Package ‘FactorCopula’

October 12, 2022

Version 0.8.1
Date 2022-06-14
Title Factor, Bi-Factor and Second-Order Copula Models
Author Sayed H. Kadhem [aut, cre],
       Aristidis K. Nikoloulopoulos [aut]
Maintainer Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Depends R (>= 3.5.0), statmod, abind, utils, polycor, VineCopula,
       matlab
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Date/Publication 2022-07-14 13:00:02 UTC

R topics documented:

FactorCopula-package .................................................. 2
discrepancy .................................................................. 3
GSS .............................................................................. 5
M2.Factor ................................................................. 6
M2.StructuredFactor ...................................................... 8
mapping ........................................................................ 10
mle.Factor ................................................................. 12
mle.StructuredFactor ..................................................... 14
PE .................................................................................. 17
rFactor ......................................................................... 18
rStructuredFactor .......................................................... 20
Select.Factor ............................................................... 22
Select.StructuredFactor .................................................. 24
Description

Estimation, model selection and goodness-of-fit of (1) factor copula models for mixed continuous and discrete data in Kadhem and Nikoloulopoulos (2021a); (2) bi-factor and second-order copula models for item response data in Kadhem and Nikoloulopoulos (2021b).

Details

This package contains R functions for:

- diagnostics based on semi-correlations (Kadhem and Nikoloulopoulos, 2021a,b; Joe, 2014) to detect tail dependence or tail asymmetry;
- diagnostics to show that a dataset has a factor structure based on linear factor analysis (Kadhem and Nikoloulopoulos, 2021a,b; Joe, 2014);
- estimation of the factor copula models in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015), and Kadhem and Nikoloulopoulos (2021a, 2021b);
- model selection of the factor copula models in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015) and Kadhem and Nikoloulopoulos (2021a, 2021b) using the heuristic algorithms in Kadhem and Nikoloulopoulos (2021a, 2021b) that automatically selects the bivariate parametric copula families that link the observed to the latent variables;
- goodness-of-fit of the factor copula models in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015) and Kadhem and Nikoloulopoulos (2021a, 2021b) using the \( M_2 \) statistic (Maydeu-Olivares and Joe, 2006). Note that the continuous and count data have to be transformed to ordinal.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Discrepancy Diagnostics to detect a factor dependence structure

Description

The diagnostic method in Joe (2014, pages 245-246) to show that each dataset has a factor structure based on linear factor analysis. The correlation matrix $R_{\text{observed}}$ has been obtained based on the sample correlations from the bivariate pairs of the observed variables. These are the linear (when both variables are continuous), polychoric (when both variables are ordinal), and polyserial (when one variable is continuous and the other is ordinal) sample correlations among the observed variables. The resulting $R_{\text{observed}}$ is generally positive definite if the sample size is not small enough; if not one has to convert it to positive definite. We calculate various measures of discrepancy between $R_{\text{observed}}$ and $R_{\text{model}}$ (the resulting correlation matrix of linear factor analysis), such as the maximum absolute correlation difference $D_1 = \max |R_{\text{model}} - R_{\text{observed}}|$, the average absolute correlation difference $D_2 = \text{avg} |R_{\text{model}} - R_{\text{observed}}|$, and the correlation matrix discrepancy measure $D_3 = \log(\det(R_{\text{model}})) - \log(\det(R_{\text{observed}})) + \text{tr}(R_{\text{model}}R_{\text{observed}})^{-1} - d$.

Usage

discrepancy(cormat, n, f3)

Arguments

cormat $R_{\text{observed}}$
n Sample size.
f3 If TRUE, then the linear 3-factor analysis is fitted.

Value

A matrix with the calculated discrepancy measures for different number of factors.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>
References


Examples

#-------------------------------------------------
# PE Data
#------------------ -----------------
data(PE)
#correlation
continuous.PE1 <- -PE[,1]
continuous.PE <- cbind(continuous.PE1, PE[,2])
u.PE <- apply(continuous.PE, 2, rank)/(nrow(PE)+1)
z.PE <- qnorm(u.PE)
categorical.PE <- data.frame(apply(PE[, 3:5], 2, factor))
nPE <- cbind(z.PE, categorical.PE)

#-------------------------------------------------
# Discrepancy measures---------------------------------
#-------------------------------------------------
cormat.PE <- as.matrix(polycor::hetcor(nPE, std.err=FALSE))
out.PE = discrepancy(cormat.PE, n = nrow(nPE), f3 = FALSE)

