Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Pair Copula Constructions for Multivariate Discrete Data

Anastasios Panagiotelis , Claudia Czado & Harry Joe

To cite this article: Anastasios Panagiotelis , Claudia Czado & Harry Joe (2012) Pair Copula Constructions for Multivariate Discrete Data, Journal of the American Statistical Association, 107:499, 1063-1072, DOI: 10.1080/01621459.2012.682850

To link to this article: https://doi.org/10.1080/01621459.2012.682850

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Pair Copula Constructions for Multivariate Discrete Data

Anastasios PANAGIOTELIS, Claudia CZADO, and Harry JOE

Multivariate discrete response data can be found in diverse fields, including econometrics, finance, biometrics, and psychometrics. Our contribution, through this study, is to introduce a new class of models for multivariate discrete data based on pair copula constructions (PCCs) that has two major advantages. First, by deriving the conditions under which any multivariate discrete distribution can be decomposed as a PCC, we show that discrete PCCs attain highly flexible dependence structures. Second, the computational burden of evaluating the likelihood for an m-dimensional discrete PCC only grows quadratically with m. This compares favorably to existing models for which computing the likelihood either requires the evaluation of 2^m terms or slow numerical integration methods. We demonstrate the high quality of inference function for margins and maximum likelihood estimates, both under a simulated setting and for an application to a longitudinal discrete dataset on headache severity. This article has online supplementary material.

KEY WORDS: D-vine; Inference function for margins; Longitudinal data; Model selection; Ordered probit regression.

1. INTRODUCTION

Over the last 20 years, copula models (for an introduction, see Joe 1997 and Nelsen 2006 and references therein) have been extensively applied to a variety of areas. These include, but are not limited to, finance, marketing, healthcare utilization, survival analysis, genetics, and hydrology—publications in these areas can be found through asearch engine, such as Web of Knowledge (*http://wokinfo.com/*). The theorem of Sklar (1959) underpins copula modeling and states that there is a copula C such that

$$F(y_1,\ldots,y_m) = C(F_1(y_1),\ldots,F_m(y_m)) \qquad (1.1)$$

if F is the joint cumulative distribution function (cdf) of an m-dimensional random vector $\mathbf{Y} := (Y_1, Y_2, \ldots, Y_m)'$ with marginal distribution functions F_j for $j = 1, 2, \ldots, m$. The copula C is unique for continuous $\mathbf{Y}$, but is only unique over the Cartesian product of the ranges of the marginal distribution functions when $\mathbf{Y}$ is discrete (for more on this issue, see Genest and Nešlehová 2007). This nonuniqueness does not preclude the use of parametric copulas for modeling discrete data (e.g., see Joe 1997; Song, Li, and Yuan 2009). Three desirable properties of a copula are: (1) *flexibility* to model multivariate data exhibiting a wide variety of dependence structures, (2) straightforward and computationally feasible *estimation* of model parameters, and (3) applicability to *higher-dimensional* problems via an ability to introduce sparsity. To date, existing approaches for copula modeling of discrete data fail to adequately satisfy at least one of these three properties.

For copulas with a closed analytical form, the probability mass function (pmf) is obtained by taking 2^m finite differences of the copula:

$$\begin{aligned}\Pr(\mathbf{Y}=\mathbf{y}) &= \sum_{i_1=0,1}\cdots\sum_{i_m=0,1}(-1)^{i_1+\cdots+i_m} \\ &\quad \Pr(Y_1 \le y_1 - i_1,\ldots,Y_m \le y_m - i_m) \\ &= \sum_{i_1=0,1}\cdots\sum_{i_m=0,1}(-1)^{i_1+\cdots+i_m} C \\ &\quad \times (F_1(y_1-i_1),\ldots,F_m(y_m-i_m)); \qquad (1.2)\end{aligned}$$

for notational convenience, we assume here, and throughout the article where $\mathbf{Y}$ is discrete, that $\mathbf{Y} \in \mathbb{N}^m$, where $\mathbb{N}$ is the set of natural numbers. We note, however, that the models and algorithms discussed in this article can be modified and applied to variables with any discrete domain. Maximum likelihood estimation is infeasible for high-dimensional data since the computational burden of evaluating the pmf grows exponentially with m. As a result, many applications of copula modeling to discrete data have been low-dimensional and used copulas that are fast to compute (see Nikoloulopoulos and Karlis 2008, Nikoloulopoulos and Karlis 2009, and Li and Wong 2011). The copulas used in these articles, however, achieve limited dependence characteristics.

Elliptical copulas have a pmf given by the following multidimensional integral:

$$\begin{aligned}&\Pr(Y_1 = y_1,\ldots,Y_m = y_m) \\ &= \int_{\Phi_1^{-1}(F_1^-)}^{\Phi_1^{-1}(F_1^+)} \cdots \int_{\Phi_1^{-1}(F_m^-)}^{\Phi_1^{-1}(F_m^+)} \phi_m(\psi_1,\ldots,\psi_m;\Gamma)\,d\psi_1\ldots d\psi_m\,. \end{aligned} \qquad (1.3)$$

Here, $\phi_m(\cdot,\Gamma)$ denotes the probability density function (pdf) of an m-dimensional elliptical distribution with location 0 and scale given by the correlation matrix Γ, Φ_1^{-1} denotes the inverse cdf of the univariate margins of the same elliptical distribution, $F_j^+ := \Pr(Y_j \le y_j)$ and $F_j^- := \Pr(Y_j \le y_j - 1)$.

Anastasios Panagiotelis is Lecturer, Department of Econometrics and Business Statistics, Monash University, Victoria 3145, Australia (E-mail: *anastasios.panagiotelis@monash.edu*). Claudia Czado is Professor, Zentrum Mathematik, Technische Universität München, Munich, Germany (E-mail: *cczado@ma.tum.de*). Harry Joe is Professor, Department of Statistics, University of British Columbia, Vancouver, BC, V6T 1Z2 (E-mail: *harry.joe@ubc.ca*).
Anastasios Panagiotelis acknowledges the support of the Alexander von Humboldt Foundation, Claudia Czado is partially supported by the German Research Foundation grant (CZ86_1_3), and Harry Joe is supported by an NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery grant. The numerical computations were performed on a Linux cluster supported by a DFG (Deutsche Forschungsgemeinschaft: German Research Foundation) grant INST 95/919-1 FUGG. The authors also acknowledge the Associate Editor and an anonymous referee for helpful comments.

Journal of the American Statistical Association
September 2012, Vol. 107, No. 499, Theory and Methods
DOI: 10.1080/01621459.2012.682850

Elliptical copula models for discrete data can be estimated by maximum likelihood, where the integral in Equation (1.3) is approximated numerically or by Bayesian methods [Smith and Khaled (forthcoming) and references therein]. However, these methods are computationally intensive and may not scale easily to higher dimensions. Zhao and Joe (2005) provided a faster alternative to estimate the parameters based on composite likelihood methods; however, this comes at the cost of statistical efficiency when compared with maximum likelihood estimation.

Our contribution is to develop a new discrete analogue to vine pair copula constructions (PCCs) that satisfies the three desirable properties outlined above. First, vine PCCs are highly flexible since any multivariate discrete distribution can be decomposed as a vine PCC, under a set of conditions outlined in Section 3. In practice, for approximations when true data-generating processes are unknown, highly flexible models can be constructed by using asymmetric copulas as bivariate building blocks. We demonstrate this flexibility by comparing the joint probabilities of multivariate distributions constructed using different pair copulas and by demonstrating that a discrete vine PCC gives a better approximation to a discretized maximum of Gaussian autoregressions compared with the Gaussian copula. Second, since computation of the pmf for a discrete vine PCC only requires $2m(m-1)$ bivariate copula function evaluations, compared with 2^m multivariate copula evaluations for the finite-difference approach, estimation by maximum likelihood or by a two-step inference function for margins (IFM) approach (Joe 1997), and the bootstrapping of standard errors under both these methods are feasible even for higher dimensions. These methods are either computationally or statistically more efficient than methods used for the feasible estimation of competing models. Third, model selection techniques can be used to identify conditional independence enabling the construction of more parsimonious vine PCC models. Given this ability to exploit sparsity in a vine framework, as well as the computational advantages outlined above, our approach has the potential to be used in truly high-dimensional settings.

The following is a brief summary of the remainder. Section 2 reviews the literature on vine PCCs for continuous margins, providing background for the discrete vine PCC introduced in Section 3. Section 4 discusses estimation and inference by comparing different methods in simulated examples. An application to intraday data on headache severity for a pooled dataset of 135 patients is covered in Section 5. In Section 6, we review the major contributions and point to potential areas of future research.

