we draw instead first u^N_2 from a uniform on $(F_2(x^-_2), F_2(x_2))$

$$\alpha(0, x_2) = \frac{C_{1|2}(F_1(0)|u^N_2) - C_{1|2}(F_1(0^-)|u^N_2)}{C_{1|2}(F_1(0)|u^O_2) - C_{1|2}(F_1(0^-)|u^O_2)}.$$

Even though the ratios are different, both procedures will result in a draw from $f(u_1, u_2|\boldsymbol{x})$.

3.3 Mixtures of Archimedean and Gaussian copulas

This section applies the previous results to the family of mixtures of Archimedean and Gaussian copulas. Working with mixtures of copulas provides a simple and yet rich and flexible modeling framework because mixtures of copulas are copulas themselves,

We are particularly interested in having a mixture of three components, two Archimedean copulas, the Clayton copula (C_{Cl}), the Gumbel copula (C_{Gu}) and a Gaussian copula (C_G) component, with corresponding densities c_{Cl}, c_{Gu} and c_G. We later apply this mixture to model the dependence between individual income distributions over 13 years. The density of this 3-component mixture of copulas is

$$c_{mix}\left(\boldsymbol{u};\Gamma,\theta_{Cl},\theta_{Gu},w_1,w_2\right)=w_1c_G\left(\boldsymbol{u};\Gamma\right)+w_2c_{Cl}\left(\boldsymbol{u};\theta_{Cl}\right)+w_3c_{Gu}\left(\boldsymbol{u};\theta_{Gu}\right), \qquad (4)$$

where w_1, w_2, and $w_3 = 1 - w_1 - w_2$ are the mixture weights, and Γ, θ_{Cl}, and θ_{Gu} are respectively the dependence parameters of the Gaussian, Clayton, and Gumbel copulas. Such a mixture of copula models has the additional flexibility of being able to capture lower and upper tail dependence. We will use a Bayesian approach to estimate the copula parameters and for simplicity, and without loss of generality, we follow Joe (2014) and use empirical CDF's to model the marginal distributions.

Let the parameter w_k denote the probability that the i-th observation comes from the k-th component in the mixture. Let $\boldsymbol{d}_i = (d_{i1}, d_{i2}, d_{i3})'$ be the vector of latent indicator variables such that $d_{ik} = 1$ when the i-th observation comes from the k-th component in the mixture. These indicator variables identify the component of the copula model defined in Eq. (4) to which the observation $\boldsymbol{y}_i$ belongs. Then,

$$\Pr\left(d_{ik}=1|\boldsymbol{w}\right)=w_k, \qquad (5)$$

with $w_k > 0$ and $\sum_{k=1}^{3} w_k = 1$.

Given the information on the n independent sample observations $\boldsymbol{y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)'$

and $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})$, and by using Bayes rule, the joint posterior density is obtained as

$$p\left(\boldsymbol{w}, \boldsymbol{d}, \Gamma, \theta_{Cl}, \theta_{Gu} | \boldsymbol{y}\right) \propto p\left(\boldsymbol{y} | \boldsymbol{w}, \boldsymbol{d}, \Gamma, \theta_{Cl}, \theta_{Gu}\right) p\left(\boldsymbol{d} | \boldsymbol{w}, \Gamma, \theta_{Cl}, \theta_{Gu}\right) p\left(\boldsymbol{w}\right) p\left(\Gamma\right) p\left(\theta_{Cl}\right) p\left(\theta_{Gu}\right) \tag{6}$$

with

$$p\left(\boldsymbol{y} | \boldsymbol{w}, \boldsymbol{d}, \Gamma, \theta_{Cl}, \theta_{Gu}\right) = \prod_{i=1}^{n} \left[c_G\left(\boldsymbol{u}; \Gamma\right)\right]^{d_{i1}} \left[c_{Cl}\left(\boldsymbol{u}; \theta_{Cl}\right)\right]^{d_{i2}} \left[c_{Gu}\left(\boldsymbol{u}; \theta_{Gu}\right)\right]^{d_{i3}},$$

and

$$p\left(\boldsymbol{d} | \boldsymbol{w}, \Gamma, \theta_{Cl}, \theta_{Gu}\right) = p\left(\mathbf{d} | \boldsymbol{w}\right) = \prod_{i=1}^{n} \prod_{k=1}^{K} w_k^{d_{ik}} = \prod_{k=1}^{K} w_k^{n_k}, \tag{7}$$

where $n_k = \sum_{i=1}^{n} I\left(d_{ik} = 1\right)$ and $I\left(d_{ik} = 1\right)$ is an indicator variable which is equal 1 if observation i belongs to the k-th component of the copula mixture model, and is 0 otherwise. We use a Dirichlet prior for $\boldsymbol{w}$, $p\left(\boldsymbol{w}\right) = Dirichlet\left(\boldsymbol{\phi}\right)$, which is defined as

$$p\left(\boldsymbol{w}\right) \propto w_1^{\phi_1 - 1} ... w_3^{\phi_3 - 1}. \tag{8}$$

The Dirichlet distribution is a common choice in Bayesian mixture modeling since it is a conjugate of the multinomial distribution (Diebold and Robert, 1994) . We use the gamma density $G(\alpha, \beta)$ as the prior distribution for θ_{Cl} and θ_{Gu}. The hyper-parameters in the prior probability density functions (PDFs) are chosen so that the priors are uninformative. We use a Metropolis within Gibbs sampling algorithm to draw observations from the joint posterior PDF defined in Eq. (6) and use the resulting MCMC draws to estimate the quantities required for inference. The relevant conditional posterior PDFs are now specified.

