Copula Modeling: An Introduction for Practitioners

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

the essence of knowledge

Boston – Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is P. K. Trivedi and D. M. Zimmer, Copula Modeling: An Introduction for Practitioners, Foundations and Trends® in Econometrics, vol 1, no 1, pp 1–111, 2005

Printed on acid-free paper

ISBN: 978-1-60198-020-5

Foundations and Trends® in Econometrics

Volume 1 Issue 1, 2005

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.

Foundations and Trends® in
Econometrics
Vol. 1, No 1 (2005) 1–111

DOI: 10.1561/0800000005

Copula Modeling: An Introduction for Practitioners*

Pravin K. Trivedi[1] **and David M. Zimmer**[2]

[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

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

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 (1999) 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

Amemiya, T. (1973), 'Regression analysis when the dependent variable is truncated normal'. *Econometrica* **41**, 997–1016.

Amemiya, T. (1985), *Advanced Econometrics*. Cambridge: Harvard University Press.

Ané, T. and C. Kharoubi (2003), 'Dependence structure and risk measure'. *Journal of Business* **76**, 411–438.

Armstrong, M. (2003), 'Copula catalogue. Part 1: Bivariate Archimedean copulas'. Unpublished paper available at http://www.cerna.ensmp.fr.

Armstrong, M. and A. Galli (2002), 'Sequential nongaussian simulations using the FGM copula'. http://www.cerna.ensmp.fr/Documents/MA-AG-WPCopula.pdf.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Ribouletm, and T. Roncalli (2000), 'Copulas for finance: A reading guide and some applications'. Unpublished Manuscript, London: Financial Econometrics Research Centre, City University Business School.

Boyer, B. H., M. S. Gibson, and M. Loretan (1999), 'Pitfalls in tests for changes in correlations'. Federal Reserves Board, IFS Discussion Paper No 597R.

Brown, E. and H. Lankford (1992), 'Gifts of money and gifts of time: Estimating the effects of tax prices and available time'. *Journal of Public Economics* **47**, 321–341.

Cameron, A. C., T. Li, P. K. Trivedi, and D. M. Zimmer (2004), 'Modeling the differences in counted outcomes using bivariate copula models: With application to mismeasured counts'. *Econometrics Journal* **7**, 566–584.

Cameron, A. C. and P. K. Trivedi (1998), *Regression Analysis of Count Data*, Econometric Society Monographs 30. New York: Cambridge University Press.

Cameron, A. C., P. K. Trivedi, F. Milne, and J. Piggott (1988), 'A microeconomic model of the demand for health care and health insurance in Australia'. *Review of Economic Studies* **55**, 85–106.

Chambers, J. M., C. L. Mallows, and B. W. Stuck (1976), 'A method for simulating stable random variables'. *Journal of the American Statistical Association* **71**, 340–344.

Chen, X. and Y. Fan (2006), 'Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification'. *Journal of Econometrics*.

Cherubini, U., E. Luciano, and W. Vecchiato (2004), *Copula Methods in Finance*. New York: John Wiley.

Chib, S. (1992), 'Bayes regression for the Tobit censored regression model'. *Journal of Econometrics* **51**, 79–99.

Chib, S. and R. Winkelmann (2001), 'Markov chain Monte Carlo analysis of correlated count data'. *Journal of Business and Economic Statistics* **19**, 428–435.

Clayton, D. G. (1978), 'A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence'. *Biometrika* **65**, 141–151.

Clayton, D. G. and J. Cuzick (1985), 'Multivariate generalizations of the proporational hazards model'. *Journal of Royal Statistical Society Series B* **34**, 187–220.

Cont, R. (2001), 'Empirical proeprties of asset returns: Stylized facts and statistical issues'. *Quantitative Finance* **1**, 223–236.

Cook, R. D. and M. E. Johnson (1981), 'A family of distributions for modelling non-elliptically symmetric multivariate data'. *Journal of Royal Statistical Society B* **43**(2), 210–218.

de la Peña, V. H., R. Ibragimov, and S. Sharakhmetov (2003), 'Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series'. 2nd Erich L. Lehmann Symposium – Optimality, IMS Lecture Notes – Monograph Series, (J. Rojo, Ed.), In Press.

Deb, P., P. K. Trivedi, and P. Varangis (1996), 'Excess Co-movement in commodity prices reconsidered'. *Journal of Applied Econometrics* **11**, 275–291.

Denuit, M. and P. Lambert (2005), 'Constraints on concordance measures in bivariate discrete data'. *Journal of Multivariate Analysis* **93**, 40–57.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Drouet-Mari, D. and S. Kotz (2001), *Correlation and Dependence*. London: Imperial College Press.