2. PCCs IN THE CONTINUOUS-DATA CASE

Throughout the remainder, we will use the following notations. Pmfs and pdfs are denoted by f_Z and cdfs by F_Z, with the subscript in all cases denoting a random variable or vector. If $\mathbf{Z} := (Z_1, Z_2, \ldots, Z_m)'$ is a vector, marginal and conditional distributions are also abbreviated as $F_j := F_{Z_j}$ or $f_j := f_{Z_j}$, $F_{jk} := F_{Z_j,Z_k}$, $F_{j|k} := F_{Z_j|Z_k}$, etc. No subscript on f or F indicates the full multivariate distribution.

We now outline the steps required to decompose a high-dimensional density into a representation that is the product of bivariate pair copula building blocks; more details are available in Czado (2010). For a continuous random vector $\mathbf{X} := (X_1, X_2, \ldots, X_m)$ with realization $\mathbf{x} := (x_1, x_2, \ldots, x_m)$, a PCC is derived by starting with the following decomposition:

$$f(x_1, \ldots, x_m) = f_{1|2,\ldots,m}(x_1|x_2, \ldots, x_m) \times f_{2|3,\ldots,m}(x_2|x_3, \ldots, x_m), \ldots, f_m(x_m). \tag{2.1}$$

Each term on the right-hand side is of the general form $f_{X_j|\mathbf{V}}(x_j|\mathbf{v})$, where x_j is a scalar element of $\mathbf{x}$ and $\mathbf{V}$ is a subset of $\mathbf{X}$. Using Sklar's theorem and assuming that multivariate and marginal densities exist, we can show that any such univariate conditional density can be decomposed into the product of a bivariate copula density and another univariate conditional density, with the conditioning set reduced by one element. Letting V_h be any scalar element of $\mathbf{V}$ and $\mathbf{V}_{\backslash h}$ its complement, with X_j not an element of $\mathbf{V}$,

$$f_{X_j|\mathbf{V}} = f_{X_j,V_h|\mathbf{V}_{\backslash h}} / f_{V_h|\mathbf{V}_{\backslash h}}. \tag{2.2}$$

Differentiating Equation (1.1) for $X_j, V_h|\mathbf{V}_{\backslash h}$ and applying to the numerator leads to

$$\begin{aligned} f_{X_j|\mathbf{V}} &= \left\{c_{X_j,V_h|\mathbf{V}_{\backslash h}}\left(F_{X_j|\mathbf{V}_{\backslash h}}, F_{V_h|\mathbf{V}_{\backslash h}}\right) f_{X_j|\mathbf{V}_{\backslash h}} f_{V_h|\mathbf{V}_{\backslash h}}\right\} / f_{V_h|\mathbf{V}_{\backslash h}} \\ &= c_{X_j,V_h|\mathbf{V}_{\backslash h}}\left(F_{X_j|\mathbf{V}_{\backslash h}}, F_{V_h|\mathbf{V}_{\backslash h}}\right) f_{X_j|\mathbf{V}_{\backslash h}}, \end{aligned} \tag{2.3}$$

where

$$c_{X_j,V_h|\mathbf{V}_{\backslash h}}(u_1, u_2) = \frac{\partial^2 C_{X_j,V_h|\mathbf{V}_{\backslash h}}(u_1, u_2)}{\partial u_1 \partial u_2} \tag{2.4}$$

is a pair copula density for X_j and V_h, conditional on $\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h}$. This expression can be applied recursively to each term in (2.1) until the multivariate density is decomposed into the product of $m(m-1)/2$ bivariate copulas and the product of the marginal densities. The arguments of the pair copulas are conditional distribution functions and can be evaluated (see Joe 1996) using expressions of the form:

$$F_{X_j|V_h,\mathbf{V}_{\backslash h}}(x_j|v_h, \mathbf{v}_{\backslash h}) = \frac{\partial C_{X_j,V_h|\mathbf{V}_{\backslash h}}(F_{X_j|\mathbf{V}_{\backslash h}}(x_j|\mathbf{v}_{\backslash h}), F_{V_h|\mathbf{V}_{\backslash h}}(v_h|\mathbf{v}_{\backslash h}))}{\partial F_{V_h|\mathbf{V}_{\backslash h}}(v_h|\mathbf{v}_{\backslash h})}. \tag{2.5}$$

Since V_h can be any element of $\mathbf{V}$ in (2.2), there are many ways in which a density can be decomposed in this manner. However, it is practical to find a set of decompositions whereby the arguments of the copula densities can be computed using pair copulas that are defined elsewhere in the PCC so that the expression in (2.5) can be used. The different ways in which a density can be decomposed to satisfy this condition were organized and summarized by Bedford and Cooke (2001, 2002) in terms of graphical models called vines. A specific family, known as the D-vine, will be covered in detail in Section 3.

Kurowicka and Cooke (2006) and Aas et al. (2009) were the first to exploit this decomposition to construct a statistical model where each bivariate building block follows some parametric copula. This imposes a restriction that the functional form of each conditional bivariate copula does not depend on the values of the conditioning set, although the values of these bivariate conditional copulas will still depend on the conditioning set through their arguments. An overview of recent developments in the vine copula literature in the continuous-data case can be found in Kurowicka and Joe (2011).

3. DISCRETE VINE PCCs

3.1 Discrete Vine PCCs

In this section, we introduce vine PCCs for discrete margins. Using similar notation to Section 2, we can decompose a pmf as

$$\begin{aligned}\Pr(Y_1 = y_1, \ldots, Y_m = y_m) &= \Pr(Y_1 = y_1 | Y_2 = y_2, \ldots, Y_m = y_m) \\ &\quad \times \Pr(Y_2 = y_2 | Y_3 = y_3, \ldots, Y_m = y_m) \times \cdots \\ &\quad \times \Pr(Y_m = y_m).\end{aligned} \tag{3.1}$$

This expression is the discrete analogue to the decomposition in Equation (2.1). Each term on the right-hand side of Equation (3.1) has the form $\Pr(Y_j = y_j | \mathbf{V} = \mathbf{v})$, where y_j is a scalar element of $\mathbf{y}$ and $\mathbf{v}$ is a subset of $\mathbf{y}$. In a similar fashion to the continuous-data case, we choose a single element of $\mathbf{v}$ and obtain the following discrete analogue to Equation (2.2):

$$\Pr(Y_j = y_j | \mathbf{V} = \mathbf{v}) = \frac{\Pr(Y_j = y_j, V_h = v_h | \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})}{\Pr(V_h = v_h | \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})} \tag{3.2}$$

$$\begin{aligned}&= \left\{ \sum_{i_j=0,1} \sum_{i_h=0,1} (-1)^{i_j+i_h} \right. \\ &\quad \left. \times \Pr(Y_j \le y_j - i_j, V_h \le v_h - i_h | \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h}) \right\} \\ &\quad \Big/ \{\Pr(V_h = v_h | \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})\}.\end{aligned} \tag{3.3}$$

The bivariate conditional probability in the numerator of (3.3) can be expressed in terms of a copula so that (3.3) becomes

$$\begin{aligned}&= \left\{ \sum_{i_j=0,1} \sum_{i_h=0,1} (-1)^{i_j+i_h} C_{Y_j, V_h | \mathbf{V}_{\setminus h}} \big(F_{Y_j | \mathbf{V}_{\setminus h}}(y_j - i_j | \mathbf{v}_{\setminus h}), \right. \\ &\quad \left. F_{V_h | \mathbf{V}_{\setminus h}}(v_h - i_h | \mathbf{v}_{\setminus h}) \big) \right\} \Big/ \{\Pr(V_h = v_h | \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})\},\end{aligned} \tag{3.4}$$

where $F_{A|B}(a|b)$ is generic notation for the distribution function of $\Pr(A \le a | B = b)$. Equation (3.4) is a discrete analogue to (2.3) since it can be applied recursively to (3.1) to decompose a multivariate pmf into bivariate copula building blocks. Finally, the arguments of the copula functions in Equation (3.4) have the form of:

$$\begin{aligned}F_{Y_j | V_h, \mathbf{V}_{\setminus h}}(y_j | v_h, \mathbf{v}_{\setminus h}) &= \big[C_{Y_j, V_h | \mathbf{V}_{\setminus h}} \big(F_{Y_j | \mathbf{V}_{\setminus h}}(y_j | \mathbf{v}_{\setminus h}), F_{V_h | \mathbf{V}_{\setminus h}}(v_h | \mathbf{v}_{\setminus h}) \big) \\ &\quad - C_{Y_j, V_h | \mathbf{V}_{\setminus h}} \big(F_{Y_j | \mathbf{V}_{\setminus h}}(y_j | \mathbf{v}_{\setminus h}), F_{V_h | \mathbf{V}_{\setminus h}}(v_h - 1 | \mathbf{v}_{\setminus h}) \big) \big] \\ &\quad / \{\Pr(Z_k = z_k | \mathbf{Z}_{\setminus k} = \mathbf{z}_{\setminus k});\end{aligned} \tag{3.5}$$

this is a discrete analogue to (2.5). For illustration, we now outline such a decomposition in detail for the three-dimensional case.