The conditional posterior probability that the ith observation comes from the kth component in the copula mixture model is

$$p\left(d_{ik} | \boldsymbol{w}, \Gamma, \theta_{Cl}, \theta_{Gu}, \boldsymbol{y}\right) = \frac{p_{ik}}{p_{i1} + ... + p_{i3}}, \tag{9}$$

where $p_{i1} = w_1 c_G(\boldsymbol{u}; \Gamma)$, $p_{i2} = w_2 c_{Cl}(\boldsymbol{u}; \theta_{Cl})$, and $p_{i3} = w_3 c_{Gu}(\boldsymbol{u}; \theta_{Gu})$ for $k = 1, 2, 3$. The conditional posterior PDF for the mixture weights $\boldsymbol{w}$ is the Dirichlet PDF

$$p(\boldsymbol{w}|\mathbf{d}, \Gamma, \theta_{Cl}, \theta_{Gu}, \boldsymbol{y}) = D(\boldsymbol{\phi} + \boldsymbol{n}), \tag{10}$$

where $\mathbf{n} = (n_1, ..., n_k)'$ and $\boldsymbol{\phi} = (\phi_1, ..., \phi_K)'$. The conditional posterior PDF for the Gaussian copula parameter matrix Γ is

$$p(\Gamma|\boldsymbol{y}, \mathbf{d}, \theta_{Cl}, \theta_{Gu}, \boldsymbol{w}) = \prod_{i \in d_{i1}=1} c_G(\boldsymbol{u}; \Gamma)\, p(\Gamma). \tag{11}$$

The conditional posterior PDF for the Clayton copula parameter θ_{Cl} is

$$p(\theta_{Cl}|\boldsymbol{y}, \mathbf{d}, \Gamma, \theta_{Gu}, \boldsymbol{w}) = \prod_{i \in d_{i1}=2} c_{Cl}(\boldsymbol{u}; \theta_{Cl})\, p(\theta_{Cl}). \tag{12}$$

The conditional posterior PDF for the Gumbel copula parameter θ_{Gu} is

$$p(\theta_{Gu}|\boldsymbol{y}, \mathbf{d}, \Gamma, \theta_{Cl}, \boldsymbol{w}) = \prod_{i \in d_{i1}=3} c_{Gu}(\boldsymbol{u}; \theta_{Gu})\, p(\theta_{Gu}). \tag{13}$$

Generating the conditional posterior density for θ_{Cl} and θ_{Gu} is not straightforward since the conditional posterior densities for both θ_{Cl} and θ_{Gu} are not in a recognizable form. We use a random walk Metropolis algorithm to draw from the conditional posterior densities of both θ_{Cl} and θ_{Gu}. The generation of the Gaussian copula matrix parameter Γ is more complicated and is explained in the next section.

The full MCMC sampling scheme is,

1. Set the starting values for $\boldsymbol{w}^{(0)}$, $\Gamma^{(0)}$, $\theta_{Cl}^{(0)}$, and $\theta_{Gu}^{(0)}$
2. Generate $(\boldsymbol{w}^{(t+1)}|\mathbf{d}^{(t)}, \Gamma^{(t)}, \theta_{Cl}^{(t)}, \theta_{Gu}^{(t)}, \boldsymbol{y})$ from Eq. (10)
3. Generate $\left(\Gamma^{(t+1)}|\boldsymbol{y}, \mathbf{d}^{(t+1)}, \theta_{Cl}^{(t)}, \theta_{Gu}^{(t)}, \boldsymbol{w}^{(t+1)}\right)$ from Eq. (11)
4. Generate $\left(\theta_{Cl}^{(t+1)}|\boldsymbol{y}, \mathbf{d}^{(t+1)}, \Gamma^{(t+1)}, \theta_{Gu}^{(t)}, \boldsymbol{w}^{(t+1)}\right)$ from Eq. (12)

5. Generate $\left(\theta_{Gu}^{(t+1)}|\boldsymbol{y}, \mathbf{d}^{(t+1)}, \Gamma^{(t+1)}, \theta_{Cl}^{(t+1)}, \boldsymbol{w}^{(t+1)}\right)$ from Eq. (13)

6. Set $t = t + 1$ and return to step 2.

Appendix S1 gives further details of the sampling scheme. In particular, it describes how to write the Gaussian, Gumbel and Clayton copulas and their densities and how to sample from them. It also details how to sample the correlation parameters of the Gaussian copula and summarizes how the one-margin at a time latent variable simulation works.

4 Application to Individual Income Dynamics

Longitudinal or panel datasets, such as the Panel Study of Income Dynamics (PSID), the British Household Panel Survey (BHPS), and the Household Income and Labour Dynamics Survey in Australia (HILDA) are increasingly used for assessing income inequality, mobility, and poverty over time. The income data from these surveys for different years are correlated due to the nature of panel studies. For such correlated samples, the standard approach of fitting univariate models to income distributions for different years may give rise to misleading results. The univariate approach treats the income distribution over different years as independent and ignores the dependence structure between incomes for different years. It does not take into account that those who earned a high income in one year are more likely to earn a high income in subsequent years and vice versa. A common way to address this problem is to use a multivariate income distribution model that takes into account the dependence between incomes for different years.

The presence of dependence in a sample of incomes from panel datasets has rarely been addressed in the past. Only recently, Vinh et al. (2010) proposed using bivariate copulas to model income distributions for two different years, using maximum likelihood estimation. However, in their applications, they do not take into account the point mass at zero income. Our methodology is more general than Vinh et al.

ACCEPTED MANUSCRIPT

(2010). We estimate a panel of incomes from 2001 to 2013 using a finite mixture of Gaussian, Clayton, and Gumbel copulas while taking into account the point mass occurring at zero incomes. Once the parameters for the multivariate income distribution have been estimated, values for various measures of inequality, mobility, and poverty can be obtained. Our methodology is Bayesian which enables us to estimate the posterior densities of the parameters of the copula models and the inequality, mobility, and poverty measures. In this example, we consider the Shorrocks (1978b) and Foster (2009) indices for illustration purposes. Other inequality, mobility, and poverty indices can be estimated similarly. See for example Bonhomme and Robin (2009) for a recent study on income mobility dynamics.