#-------------------------------------------------
#-------------------------------------------------
# GSS Data
#------------------ -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME, AGE)
continuous.GSS <- apply(continuous.GSS, 2, rank)/(nrow(GSS)+1)
z.GSS <- qnorm(continuous.GSS)
ordinal.GSS <- cbind(DEGREE, PINCOME, PDEGREE)
count.GSS <- cbind(CHILDREN, PCHILDREN)

# Transforming the count variables to ordinal
# count1 : CHILDREN
count1 = count.GSS[,1]
count1[count1 > 3] = 3

# count2: PCHILDREN
count2 = count.GSS[,2]
count2[count2 > 7] = 7

# Combining both transformed count variables
nncount.GSS = cbind(count1, count2)

# Combining ordinal and transformed count variables
gss <- cbind(ordinal.GSS, ncount.GSS)
gss <- data.frame(apply(gss, 2, factor))
# combining continuous and categorical variables
nGSS = cbind(z.GSS, gss)
#-------------------------------------------------
# Discrepancy measures----------------------------
#-------------------------------------------------
#correlation matrix for mixed data
cormat.GSS <- as.matrix(polycor::hetcor(nGSS, std.err=FALSE))
#discrepancy measures
out.GSS = discrepancy(cormat.GSS, n = nrow(nGSS), f3 = TRUE)

GSS

The 1994 General Social Survey

Description
Hoff (2007) analysed seven demographic variables of 464 male respondents to the 1994 General Social Survey. Of these seven, two were continuous (income and age of the respondents), three were ordinal with 5 categories (highest degree of the survey respondent, income and highest degree of respondent’s parents), and two were count variables (number of children of the survey respondent and respondent’s parents).

Usage
data(GSS)

Format
A data frame with 464 observations on the following 7 variables:
INCOME Income of the respondent in 1000s of dollars, binned into 21 ordered categories.
DEGREE Highest degree ever obtained (0:None, 1:HS, 2:Associates, 3:Bachelors, 4:Graduate).
CHILDREN Number of children of the survey respondent.
PINCOME Financial status of respondent’s parents when respondent was 16 (on a 5-point scale).
PDEGREE Highest degree of the survey respondent’s parents (0:None, 1:HS, 2:Associates, 3:Bachelors, 4:Graduate).
PCHILDREN Number of children of the survey respondent’s parents - 1.
AGE Age of the respondents in years.

Source
Description

The limited information $M_2$ statistic (Maydeu-Olivares and Joe, 2006) of factor copula models for mixed continuous and discrete data.

Usage

M2.1F(tcontinuous, ordinal, tcount, cpar, copF1, gl)
M2.2F(tcontinuous, ordinal, tcount, cpar, copF1, copF2, gl, SpC)

Arguments

tcontinuous $n \times d_1$ matrix with the transformed continuous to ordinal response data, where $n$ and $d_1$ is the number of observations and transformed continuous variables, respectively.

ordinal $n \times d_2$ matrix with the ordinal response data, where $n$ and $d_2$ is the number of observations and ordinal variables, respectively.

tcount $n \times d_3$ matrix with the transformed count to ordinal response data, where $n$ and $d_3$ is the number of observations and transformed count variables, respectively.

cpar A list of estimated copula parameters.

copF1 $(d_1 + d_2 + d_3)$-vector with the names of bivariate copulas that link the each of the observed variables with the 1st factor. Choices are “bvn” for BVN, “bvt” with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1joe” for 1-reflected Joe, “2joe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.

copF2 $(d_1 + d_2 + d_3)$-vector with the names of bivariate copulas that link the each of the observed variables with the 2nd factor. Choices are “bvn” for BVN, “bvt” with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1joe” for 1-reflected Joe, “2joe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.

gl Gauss legendre quadrature nodes and weights.

SpC Special case for the 2-factor copula model with BVN copulas. Select a bivariate copula at the 2nd factor to be fixed to independence. e.g. “SpC = 1” to set the first copula at the 2nd factor to independence.
Details
The $M_2$ statistic has been developed for goodness-of-fit testing in multidimensional contingency tables by Maydeu-Olivares and Joe (2006). Nikoloulopoulos and Joe (2015) have used the $M_2$ statistic to assess the goodness-of-fit of factor copula models for ordinal data. We build on the aforementioned papers and propose a methodology to assess the overall goodness-of-fit of factor copula models for mixed continuous and discrete responses. Since the $M_2$ statistic has been developed for multivariate ordinal data, we propose to first transform the continuous and count variables to ordinal and then calculate the $M_2$ statistic at the maximum likelihood estimate before transformation.