3.2 A Three-Dimensional Illustration

For $m = 3$, we decompose the pmf of the joint density,

$$\begin{aligned}\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) &= \Pr(Y_1 = y_1 |, Y_2 = y_2, Y_3 = y_3) \\ &\quad \times \Pr(Y_3 = y_3 |, Y_2 = y_2) \times \Pr(Y_2 = y_2).\end{aligned} \tag{3.6}$$

Applying (3.4) to the first term on the right-hand side with $Y_j = Y_1$, $V_h = Y_3$, $V_{\setminus h} = Y_2$ gives

$$\begin{aligned}\Pr(Y_1 = y_1 | Y_2 = y_2, Y_3 = y_3) &= \left\{ \sum_{i_1=0,1} \sum_{i_3=0,1} (-1)^{i_1+i_3} C_{13|2}(F_{1|2}(y_1 - i_1 | y_2), \right. \\ &\quad \left. F_{3|2}(y_3 - i_3 | y_2)) \right\} \Big/ \{\Pr(Y_3 = y_3 | Y_2 = y_2)\}.\end{aligned} \tag{3.7}$$

Using (3.5), the first argument of the copula function in the numerator of (3.7) is

$$F_{1|2}(y_1 - i_1 | y_2) = \frac{C_{12}(F_1(y_1 - i_1), F_2(y_2)) - C_{12}(F_1(y_1 - i_1), F_2(y_2 - 1))}{\Pr(Y_2 = y_2)}, \tag{3.8}$$

while the second argument can be expressed as

$$F_{3|2}(y_3 - i_3 | y_2)) = \frac{C_{23}(F_2(y_2), F_3(y_3 - i_3)) - C_{23}(F_2(y_2 - 1), F_3(y_3 - i_3))}{\Pr(Y_2 = y_2)}. \tag{3.9}$$

Since the denominator of (3.7) cancels with the second term on the right-hand side of (3.6), the full expression for the pmf of this three-dimensional vine PCC is given by

$$\begin{aligned}\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) &= \left\{ \sum_{i_1=0,1} \sum_{i_3=0,1} (-1)^{i_1+i_3} C_{13|2} \right. \\ &\quad \left(\frac{C_{12}(F_1(y_1 - i_1), F_2(y_2)) - C_{12}(F_1(y_1 - i_1), F_2(y_2 - 1))}{F_2(y_2) - F_2(y_2 - 1)}, \right. \\ &\quad \left. \left. \frac{C_{23}(F_2(y_2), F_3(y_3 - i_3)) - C_{23}(F_2(y_2 - 1), F_3(y_3 - i_3))}{F_2(y_2) - F_2(y_2 - 1)} \right) \right\} \\ &\quad [F_2(y_2) - F_2(y_2 - 1)].\end{aligned} \tag{3.10}$$

3.3 Discussion of Discrete Vines

As was the case for continuous data, the decomposition can motivate flexible statistical models by assuming that a single copula can be specified for the conditional distribution of $Y_j, V_h | \mathbf{V}_{\setminus h}$ for all possible values of $\mathbf{V}_{\setminus h}$. Since Sklar's theorem states that the copula function is not unique in the discrete case, this assumption is potentially less restrictive in the discrete case than in the continuous-data case. We now investigate in more detail the restrictions that this assumption imposes.

For $\mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h}$, Sklar's theorem defines a unique copula with a domain given by the Cartesian product of the ranges of the cdf's of $Y_j|\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h}$ and the range of the cdf's of $V_h|\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h}$. If a single copula, denoted $C_{Y_j,V_h|\mathbf{V}_{\backslash h}}$, exists over all possible values of $\mathbf{V}_{\backslash h}$, then it will be unique over the union of these domains. Let $a_{(1)} < a_{(2)} < \cdots < a_{(\kappa_1)}$ be, in increasing order, the distinct points corresponding to the ranges of $F_{Y_j|\mathbf{V}_{\backslash h}}$, and let $b_{(1)} < b_{(2)} < \cdots < b_{(\kappa_2)}$ correspond to the ranges of $F_{V_h|\mathbf{V}_{\backslash h}}$. Also, let $a_{(0)} = b_{(0)} = 0$ and $a_{(\kappa_1+1)} = b_{(\kappa_2+1)} = 1$. Letting $a_{y_j|\mathbf{v}_{\backslash h}} := \Pr(Y_j \le y_j|\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h})$, $b_{v_h|\mathbf{v}_{\backslash h}} := \Pr(V_h \le v_h|\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h})$, and $p_{y_j,v_h|\mathbf{v}_{\backslash h}} := \Pr(Y_j \le y_j, V_h \le v_h|\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h})$, the constraints

$$C_{Y_j,V_h|\mathbf{V}_{\backslash h}}\left(a_{y_j|\mathbf{v}_{\backslash h}}, b_{v_h|\mathbf{v}_{\backslash h}}\right) = p_{y_j,v_h|\mathbf{v}_{\backslash h}}, \tag{3.11}$$

$$C_{Y_j,V_h|\mathbf{V}_{\backslash h}}\left(a_{y_j|\mathbf{v}_{\backslash h}}, 1\right) = a_{y_j|\mathbf{v}_{\backslash h}}, \quad C_{Y_j,V_h|\mathbf{V}_{\backslash h}}\left(1, b_{v_h|\mathbf{v}_{\backslash h}}\right) = b_{v_h|\mathbf{v}_{\backslash h}}, \tag{3.12}$$

must be satisfied for all y_j, v_h, and $\mathbf{v}_{\backslash h}$. This leads to 3κ constraints, where κ is the product of the cardinalities of the sets of possible values for Y_j, V_h, and $\mathbf{V}_{\backslash h}$.

Theorem 1 (Existence of a constant conditional copula). Consider the conditional distributions of $(Y_j, V_h)|\mathbf{V}_{\backslash h} = \mathbf{v}_{\backslash h}$ over all possible values of $\mathbf{v}_{\backslash h}$. Let $a_{(1)} < \cdots < a_{(\kappa_1)}$ be the distinct points corresponding to the ranges of $F_{Y_j|\mathbf{V}_{\backslash h}}(\cdot|\mathbf{v}_{\backslash h})$, and let $b_{(1)} < \cdots < b_{(\kappa_2)}$ similarly correspond to the ranges of $F_{V_h|\mathbf{V}_{\backslash h}}(\cdot|\mathbf{v}_{\backslash h})$. A bivariate copula $C = C_{Y_j,V_h|\mathbf{V}_{\backslash h}}$, constant over $\mathbf{v}_{\backslash h}$, exists if a solution to the 3κ constraints in (3.11) and (3.12) exists where all R_{jk} are nonnegative. Here, $R_{jk} := C(a_{(j)}, b_{(k)}) + C(a_{(j-1)}, b_{(k)}) - C(a_{(j)}, b_{(k-1)}) + C(a_{(j-1)}, b_{(k-1)})$ represent $(\kappa_1 + 1)(\kappa_2 + 1)$ unknowns.

Proof. The nonnegative solution is necessary because R_{jk} must correspond to probabilities if C is a copula. Given that a function C satisfies the nonnegative rectangle condition on a rectangular grid such as the above, a simple extension to a valid copula is one that is piecewise uniform over the rectangles $[a_{(j-1)}, a_{(j)}] \times [b_{(k-1)}, b_{(k)}]$. A proof of this piecewise uniform argument can be found on page 14 of Joe (1997). □

Linear programming or a linear system reduction can be used to determine whether there is a nonnegative solution. Note that since there are $(m-1)(m-2)/2$ potential ways to choose Y_j, V_h, and $\mathbf{V}_{\backslash h}$, κ_1, κ_2, and κ will depend on the set of conditional distributions being considered. This leads to the following corollary:

Corollary 1 (Existence of a vine PCC representation for an arbitrary multivariate discrete distribution). A multivariate discrete distribution can be expressed as a vine PCC if a decomposition exists where every set of bivariate conditional distributions in this vine has a constant conditional copula, with nonnegative rectangle variables satisfying constraints of the forms (3.11) and (3.12).