This paper uses unit record data from the Household, Income, and Labour Dynamics in Australia (HILDA) survey. The HILDA project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

Although a number of income related variables are available, we use the imputed income series _WSCEI in this example. This variable contains the average individual weekly wage and salary incomes from all paid employment over the period considered. It is reported before taxation and governmental transfers. The income data were also adjusted to account for the effects of inflation using the Consumer Price Index data obtained from the Australian Bureau of Statistics, which is based in 2010 dollars. From these data, a dependence sample was constructed by establishing whether a particular individual had recorded an income in all the years. Individuals who only recorded incomes in some of the years being considered were removed. In addition, we also focus our attention on individuals who are in the labor force (both employed and unemployed). We found that 1745 individuals recorded an income for all 13 years. Table 1 summarizes the distributions of real individual disposable income in

Australia for the years 2001 - 2013 and shows that all income distributions exhibit positive skewness and fat long right tails typical of income distributions. If the ordering of the distributions is judged on the basis of the means or the medians, the population becomes better off as it moves from 2001 to 2013, except between the period 2006 and 2007. These effects are also confirmed by Figures S1 to S3 in Appendix S4

4.1 Foster's (2009) Chronic Poverty Measures

The measurement of chronic income poverty is important because it focuses on those whose lack of income stops them from obtaining the "minimum necessities of life" for much of their life course. Let $z \in \mathbb{R}^+$ be the poverty line. It is the level of income/wages which is just sufficient for someone to be able to afford the minimum necessities of life. For every $i = 1, ..., n$ and $t = 1, ..., T$, the row vector $\mathbf{y}_i = (y_{i1}, ..., y_{iT})$ contains individual i's incomes across time and the column vector $\mathbf{y}_{*t} = (y_{1t}, ..., y_{nt})'$ contains the income distribution at period t.

The measurement of chronic poverty is split into two steps: an "identification" step and an aggregation step. The identification function $\rho(\mathbf{y}_i; z)$ indicates that individual i is in chronic poverty when $\rho(\mathbf{y}_i; z) = 1$, while $\rho(\mathbf{y}_i; z) = 0$ otherwise. Foster (2009) proposed an identification method that counts the number of periods of poverty experienced by a particular individual, $y_{it} < z$, and then expressed it as a fraction d_i of the T periods. The identification function $\rho_\tau(\mathbf{y}_i; z) = 1$ if $d_i \geq \tau$ and $\rho_\tau(\mathbf{y}_i; z) = 0$ if $d_i < \tau$.

The aggregation step combines the information on the chronically poor people to obtain an overall level of chronic poverty in a given society. We use the extension of univariate Foster, Greer and Thorbecke (FGT) poverty indices of Foster et al. (1984). These are given by

$$FGT^\alpha(z) = \frac{1}{n}\sum_{i=1}^{n} g_i^\alpha \ ,$$

 ACCEPTED MANUSCRIPT

Table 1: Descriptive statistics for real individual wages for Australia for the years 2001 - 2013

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Mean	684	734	766	819	874	923	783	1067	1105	1128	1188	1215	1245
Median	616	673	712	753	803	852	702	969	1003	1048	1051	1101	1100
Std. dev.	551	591	568	645	674	668	694	788	825	869	990	916	950
skewness	2.1	3.0	2.1	3.7	2.9	1.5	2.0	2.0	1.8	2.0	3.7	1.9	1.5
kurtosis	15.7	26.6	16.0	40.2	27.1	8.7	11.1	12.7	11.7	15.7	37.5	13.0	7.4

ACCEPTED MANUSCRIPT

where $g_i^\alpha = 0$ if $y_i > z$ and $g_i^\alpha(z) = \left(\frac{z-y_i}{z}\right)^\alpha$ if $y_i \leq z$, and α measures inequality aversion. The FGT measure when $\alpha = 0$ is called the headcount ratio, when $\alpha = 1$ it is called the poverty gap index and when $\alpha = 2$ it is called the poverty severity index. Foster (2009) proposed duration adjusted FGT poverty indices: duration adjusted headcount ratio and duration adjusted poverty gap. Following Foster (2009), we define the normalized gap matrix as $G^\alpha(z) := [g_{it}^\alpha(z)]$, where $g_{it}^\alpha(z) = 0$ if $y_{it} > z$ and $g_{it}^\alpha(z) = \left(\frac{z-y_{it}}{z}\right)^\alpha$ if $y_{it} \leq z$. Then, identification is incorporated into the censored matrix $G^\alpha(z,\tau) = [g_{it}^\alpha(z,\tau)]$, where $g_{it}^\alpha(z,\tau) = g_{it}^\alpha(z)\rho_\tau(\mathbf{y}_i; z)$. The entries for the non-chronically poor are censored to zero, while the entries for the chronically poor are left unchanged. When $\alpha = 0$, the measure becomes the duration adjusted headcount ratio and is the mean of $G^0(z,\tau)$; when $\alpha = 1$, the measure becomes the duration adjusted poverty gap, and is given by the mean of $G^1(z,\tau)$.

4.2 Shorrocks (1978a) Income Mobility Measures

The measurement of income mobility focuses on how individuals' income changes over time. Many mobility measures have been developed and applied to empirical data to describe income dynamics; see Shorrocks (1978b), Shorrocks (1978a), Formby et al. (2004), Dardanoni (1993), Fields and Ok (1996), Maasoumi and Zandvakili (1986), and references therein. However, statistical inference on income mobility has been largely neglected in the literature. Only recently, some researchers have developed statistical inference procedures for the measurement of income mobility (Biewen, 2002; Maasoumi and Trede, 2001; Formby et al., 2004). Here, we show that our approach can be used to obtain the posterior densities of mobility measures which can then be used for making inference on income mobility.

Shorrocks (1978b) proposed a measure of income mobility that is based on transition matrices. Following Formby et al. (2004), we consider the joint distribution of two income variables y_1 and y_2 with a continuous CDF $F(y_1, y_2)$. This distribution captures all the transitions between y_1 and y_2. In this application, we consider the mobility between two points in time. The movement between y_1 and y_2 is described by a transition matrix. To form the the transition matrix from $F(y_1, y_2)$, we need to determine the number of, and boundaries between, income classes. Suppose there are m classes in each of the income variables and the boundaries of these classes are $0 < \tau_1^{y_1} < ... < \tau_{m-1}^{y_1} < \infty$ and $0 < \tau_1^{y_2} < ... < \tau_{m-1}^{y_2} < \infty$. The resulting transition matrix is denoted $P = [p_{ij}]$. Each element p_{ij} is a conditional probability that an individual moves to class j of income y_2 given that they are initially in class i with income y_1. It is defined as

$$p_{ij} := \frac{\Pr\left(\tau_{i-1}^{y_1} \leq y_1 < \tau_i^{y_1} \text{ } and \text{ } \tau_{j-1}^{y_2} \leq y_2 < \tau_j^{y_2}\right)}{\Pr\left(\tau_{i-1}^{y_1} \leq y_1 < \tau_i^{y_1}\right)},$$

where $\Pr\left(\tau_{i-1}^{y_1} \leq y_1 < \tau_i^{y_1}\right)$ is the probability that an individual falls into income class i of y_1.

