

Using R vine copulas to explain dependencies of health care costs

Daniel Kraus^{*}

daniel.kraus@tum.de

Using R **vine copulas** (vine-copula.org) to explain the dependencies of insurance data has become very popular. High dimensional multivariate data sets with **complex dependence structures** (e.g. high tail dependence) can be modeled using only bivariate copulas by a **pair copula construction**. The data set we consider contains the total costs of ambulant, stationary and dental treatments for the years 2005 to 2007 for a large number of insured individuals. Further, for each person demographic variables such as age, gender and ZIP code are provided. Using nine-dimensional R vine copulas we investigate the dependence structure of the three different cost types between the three years. The likelihood of several R vine models with different dependence structures is compared. Finally, we fit a model containing only the costs from 2005 and 2006 and use it to **predict the health care costs** of the year 2007 and compare the predictions with the observed results.

This is joint work with Claudia Czado.

^{*}Technische Universität München, Zentrum Mathematik, Lehrstuhl für Mathematische Statistik, Parkring, 85748 Garching-Hochbrück, GERMANY.