To investigate this further, we consider the case where Y_j, V_h, and $V_{\backslash h}$ are all binary scalars. Let $\Pr(Y_j \le 0|V_{\backslash h} = 0) := a_{0|0} < a_{0|1} := \Pr(Y_j \le 0|V_{\backslash h} = 1)$ and $\Pr(V_h \le 0|V_{\backslash h} = 0) := b_{0|0} < b_{0|1} := \Pr(V_h \le 0|V_{\backslash h} = 1)$. Dropping the subscript on the copula for convenience, also let $\Pr(Y_j \le 0, V_h \le 0|V_{\backslash h} = 0) := C(a_{0|0}, b_{0|0}) := p_{0,0|0} < p_{0,0|1} := C(a_{0|1}, b_{0|1}) = \Pr(Y_j \le 0, V_h \le 0|V_{\backslash h} = 1)$. In this case, a necessary condition for a nonnegative rectangle solution to the problem posed above exists as long as $p_{0,0|1} - p_{0,0|0}$ does not exceed $b_{0|1} - b_{0|0} + a_{0|1} - a_{0|0}$; this follows by applying upper bounds to the probabilities of the two rectangular regions whose union is $([0, a_{0|1}] \times [0, b_{0|1}]) \backslash ([0, a_{0|0}] \times [0, b_{0|0}])$. However, if $C(a_{0|0}, b_{0|0})$ is set to its lower bound of $\max\{0, a_{0|0} + b_{0|0} - 1\}$ and $C(a_{0|1}, b_{0|1})$ is set to its upper bound of $\min\{a_{0|1}, b_{0|1}\}$, and $p_{0,0|1} - p_{0,0|0} = \min\{a_{0|1}, b_{0|1}\} - \max\{0, a_{0|0} + b_{0|0} - 1\} > b_{0|1} - b_{0|0} + a_{0|1} - a_{0|0}$, then no copula exists to describe the conditional dependence between Y_j and V_h for $V_{\backslash h} = 0$ and $V_{\backslash h} = 1$; this can happen if the differences $b_{0|1} - b_{0|0}$ and $a_{0|1} - a_{0|0}$ are small enough. Therefore, in both this specific case and more general cases, one must be careful to use a vine decomposition where the strength of dependence of the conditional bivariate distribution does not vary greatly over different values of the conditioning set.

Since the true data-generating process is never known for real data, the question of finding the vine PCC decomposition for a given multivariate distribution is rarely encountered in practice. However, the above provides some guidance to the very practical question of model selection in a vine PCC framework. Since a large number of vine PCCs is possible, one must find a structure with fairly constant dependence over different values of the conditioning set. One way to do this is to identify conditional independence among pairs where present. Alternatively, there may be some intuitive reason for selecting a specific vine—one such example being a D-vine for longitudinal data, to which we now turn our attention.

3.4 Discrete D-Vine

For the remainder of the article, we focus on a specific vine, known as a D-vine, which is depicted in Figure 1. An algorithm for computing the pmf of a discrete D-vine of arbitrary dimension is outlined in the Appendix.

Any vine is characterized by $m - 1$ trees, denoted T_j, for $j = 1, \ldots, m - 1$. The jth tree is made up of nodes, denoted N_j, and edges that join these nodes, denoted E_j. In the first tree of a D-vine, the margins are ordered, and the edges simply join adjacent nodes, yielding $E_1 := \{12, 23, 34, 45\}$. The edges on the first tree become the nodes on the second tree and, in general, $N_{j+1} := E_j$. The edges of trees $T_2, \ldots, T_{m-1}$ also connect adjacent nodes. Any element shared by two nodes will be in the conditioning set of the edge joining them. For example, the edge joining node 12 and node 23 is 13|2, while the edge joining 24|3 and 35|4, will be 25|34. The pair copulas that make

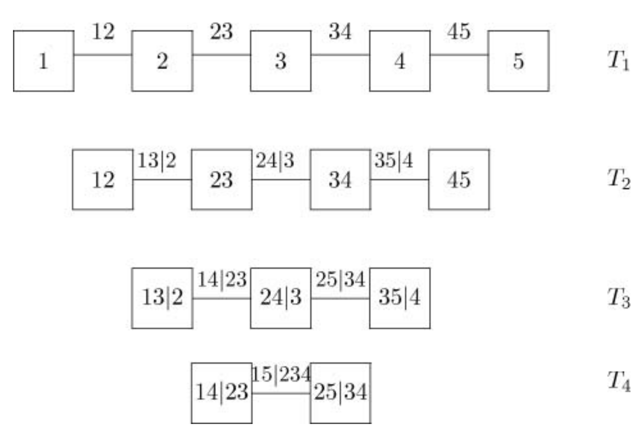

Figure 1. A five-dimensional D-vine, with T_j denoting the jth tree for $j = 1, 2, 3, 4$.

Table 1. Summary of parameter values for the five-dimensional D-vine with Bernoulli margins discussed in Section 3.5 and used for the simulation study in Section 4.2

Case	Marginal probabilities	Dependence	Model	Pair copulas
1	*Low*: $\Pr(Y_j = 0) = 0.3$	*Low*: Kendall's $\tau = 0.3, 0.2, 0.1, 0.05$	1A	All Gaussian
	for $j = 1, \ldots, 5$	for pair copulas corresponding	1B	All MTCJ/Clayton
		to trees $T_1, \ldots, T_4$, respectively	1C	All Gumbel
2	*High*: $\Pr(Y_j = 0) = 0.7$	*Low*: Kendall's $\tau = 0.3, 0.2, 0.1, 0.05$	2A	All Gaussian
	for $j = 1, \ldots, 5$	for pair copulas corresponding	2B	All MTCJ/Clayton
		to trees $T_1, \ldots, T_4$, respectively	2C	All Gumbel
3	*Low*: $\Pr(Y_j = 0) = 0.3$	*High*: Kendall's $\tau = 0.7, 0.4, 0.3, 0.2$	3A	All Gaussian
	for $j = 1, \ldots, 5$	for pair copulas corresponding	3B	All MTCJ/Clayton
		to trees $T_1, \ldots, T_4$, respectively	3C	All Gumbel
4	*High*: $\Pr(Y_j = 0) = 0.7$	*High*: Kendall's $\tau = 0.7, 0.4, 0.3, 0.2$	4A	All Gaussian
	for $j = 1, \ldots, 5$	for pair copulas corresponding	4B	All MTCJ/Clayton
		to trees $T_1, \ldots, T_4$, respectively	4C	All Gumbel

up the corresponding PCC are simply indicated by the edges of the entire vine $\{E_1, E_2, \ldots, E_{m-1}\}$.

3.5 Flexibility of the Model

A major advantage of vine PCCs is that a wide variety of dependence structures can be modeled by selecting different copula families as building blocks. Although many dependence concepts such as tail dependence degenerate in the discrete case, selecting different bivariate copula families in a discrete D-vine has a substantial impact on the joint probabilities of the multivariate distribution. To demonstrate this point, we consider a five-dimensional D-vine with Bernoulli margins. We consider four cases, summarized in Table 1, and for each case, set all pair copulas to Gaussian, all pair copulas to MTCJ/Clayton,[2] and all pair copulas to Gumbel, giving 12 models overall. Since we are primarily interested in studying the effect of choosing different bivariate copula families in isolation, we keep the marginal probabilities and dependence as measured by Kendall's τ of the copula constant within each case. However, we provide an important caveat that the Kendall's τ of discrete data is corrected for ties and will depend on the marginal probabilities. Consequently, in contrast to the continuous-data case, the Kendall's τ of the data will not be equivalent to the Kendall's τ of the bivariate copulas (see Denuit and Lambert 2005).[3]

The joint probabilities at four possible realizations of the data are summarized for the 12 different models in Table 2. These results show how in applications where both the marginal and the joint probabilities of zeros in the data are relatively high, it may be more appropriate to use Gumbel copulas than MTCJ/Clayton copulas. We note that quantities that depend on these joint probabilities, such as log odds ratios, would also vary when different copula families are chosen as building blocks. Furthermore, there is a similar pattern when discrete random variables take more than two values; for example, different choices of pair copulas can affect how joint tail probabilities differ for the multivariate ordered probit model. The results in Table 2 demonstrate the flexibility of D-vine PCCs but also beg the question of how pair copula building blocks should be selected. We turn our attention to this issue in Section 5.