A mobility measure $M(P)$ can be defined as a function of the transition matrix P. We say that a society with transition matrix P_1 is more mobile than one with transition matrix P_2, according to mobility measure $M(P)$, if and only if $M(P_1) > M(P_2)$. We consider a mobility measure developed by Shorrocks (1978b) and defined as

$$M_1(P) := \frac{m - \sum_{i=1}^{m} p_{ii}}{m-1};$$

M_1 measures the average probability across all classes that an individual will leave his initial class in the next period.

4.3 Empirical Analysis

This section discusses the results from the analysis of the real individual wages data after estimating the proposed multivariate income distribution model using a Bayesian approach. The univariate income distribution is usually modeled using Dagum or Singh-Maddala distributions (Kleiber, 1996). In this example, the marginal income distribution is modeled using the empirical distribution function, for simplicity. It is straightforward to extend the MCMC sampling scheme in Section 3 to estimate both marginal and joint parameters as in Pitt et al. (2006) and Smith and Khaled (2012).

First, we present the model selection results and the estimated parameters of the copula models. To select the best copula model, we use the DIC_3 criterion of Celeux et al. (2006) and the cross-validated log predictive score (LPDS) (Good, 1952; Geisser, 1980). The DIC_3 criterion is defined as

$$DIC_3 := -4\mathbb{E}_\theta \left(\log p\left(\boldsymbol{y}|\theta\right)|\boldsymbol{y}\right) + 2\log \widehat{p}\left(\boldsymbol{y}\right),$$

where $\widehat{p}\left(\boldsymbol{y}\right) = \prod_{i=1}^{n} \widehat{p}\left(y_i\right)$. We next define the B-fold cross-validated LPDS. Suppose that the dataset $\mathcal{D}$ is split into roughly B equal parts $\mathcal{D}_1, ..., \mathcal{D}_B$. Then, the $B-$fold cross validated LPDS is defined as

$$LPDS\left(\widehat{p}\right) := \sum_{j=1}^{B} \sum_{y_j \in \mathcal{D}_j} \log \widehat{p}\left(\mathbf{y}_j|\mathcal{D} \setminus \mathcal{D}_j\right).$$

In our work we take $B = 5$. Table 2 shows that the best model, according to both criteria, is the mixture of Gaussian, Clayton, and Gumbel copulas. We estimate the best model with an initial burnin period of 10000 iterates and a Monte Carlo sample of 10000 iterates. Next, we use the iterates from the best model to estimate transition probabilities from 0 to positive wages and from positive wages to zero, Spearman's correlation coefficient, and the mobility and poverty measures, by averaging over the posterior distribution of the parameters.

Table 3 shows some of the estimated parameters and corresponding 95% credible intervals for the chosen copula mixture model. The parameters and their 95% credible intervals are quite tight, indicating that the parameters are well estimated. It is clear that there are significant differences in the estimated parameters by taking into account the point mass at zero wages compared to the parameters estimated by not taking into account this point mass. The estimated mixture weight parameters show that the Gaussian copula has the highest weight, followed by the Clayton and Gumbel copulas. As the weight of the Clayton copula is higher than of the Gumbel copula, it implies that there are more people with lower tail dependence than upper tail dependence. This may coincide with a relatively higher degree of income mobility amongst high income earners.

Table 2: Model Selection of the copula to model 13 years of income distribution with point mass at zero incomes

Model	DIC_3	LPDS-CV
Clayton	-1.21×10^4	6.03×10^3
Gumbel	-1.75×10^4	4.95×10^3
Gaussian	-2.13×10^4	4.29×10^4
Mixture (Gaussian, Clayton)	-2.86×10^4	5.63×10^4
Mixture (Gaussian, Gumbel)	-2.83×10^4	5.54×10^4
Mixture (Clayton, Gumbel)	-1.68×10^4	3.31×10^4
Mixture (Gaussian, Clayton, Gumbel)	$-2.89 \times 10^{4\star}$	$5.68 \times 10^{4\star}$

Table 3: Some of the estimated parameters of the mixture of the Gaussian, Gumbel and Clayton copulas to model 13 years of income distributions. The 95% credible intervals are in brackets

Parameters	Copula (Point Mass)	Copula (No Point Mass)
θ_{Cl}	0.15 (0.12,0.18)	0.33 (0.29,0.37)
θ_{Gu}	1.94 (1.84,2.06)	2.33 (2.23,2.45)
w_1	0.66 (0.64,0.69)	0.62 (0.60,0.65)
w_2	0.21 (0.19,0.24)	0.23 (0.21,0.26)

Tables S1 and S2 in Appendix S4 present the estimates of the transition probabilities from 0 to positive wages and from positive to 0 wages. The estimates of the transition probabilities seem to be close to their sample (non-parametric) counterparts. The estimates of transition probabilities from 0 to positive wages are similar (0.39-0.49) in the period from 2001-2006. Similarly, the estimates are similar in the period 2008-2013 (0.34-0.38). However, there are higher estimates for the period 2006-2007 and 2007-2008 (0.83 and 0.87, respectively). Similar results are observed for the transition probabilities from positive to zero wages. The estimates of the transition probabilities are roughly the same between the periods 2001-2006 and

2008-2013. There are higher estimates for the period 2006-2007 and 2007-2008. This phenomenon may indicate that there is very high income mobility between 2006-2007 and 2007-2008. Note that the model that does not take into account the point masses at zero cannot give us the estimates of the transition probabilities.

Tables 4 and 5 show the estimate of Spearman's rho dependence and Shorrocks (1978b) mobility measure. We can see from these two measures that there are very high values of the mobility measure and very low values of Spearman's rho dependence measure between 2006-2007 and 2007-2008. This confirms our previous analysis that in the period 2006-2008 there is very high mobility between income earners. Table 6 shows the estimates of Foster's chronic poverty measures: duration adjusted headcount ratio and duration adjusted poverty gap. The two measures indicate that chronic poverty is significantly lower in the 2007-2013 period compared to the 2001-2006 period. The standard of living in Australia is higher in the period 2007-2013 compared to the period 2001-2006. Furthermore, we can see that the estimates of Spearman's rho dependence, mobility, and chronic measures are different between the estimates that take into account the point masses and the estimates that do not take into account the point masses at zero wages. Figure 1 shows the posterior densities of duration adjusted headcount ratio for the years 2007-2013 for the two estimates. The figure shows that the posterior densities almost do not overlap, indicating that the two estimates are significantly different. Therefore, whenever the point masses are present, it is strongly recommended to incorporate them into the model to guard against biased estimates.