Durbin, J. and A. S. Stuart (1951), 'Inversions and rank correlations'. *Journal of Royal Statistical Society Series B* **2**, 303–309.

Durrleman, V., A. Nikeghbali, and T. Roncalli (2000), 'How to get bounds for distribution convolution? A simulation study and an application to risk management'. GRO Credit Lyonnais, working paper.

Embrechts, P., A. McNeil, and D. Straumann (2002), 'Correlation and dependence in risk management: Properties and pitfalls'. In: M. A. H. Dempster (ed.): *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press, pp. 176–223.

Fang, K.-T. and Y.-T. Zhang (1990), *Generalized Multivariate Analysis*. Berlin and New York: Springer-Verlag.

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

Frees, E., J. Carriere, and E. Valdez (1996), 'Annuity valuation with dependent mortality'. *Journal of Risk and Insurance* **63**, 229–261.

Frees, E. W. and E. A. Valdez (1998), 'Understanding relationships using copulas'. *North American Actuarial Journal* **2**(1), 1–26.

Genest, C. and J. Mackay (1986), 'The joy of copulas: Bivariate distributions with uniform marginals'. *The American Statistician* **40**, 280–283.

Genest, C., Q. Molina, and R. Lallena (1995), 'De L'impossibilité de Constuire des lois a Marges Multidimensionnelles Données a Partir de Copules'. *Comptes Rendus de l'Académie des Sciences Serie I, Mathematique* **320**, 723–726.

Genest, C. and L. Rivest (1993), 'Statistical inference procedures for bivariate Archimedean copulas'. *Journal of the American Statistical Association* **88**(423), 1034–1043.

Genius, M. and E. Strazzera (2004), 'The copula approach to sample selection modelling: An application to the recreational value of forests'. Working paper, The Fondazione Eni Enrico Mattei.

Georges, P., A.-G. Lamy, A. Nicolas, G. Quibel, and T. Roncalli (2001), 'Multivariate survival modeling: A unified approach with copulas'. Unpublished paper, France: Groupe de Recherche Opérationnelle Crédit Lyonnais.

Gourieroux, C. and A. Monfort (1996), *Simulation Based Econometric Methods*. New York: Oxford University Press.

Greene, W. H. (1990), 'Multiple roots of the Tobit log-likelihood'. *Journal of Econometrics* **46**, 365–380.

Gumbel, E. J. (1960), 'Distributions des Valeurs Extremes en Plusieurs Dimensions'. *Publications de l'Institute de Statistíque de l'Université de Paris* **9**, 171–173.

Heckman, J. (1976), 'The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models'. *Annals of Economics and Social Measurement* **5**, 475–492.

Heckman, J. J. and B. E. Honoré (1989), 'The identifiability of the competing risks model'. *Biometrika* **76**, 325–330.

Heinen, A. and E. Rengifo (2003a), 'Comovements in trading activity: A multivariate autoregressive model of time series count data using copulas'. CORE Discussion Paper.

Heinen, A. and E. Rengifo (2003b), 'Multivariate modelling of time series count data: An autoregressive conditional Poisson model'. CORE Discussion Paper.

Hoeffding, W. (1940), 'Scale-invariant correlation theory'. In: N. I. Fisher and P. K. Sen (eds.): *The Collected Works of Wassily Hoeffding*. New York: Springer-Verlag, pp. 57–107.

Hoeffding, W. (1941), 'Scale-invariant correlation measures for discontinuous distributions'. In: N. I. Fisher and P. K. Sen (eds.): *The Collected Works of Wassily Hoeffding*. New York: Springer-Verlag, pp. 109–133.

Hougaard, P. (1986), 'A class of multivariate failure time distributions'. *Biometrika* **73**, 671–678.

Hougaard, P. (1987), 'Modeling multivariate survival'. *Scandinavian Journal of Statistics* **14**, 291–304.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*. New-York: Springer-Verlag.

Hougaard, P., B. Harvald, and N. V. Holm (1992), 'Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930'. *Journal of the American Statistical Association* **87**, 17–24.

Hu, L. (2004), 'Dependence patterns across financial markets: A mixed copula approach'. The Ohio State University, working paper.

Huang, H.-C. (1999), 'Estimation of the SUR Tobit model via the MCECM algorithm'. *Economics Letters* **64**, 25–30.

Huang, H.-C., F. A. Sloan, and K. W. Adamache (1987), 'Estimation of seemingly unrelated Tobit regressions via the EM algorithm'. *Journal of Business and Economic Statistics* **5**, 425–430.

Husler, J. and R. Reiss (1989), 'Maxima of normal random vectors: Between independence and complete dependence'. *Statistics and Probability Letters* **7**, 283–286.