To further investigate the flexibility of vine PCCs, and to motivate the use of asymmetric copulas, we consider the following discretized maximum of Gaussian autoregressions. We let $Z_{i1}, Z_{i2}, \ldots, Z_{im}$ for $i = 1, \ldots, I$ be I independent AR(1) processes of length m with lag 1 correlation ρ and Gaussian errors. We define $W_j := \max\{Z_{1j},$

Table 2. Some joint probabilities for a five-dimensional D-vine with Bernoulli margins

Y	All Gaussian	All MTCJ	All Gumbel	Independent
	Case 1: low $\Pr(Y_j = 0)$, low dependence			
00000	0.0377	0.0482	0.0319	0.0024
01010	0.0097	0.0102	0.0097	0.0132
10101	0.0157	0.0128	0.0173	0.0309
11111	0.3185	0.3835	0.2920	0.1681
	Case 2: high $\Pr(Y_j = 0)$, low dependence			
00000	0.3185	0.2672	0.3603	0.1681
01010	0.0157	0.0190	0.0144	0.0309
10101	0.0097	0.0099	0.0105	0.0132
11111	0.0377	0.0261	0.0437	0.0024
	Case 3: low $\Pr(Y_j = 0)$, high dependence			
00000	0.1648	0.1839	0.1683	0.0024
01010	0.0028	0.0026	0.0028	0.0132
10101	0.0037	0.0013	0.0047	0.0309
11111	0.5366	0.6267	0.5028	0.1681
	Case 4: high $\Pr(Y_j = 0)$, high dependence			
00000	0.5366	0.4553	0.5899	0.1681
01010	0.0037	0.0054	0.0032	0.0309
10101	0.0028	0.0023	0.0040	0.0132
11111	0.1648	0.1610	0.1706	0.0024

NOTE: The leftmost column describes the realization of **Y**; for example, "01010" denotes $Y_1 = 0$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, and $Y_5 = 0$. Here, "low $\Pr(Y_j = 0)$" refers to the case where $\Pr(Y_j = 0) = 0.3$ for all j, "high $\Pr(Y_j = 0)$" refers to the case where $\Pr(Y_j = 0) = 0.7$ for all j, "low dependence" refers to the case where Kendall's $\tau = 0.3, 0.2, 0.1, 0.05$ for pair copulas corresponding to the first, second, third, and fourth tree, respectively, and "high dependence" refers to the case where Kendall's $\tau = 0.7, 0.4, 0.3, 0.2$ for pair copulas corresponding to the first, second, third, and fourth tree, respectively. In the rightmost column, the joint probabilities for independent margins are given for comparison.

[2]Although the MTCJ copula should be attributed to Mardia, Takahashi, Cook, and Johnson (see Joe, Li, and Nikoloulopoulos 2010 for a detailed discussion), the bivariate version is more commonly referred to as the Clayton copula. We use the name MTCJ/Clayton throughout the article as a compromise.

[3]We thank a referee for pointing this out to us.

$Z_{2j}, \ldots, Z_{Ij}\}$ for $j = 1, \ldots, m$. Finally, we let $Y_j = k$ if and only if $\alpha_{k-1} \le W_j \le \alpha_k$ for $k = 1, \ldots, K$, where $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_{K-1} < \alpha_K = \infty$ are cutpoints chosen so that Y_j are marginally uniform. The multivariate discrete pmf from this construction involves multidimensional integrals but all joint probabilities can be computed via Monte Carlo simulation. It can also be approximated by a discrete D-vine based on parametric pair copulas with tail asymmetry or by a discretized normal distribution, which we use as a benchmark. Approximations can be found by finding parameters that minimize the Kullback–Leibler divergence between the D-vine model and the actual multivariate discrete distribution. Setting $m = 4$, $\rho = 0.95$, $I = 100$, and $K = 3$, the minimized Kullback–Leibler divergence when using a D-vine with all bivariate Gumbel copulas is 0.012, with a maximum absolute deviation of 0.008 in the probabilities, while the corresponding results for a discretized multivariate normal are 0.020 for Kullback–Leibler divergence and 0.026 for maximum absolute deviation in the probabilities. Also, the maximum absolute deviation of the bivariate marginal probabilities is 0.008 for the D-vine compared with 0.024 for discretized multivariate normal. Similar results are obtained for different values of m, I, and K, highlighting the flexibility of using bivariate asymmetric copulas in a vine PCC compared with elliptical alternatives. By considering a discrete D-vine with options other than the Gumbel copula, the Kullback–Leibler divergence can be decreased further. In addition, even bigger differences in the Kullback–Leibler divergence (discrete D-vine vs. discretized multivariate normal) were found for a multivariate discrete distribution based on discretizing a continuous-state Markov chain $(W_1, \ldots, W_m)$, where the transition function $F_{W_j|W_{j-1}}(\cdot|w)$ is based a copula with upper-tail dependence.

4. ESTIMATION AND SIMULATION STUDY

4.1 Estimation

Evaluating the multivariate pmf of a discrete vine PCC at a single point requires that each of the $m(m-1)/2$ bivariate copulas is evaluated at four different values for a total of $2m(m-1)$ evaluations of bivariate copula functions. Estimation can be carried out using two popular approaches: two-step IFM and full likelihood; this might not be feasible when evaluation of the pmf grows exponentially with m.

The IFM approach is covered in Joe (1997). Let the jth univariate margin be $F_j(\cdot; \boldsymbol{\lambda}_j)$, with univariate parameter $\boldsymbol{\lambda}_j$. Suppose we observe a sample $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})'$ for $i = 1, 2, \ldots, n$. In the first step, maximum likelihood estimates (MLEs) of the univariate parameters, denoted $\hat{\boldsymbol{\lambda}}_j$, are obtained for each j, ignoring dependence between the variables. In the second step, the copula parameters, denoted $\boldsymbol{\theta}$, are estimated by maximum likelihood, but with the univariate parameters fixed at $\hat{\boldsymbol{\lambda}} := (\hat{\boldsymbol{\lambda}}_1', \ldots, \hat{\boldsymbol{\lambda}}_m')'$. This corresponds to solving the vector of inference or estimating functions $\sum_{i=1}^n g(\mathbf{y}_i, \boldsymbol{\eta}) = 0$ for $\boldsymbol{\eta} := (\hat{\boldsymbol{\lambda}}', \boldsymbol{\theta}')'$, where

$$g(\mathbf{y}_i, \boldsymbol{\eta}) = \left(\frac{\partial L_1(\boldsymbol{\lambda}_1; y_{i1})}{\partial \boldsymbol{\lambda}_1}, \ldots, \frac{\partial L_m(\boldsymbol{\lambda}_m; y_{im})}{\partial \boldsymbol{\lambda}_m}, \frac{\partial L\left((\hat{\boldsymbol{\lambda}}, \boldsymbol{\theta}); \mathbf{y}_i\right)}{\partial \boldsymbol{\theta}} \right)'. \tag{4.1}$$

Here, $L(\boldsymbol{\eta}; \mathbf{y}_i)$ denotes the contribution of the ith observation to the full multivariate likelihood and $L_j(\boldsymbol{\lambda}_j; y_{ij})$ denotes the contribution of the ith observation to the jth univariate likelihood. From the theory of estimating equations, the asymptotic distribution of $n^{-1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ is normal, with covariance given by the inverse Godambe information matrix, $V = D_g^{-1} M_g (D_g^{-1})'$, where $D_g = \mathrm{E}\,[\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \partial g(\mathbf{y}_i, \boldsymbol{\eta})/\partial \boldsymbol{\eta}']$ and $M_g = \mathrm{E}[\lim_{n\to\infty} n^{-1} \sum_{i=1}^n g(\mathbf{y}_i, \boldsymbol{\eta}) g'(\mathbf{y}_i, \boldsymbol{\eta})]$. An algorithm for evaluating the Hessian matrix $\sum_{i=1}^n \partial^2 L\big((\hat{\boldsymbol{\lambda}}, \boldsymbol{\theta}); \mathbf{y}_i\big)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ for a discrete vine PCC has been developed and can be used to consistently estimate this asymptotic variance. However, there is also strong theoretical justification for using methods such as the jackknife (Joe 1997) or bootstrap (Hu and Kalbfleisch 2000) to estimate V. Since our likelihood is fast to compute, we use the bootstrap in Section 5.

Another estimation approach is the joint estimation of the univariate and copula parameters by maximum likelihood. To compute the standard errors of MLEs analytically, the cross-derivatives of the likelihood with respect to univariate and copula parameters must be obtained. These derivatives depend on the univariate models used and will usually be tedious to compute. We prefer to use the bootstrap, which provides a more general method for obtaining standard errors and, due to the form of the likelihood, is not computationally too burdensome.

