Copula Modeling:
An Introduction for Practitioners

Full text available at: http://dx.doi.org/10.1561/0800000005
Copula Modeling: An Introduction for Practitioners

Pravin K. Trivedi

Department of Economics, Indiana University
Wylie Hall 105
Bloomington, IN 47405
trivedi@indiana.edu

David M. Zimmer

Western Kentucky University, Department of Economics
1906 College Heights Blvd.
Bowling Green, KY 42101
dmzimmer@gmail.com
formerly at U.S. Federal Trade Commission

Full text available at: http://dx.doi.org/10.1561/0800000005
Editorial Scope

Foundations and Trends® in Econometrics will publish survey and tutorial articles in the following topics:

- Identification
- Model Choice and Specification Analysis
- Non-linear Regression Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Dynamic Specification
- Inference and Causality
- Continuous Time Stochastic Models
- Modeling Non-linear Time Series
- Unit Roots
- Cointegration
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Interactions-based Models
- Duration Models
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians
Foundations and Trends® in Econometrics, 2005, Volume 1, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.
Copula Modeling: An Introduction for Practitioners*

Pravin K. Trivedi¹ and David M. Zimmer²

¹ Department of Economics, Indiana University, Wylie Hall 105, Bloomington, IN 47405, trivedi@indiana.edu
² Western Kentucky University, Department of Economics, 1906 College Heights Blvd., Bowling Green, KY 42101, dmzimmer@gmail.com formerly at U.S. Federal Trade Commission

Abstract

This article explores the copula approach for econometric modeling of joint parametric distributions. Although theoretical foundations of copulas are complex, this text demonstrates that practical implementation and estimation are relatively straightforward. An attractive feature of parametrically specified copulas is that estimation and inference are based on standard maximum likelihood procedures, and thus copulas can be estimated using desktop econometric software. This represents a substantial advantage of copulas over recently proposed simulation-based approaches to joint modeling.

*The authors are grateful to the Editor Bill Greene and an anonymous reviewer for helpful comments and suggestions for improvement, but retain responsibility for the contents of the present text.
Contents

1 Introduction 1

2 Copulas and Dependence 7
  2.1 Basics of Joint Distributions 7
  2.2 Copula Functions 9
  2.3 Some Common Bivariate Copulas 14
  2.4 Measuring Dependence 19
  2.5 Visual Illustration of Dependence 27

3 Generating Copulas 33
  3.1 Method of Inversion 34
  3.2 Algebraic Methods 36
  3.3 Mixtures and Convex Sums 37
  3.4 Archimedean copulas 42
  3.5 Extensions of Bivariate Copulas 45

4 Copula Estimation 55
  4.1 Copula Likelihoods 56
  4.2 Two-Step Sequential Likelihood Maximization 59
  4.3 Copula Evaluation and Selection 63
  4.4 Monte Carlo Illustrations 68
  4.5 Empirical Applications 74
1

Introduction

This article explores the copula approach for econometric modeling of joint parametric distributions. Econometric estimation and inference for data that are assumed to be multivariate normal distributed are highly developed, but general approaches for joint nonlinear modeling of nonnormal data are not well developed, and there is a frequent tendency to consider modeling issues on a case-by-case basis. In econometrics, nonnormal and nonlinear models arise frequently in models of discrete choice, models of event counts, models based on truncated and/or censored data, and joint models with both continuous and discrete outcomes.

Existing techniques for estimating joint distributions of nonlinear outcomes often require computationally demanding simulation-based estimation procedures. Although theoretical foundations of copulas are complex, this text demonstrates that practical implementation and estimation is relatively straightforward. An attractive feature of parametrically specified copulas is that estimation and inference are based on standard maximum likelihood procedures, and thus copulas can be estimated using desktop econometric software such as Stata, Limdep, or
2 Introduction

SAS. This represents a substantial advantage of copulas over recently proposed simulation-based approaches to joint modeling.

Interest in copulas arises from several perspectives. First, econometricians often possess more information about marginal distributions of related variables than their joint distribution. The copula approach is a useful method for deriving joint distributions given the marginal distributions, especially when the variables are nonnormal. Second, in a bivariate context, copulas can be used to define nonparametric measures of dependence for pairs of random variables. When fairly general and/or asymmetric modes of dependence are relevant, such as those that go beyond correlation or linear association, then copulas play a special role in developing additional concepts and measures. Finally, copulas are useful extensions and generalizations of approaches for modeling joint distributions and dependence that have appeared in the literature.

According to Schweizer (1991), the theorem underlying copulas was introduced in a 1959 article by Sklar written in French; a similar article written in English followed in 1973 (Sklar, 1973). Succinctly stated, copulas are functions that connect multivariate distributions to their one-dimensional margins. If $F$ is an $m$-dimensional cumulative distribution function (cdf) with one-dimensional margins $F_1, \ldots, F_m$, then there exists an $m$-dimensional copula $C$ such that $F(y_1, \ldots, y_m) = C(F_1(y_1), \ldots, F_m(y_m))$. The case $m = 2$ has attracted special attention.