Hutchinson, T. P. and C. D. Lai (1990), *Continuous Bivariate Distributions, Emphasising Applications*. Sydney, Australia: Rumsby.

Joe, H. (1990), 'Families of min-stable multivariate exponential and multivariate extreme value distributions'. *Statistics and Probability Letters* **9**, 75–81.

Joe, H. (1993), 'Parametric families of multivariate distributions with given margins'. *Journal of Multivariate Analysis* **46**, 262–282.

Joe, H. (1994), 'Multivariate extreme-value distributions with applications to environmental data'. *Canadian Journal of Statistics* **22**(1), 47–64.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Junker, M. and A. May (2005), 'Measurement of aggregate risk with copulas'. *Econometrics Journal* **8**, 428–454.

Kamakura, W. A. and M. Wedel (2001), 'Exploratory Tobit factor analysis for multivariate censored data'. *Multivariate Behavioral Research* **36**, 53–82.

Kimeldorf, G. and A. R. Sampson (1975), 'Uniform representations of bivariate distributions'. *Communications in Statistics* **4**, 617–627.

Klugman, S. A. and R. Parsa (2000), 'Fitting bivariate loss distributions with copulas'. *Insurance: Mathematics and Economics* **24**, 139–148.

Lee, L. (1983), 'Generalized econometric models with selectivity'. *Econometrica* **51**, 507–512.

Li, D. X. (2000), 'On default correlation: A copula function approach'. *Journal of Fixed Income* **9**, 43–54.

Machado, J. A. F. and J. M. C. Santos Silva (2005), 'Quantiles for counts'. *Journal of the American Statistical Association* **100**, 1226–1237.

Marshall, A. (1996), 'Copulas, marginals, and joint distributions'. In: L. Ruschendorf, B. Schweizer, and M. D. Taylor (eds.): *Distributions with Fixed Marginals and Related Topics*. Hayward, CA: Institute of Mathematic Statistics, pp. 213–222.

Marshall, A. W. and I. Olkin (1967), 'A multivariate exponential distribution'. *Journal of the American Statistical Association* **62**, 30–44.

Marshall, A. W. and I. Olkin (1988), 'Families of multivariate distributions'. *Journal of the American Statistical Association* **83**, 834–841.

Marshall, A. W. and I. Olkin (1990), 'Multivariate distributions generated from mixtures of convolution and product families'. In: H. W. Block, A. R. Sampson, and T. H. Savits (eds.): *Topics in Statistical*

Dependence, Vol. 16 of *IMS Lecture Notes-Monograph Series*. pp. 371–393.

Meester, S. and J. MacKay (1994), 'A parametric model for cluster correlated categorical data'. *Biometrics* **50**, 954–963.

Meng, X. and D. B. Rubin (1996), 'Efficient methods for estimating and testing seemingly unrelated regressions in the presence of latent variables and missing observations'. In: D. A. Berry, K. M. Chaloner, and J. K. Geweke (eds.): *Bayesian Analysis in Statistics and Econometrics*. John Wiley & Sons, Inc, pp. 215–227.

Miller, D. J. and W. H. Liu (2002), 'On the recovery of joint distributions from limited information'. *Journal of Econometrics* **107**, 259–274.

Morgenstern, D. (1956), 'Einfache Beispiele Zweidimensionaler Verteilungen'. *Mitteilingsblatt fur Mathematische Statistik* **8**, 234–235.

Munkin, M. and P. K. Trivedi (1999), 'Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with application'. *Econometric Journal* **1**, 1–21.

Nelsen, R. B. (1991), 'Copulas and association'. In: G. Dall'Aglio, S. Kotz, and G. Salinetti (eds.): *Advances in Probability Distributions with Given Marginals*. Dordrecht: Kluwer Academic Publishers, pp. 31–74.

Nelsen, R. B. (2006), *An Introduction to Copulas*. 2nd edition. New York: Springer.

Oakes, D. (1982), 'A model for association in bivariate survival data'. *Journal of Royal Statistical Society B* **44**, 414–422.

Patton, A. (2005a), 'Modelling asymmetric exchange rate dependence'. *International Economic Review*. Forthcoming.

Patton, A. (2005b), 'Estimation of multivariate models for time series of possibly different lengths'. *Journal of Applied Econometrics*. forthcoming.

Pitt, M., D. Chan, and R. Kohn (2006), 'Efficient Bayesian inference for Gaussian copula regression'. *Biometrika* **93**, 537–554.

Prieger, J. (2002), 'A flexible parametric selection model for non-normal data with application to health care usage'. *Journal of Applied Econometrics* **17**, 367–392.

Prokhorov, A. and P. Schmidt (2006), 'Robustness, redundancy, and validity of copulas in likelihood models'. Working Paper, Michigan State University.