4.2 Bernoulli Margins

In this section, we describe a simulation study to demonstrate our proposed methods for Bernoulli margins. We simulated 100 replications of data, each with a sample size of 300, from each of the 12 models summarized in Table 1. The algorithm for data simulation is outlined in more detail in the online supplement. Although we have 15 model parameters, there are only 32 different realizations of $\mathbf{Y}$ that can be observed, and since our sample size is quite small, in most simulated datasets, only around 20 unique values of $\mathbf{Y}$ are observed. This study, therefore, represents a highly challenging environment for estimating the discrete D-vine model. For each parameter and each model, we evaluate the parameter estimates, the relative bias, and the root mean square error averaged over 100 replications. We also compute the proportion of replications where the true parameter value falls inside a 95% bootstrapped confidence interval, but note that similar estimates could be obtained using analytical standard errors. All results are obtained using both two-step IFM and joint MLEs.

Table 3 presents results only for model 1A, but they are representative of the other models. Although joint MLEs are asymptotically efficient compared with two-step IFM, the differences summarized in Table 3, which are based on finite samples, are fairly small. This is encouraging since two-step IFM estimation is faster, particularly for more complicated marginal models. We also observe that estimating the copula parameters becomes more difficult for higher-order trees, as indicated by higher bias and root mean square error. A potential solution to this problems is to truncate the D-vine, which we consider next.

4.3 Poisson Margins, High-Dimensional

In this section, to demonstrate the applicability of our model to high dimensions, we conduct one further simulation by

Table 3. Mean, relative bias (Rel. bias), root mean square error (RMSE), and coverage of bootstrapped standard errors (Cov) for both two-step inference function for margins (IFM) and joint maximum likelihood estimates (MLEs) for model 1A described in Table 1

		Two-step IFM				Joint MLEs			
Parameter	True	Mean	Rel. bias	RMSE	Cov	Mean	Rel. bias	RMSE	Cov
τ_{12}	0.3	0.2901	−0.0330	0.0618	0.91	0.2971	−0.0096	0.0609	0.95
τ_{23}	0.3	0.2926	−0.0245	0.0614	0.90	0.2996	−0.0013	0.0597	0.94
τ_{34}	0.3	0.2850	−0.0500	0.0601	0.93	0.2921	−0.0262	0.0586	0.97
τ_{45}	0.3	0.2943	−0.0188	0.0629	0.95	0.3015	0.0050	0.0622	0.97
$\tau_{13\|2}$	0.2	0.1982	−0.0088	0.0575	0.94	0.2004	0.0019	0.0597	0.95
$\tau_{24\|3}$	0.2	0.1956	−0.0220	0.0657	0.91	0.1991	−0.0044	0.0670	0.94
$\tau_{35\|4}$	0.2	0.2001	0.0005	0.0715	0.86	0.2002	0.0009	0.0738	0.88
$\tau_{14\|23}$	0.1	0.1106	0.1059	0.0661	0.93	0.1008	0.0083	0.0663	0.96
$\tau_{25\|34}$	0.1	0.1003	0.0030	0.0615	0.96	0.0894	−0.1058	0.0634	0.96
$\tau_{15\|234}$	0.05	0.0632	0.2635	0.0601	0.95	0.0460	−0.0793	0.0663	0.96
$\Pr(Y_1 = 0)$	0.3	0.2945	−0.0184	0.0263	0.96	0.2944	−0.0188	0.0263	0.96
$\Pr(Y_2 = 0)$	0.3	0.2988	−0.0039	0.0237	0.99	0.2988	−0.0041	0.0237	0.99
$\Pr(Y_3 = 0)$	0.3	0.3022	0.0073	0.0252	0.96	0.3021	0.0070	0.0252	0.97
$\Pr(Y_4 = 0)$	0.3	0.2994	−0.0019	0.0304	0.91	0.2994	−0.0019	0.0304	0.91
$\Pr(Y_5 = 0)$	0.3	0.3039	0.0129	0.0307	0.92	0.3038	0.0125	0.0307	0.91

NOTE: Results are averaged over 100 replications of data, each having sample size 300. Coverage refers to the proportion of simulations where a 95% bootstrapped confidence interval contains the true parameter value.

generating a single dataset of sample size 1000 from a 20-dimensional model with Poisson margins, each with mean 10. This model has 181 parameters. It is advantageous in such large applications to uncover a more parsimonious dependence structure by identifying whether a subset of pair copula building blocks are equivalent to the independence copula. In the special case where independence copulas are used in trees $l + 1, \ldots, m - 2, m - 1$, we say that the D-vine is "truncated after the lth tree" [compare with Brechmann, Czado, and Aas (2012)] or, alternatively, that the level of truncation is l. When a D-vine is used to model intraday data, then truncation after the lth tree corresponds to an l-order intraday Markov structure (Smith et al. 2010).

In our 20-dimensional example, we set the pair copulas in the first two trees to Gaussian copulas with Kendall's $\tau = 0.3$ and truncate after the second tree. We estimate the model assuming a level of truncation at 1, 2, 3, and 4 as well as estimate the full model. We compute Akaike's information criterion (AIC) for each of these models and select the level of truncation by choosing the model with the lowest AIC. This model selection strategy has been shown to be an effective "quick-and-dirty" algorithm in the continuous-data case by Brechmann, Czado, and Aas (2012). The results of this procedure are summarized in Table 4.

The AIC improves drastically when going from estimation with truncation after one tree to estimation with truncation after two trees (the true model) and deteriorates slowly thereafter. To test the robustness of this result, we simulated 30 replications of the data and truncation after two vines was selected every time. Table 4 also demonstrates the reduction in computing time that can be achieved by truncation. We also note that even the full model takes roughly 8 hr to estimate on a 8 AMD Opetron 8384 Quad-Core CPU. This is relatively short, considering that an alternative based on taking finite differences would require 1000×2^{20} evaluations of the closed form of a multivariate copula to compute the likelihood function just once, while techniques based on computing rectangle probabilities would require the evaluation of a 20-dimensional integral.

5. APPLICATION TO INTRADAY HEADACHE SEVERITY DATA

In this section, we consider an application of our models and methods to a survey of sufferers of chronic headaches. The full pooled dataset includes 5609 daily observations from 135 respondents, who keep a log of the severity of their headaches, allocating a score ranging from 0 for no headache to 5 for an intense headache. The number of days reported by each respondent varies from 7 to 381 days. Data were recorded at four different times of the day, motivating a four-dimensional D-vine PCC model, with an intuitive ordering of the margins (morning–afternoon–evening–night) that allows intraday dependence in headache severity to be investigated. Covariates that measure both individual attributes of the patients and important meteorological variables were also available, motivating ordered probit regressions as marginal models. Full details on

Table 4. AIC and computing time for 20-dimensional simulated dataset with Poisson margins with mean 10 discussed in Section 4.3

Level of truncation	AIC	Computing time (seconds)
1	98,447	2759
2	**94,341**	**6975**
3	94,356	9450
4	94,369	11,924
19	94,490	32,252

NOTE: The pair copulas on the first two trees are Gaussian with Kendall's $\tau = 0.3$, while all other pair copulas are the independence copula. The sample size is 1000, and similar results were obtained for 30 replications of data generated from this model.

Table 5. Summary of estimates of copula parameters for the headache application in Section 5, with 95% bootstrapped confidence intervals also included

Parameter	Pair copula	MLEs	Lower 95% CI	Upper 95% CI
θ_{MA}	Gaussian	0.664 (0.44)	0.549 (0.36)	0.733 (0.49)
θ_{AE}	Gaussian	0.662 (0.46)	0.569 (0.37)	0.760 (0.51)
θ_{EN}	MTCJ/Clayton	2.247 (0.53)	1.546 (0.44)	3.019 (0.60)
$\theta_{\mathrm{ME}\mid\mathrm{A}}$	MTCJ/Clayton	0.423 (0.17)	0.164 (0.08)	0.653 (0.25)
$\theta_{\mathrm{AN}\mid\mathrm{E}}$	Gaussian	0.439 (0.28)	0.262 (0.17)	0.508 (0.33)
$\theta_{\mathrm{MN}\mid\mathrm{AE}}$	Gaussian	0.244 (0.16)	0.092 (0.06)	0.347 (0.22)

NOTE: Parentheses are for estimated Kendall's τ of the copula and the corresponding lower/upper confidence intervals (CIs). Here, "M" denotes morning, "A" afternoon, "E" evening, and "N" night, so $\theta_{\mathrm{MN}\mid\mathrm{AE}}$, for example, describes the dependence between headache severity in the morning and night, conditional on headache severity in the afternoon and evening.

the covariates used can be found in the online supplement and in Varin and Czado (2010).

Since we are mainly interested in intraday dependence rather than dependence between different days, we only use observations from Mondays and Thursdays. We assume that dependence in headache severity is negligible over periods longer than at least 3 days, and therefore, we only investigate dependence within a single day rather than across different days. After removing all observations where at least some entries are missing, we obtain a reduced sample of 621 four-dimensional observations. Since our emphasis here is primarily demonstrative, we do not include random effects in the model to capture individual heterogeneity that is unexplained by the covariates; however, we recognize that this is an interesting potential extension to the model.