The term copula was introduced by Sklar (1959). However, the idea of copula had previously appeared in a number of texts, most notably in Hoeffding (1940, 1941) who established best possible bounds for these functions and studied measures of dependence that are invariant under strictly increasing transformations. Relationships of copulas to other work is described in Nelsen (2006).

Copulas have proved useful in a variety of modeling situations. Several of the most commonly used applications are briefly mentioned:

- Financial institutions are often concerned with whether prices of different assets exhibit dependence, particularly in the tails of the joint distributions. These models typically assume that asset prices have a multivariate normal
distribution, but Ané and Kharoubi (2003) and Embrechts et al. (2002) argue that this assumption is frequently unsatisfactory because large changes are observed more frequently than predicted under the normality assumption. Value at Risk (VaR) estimated under multivariate normality may lead to underestimation of the portfolio VaR. Since deviations from normality, e.g., tail dependence in the distribution of asset prices, greatly increase computational difficulties of joint asset models, modeling based on a copula parameterized by nonnormal marginals is an attractive alternative; see Bouyé et al. (2000), Klugman and Parsa (2000).

- Actuaries are interested in annuity pricing models in which the relationship between two individuals’ incidence of disease or death is jointly related (Clayton 1978). For example, actuaries have noted the existence of a “broken heart” syndrome in which an individual’s death substantially increases the probability that the person’s spouse will also experience death within a fixed period of time. Joint survivals of husband/wife pairs tend to exhibit nonlinear behavior with strong tail dependence and are poorly suited for models based on normality. These models are prime candidates for copula modeling.

- Many microeconometric modeling situations have marginal distributions that cannot be easily combined into joint distributions. This frequently arises in models of discrete or limited dependent variables. For example, Munkin and Trivedi (1990) explain that bivariate distributions of discrete event counts are often restrictive and difficult to estimate. Furthermore, joint modeling is especially difficult when two related variables come from different parametric families. For example, one variable might characterize a multinomial discrete choice and another might measure an event count. As there are few, if any, parametric joint distributions based on marginals from different families, the copula approach provides a general and straightforward approach for constructing joint distributions in these situations.
• In some applications, a flexible joint distribution is part of a larger modeling problem. For example, in the linear self-selection model, an outcome variable, say income, is only observed if another event occurs, say labor force participation. The likelihood function for this model includes a joint distribution for the outcome variable and the probability that the event is observed. Usually, this distribution is assumed to be multivariate normal, but Smith (2003) demonstrates that for some applications, a flexible copula representation is more appropriate.

Several excellent monographs and surveys are already available, particularly those by Joe (1997) and Nelsen (2006). Schweizer and Sklar (1983, ch. 6, provide a mathematical account of developments on copulas over three decades. Nelsen (1991) focuses on copulas and measures of association. Other surveys take a contextual approach. Frees and Valdez (1998) provide an introduction for actuaries that summarizes statistical properties and applications and is especially helpful to new entrants to the field. Georges et al. (2001) provide a review of copula applications to multivariate survival analysis. Cherubini et al. (2004) focus on financial applications, but they also provide an excellent coverage of copula foundations for the benefit of a reader who may be new to the area. For those whose main concern is with modeling dependence using copulas, Embrechts et al. (2002) provide a lucid and thorough coverage.

In econometrics there is a relatively small literature that uses copulas in an explicit manner. Miller and Liu (2002) mention the copula method in their survey of methods of recovering joint distributions from limited information. Several texts have modeled sample selection using bivariate latent variable distributions that can be interpreted as specific examples of copula functions even though the term copula or copula properties are not explicitly used; see Lee (1983), Prieger (2002) and van Ophem (1999, 2000). However, Smith (2003) explicitly uses the (Archimedean) copula framework to analyze the self-selection problem. Similarly for the case of joint discrete distributions, a number of studies that explore models of correlated count variables, without explicitly
using copulas, are developed in Cameron et al. (1988), Munkin and Trivedi (1999), and Chib and Winkelmann (2001). Cameron et al. (2004), use the copula framework to analyze the empirical distribution of two counted measures of the same event. Zimmer and Trivedi (2006) use a trivariate copula framework to analyze a selection model with counted outcomes. In financial econometrics and time series analysis, the copula approach has attracted considerable attention recently. Bouyé et al. (2000) and Cherubini et al. (2004) cover many issues and financial applications. A central issue is on the nature of dependence and hence the interpretation of a copula as a dependence function dominates. See Patton (forthcoming) for further discussion of copulas in time series settings.