Table 4: Estimates of the Spearman rho dependence measure of the mixture of the Gaussian, Gumbel and Clayton copulas and 95% credible intervals (in brackets)

Period	Copula (Point Mass)	Copula (No Point Mass)
2001-2002	0.703 (0.684,0.722)	0.740 (0.723,0.757)
2002-2003	0.719 (0.700,0.737)	0.743 (0.726,0.759)
2003-2004	0.721 (0.702,0.739)	0.743 (0.727,0.759)
2004-2005	0.723 (0.7040,0.741)	0.747 (0.730,0.763)
2005-2006	0.727 (0.708,0.745)	0.750 (0.733,0.766)
2006-2007	0.020 (−0.028,0.068)	0.030 (−0.020,0.086)
2007-2008	0.025 (−0.023,0.073)	0.037 (−0.013,0.093)
2008-2009	0.725 (0.706,0.744)	0.7500 (0.733,0.766)
2009-2010	0.735 (0.716,0.753)	0.758 (0.741,0.775)
2010-2011	0.740 (0.720,0.758)	0.764 (0.747,0.781)
2011-2012	0.737 (0.718,0.755)	0.762 (0.745,0.778)
2012-2013	0.733 (0.714,0.752)	0.759 (0.742,0.776)

Table 5: Estimates of Shorrocks (1978a) Mobility Measure ($m = 5$) of the mixture of the Gaussian, Gumbel and Clayton copulas

Period	Non-Parametric	Copula (Point Mass)	Copula (No Point Mass)
2001-2002	0.414 (0.367,0.466)	0.569 (0.549,0.588)	0.518 (0.501,0.534)
2002-2003	0.411 (0.361,0.461)	0.526 (0.508,0.543)	0.499 (0.484,0.516)
2003-2004	0.366 (0.324,0.409)	0.500 (0.483,0.516)	0.479 (0.463,0.495)
2004-2005	0.380 (0.341,0.418)	0.489 (0.473,0.506)	0.465 (0.450,0.480)
2005-2006	0.392 (0.352,0.427)	0.484 (0.468,0.5000)	0.459 (0.444,0.475)
2006-2007	0.996 (0.974,1.019)	0.969 (0.957,0.980)	0.918 (0.878,0.938)
2007-2008	0.987 (0.959,1.015)	0.933 (0.921,0.945)	0.885 (0.843,0.906)
2008-2009	0.411 (0.384,0.441)	0.510 (0.493,0.526)	0.480 (0.465,0.495)
2009-2010	0.380 (0.350,0.409)	0.500 (0.482,0.516)	0.465 (0.449,0.481)
2010-2011	0.381 (0.351,0.411)	0.481 (0.463,0.500)	0.440 (0.424,0.456)
2011-2012	0.380 (0.353,0.405)	0.492 (0.475,0.510)	0.453 (0.437,0.469)
2012-2013	0.365 (0.339,0.395)	0.517 (0.499,0.536)	0.475 (0.458,0.493)

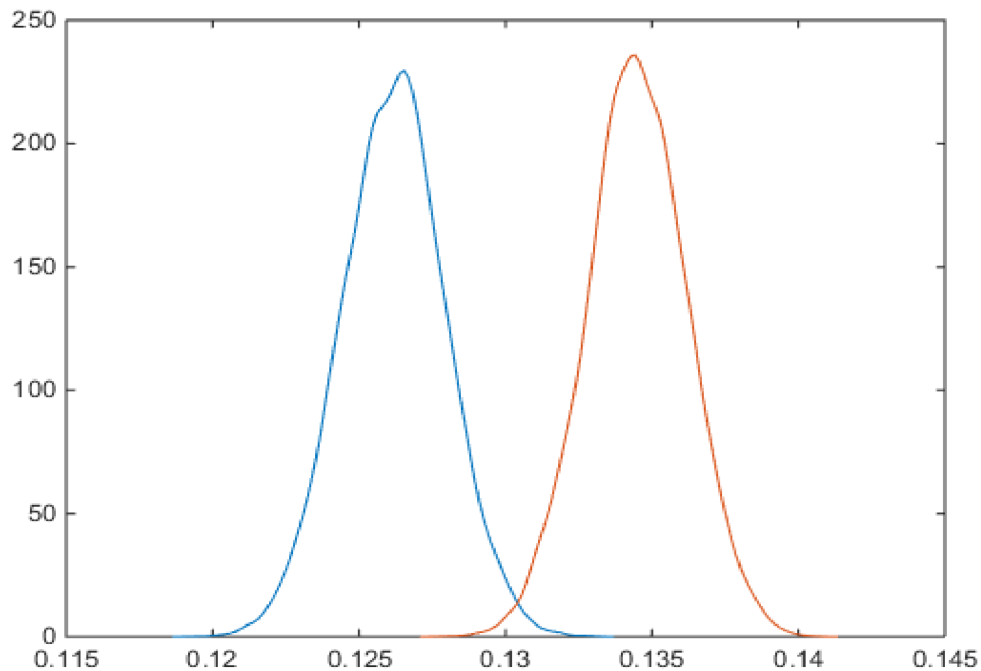
ACCEPTED MANUSCRIPT

Table 6: Estimates of Foster's chronic poverty measure of the mixture of the Gaussian, Gumbel and Clayton copulas with 95% credible intervals (in brackets)

Measure	Period	Non-Parametric	Copula (Point Mass)	Copula (No Point Mass)
adj. headcount	2001-2006	0.211 (0.193,0.229)	0.192 (0.187,0.197)	0.201 (0.197,0.205)
adj. headcount	2007-2013	0.135 (0.120,0.149)	0.126 (0.123,0.130)	0.135 (0.131,0.138)
adj. poverty gap	2001-2006	0.137 (0.123,0.150)	0.134 (0.130,0.137)	0.141 (0.137,0.144)
adj. poverty gap	2007-2013	0.108 (0.095,0.119)	0.104 (0.101,0.107)	0.111 (0.108,0.114)

ACCEPTED MANUSCRIPT

Figure 1: Estimated headcount posterior densities based on including (left density-blue line) and not including point masses (right density-orange line) at 0 (2007-2013)



5 Conclusion and discussion

The paper shows how to define and derive the density of the observations of a general copula model when some of the marginals are discrete, some are continuous *and* some are a mixture of discrete and continuous components. This is done by carefully defining the likelihood as the density of the observations with respect to a mixed measure and allows us to define the likelihood for general copula models and hence carry out likelihood based inference. Our work extends in a very general way the current literature on likelihood based inference which focuses on copulas where each marginal is either discrete or continuous. The inference in the paper is Bayesian and we show how to construct an efficient MCMC scheme to estimate functionals of the posterior distribution. Although our discussion and examples focus on Gaussian and Archimedean copulas, our treatment is quite general and can be applied as long as it is possible to compute certain marginal and conditional copulas either in closed-form or numerically.