Value
A list containing the following components:

- **$M_2$**: The $M_2$ statistic which has a null asymptotic distribution that is $\chi^2$ with $s - q$ degrees of freedom, where $s$ is the number of univariate and bivariate margins that do not include the category 0 and $q$ is the number of model parameters.
- **df**: $s - q$.
- **p-value**: The resultant $p$-value.

Author(s)
Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples
```r
# Setting quadreture points
nq <- 25
gl <- gauss.quad.prob(nq)
# PE Data
#------------------ -----------------
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE2 = PE[,2]
continuous.PE <- cbind(continuous.PE1, continuous.PE2)
```
M2.StructuredFactor

**Goodness-of-fit of bi-factor and second-order copula models for item response data**

**Description**

The limited information $M_2$ statistic (Maydeu-Olivares and Joe, 2006) of bi-factor and second-order copula models for item response data.

**Usage**

```r
M2Bifactor(y, cpar, copnames1, copnames2, gl, ngrp, grpsize)
M2Second_order(y, cpar, copnames1, copnames2, gl, ngrp, grpsize)
```

**Arguments**

- **y**: $n \times d$ matrix with the ordinal response data, where $n$ and $d$ is the number of observations and variables, respectively.
- **cpar**: A list of estimated copula parameters.
- **copnames1**: For the bi-factor copula: $d$-vector with the names of bivariate copulas that link each of the observed variables with the common factor. For the second-order factor copula: $G$-vector with the names of bivariate copulas that link the each of the group-specific factors with the common factor, where $G$ is the number of groups of items. Choices are “bvn” for BVN, “bvt$\nu$” with $\nu = \{2, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.
copnames2  

For the bi-factor copula: \( d \)-vector with the names of bivariate copulas that link the each of the observed variables with the group-specific factor. For the second-order factor copula: \( d \)-vector with the names of bivariate copulas that link the each of the observed variables with the group-specific factor. Choices are “bvn” for BVN, “bvt\( \nu \)” with \( \nu = \{2, \ldots, 9\} \) degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.

gl  

Gauss legendre quadrature nodes and weights.

ngrp  

number of non-overlapping groups.

grpsize  

vector indicating the size for each group, e.g., c(4,4,4) indicating four items in all three groups.

Details

The \( M_2 \) statistic has been developed for goodness-of-fit testing in multidimensional contingency tables by Maydeu-Olivares and Joe (2006). We use the \( M_2 \) to assess the overall fit for the bi-factor and second-order copula models for item response data (Kadhem & Nikoloulopoulos, 2021).

Value

A list containing the following components:

\( M_2 \)  

The \( M_2 \) statistic which has a null asymptotic distribution that is \( \chi^2 \) with \( s - q \) degrees of freedom, where \( s \) is the number of univariate and bivariate margins that do not include the category 0 and \( q \) is the number of model parameters.

\( df \)  

\( s - q \).

\( p \)-value  

The resultant \( p \)-value.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples

```r
#------------------------------------------------
# Setting quadreture points
nq <- 15
g1 <- gauss.quad.prob(nq)
#------------------------------------------------
```
# TAS Data

```r
#using a subset of the data

#group1: 1,3,6,7,9,13,14
g1=c(1,3,6)

#group2: 2,4,11,12,17

g2=c(2,4,11)

#group3: 5,8,10,15,16,18,19,20

g3=c(5,8,10)

#Rearrange items within testlets

set.seed(123)
i=sample(1:nrow(TAS),500)
ydat=TAS[i,c(g1,g2,g3)]

d=ncol(ydat);d

#size of each group

g1=length(g1)
g2=length(g2)
g3=length(g3)

#number of groups

ngrp=length(c(g1,g2,g3))
```

```r
#M2

#BI-FACTOR

tauX0 = c(0.49,0.16,0.29,#0.09,0.47,0.49,0.30,
0.46,0.41,0.33,#0.29,0.24,
0.10,0.16,0.14),#0.12,0.03,0.03,0.10,0.10)

tauXg = c(0.09,0.37,0.23,#0.53,0.24,0.32,0.27,
0.53,0.58,0.20,#0.23,0.25,0.34,0.33,
0.30,0.19,0.24),#0.29,0.43,0.26)

copX0 = rep("bvt2", d)
copXg = c(rep("rgum", g1), rep("bvt3", g2+g3))

#converting taus to cpars

cparX0=mapply(function(x,y) tau2par(x,y),x=copX0,y=tauX0)
cparXg=mapply(function(x,y) tau2par(x,y),x=copXg,y=tauXg)
cpar=c(cparX0,cparXg)

m2_Bifactor = M2Bifactor(y=ydat, cpar, copX0, copXg, g1, ngrp, gsize)
```

---

**mapping**  
Mapping of Kendall’s tau and copula parameter
mapping

Description
Bivariate copulas: mapping of Kendall’s tau and copula parameter.