We consider using the Gaussian, MTCJ/Clayton, and Gumbel copulas as pair copulas in the vine PCC. Since there are only six pair copulas in a four-dimensional model, it is feasible to estimate all $3^6 = 729$ possible models using MLEs and select the model with the highest maximized log-likelihood value as the "best" model. This model has a maximized log-likelihood value of -2608 and is achieved by a combination of pair copulas summarized in Table 5. Since all models have the same number of parameters, this corresponds to choosing the model according to information criteria such as the AIC and BIC (Bayesian information criterion).

To further investigate the improvement in the goodness of fit of our selected model, we conduct tests given in Vuong (1989) and Clarke (2007) for nonnested models. Both test the null hypothesis that the two models are equally valid against the alternative that the model with the higher log-likelihood is to be preferred over the other. Two separate model comparisons were made. First, we compare the "best" model with the model that had the lowest maximized value of the log-likelihood of -2977. Here, both the Vuong and the Clarke test give p-values < 0.0001, thus rejecting the null hypothesis that both models are equally valid. Second, we compare the "best" model with a model constructed using only Gaussian pair copulas, which, with a maximized log-likelihood of -2620, is the eighth best model overall. We note here that this model is numerically close but not exactly equivalent to a standard multivariate ordered probit model. When comparing the "best" model with the vine composed of only Gaussian pair copula building blocks, the Vuong and the Clarke tests yield p-values of 0.32 and < 0.0001, respectively. Overall, these comparisons provide evidence that the correct choice of pair copulas can significantly improve in-sample fit in a statistical sense. The correct choice of pair copulas will also improve out-of-sample prediction, and evidence of this can be found in a validation study provided in the online supplement.

Also reported in Table 5 are the estimates of the copula parameters with the bounds of a 95% bootstrapped confidence interval (the same results reparameterized in terms of Kendall's τ of the copula are given in parentheses). As expected, the dependence is strongest in the first tree of the D-vine. Interestingly enough, although the dependence in the last tree is the weakest, the copula parameter $\theta_{\mathrm{MN}\,|\,\mathrm{AE}}$ is significantly different from zero and there is no apparent advantage to truncating the D-vine for this model.

Table 6 summarizes the marginal parameter estimates for the ordered probit regression. The statistically significant coefficients at a 95% level of significance are highlighted in bold. In

Table 6. Summary of marginal parameter estimates for the D-vine with ordered probit margins in Section 5, based on joint estimation with pair copulas chosen as in Table 5

	Morning	Afternoon	Evening	Night
β_1	−0.27	0.16	0.31	0.27
β_2	**0.96**	0.31	0.39	0.32
β_3	**1.06**	**1.13**	**1.32**	**1.49**
β_4	0.05	−0.05	0.31	0.01
β_5	−0.35	−0.28	**−0.46**	−0.38
β_6	−0.00	**0.73**	0.24	0.20
β_7	**1.25**	**1.46**	**0.80**	**1.01**
β_8	**0.44**	**0.49**	**0.50**	**0.63**
β_9	0.01	−0.02	0.01	0.00
β_{10}	0.00	0.00	−0.00	−0.00
β_{11}	**0.05**	**0.04**	0.03	0.03
β_{12}	**3.43**	**2.92**	1.69	1.93
β_{13}	**4.37**	2.57	1.10	3.56
β_{14}	**−0.06**	**−0.05**	−0.02	−0.03
β_{15}	**−0.08**	−0.04	−0.02	**−0.06**
β_{16}	−0.12	0.12	0.14	−0.12
β_{17}	−0.17	−0.19	0.17	−0.16
β_{18}	0.38	0.20	0.12	0.22
β_{19}	−0.14	0.01	0.24	0.06

NOTE: The covariates that correspond to each of the coefficients (β's) are described in the online supplement. The figures in bold have 95% bootstrapped confidence intervals that do not contain 0.

general, a major advantage of taking dependence into account in regression modeling is that more efficient estimates are obtained, which in turn can lead to more powerful significance tests on the coefficients. To examine this issue further, we estimated the marginal models assuming intraday independence and compared these estimates with the estimates obtained by using a copula. The coefficient estimates and, in particular, the coefficients of β_5 in the morning and β_{15} in the night were significant for the copula model but not when independence was assumed. Overall, this demonstrates that our model, in addition to uncovering interesting dependence properties, can also have an impact on the inferences made on marginal coefficients in regression modeling.

6. CONCLUSIONS

Our broad aim of this article was to develop a copula modeling framework for multivariate discrete data that is flexible, easy to estimate, and applicable in high dimensions. We developed a discrete analogue to vine PCCs and derived the conditions that need to be satisfied so that any discrete multivariate distribution can be decomposed as a vine PCC where given. The flexibility of using asymmetric pair copulas as building blocks was demonstrated, and a fast algorithm for computing the pmf of a D-vine PCC with discrete margins was developed.

Although we feel that our contribution is a major step forward in the modeling of multivariate discrete data, model selection remains an important open question. In particular, algorithms must be found for selecting both vine structures and the parametric copulas used as building blocks, with a particular emphasis on identifying conditional independence where present. The general strategy here could be either to use fast simple diagnostics or, alternatively, in the spirit of Bayesian model selection, to sample from the model space using reversible-jump Markov chain Monte Carlo.

Overall, our contribution is a novel approach for modeling multivariate discrete data that is general, flexible, fast, and can be extended to higher dimensions. We pose several new research questions but simultaneously provide direction toward possible solutions. For these reasons, we believe that vine copula approaches have great potential as an area for future research and represent the way forward in the modeling of discrete multivariate data.

APPENDIX: ALGORITHM FOR COMPUTING LIKELIHOOD

This appendix has an algorithm to compute the likelihood of a D-vine; the algorithm can be generalized to C-vines and regular vines. Let $\mathbf{Y} := (Y_1, Y_2, \ldots, Y_m)'$ be a discrete-valued random m-vector. For simplicity and without loss of generality, we assume $\mathbf{Y} \in \mathbb{N}^m$. For a vector of integers $\mathbf{i}$, let $\mathbf{Y}_\mathbf{i} = (Y_i : i \in \mathbf{i})$. For notation, let $F^+_{g|\mathbf{i}} := \Pr(Y_g \le y_g | \mathbf{Y}_\mathbf{i} = \mathbf{y}_i)$, $F^-_{g|\mathbf{i}} := \Pr(Y_g \le y_g - 1 | \mathbf{Y}_\mathbf{i} = \mathbf{y}_\mathbf{i})$, and $f_{g|\mathbf{i}} := \Pr(Y_g = y_g | \mathbf{Y}_\mathbf{i} = \mathbf{y}_\mathbf{i})$. Note that $\mathbf{i}$ empty corresponds to marginal probabilities. For the copula functions evaluated at four different values, we let

$$C^{++}_{gj|\mathbf{i}} := C_{gj|\mathbf{i}}\Big(F^+_{g|\mathbf{i}}, F^+_{j|\mathbf{i}}\Big), \quad C^{+-}_{gj|\mathbf{i}} := C_{gj|\mathbf{i}}\Big(F^+_{g|\mathbf{i}}, F^-_{j|\mathbf{i}}\Big),$$
$$C^{-+}_{gj|\mathbf{i}} := C_{gj|\mathbf{i}}\Big(F^-_{g|\mathbf{i}}, F^+_{j|\mathbf{i}}\Big), \quad C^{--}_{gj|\mathbf{i}} := C_{gj|\mathbf{i}}\Big(F^-_{g|\mathbf{i}}, F^-_{j|\mathbf{i}}\Big).$$