We believe that our article can be extended in the following directions. First, using our definition of the likelihood can also enable maximum likelihood type inference using, for example, simulated EM or simulated maximum likelihood. Second, our approach can be applied to copulas based on pair-copula constructions (e.g. Aas et al., 2009) or vine copulas (e.g. Bedford and Cooke, 2002) because our methods apply to arbitrary copulas with the only requirement that it is possible to write down several conditional marginal copulas and copula densities and being able to compute those either analytically or numerically. See Smith and Khaled (2012) for an application of their approach to vine copulas.

Third, by using pseudo marginal methods (e.g. Andrieu et al., 2010), our methodology can also be extended to the case where the likelihood of the copula can only be estimated unbiasedly, rather than evaluated. We leave all such extensions to future work.

Our article illustrates the methodology and algorithms by applying them to estimate a multivariate income dynamics model. Examples of further possible applications arise from any setup where one or more of the following variables are present: wages (where there is a point mass at the minimum wage) individual sales figures, where there is a point mass at 0 (many individuals deciding not to purchase) and a smooth distribution above that point (corresponding to a continuum of price figures). Another interesting potential application is to extend the general truncated/censored variable models in econometrics to a copula framework, e.g., for multivariate tobit and sample selection models.

Acknowledgement

We would like to thank two anonymous referees and the associate editor for suggestions that helped improve the clarity of the paper. The research of David Gunawan and Robert Kohn was partially supported by an Australian Research Council Discovery Grant DP150104630.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and economics*, 44, 182–198.

Andrieu, C., Doucet, A., and Holenstein, R. (2010), "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society, Series B*, 72, 1–33.

Bedford, T. and Cooke, R. M. (2002), "Vines: A new graphical model for dependent random variables," *Annals of Statistics*, 1031–1068.

Biewen, M. (2002), "Bootstrap Inference for Inequality, Mobility, and Poverty Measurement," *Journal of Econometrics*, 108, 317–342.

Bonhomme, S. and Robin, J.-M. (2009), "Assessing the equalizing force of mobility using short panels: France, 1990–2000," *The Review of Economic Studies*, 76, 63–92.

Brechmann, E., Czado, C., and Paterlini, S. (2014), "Flexible dependence modeling of operational risk losses and its impact on total capital requirements," *Journal of Banking & Finance*, 40, 271–285.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006), "Deviance Information Criteria for Missing Data Models," *Bayesian Analysis*, 1, 651–674.

Dardanoni, V. (1993), "On Measuring Social Mobility," *Journal of Economic Theory*, 61, 372–394.

De Leon, A. R. and Chough, K. C. (2013), *Analysis of Mixed Data: Methods & Applications*, CRC Press.

de Leon, A. R. and Wu, B. (2011), "Copula-based regression models for a bivariate mixed discrete and continuous outcome," *Statistics in Medicine*, 30, 175–185.

 ACCEPTED MANUSCRIPT

Diebold, J. and Robert, C. P. (1994), "Estimation of Finite Mixture Distributions through Bayesian Sampling," *Journal of Royal Statistician Society Series B*, 56, 363–375.

Durante, F. and Sempi, C. (2015), *Principles of copula theory*, CRC Press.

Fan, Y. and Patton, A. J. (2014), "Copulas in econometrics," *Annu. Rev. Econ.*, 6, 179–200.

Fields, G. and Ok, E. A. (1996), "The Meaning and Measurement of Income Mobility," *Journal of Economic Theory*, 71, 349–377.

Formby, J. P., Smith, W. J., and Zheng, B. (2004), "Mobility Measurement, Transition Matrices, and Statistical Inference," *Journal of Econometrics*, 120, 181–205.

Foster, J., Joel, G., and Eric, T. (1984), "A class of decomposable poverty Measures," *Econometrica*, 52, 761–765.

Foster, J. E. (2009), *A Class of Chronic Poverty Measures*, Oxford: Oxford University Press, poverty dynamics: interdisciplinary perspectives: 59-76 ed.

Geisser, S. (1980), "Discussion of Sampling and Bayes Inference in Scientific Modeling and Robustness by G. E. P. Box," *Journal of the Royal Statistical Society Series A*, 143, 416–417.

Genest, C. and Neslehová, J. (2007), "A primer on copulas for count data," *Astin Bulletin*, 37, 475–515.

Good, I. J. (1952), "Rational Decisions," *Journal of the Royal Statistical Society B*, 14, 107–114.

He, J., Li, H., Edmondson, A. C., Rader, D. J., and Li, M. (2012), "A Gaussian copula approach for the analysis of secondary phenotypes in case–control genetic association studies," *Biostatistics*, 13, 497–508.

Hoff, P. D. (2007), "Extending the rank likelihood for semiparametric copula estimation," *The Annals of Applied Statistics*, 265–283.

Jiryaie, F., Withanage, N., Wu, B., and de Leon, A. (2016), "Gaussian copula distributions for mixed data, with application in discrimination," *Journal of Statistical Computation and Simulation*, 86, 1643–1659.

Joe, H. (2014), *Dependence modeling with copulas*, CRC Press.

Kleiber, C. (1996), "Dagum and Singh Maddala income distributions," *Economics Letters*, 53, 265–268.

Maasoumi, E. and Trede, M. (2001), "Comparing Income Mobility in Germany and the United States using Generalized Entropy Measures," *Review of Economics and Statistics*, 83, 551–559.