Usage
par2tau(copulaname, cpar)
tau2par(copulaname, tau)

Arguments
cpar Copula parameter(s).
tau Kendall’s tau.

Value
Kendall’s tau or copula parameter.

References

Examples

# 1-param copulas
#BVN copula
cpar.bvn = tau2par("bvn", 0.5)
tau.bvn = par2tau("bvn", cpar.bvn)

#Frank copula
cpar.frk = tau2par("frk", 0.5)
tau.frk = par2tau("frk", cpar.frk)

#Gumbel copula
cpar.gum = tau2par("gum", 0.5)
tau.gum = par2tau("gum", cpar.gum)

#Joe copula
mle.Factor

Maximum likelihood estimation of factor copula models for mixed data

Description

We use a two-stage estimation approach toward the estimation of factor copula models for mixed continuous and discrete data.

Usage

mle1factor(continuous, ordinal, count, copF1, gl, hessian, print.level)
mle2factor(continuous, ordinal, count, copF1, copF2, gl, hessian, print.level)
mle2factor.bvn(continuous, ordinal, count, copF1, copF2, gl, SpC, print.level)

Arguments

continuous  \( n \times d_1 \) matrix with the continuous response data, where \( n \) and \( d_1 \) is the number of observations and continuous variables, respectively.

ordinal  \( n \times d_2 \) matrix with the ordinal response data, where \( n \) and \( d_2 \) is the number of observations and ordinal variables, respectively.

count  \( n \times d_3 \) matrix with the count response data, where \( n \) and \( d_3 \) is the number of observations and count variables, respectively.

copF1  \( (d_1 + d_2 + d_3) \)-vector with the names of bivariate copulas that link the each of the observed variables with the 1st factor. Choices are “bvn” for BVN, “bvt\( \nu \)” with \( \nu = \{1, \ldots, 9\} \) degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1rjoe” for 1-reflected Joe, “2rjoe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.

g1 Gauss legendre quadrature nodes and weights.

SpC Special case for the 2-factor copula model with BVN copulas. Select a bivariate copula at the 2nd factor to be fixed to independence. e.g. “SpC = 1” to set the first copula at the 2nd factor to independence.

hessian If TRUE, the hessian of the negative log-likelihood is calculated during the minimization process.

print.level Determines the level of printing which is done during the minimization process; same as in nlm.

Details

Estimation is achieved by maximizing the joint log-likelihood over the copula parameters with the univariate parameters/distributions fixed as estimated at the first step of the proposed two-step estimation approach.

Value

A list containing the following components:

cutpoints The estimated univariate cutpoints (fitting the univariate probit model).
negbinest The estimated univariate parameters for the count responses (fitting the negative binomial distribution).
loglik The maximized joint log-likelihood.
cpar Estimated copula parameters in a list form.
taus The estimated copula parameters in Kendall’s tau scale.
SEs The SEs of the Kendall’s tau estimates.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples

# Setting quadrature points
nq <- 25
gl <- gauss.quad.prob(nq)

# PE Data
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE2 = PE[,2]
continuous.PE <- cbind(continuous.PE1, continuous.PE2)
categorical.PE <- PE[, 3:5]

## Estimation
## One-factor copula model

cop1f.PE <- c("joe", "joe", "rjoe", "joe", "gum")
est1factor.PE <- mle1factor(continuous.PE, categorical.PE,
                         count=NULL, copFl=cop1f.PE, gl, hessian = TRUE)
est1factor.PE

# GSS Data
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME, AGE)
ordinal.GSS <- cbind(DEGREE, PINCOME, PDEGREE)
count.GSS <- cbind(CHILDREN, PCHILDREN)

## Estimation
## One-factor copula model

cop1f.GSS <- c("joe","2rjoe","bvt3","bvt3",
               "rgum","2rjoe","2rgum")
est1factor.GSS <- mle1factor(continuous.GSS, ordinal.GSS,
                         count.GSS, copFl = cop1f.GSS, gl, hessian = TRUE)

---

*mle.StructuredFactor*  
Maximum likelihood estimation of the bi-factor and second-order copula models for item response data
Description

We approach the estimation of the bi-factor and second-order copula models for item response data with the IFM method of Joe (2005).