Schweizer, B. (1991), 'Thirty years of copulas'. In: G. Dall'Aglio, S. Kotz, and G. Salinetti (eds.): *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*. The Netherlands: Kluwer Academic Publishers.

Schweizer, B. and A. Sklar (1983), *Probabilistic Metric Spaces*. New York: North Holland.

Schweizer, B. and E. F. Wolff (1981), 'On nonparametric measures of dependence for random variables'. *Annals of Statistics* **9**, 870–885.

Sklar, A. (1973), 'Random variables, joint distributions, and copulas'. *Kybernetica* **9**, 449–460.

Skrondal, A. and S. Rabe-Hesketh (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Smith, M. (2003), 'Modeling selectivity using archimedean copulas'. *Econometrics Journal* **6**, 99–123.

Smith, M. (2005), 'Using copulas to model switching regimes with an application to child labour'. *Economic Record* **81**, S47–S57.

Song, P. X. (2000), 'Multivariate dispersion models generated from Gaussian copula'. *Scandinavian Journal of Statistics* **27**, 305–320.

Song, P. X. K., Y. Fan, and J. D. Kalbfleisch (2005), 'Maximization by parts in likelihood inference'. *Journal of the American Statistical Association* **100**, 1145–1158.

Stevens, W. L. (1950), 'Fiducial Limits of the parameter of a discontinuous distribution'. *Biometrika* **37**, 117–129.

Train, K. E. (2003), *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.

Van Ophem, H. (1999), 'A general method to estimate correlated discrete random variables'. *Econometric Theory* **15**, 228–237.

Van Ophem, H. (2000), 'Modeling selectivity in count data models'. *Journal of Business and Economic Statistics* **18**, 503–511.

Vuong, Q. (1989), 'Likelihood ratio test for model selection and nonnested hypotheses'. *Econometrica* **57**, 307–333.

Wales, T. J. and A. D. Woodland (1983), 'Estimation of consumer demand systems with binding non-negativity constraints'. *Journal of Econometrics* **21**, 263–285.

Wang, W. (2003), 'Estimating the association parameter for copula models under dependent censoring'. *Journal of the Royal Statistical Society B* **65**, 257–273.

Zimmer, D. M. and P. K. Trivedi (2006), 'Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand'. *Journal of Business and Economic Statistics* **24**, 63–76.

A

Copulas and Random Number Generation

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.

$$\begin{aligned} c_{u_1}(u_2) &= \Pr[U_2 \le u_2 | U_1 \le u_1] \\ &= \lim_{\Delta u_1 \to 0} \frac{C(u_1 + \Delta u_1, u_2) - C(u_1, u_2)}{\Delta u_1} \\ &= \frac{\partial}{\partial u_1} C(u_1, u_2). \end{aligned}$$

By Theorem 2.2.7 in Nelsen (2006), a nondecreasing function $c_{u_1}(u_2)$ exists almost everywhere in the unit interval.[1]

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.

Copula	Conditional copula
Clayton	$u_2 = \left(v_1^{-\theta}\left(v_2^{-\theta/(\theta+1)} - 1\right) + 1\right)^{-1/\theta}$
Frank	$u_2 = -\frac{1}{\theta}\log\left(1 + \frac{v_2(1-e^{-\theta})}{v_2(e^{-\theta v_1} - 1) - e^{-\theta v_1}}\right)$
FGM	$u_2 = 2v_2/\left(\sqrt{B} - A\right)$
	$A = \theta(2u_1 - 1); B = [1 - \theta(2u_1 - 1)]^2 + 4\theta v_2\,(2u_1 - 1)$

[1] See Example 2.20 in Nelsen (2006: 41–42) which gives the algorithm for drawing from $C(u_1,u_2) = u_1u_2/(u_1 + u_2 - u_1u_2)$.

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 Φ 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 $C_2(u_2|v_1)$ which requires an iterative solution, which is computationally expensive for applications with many simulated draws. Marshall and Olkin (1988) 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 γ 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 γ from a positive stable distribution $PS(\alpha, 1)$. This is accomplished using the following algorithm by Chambers et al. (1976).

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

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

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

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

Then γ 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 $u_1 > S$, replace (i) $Y_1 \leftarrow Y_1 + 1$, (ii) $P_0 \leftarrow \mu_1 P_0/Y_1$, (iii) $S \leftarrow S + P_0$. This process continues until $u_1 < S$.

These steps produce a simulated variable Y_1 with Poisson distribution with mean λ_1. To obtain draws of the second Poisson variable Y_2, replace u_1 and μ_1 with u_2 and μ_2 and repeat the steps above. Then the pair (Y_1, Y_2) are jointly distributed Poisson variables with means μ_1 and μ_2.