Maasoumi, E. and Zandvakili, S. (1986), "A class of Generalized Measures of Mobility with Applications," *Economics Letters*, 22, 97–102.

Panagiotelis, A., Czado, C., and Joe, H. (2012), "Pair copula constructions for multivariate discrete data," *Journal of the American Statistical Association*, 107, 1063–1072.

Patton, A. J. (2009), "Copula–based models for financial time series," in *Handbook of financial time series*, Springer, pp. 767–785.

Pitt, M., Chan, D., and Kohn, R. (2006), "Efficient Bayesian Inference for Gaussian Copula Regression Models," *Biometrika*.

Shen, C. and Weissfeld, L. (2006), "A copula model for repeated measurements with non-ignorable non-monotone missing outcome," *Statistics in medicine*, 25, 2427–2440.

Shorack, G. R. (2000), *Probability for statisticians*, Springer.

Shorrocks, A. F. (1978a), "Income Inequality and Income Mobility," *Journal of Economic Theory*, 19, 376–393.

— (1978b), "The Measurement of Mobility," *Econometrica*, 46, 1013–1024.

Smith, M. and Khaled, M. A. (2012), "Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation," *Journal of the American Statistical Association*.

Song, P. X.-K., Li, M., and Yuan, Y. (2009), "Joint regression analysis of correlated data using Gaussian copulas," *Biometrics*, 65, 60–68.

Stöber, J., Hong, H. G., Czado, C., and Ghosh, P. (2015), "Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses," *Computational Statistics & Data Analysis*, 88, 28–39.

Trivedi, P. K. and Zimmer, D. M. (2007), *Copula modeling: an introduction for practitioners*, Now Publishers.

Vinh, A., Griffiths, W. E., and Chotikapanich, D. (2010), "Bivariate Income Distribution for Assessing Inequality and Poverty Under Dependent Samples," *Economic Modelling*, 27, 6, 1473–1483.

Zilko, A. A. and Kurowicka, D. (2016), "Copula in a Multivariate Mixed Discrete-continuous Model," *Comput. Stat. Data Anal.*, 103.

Appendix A Difference operator notation

Since the difference operator notation can be easily confusing, it is useful to adopt the convention below. The notation has two components:

1. Whenever the $\triangle$ operators are applied to a function, an indexing is used to make the domain of the function clear.

2. A dot marks the position of the variables that are being differenced.

Here are some examples to illustrate the use of that notation.

- Consider a function $g(x)$ where x is a scalar. Then $\triangle_a^b g_x(\cdot)$ defines

$$\triangle_a^b g_x(\cdot) := g(b) - g(a)$$

- Consider a function $g(x, y)$ where both x and y are scalars. By $\triangle_a^b g_{x,y}(\cdot, z)$ we mean that the differencing is only applied to x while the second argument is fixed at $y = z$, that is

$$\triangle_a^b g_{x,y}(\cdot, z) := g(b, z) - g(a, z)$$

- Consider a function $g(\boldsymbol{x})$ where $\boldsymbol{x}$ is two-dimensional. By $\triangle_{\boldsymbol{a}}^{\boldsymbol{b}} g_{\boldsymbol{x}}(\cdot)$, we mean

$$\begin{aligned} \triangle_{\boldsymbol{a}}^{\boldsymbol{b}} g_{\boldsymbol{x}}(\cdot) &= \triangle_{a_1}^{b_1} \triangle_{a_2}^{b_2} g_{\boldsymbol{x}}(\cdot) \\ &= \triangle_{a_1}^{b_1} (g_{x_1,x_2}(\cdot, b_2) - g_{x_1,x_2}(\cdot, a_2)) \\ &= g(b_1, b_2) - g(b_1, a_2) - g(a_1, b_2) + g(a_1, a_2) \end{aligned}$$

- Consider a function $g(\boldsymbol{x}, \boldsymbol{y})$. If the differencing is applied to $\boldsymbol{y}$ and not $\boldsymbol{x}$, and if $\boldsymbol{y}$ is two-dimensional, then $\triangle_{\boldsymbol{a}}^{\boldsymbol{b}} g_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x}, \cdot)$ means

$$\begin{aligned} \triangle_{\boldsymbol{a}}^{\boldsymbol{b}} g_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x}, \cdot) &:= \triangle_{a_1}^{b_1} \triangle_{a_2}^{b_2} g_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x}, \cdot) \\ &= \triangle_{a_1}^{b_1} (g_{\boldsymbol{x},y_1,y_2}(\boldsymbol{x}, \cdot, b_2) - g_{\boldsymbol{x},y_1,y_2}(\boldsymbol{x}, \cdot, b_2)) \\ &= g(\boldsymbol{x}, b_1, b_2) - g(\boldsymbol{x}, a_1, b_2) - g(\boldsymbol{x}, a_2, b_1) + g(\boldsymbol{x}, a_1, a_2) \end{aligned}$$

ACCEPTED MANUSCRIPT

Appendix B Deriving the likelihood and the conditional density

This appendix deals with densities defined with respect to mixed measures. Such densities are formally defined by Radon-Nikodym derivatives. In particular, we obtain the joint density Eq. (1) of $\boldsymbol{X}$ and $\boldsymbol{U}$ and the corresponding mixed measure. We then show how to obtain the closed-form formulas for the densities Eq. (2) and Eq. (3), and their corresponding mixed measures, from the density Eq. (1).

We need the following three elementary lemmas to obtain the results. They are likely to be known in the literature, but we include their proofs for completeness.