Usage

\begin{verbatim}
mleBifactor(y, copnames1, copnames2, gl, ngrp, grpsize, hessian, print.level)
mleSecond_order(y, copnames1, copnames2, gl, ngrp, grpsize, hessian, print.level)
\end{verbatim}

Arguments

\begin{itemize}
  \item \texttt{y} \hspace{1cm} n \times d matrix with the item response data, where \( n \) and \( d \) is the number of observations and variables, respectively.
  \item \texttt{copnames1} \hspace{1cm} For the bi-factor copula: \( d \)-vector with the names of bivariate copulas that link the each of the oberved variables with the common factor. For the second-order factor copula: \( G \)-vector with the names of bivariate copulas that link the each of the group-specific factors with the common factor, where \( G \) is the number of groups of items. Choices are “bvn” for BVN, “bvt\( \nu \)” with \( \nu = \{1, \ldots, 9\} \) degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.
  \item \texttt{copnames2} \hspace{1cm} For the bi-factor copula: \( d \)-vector with the names of bivariate copulas that link the each of the oberved variables with the group-specific factor. For the second-order factor copula: \( d \)-vector with the names of bivariate copulas that link the each of the oberved variables with the group-specific factor. Choices are “bvn” for BVN, “bvt\( \nu \)” with \( \nu = \{1, \ldots, 9\} \) degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.
  \item \texttt{gl} \hspace{1cm} Gauss legendre quadrature nodes and weights.
  \item \texttt{ngrp} \hspace{1cm} number of non-overlapping groups.
  \item \texttt{grpsize} \hspace{1cm} vector indicating the size for each group, e.g., \texttt{c(4,4,4)} indicating four items in all three groups.
  \item \texttt{hessian} \hspace{1cm} If TRUE, the hessian of the negative log-likelihood is calculated during the minimization process.
  \item \texttt{print.level} \hspace{1cm} Determines the level of printing which is done during the minimization process; same as in \texttt{nlm}.
\end{itemize}

Details

Estimation is achieved by maximizing the joint log-likelihood over the copula parameters with the univariate cutpoints fixed as estimated at the first step of the proposed two-step estimation approach.
Value

A list containing the following components:

- **cutpoints**: The estimated univariate cutpoints (fitting the univariate probit model).
- **taus**: The estimated copula parameters in Kendall's tau scale.
- **SEs**: The SEs of the Kendall's tau estimates.
- **loglik**: The maximized joint log-likelihood.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples

```r
# Setting quadreture points
nq <- 25
gl <- gauss.quad.prob(nq)

# TAS Data
# ----------------- -----------------
# using a subset of the data
# Group1: 1,3,6,7,9,13,14
g1 <- c(1,3,6)
# Group2: 2,4,11,12,17
g2 <- c(2,4,11)
# Group3: 5,8,10,15,16,18,19,20
g3 <- c(5,8,10)
# Rearrange items within testlets
set.seed(123)
i <- sample(1:nrow(TAS),500)
ydat <- TAS[i,c(g1,g2,g3)]

d <- ncol(ydat); d
n <- nrow(ydat); n
```

Political-economic risk of 62 countries for the year 1987

Description

Quinn (2004) used 5 mixed variables, namely the continuous variable black-market premium in each country (used as a proxy for illegal economic activity), the continuous variable productivity as measured by real gross domestic product per worker in 1985 international prices, the binary variable independence of the national judiciary (1 if the judiciary is judged to be independent and 0 otherwise), and the ordinal variables measuring the lack of expropriation risk and lack of corruption.

Usage

data(PE)

Format

A data frame with 62 observations (countries) on the following 5 variables:

- BM Black-market premium.
- GDP Gross domestic product.
- IJ Independent judiciary.
- XPR Lack of expropriation risk.
- CPR Lack of corruption.

Source

Description

Simulating dependent standard uniform and ordinal response data from factor copula models.

