

# Copulas and Regression

Edward W. (Jed) Frees

July 15, 2009

## Abstract

It has now been fifty years since the introduction of copulas in 1959 by Sklar in the context of probabilistic metric spaces. Copulas are now a widely used tool in biomedical applications, finance and insurance for understanding relationships among variables whose distribution cannot be approximated by a normal curve.

In this talk, I will introduce copulas and show how they can be used in regression and panel data contexts. I will also give a number of personal examples, mainly from studies risk management and insurance, to illustrate how copulas can be used. Through this introductory material and examples from a specific field, I hope to suggest to the members of the audience how they can use copulas in their own fields of applications.

## 1 Selected References

### 1.1 Introductory Material

- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley, New York.
- Demarta, Stefano and Alexander J. McNeil (2005). The  $t$  copula and related copulas. *International Statistical Review* 73(1), 111-129.
- Frees, Edward W. and Emiliano Valdez (1998). Understanding relationships using copulas. *North American Actuarial Journal* 2, No. 1, 1-25.
- Genest, Christian, and Anne-Catherine Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12, 347-368.
- Genest, Christian, and Jock Mackay (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician* 40, 280-283.
- Joe, Harry (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, New York.
- Lindsey, J. K. and P. J. Lindsey (2006). Multivariate distributions with correlation matrices for nonlinear repeated measurements. *Computational Statistics and Data Analysis* 50, 720-732.
- Mikosch, Thomas (2005). Copulas: Tales and facts. *Extremes*.
- Nelson, Roger B (1999). *An Introduction to Copulas*, Lecture Notes in Statistics 139, Springer.
- Yan, Jun (2007). Enjoy the joy of copulas. *Journal of Statistical Software* 21(4), 121.

### 1.2 Additional Foundation Material

- Marshall, A.W. and Ingram Olkin (1988). Families of multivariate distributions. *Journal of the American Statistical Association* 83, 834-841.
- Pitt, M., D. Chan and R. Kohn (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* 93, 537-554.
- Sklar, A. (1959). Fonctions de repartition a  $n$  dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* 8, 229-231.

### 1.3 Diagnostic Checking

Berg, Daniel (2007). Copula goodness-of-fit testing: An overview and power comparison. University of Oslo.

Genest, Christian, K. Ghoudi and Louis-Paul Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82, 543-552.

Genest, Christian, Jean-François Quessy, and Bruno Rémillard (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* 33 (2), 337-366.

Genest, Christian, and Bruno Rémillard (2007). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models.

Genest, Christian, Bruno Rémillard and David Beaudoin (2009). Goodness-of-fit testing for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44 (2), 199-213.

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

Panchenko, Valentyn (2005). Goodness-of-fit for copulas. *Physica A* 355, 176-182.

Sun, Jiafeng, Edward W. Frees and Marjorie A. Rosenberg (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics* 42(2), 817-830.

Wang, Weijing and Martin T. Wells (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association* 95, 62-72.

### 1.4 Copulas with Discrete Data

Cameron, A. Colin, Li, T., Pravin K. Trivedi and David M. Zimmer (2004). Modelling the differences in counted outcomes using bivariate copula models with an application to mismeasured counts. *Econometrics Journal*, 72: 566-584.

Genest, Christian and Johanna Nešlehová (2007). A primer on copulas for count data. *Astin Bulletin* 37 (2), 475-515.

Madsen, Lisa (2009). Maximum likelihood estimation of regression parameters with spatially dependent discrete data. To appear, *The Journal of Agricultural, Biological, and Environmental Statistics*.

Zimmer, David M. and Pravin 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 (1), 63-76.

### 1.5 Additional Personal References

Frees, Edward W. (2004). *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge University Press, New York.

Frees, Edward W. (2010). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press, New York.

Frees, Edward W. and Ping Wang (2005). Credibility using copulas. *North American Actuarial Journal* 9 (2), 31-48.

Frees, Edward W. and Ping Wang (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics* 38(2), 360-373.

Frees, Edward W., Jacques Carriere and Emiliano Valdez (1996). Annuity valuation with dependent mortality. *Journal of Risk and Insurance* 63, 229-261.

Frees, Edward W. and Emiliano Valdez (2008). Hierarchical insurance claims modeling. *The Journal of the American Statistical Association* 103 (484), 1457-1469.