The purpose of this article is to provide practitioners with a useful guide to copula modeling. Special attention is dedicated to issues related to estimation and misspecification. Although our main focus is using copulas in an applied setting, particularly cross sectional microeconometric applications, it is necessary to cover important theoretical foundations related to joint distributions, dependence, and copula generation. Sections 2 and 3 primarily deal with these theoretical issues. The reader who is already familiar with the basics of copulas and dependence may wish to skip directly to Section 4, which highlights issues of estimation and presents several empirical applications. Section 5 offers concluding remarks as well as suggestions for future research. Throughout the text, various Monte Carlo experiments and simulations are used to demonstrate copula properties. Methods for generating random numbers from copulas are presented in the Appendix.
References


Frank, M. J. (1979), ‘On the simultaneous associativity of $F(x,y)$ and $x+y - F(x,y)$’. *Aequationes Math* 19, 194–226.


References


Full text available at: http://dx.doi.org/10.1561/0800000005


Simulation is a useful tool for understanding and exhibiting dependence structures of joint distributions. According to Nelsen (2006: 40), “one of the primary applications of copulas is in simulation and Monte Carlo studies.” Draws of pseudo-random variates from particular copulas can be displayed graphically, which allows one to visualize dependence properties such as tail dependence. Methods of drawing from copulas are also needed when conducting Monte Carlo experiments. This chapter presents selected techniques for drawing random variates from bivariate distributions and illustrates them with a few examples. In our experience, the appropriate method for drawing random variables depends upon which distribution is being considered; some methods are best suited for drawing variables from particular distributions. We do not claim that the methods outlined below are necessarily the “best” approaches for any given application. Rather, in our experience, the following approaches are straightforward to implement and provide accurate draws of random variates.

Random variates can be plotted to show dependence between variables. Many copula researchers rely on scatter plots to visualize
differences between various copulas (Embrechts et al., 2002). Other researchers report pdf contour plots (Smith, 2003), which are presumably easier to interpret than three-dimensional pdf graphs. Nevertheless, some researchers report combinations of pdf contour plots and three-dimensional graphs (Ané and Kharoubi, 2003), while others report all three: scatter plots, contour graphs, and three-dimensional figures (Bouyé et al., 2000). All three techniques convey the same information, so whichever presentation one chooses is essentially a matter of preference. We use scatter plots for several reasons. First, scatter plots are easier to generate than pdf contour plots or three-dimensional figures and do not require complicated graphing software. Second, random draws used to create scatter plots are also useful for generating simulated data in Monte Carlo experiments. Third, scatter plots can be easily compared to plots of real life data to assist in choosing appropriate copula functions. Finally, we feel that interpretations are more straightforward for scatter plots than they are for pdf contour plots or three-dimensional figures.

A.1 Selected Illustrations

In this section, we sketch some algorithms for making pseudo-random draws from copulas. These algorithms can be viewed as adaptations of various general methods for simulating draws from multivariate distributions.

A.1.1 Conditional sampling

For many copulas, conditional sampling is a simple method of simulating random variates. The steps for drawing from a copula are:

- Draw $u$ from standard uniform distribution.
- Set $y = F^{-1}(u)$ where $F^{-1}$ is any quasi-inverse of $F$.
- Use the copula to transform uniform variates. One such transformation method uses the conditional distribution of
$U_2$, given $u_1$.

\[
c_{u_1}(u_2) = \Pr[U_2 \leq u_2 | U_1 \leq u_1] = \lim_{\Delta u_1 \to 0} \frac{C(u_1 + \Delta u_1, u_2) - C(u_1, u_2)}{\Delta u_1} = \frac{\partial}{\partial u_1} C(u_1, u_2)
\]

By Theorem 2.2.7 in Nelsen (2006), a nondecreasing function $c_{u_1}(u_2)$ exists almost everywhere in the unit interval.

In practice, conditional sampling is performed through the following steps:

- Draw two independent random variables $(v_1, v_2)$ from $U(0, 1)$.
- Set $u_1 = v_1$.
- Set $u_2 = C_2(u_2 | u_1 = v_1) = \partial C(u_1, u_2) / \partial u_1$.

Then the pair $(u_1, u_2)$ are uniformly distributed variables drawn from the respective copula $C(u_1, u_2; \theta)$. This technique is best suited for drawing variates from the Clayton, Frank, and FGM copulas; see Armstrong and Galli (2002). The following equations show how this third step is implemented for these three different copulas (Table A.1).

### A.1.2 Elliptical sampling

Methods for drawing from elliptical distributions, such as the bivariate normal and bivariate $t$-distribution, are well-established in statistics.