Usage

```r
r1factor(n, d1, d2, categ, theta, copF1)
r2factor(n, d1, d2, categ, theta, delta, copF1, copF2)
```

Arguments

- **n**: Sample size.
- **d1**: Number of standard uniform variables.
- **d2**: Number of ordinal variables.
- **categ**: A vector of categories for the ordinal variables.
- **theta**: Copula parameters for the 1st factor.
- **delta**: Copula parameters for the 2nd factor.
- **copF1**: \((d_1 + d_2)\)-vector with the names of bivariate copulas that link the each of the observed variables with the 1st factor. Choices are “bvn” for BVN, “bvt[ν]” with \(ν = \{1, \ldots, 9\}\) degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1joe” for 1-reflected Joe, “2joe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.
- **copF2**: \((d_1 + d_2)\)-vector with the names of bivariate copulas that link the each of the observed variables with the 2nd factor. Choices are “bvn” for BVN, “bvt[ν]” with \(ν = \{1, \ldots, 9\}\) degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1joe” for 1-reflected Joe, “2joe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.

Value

Data matrix of dimension \(n \times d\), where \(n\) is the sample size, and \(d = d_1 + d_2\) is the total number of variables.

Author(s)

- Sayed H. Kadhem <s.kadhem@uea.ac.uk>
- Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>
References


Examples

```r
# One-factor copula model
n = 100
# Continuous Variables
d1 = 5
# Ordinal Variables
d2 = 3
# Categories for ordinal
categ = c(3, 4, 5)
theta = rep(2, d1+d2)
copnamesF1 = rep("gum", d1+d2)

# Simulating data
datF1 = r1factor(n, d1=d1, d2=d2, categ, theta, copnamesF1)

# Plotting continuous data
pairs(qnorm(datF1[, 1:d1]))
```

```r
# Two-factor copula model
n = 100
# Continuous Variables
d1 = 5
# Ordinal Variables
d2 = 3
# Categories for ordinal
```
categ = c(3,4,5)

# Copula parameters ---------------------------------
theta = rep(2.5, d1+d2)
delta = rep(1.5, d1+d2)

# Copula names --------------------------------------
copnamesF1 = rep("gum", d1+d2)
copnamesF2 = rep("gum", d1+d2)

#----------------- Simulating data -------------------
datF2 = r2factor(n, d1=d1, d2=d2, categ, theta, delta,
copnamesF1, copnamesF2)

#----------------- Plotting data ----------------------
pairs(qnorm(datF2[,1:d1]))

rStructuredFactor: Simulation of bi-factor and second-order copula models for item response data

Description
Simulating dependent item response data from the bi-factor and second-order copula models for item response data.

Usage
rBifactor(n, d, grpsize, categ, copnames1, copnames2, theta1, theta2)
rSecond_order(n, d, grpsize, categ, copnames1, copnames2, theta1, theta2)

Arguments

n Sample size.
d Number of observed variables/items.
grpsize vector indicating the size for each group, e.g., c(4,4,4) indicating four items in all three groups.
categ A vector of categories for the observed variables/items.
theta1 For the bi-factor model: copula parameter vector of size \( d \) for items with the common factor. For the second-order copulas: copula parameter vector of size \( G \) for the common factor and group-specific factors.
theta2 For the bi-factor model: copula parameter vector of size \( d \) for items with the group-specific factor. For the second-order copulas: copula parameter vector of size \( d \) for items with the group-specific factor.
For the bi-factor copula: $d$-vector with the names of bivariate copulas that link each of the observed variables with the common factor. For the second-order factor copula: $G$-vector with the names of bivariate copulas that link each of the group-specific factors with the common factor, where $G$ is the number of groups of items. Choices are “bv” for BV, “bvt$\nu$” with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.

For the bi-factor copula: $d$-vector with the names of bivariate copulas that link each of the observed variables with the group-specific factor.

For the second-order factor copula: $d$-vector with the names of bivariate copulas that link each of the observed variables with the group-specific factor. Choices are “bv” for BV, “bvt$\nu$” with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.

Value

Data matrix of dimension $n \times d$, where $n$ is the sample size, and $d$ is the total number of observed variables/items.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples

```r
# ---------------------------------------------------
# ---------------------------------------------------
#Sample size
n = 500

#Ordinal Variables ---------------------------------
d = 9
grpsize=c(3,3,3)
ngrp=length(grpsize)

categ = rep(3,d)

#Bi-factor copula model
```

```r
# ---------------------------------------------------
```
# Copula parameters
theta = rep(2.5, d)
delta = rep(1.5, d)

# Copula names
copulanames1 = rep("gum", d)
copulanames2 = rep("gum", d)

# Simulating data
data_Bifactor = rBifactor(n, d, grpsize, categ, copulanames1, copulanames2, theta, delta)

# Second-order copula model
# Copula parameters
theta = rep(1.5, ngrp)
delta = rep(2.5, d)

copulanames1 = rep("gum", ngrp)
copulanames2 = rep("gum", d)

data_Second_order = rSecond_order(n, d, grpsize, categ, copulanames1, copulanames2, theta, delta)

---

Select.Factor

Model selection of the factor copula models for mixed data

Description

A heuristic algorithm that automatically selects the bivariate parametric copula families that link the observed to the latent variables.