Lemma 1. *Let $F(\boldsymbol{x}, \boldsymbol{y})$ be the distribution function of an absolutely continuous random vector $(\boldsymbol{X}', \boldsymbol{Y}')'$ where $\boldsymbol{x} \in \mathbb{R}^k$ and $\boldsymbol{y} \in \mathbb{R}^p$. Then,*

$$\frac{\partial^k F(\boldsymbol{x}, \boldsymbol{y})}{\partial x_1 \cdots \partial x_k} = F(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x}),$$

where $F(\boldsymbol{y}|\boldsymbol{x})$ and $f(\boldsymbol{x})$ are respectively the distribution function of $\boldsymbol{Y}$ conditional on $\boldsymbol{X} = \boldsymbol{x}$ and the density of $\boldsymbol{X}$. Similarly, in an obvious notation,

$$\frac{\partial^p F(\boldsymbol{x}, \boldsymbol{y})}{\partial y_1 \cdots \partial y_p} = F(\boldsymbol{x}|\boldsymbol{y}) f(\boldsymbol{y}).$$

Proof. The identity comes from

$$\frac{\partial^p}{\partial y_1 \cdots \partial y_p} F(\boldsymbol{y}|\boldsymbol{x}) = f(\boldsymbol{y}|\boldsymbol{x}) = \frac{f(\boldsymbol{y}, \boldsymbol{x})}{f(\boldsymbol{x})} = \frac{\frac{\partial^{p+k} F(\boldsymbol{x}, \boldsymbol{y})}{\partial y_1 \cdots \partial y_p \partial x_1 \cdots \partial x_k}}{f(\boldsymbol{x})} = \frac{\partial^p}{\partial y_1 \cdots \partial y_p} \left(\frac{\frac{\partial^k F(\boldsymbol{x}, \boldsymbol{y})}{\partial x_1 \cdots \partial x_k}}{f(\boldsymbol{x})} \right).$$

□

The next lemma is an immediate consequence of the previous lemma.

Lemma 2. *Let $f(\boldsymbol{x}, \boldsymbol{y})$ be the density of an absolutely continuous random vector*

ACCEPTED MANUSCRIPT

$(\boldsymbol{X}', \boldsymbol{Y}')'$ *where* $\boldsymbol{x} \in \mathbb{R}^k$ *and* $\boldsymbol{y} \in \mathbb{R}^p$ *then*

$$\int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} f(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}x_1 \dots \mathrm{d}x_k = \triangle_{a_1}^{b_1} \cdots \triangle_{a_k}^{b_k} F_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|.) f_{\boldsymbol{X}}(.)$$

where $F(\boldsymbol{y}|\boldsymbol{x})$ *and* $f(\boldsymbol{x})$ *are respectively the conditional distribution function of* $\boldsymbol{Y}$ *on* $\boldsymbol{X} = \boldsymbol{x}$ *and the density of* $\boldsymbol{X}$*.*

Proof. Write the density function

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}) &= \frac{\partial^{p+k} F(\boldsymbol{x}, \boldsymbol{y})}{\partial y_1 \cdots \partial y_p \partial x_1 \cdots \partial x_k} \\ &= \frac{\partial^k}{\partial x_1 \cdots \partial x_k} \left(\frac{\partial^p F(\boldsymbol{x}, \boldsymbol{y})}{\partial y_1 \cdots \partial y_p} \right) \\ &= \frac{\partial^k}{\partial x_1 \cdots \partial x_k} (F(\boldsymbol{x}|\boldsymbol{y}) f(\boldsymbol{y})) \end{aligned}$$

where the last line follows from the previous lemma. The desired result follows by an application of the fundamental theorem of calculus. □

Lemma 3. *Suppose that* U *is uniformly distributed on* $[0, 1]$*.*

(i) Suppose that X *is a univariate random variable with CDF* $F(x)$ *that has an inverse and a density* $f(x)$*. Then,* $\mathrm{d}u \delta_{F^{-1}(u)}(\mathrm{d}x) = \delta_{F(x)}(\mathrm{d}u) f(x) \mathrm{d}x$*, where* $\mathrm{d}u, \mathrm{d}x$ *are Lebesgue measures.*

(ii) Suppose that X *is a discrete univariate random variable with support on the discrete set* $I = \{x\}$*. Then,* $\mathrm{d}u \delta_{\{F(x^-) \leq u < F(x)\}}(\mathrm{d}x) = \mathcal{I}\{u : F(x^-) \leq u < F(x)\} \mathrm{d}u \delta_I(\mathrm{d}x)$

The proofs of parts (i) and (ii) are in Section S3.

Suppose that the indices $\mathcal{M}_{\mathcal{C}}$ correspond to the continuous random variables, the indices $\mathcal{M}_{\mathcal{D}}$ to the discrete random variables and the indices $\mathcal{M}_{\mathcal{J}}$ to a mixture of discrete and continuous random variables. We define the joint density of $\boldsymbol{X}$ and $\boldsymbol{U}$

 ACCEPTED MANUSCRIPT

as

$$f(\boldsymbol{x},\boldsymbol{u}) := c(\boldsymbol{u}) \prod_{j\in\mathcal{M}_\mathcal{C}} \mathcal{I}(u_j = F_j(x_j)) \prod_{j\in\mathcal{M}_\mathcal{D}} \mathcal{I}(F_j(x_j^-) \leq u_j < F_j(x_j))\times$$
$$\prod_{j\in\mathcal{M}_\mathcal{J}} \big(\mathcal{I}(u_j = F_j(x_j)) + \mathcal{I}(F_j(x_j^-) \leq u_j < F_j(x_j))\big) \tag{14}$$

with respect to the measure

$$\mathrm{d}\boldsymbol{u} \prod_{j\in\mathcal{M}_\mathcal{C}} \delta_{F_j^{-1}(u_j)}(\mathrm{d}x_j) \prod_{j\in\mathcal{M}_\mathcal{D}} \delta_{F_j(x_j^-)\leq u_j<F_j(x_j)}(\mathrm{d}x_j)\times$$
$$\prod_{j\in\mathcal{M}_\mathcal{J}} \big(\mathcal{I}(u_j = F_j(x_j))\mathrm{d}x_j + \mathcal{I}(F_j(x_j^-) \leq u_j < F_j(x_j))\delta_{F_j(x_j^-)\leq u_j<F_j(x_j)}(\mathrm{d}x_j)\big) \tag{15}$$

Lemma 4. *(i) Eq. (1) gives the joint density of $\boldsymbol{X}$ and $\boldsymbol{U}$ at a given value $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{U} = \boldsymbol{u}$.*

(ii) Eq. (2) is the marginal density of $\boldsymbol{X}$ at $\boldsymbol{X} = \boldsymbol{x}$.

(iii) Eq. (3) is the conditional density of $\boldsymbol{U}_{\mathcal{D}(\boldsymbol{x})}$ given $\boldsymbol{X} = \boldsymbol{x}$.

Proof. Part (i) follows directly from Eq. (14) and Eq. (15). Part (ii) follows by integrating out $\boldsymbol{u}$ using Lemma 2. Part (iii) follows from Lemma 3. □