#### Table A.1 Selected conditional transforms for copula generation.

<table>
<thead>
<tr>
<th>Copula</th>
<th>Conditional copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>$u_2 = \left(\frac{v_1^{-\theta}}{v_2^{\theta}(\theta+1) - 1}\right)^{-1/\theta}$</td>
</tr>
<tr>
<td>Frank</td>
<td>$u_2 = -\frac{1}{\theta} \log \left(1 + \frac{v_2(1 - e^{-\theta})}{v_2(e^{-\theta} v_1 - 1) - e^{-\theta} v_1}\right)$</td>
</tr>
<tr>
<td>FGM</td>
<td>$u_2 = 2v_2 / \left(\sqrt{B} - A\right)$</td>
</tr>
</tbody>
</table>

$A = \theta(2u_1 - 1); B = [1 - \theta(2u_1 - 1)]^2 + 4\theta v_2 (2u_1 - 1)$

---

1. See Example 2.20 in Nelsen (2006: 41–42) which gives the algorithm for drawing from $C(u_1, u_2) = u_1 u_2 / (u_1 + u_2 - u_1 u_2)$. 

Full text available at: http://dx.doi.org/10.1561/0800000005
These same methods are used to draw values from the Gaussian copula. The following algorithm generates random variables $u_1$ and $u_2$ from the Gaussian copula $C(u_1, u_2; \theta)$:

- Generate two independently distributed $N(0,1)$ variables $v_1$ and $v_2$.
- Set $y_1 = v_1$.
- Set $y_2 = v_1 \cdot \theta + v_2 \sqrt{1 - \theta^2}$.
- Set $u_i = \Phi(y_i)$ for $i = 1, 2$ where $\Phi$ is the cumulative distribution function of the standard normal distribution.

Then the pair $(u_1, u_2)$ are uniformly distributed variables drawn from the Gaussian copula $C(u_1, u_2; \theta)$.

### A.1.3 Mixtures of powers simulation

To make draws from the Gumbel copula using conditional sampling, we need to calculate $C_2(v_2|v_1)$ which requires an iterative solution, which is computationally expensive for applications with many simulated draws. [Marshall and Olkin (1988)](http://dx.doi.org/10.1561/0800000005) suggest an alternative algorithm based on mixtures of powers. The following algorithm shows how the technique is used to generate draws from the Gumbel copula:

- Draw a random variable $\gamma$ having Laplace transformation $\tau(t) = \exp(-t^{1/\theta})$. See below for additional detail.
- Draw two independent random variables $(v_1, v_2)$ from $U(0, 1)$.
- Set $u_i = \tau \left(-\gamma^{-1} \ln v_i\right)$ for $i = 1, 2$.

Then $(u_1, u_2)$ are uniformly distributed variables drawn from the Gumbel copula.

However, to implement the first step we have to draw a random variable $\gamma$ from a positive stable distribution $PS(\alpha, 1)$. This is accomplished using the following algorithm by [Chambers et al. (1976)](http://dx.doi.org/10.1561/0800000005):

- Draw a random variable $\eta$ from $U(0, \pi)$.
- Draw a random variable $w$ from the exponential distribution with mean equal to 1.
A.1. Selected Illustrations

115

• Setting $\alpha = 1/\theta$, generate

$$z = \frac{\sin((1 - \alpha)\eta)(\sin(\alpha \eta))^{\alpha}}{\sin(\eta)^{1-\alpha}}.$$ 

• Set $\gamma = (z/w)^{(1-\alpha)/\alpha}$.

Then $\gamma$ is a randomly draw variable from a $PS(\alpha,1)$ distribution.

A.1.4 Simulating discrete variables

Methods for drawing discrete variables depend upon which type of discrete variate is desired. We focus on simulating discrete Poisson variables using a method based on Devroye’s technique of sequential search (Devroye, 1986). The algorithm is as follows:

• Draw correlated uniform random variables $(u_1, u_2)$ from a particular copula using any of the methods discussed above.
• Set the Poisson mean $= \mu_1$ such that $\Pr(Y_1 = 0) = e^{-\mu_1}$.
• Set $Y_1 = 0$, $P_0 = e^{-\mu_1}$, $S = P_0$.
• If $u_1 < S$, then $Y_1$ remains equal to 0.
• If $u_1 > S$, then proceed sequentially as follows. While $u_1 > S$,

1. $Y_1 \leftarrow Y_1 + 1$,
2. $P_0 \leftarrow \mu_1 P_0 / Y_1$,
3. $S \leftarrow S + P_0$.

This process continues until $u_1 < S$.

These steps produce a simulated variable $Y_1$ with Poisson distribution with mean $\lambda_1$. To obtain draws of the second Poisson variable $Y_2$, replace $u_1$ and $\mu_1$ with $u_2$ and $\mu_2$ and repeat the steps above. Then the pair $(Y_1, Y_2)$ are jointly distributed Poisson variables with means $\mu_1$ and $\mu_2$. 

Full text available at: http://dx.doi.org/10.1561/0800000005