Usage

```
select1F(continuous, ordinal, count, copnamesF1, gl)
select2F(continuous, ordinal, count, copnamesF1, copnamesF2, gl)
```

Arguments

- **continuous**: \( n \times d_1 \) matrix with the continuous response data, where \( n \) and \( d_1 \) is the number of observations and continuous variables, respectively.
- **ordinal**: \( n \times d_2 \) matrix with the ordinal response data, where \( n \) and \( d_2 \) is the number of observations and ordinal variables, respectively.
- **count**: \( n \times d_3 \) matrix with the count response data, where \( n \) and \( d_3 \) is the number of observations and count variables, respectively.
Select.Factor


copnamesF2


gl

Gauss legendre quadrature nodes and weights.

Details

The linking copulas at each factor are selected with a sequential algorithm under the initial assumption that linking copulas are Frank, and then sequentially copulas with non-tail quadrant independence are assigned to any of pairs where necessary to account for tail asymmetry (discrete data) or tail dependence (continuous data).

Value

A list containing the following components:

'1st factor'

The selected bivariate linking copulas for the 1st factor.

'2nd factor'

The selected bivariate linking copulas for the 2nd factor.

AIC

Akaike information criterion.

taus

The estimated copula parameters in Kendall’s tau scale.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Examples

```r
#------------------------------------------------
# Estimation
#------------------ -----------------
# Setting quadreture points
nq<-25
gl<-gauss.quad.prob(nq)
#------------------------------------------------
# PE Data
#------------------ -----------------
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE <- cbind(continuous.PE1, PE[,2])
categorical.PE <- PE[, 3:5]
#------------------ One-factor -----------------
# listing the possible copula candidates:
d <- ncol(PE)
copulasF1 <- rep(list(c("bvn", "bvt3", "bvt5", "frk", "gum", "rgum", "rjoe","joe", "1rjoe","2rjoe","1rgum","2rgum")), d)
out1F.PE <- select1F(continuous.PE, categorical.PE, count=NULL, copulasF1, gl)
```

**Select.StructuredFactor**

Model selection of the bi-factor and second-order copula models for item response data

**Description**

A heuristic algorithm that automatically selects the bivariate parametric copula families for the bi-factor and second-order copula models for item response data.

**Usage**

```r
selectBifactor(y, grpsize, copnames, gl)
selectSecond_order(y, grpsize, copnames, gl)
```

**Arguments**

- `y` \( n \times d \) matrix with the item response data, where \( n \) and \( d \) is the number of observations and variables, respectively.
grpsize  vector indicating the size for each group, e.g., c(4,4,4) indicating four items in all three groups.

copnames  A vector with the names of possible candidates of bivariate copulas to be selected for the bi-factor and second-order copula models for item response data. Choices are “bvn” for BVN, “bvtν” with ν = {1,...,9} degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel.

g1  Gauss legendre quadrature nodes and weights.

Details

The linking copulas at each factor are selected with a sequential algorithm under the initial assumption that linking copulas are BVN, and then sequentially copulas with tail dependence are assigned to any of pairs where necessary to account for tail asymmetry.

Value

A list containing the following components:

‘‘common factor’’
The selected bivariate linking copulas for the common factor (Bi-factor: copulas link items with the common factor. Second-order: copulas link group-specific factors with the common factor).

‘‘group-specific factor’’
The selected bivariate linking copulas for the items with group-specific factors.

log-likelihood  The maximized joint log-likelihood.

taus  The estimated copula parameters in Kendall’s tau scale.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples

#------------------------------------------------
# Setting quadreture points
nq <- 15
g1 <- gauss.quad.prob(nq)
#------------------------------------------------
# TAS Data
#------------------ -----------------
data(TAS)
# using a subset of the data
# group1: 1,3,6,7,9,13,14
g1 = c(1,3)
# group2: 2,4,11,12,17

g2 = c(2,4)
# group3: 5,8,10,15,16,18,19,20

g3 = c(5,8)
# Rearrange items within testlets
set.seed(123)
i = sample(1:nrow(TAS), 500)
ydat = TAS[i, c(g1, g2, g3)]

# size of each group
g1 = length(grp1)
g2 = length(grp2)
g3 = length(grp3)
grpsize = c(g1, g2, g3)

# listing possible copula candidates:
copnames = c("bvt2", "bvt3")

Bifactor_model = selectBifactor(ydat, grpsize, copnames, gl)

semicorr

### Diagnostics to detect tail dependence or tail asymmetry.

**Description**

The sample versions of the correlation $\rho_N$, upper semi-correlation $\rho^+ _N$ (correlation in the joint upper quadrant) and lower semi-correlation $\rho^- _N$ (correlation in the joint lower quadrant). These are sample linear (when both variables are continuous), polychoric (when both variables are ordinal), and polyserial (when one variable is continuous and the other is ordinal) correlations.