The pmf corresponding to a discrete D-vine is computed as

(1) For $d = 1, \ldots, m$, evaluate F^+_d, F^-_d, and $f_d = F^+_d - F^-_d$.
(2) For $d = 1, \ldots, m-1$, evaluate $C^{++}_{d,d+1} = C_{d,d+1}(F^+_d, F^+_{d+1})$, $C^{+-}_{d,d+1} = C_{d,d+1}(F^+_d, F^-_{d+1})$, $C^{-+}_{d,d+1} = C_{d,d+1}(F^-_d, F^+_{d+1})$, and $C^{--}_{d,d+1} = C_{d,d+1}(F^-_d, F^-_{d+1})$.
(3) For $d = 1, m-2$:
   (a) Evaluate $F^+_{d|d+1} = (C^{++}_{d,d+1} - C^{+-}_{d,d+1})/(f_{d+1})$, $F^-_{d|d+1} = (C^{-+}_{d,d+1} - C^{--}_{d,d+1})/(f_{d+1})$, and $f_{d|d+1} = F^+_{d|d+1} - F^-_{d|d+1}$. These are used as the first arguments for $C_{d,d+2|d+1}$ in step (c) below.
   (b) Evaluate $F^+_{d+2|d+1} = (C^{++}_{d+1,d+2} - C^{-+}_{d+1,d+2})/(f_{d+1})$, $F^-_{d+2|d+1} = (C^{+-}_{d+1,d+2} - C^{--}_{d+1,d+2})/(f_{d+1})$, and $f_{d+2|d+1} = F^+_{d+2|d+1} - F^-_{d+2|d+1}$. These are used as the second arguments for $C_{d,d+2|d+1}$ in step (c) below.
   (c) Evaluate $C^{ab}_{d,d+2|d+1}(F^a_{d|d+1}, F^b_{d+2|d+1})$ for $a, b \in +, -$.
(4) For $t = 3, \ldots, m-1$ and $d = 1, \ldots, m-t$:
   (a) For the first argument of $C_{d,d+t|d+1,\ldots,d+t-1}$ for step (c) below:
      (i) evaluate $F^+_{d|d+1,\ldots,d+t-1}$
$$= \frac{C^{++}_{d,d+t-1|d+1,\ldots,d+t-2} - C^{+-}_{d,d+t-1|d+1,\ldots,d+t-2}}{f_{d+t-1|d+1,\ldots,d+t-2}}$$
      (ii) evaluate $F^-_{d|d+1,\ldots,d+t-1}$
$$= \frac{C^{-+}_{d,d+t-1|d+1,\ldots,d+t-2} - C^{--}_{d,d+t-1|d+1,\ldots,d+t-2}}{f_{d+t-1|d+1,\ldots,d+t-2}}$$
      (iii) evaluate $f_{d|d+1,\ldots,d+t-1} = F^+_{d|d+1,\ldots,d+t-1} - F^-_{d|d+1,\ldots,d+t-1}$.
   (b) For the second argument of $C_{d,d+t|d+1,\ldots,d+t-1}$ for step (c) below:
      (i) evaluate $F^+_{d+t|d+1,\ldots,d+t-1}$
$$= \frac{C^{++}_{d+1,d+t|d+2,\ldots,d+t-1} - C^{-+}_{d+1,d+t|d+2,\ldots,d+t-1}}{f_{d+1|d+2,\ldots,d+t-1}}$$
      (ii) evaluate $F^-_{d+t|d+1,\ldots,d+t-1}$
$$= \frac{C^{+-}_{d+1,d+t|d+2,\ldots,d+t-1} - C^{--}_{d+1,d+t|d+2,\ldots,d+t-1}}{f_{d+1|d+2,\ldots,d+t-1}}$$
      (iii) evaluate $f_{d+t|d+1,\ldots,d+t-1} = F^+_{d+t|d+1,\ldots,d+t-1} - F^-_{d+t|d+1,\ldots,d+t-1}$.
   (c) Evaluate $C^{ab}_{d,d+t|d+1,\ldots,d+t-1}(F^a_{d|d+1,\ldots,d+t-1}, F^b_{d+t|d+1,\ldots,d+t-1})$ for $a, b \in +, -$.
(5) Evaluate $F^+_{1|2,\ldots,m} = (C^{++}_{1,m|2,\ldots,m-1} - C^{+-}_{1,m|2,\ldots,m-1})/(f_{m|2,\ldots,m-1})$, $F^-_{1|2,\ldots,m} = (C^{-+}_{1,m|2,\ldots,m-1} - C^{--}_{1,m|2,\ldots,m-1})/(f_{m|2,\ldots,m-1})$, and $f_{1|2,\ldots,m} = F^+_{1|2,\ldots,m} - F^-_{1|2,\ldots,m}$.
(6) The pmf is given by $f_{12} \prod_{i=3}^{m} f_{i|1,\ldots,i-1}$.

SUPPLEMENTARY MATERIALS

The supplementary materials include an algorithm for generating from a discrete D-vine and additional materials on the application to headache severity data including an out of sample validation.

[Received May 2011. Revised March 2012.]

REFERENCES

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), "Pair-Copula Constructions of Multiple Dependence," *Insurance, Mathematics and Economics*, 44, 182–198. [1064]

Bedford, T., and Cooke, R. (2001), "Probability Density Decomposition for Conditionally Dependent Random Variables," *The Annals of Mathematics and Artificial Intelligence*, 32, 245–268. [1064]

——— (2002), "Vines—A New Graphical Model for Dependent Random Variables," *The Annals of Statistics*, 30, 1031–1068. [1064]

Brechmann, E., Czado, C., and Aas, K. (2012), "Truncated and Simplified Regular Vines in High Dimensions With Application to Financial Data," *Canadian Journal of Statistics*, 40, 68–85. [1069]

Clarke, K. (2007), "A Simple Distribution Free Test for Non-Nested Model Selection," *Political Analysis*, 13, 347–363. [1070]

Czado, C. (2010), "Pair-Copula Constructions of Multivariate Copulas," in *Workshop on Copula Theory and Its Applications*, eds. P. Jaworki, F. Durante, W. Härdle, and W. Rychlik, New York: Springer, pp. 93–109. [1064]

Denuit, M., and Lambert, P. (2005), "Constraints on Concordance Measures in Bivariate Discrete Data," *Journal of Multivariate Analysis*, 93, 40–57. [1067]

Genest, C., and Nešlehová, J. (2007), "A Primer on Copulas for Count Data," *The Astin Bulletin*, 37, 475–515. [1063]

Hu, F., and Kalbfleisch, J. (2000), "The Estimating Function Bootstrap," *Canadian Journal of Statistics*, 28, 449–481. [1068]

Joe, H. (1996), "Families of m-Variate Distributions With Given Margins and $m(m-1)/2$ Bivariate Dependence Parameters," in *Distributions With Fixed Marginals and Related Topics*, eds. L. Rüschendorf, B. Schweizer, and M. Taylor, Hayward, CA: Institute of Mathematical Statistics, pp. 20–141. [1064]

——— (1997), *Multivariate Models and Dependence Concepts, Monographs on Statistics and Applied Probability*, London: Chapman & Hall. [1063,1064,1066,1068]

Joe, H., Li, H., and Nikoloulopoulos, A. (2010), "Tail Dependence Functions and Vine Copulas," *Journal of Multivariate Analysis*, 101, 252–270. [xxxx]

Kurowicka, D., and Cooke, R. (2006), *Uncertainty Analysis With High Dimensional Dependence Modelling* (Wiley Series in Probability and Statistics), Chichester: Wiley. [1064]

Kurowicka, D., and Joe, H. (eds.) (2011), *Dependence Modeling: Vine Copula Handbook*, London: World Scientific Publishing Company. [1064]

Li, J., and Wong, W. (2011), "Two-Dimensional Toxic Dose and Multivariate Logistic Regression, With Application to Decompression Sickness," *Biostatistics*, 12, 143–155. [1063]

Nelsen, R. (2006), *An Introduction to Copulas* (Springer Series in Statistics) (2nd ed.), New York: Springer. [1063]

Nikoloulopoulos, A., and Karlis, D. (2008), "Multivariate Logit Copula Model With an Application to Dental Data," *Statistics in Medicine*, 27, 6393–6406. [1063]

——— (2009), "Finite Normal Mixture Copulas for Multivariate Discrete Data Modeling," *Journal of Statistical Planning and Inference*, 139, 3878–3890. [1063]

Sklar, A. (1959), "Fonctions de Répartition à n Dimensions et Leur Marges [Distributional Functions to n Dimensions and Their Margins]," *Publications of the Institute of Statistics, University of Paris*, 8, 229–231. [1063]

Smith, M., and Khaled, M. (2012), "Estimation of Copula Models With Discrete Margins," *Journal of the American Statistical Association*, 107, 290–303. [1064]

Smith, M., Min, A., Almeida, C., and Czado, C. (2010), "Modelling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence," *Journal of the American Statistical Association*, 105, 1467–1479. [1069]

Song, P. X-K, Li, M., and Yuan, Y. (2009), "Joint Regression Analysis of Correlated Data Using Gaussian Copulas," *Biometrics*, 65, 60–68. [1063]

Varin, C., and Czado, C. (2010), "A Mixed Autoregressive Probit Model for Ordinal Longitudinal Data," *Biostatistics*, 11, 127–138. [1070]

Vuong, Q. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333. [1070]

Zhao, Y., and Joe, H. (2005), "Composite Likelihood Estimation in Multivariate Data Analysis," *Canadian Journal of Statistics*, 33, 335–356. [1064]