**Usage**

`semicorr(dat, type)`

**Arguments**

- `dat`: Data frame of mixed continuous and ordinal data.
- `type`: A vector with 1’s for the location of continuous data and 2’s for the location of ordinal data.

**Value**

A matrix containing the following components for `semicorr()`:

- `rho`: $\rho_N$.
- `lrho`: $\rho^- _N$.
- `urho`: $\rho^+ _N$. 
Examples

```
#------------------------------------------------
# PE Data
#------------------ -----------------
data(PE)
#correlation
continuous.PE1 <- -PE[,1]
continuous.PE <- cbind(continuous.PE1, PE[,2])
categorical.PE <- data.frame(apply(PE[, 3:5], 2, factor))
nPE <- cbind(continuous.PE, categorical.PE)

#------------------------------------------------
# Semi-correlations-------------------------------
#-------------------------------------------------
# Exclude the dichotomous variable
sem.PE = nPE[, -3]
semicorr.PE = semicorr(dat = sem.PE, type = c(1,1,2,2))

#------------------------------------------------
#------------------------------------------------
# GSS Data
#------------------ -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME, AGE)
ordinal.GSS <- cbind(DEGREE, PINCOME, PDEGREE)
count.GSS <- cbind(CHILDREN, PCHILDREN)

# Transforming the COUNT variables to ordinal
# count1 : CHILDREN
count1 = count.GSS[, 1]
count1[count1 > 3] = 3

# count2: PCHILDREN
count2 = count.GSS[, 2]
count2[count2 > 7] = 7

# Combining both transformed count variables
ncount.GSS = cbind(count1, count2)
```
# Combining ordinal and transformed count variables
categorical.GSS <- cbind(ordinal.GSS, ncount.GSS)
categorical.GSS <- data.frame(apply(categorical.GSS, 2, factor))

# combining continuous and categorical variables
nGSS = cbind(continuous.GSS, categorical.GSS)

# Semi-correlations
semicorr.GSS = semicorr(dat = nGSS, type = c(1, 1, rep(2,5)))

---

**Toronto Alexithymia Scale (TAS)**

**Description**

The Toronto Alexithymia Scale is the most utilized measure of alexithymia in empirical research and is composed of $d = 20$ items that can be subdivided into $G = 3$ non-overlapping groups: $d_1 = 7$ items to assess difficulty identifying feelings (DIF), $d_2 = 5$ items to assess difficulty describing feelings (DDF) and $d_3 = 8$ items to assess externally oriented thinking (EOT). Students were 17 to 25 years old and 58% of them were female and 42% were male. They were asked to respond to each item using one of $K = 5$ categories: “1 = completely disagree”, “2 = disagree”, “3 = neutral”, “4 = agree”, “5 = completely agree”.

**Usage**

data(TAS)

**Format**

A data frame with 1925 observations on the following 20 items:

- **DIF** items: 1,3,6,7,9,13,14.
- **DDF** items: 2,4,11,12,17.
- **EOT** items: 5,8,10,15,16,18,19,20.

**Source**


transformation  

Continuous/count to ordinal responses

Description
Transforming a continuous/count to ordinal variable with $K$ categories.

Usage

```
continuous2ordinal(continuous, categ)
count2ordinal(count, categ)
```

Arguments

- **continuous**: Matrix of continuous data.
- **count**: Matrix of count data.
- **categ**: The number of categories $K$.

Examples

```
# PE Data
data(PE)
continuous.PE <- PE[, 1:2]
# Transforming the continuous to ordinal data:
tcontinuous = continuous2ordinal(continuous.PE, categ=5)
table(tcontinuous)
# Transforming the count to ordinal data:
set.seed(12345)
count.data = rpois(1000, 3)
tcount = count2ordinal(count.data, 5)
table(tcount)
```

Vuong.Factor  

Vuong’s test for the comparison of factor copula models

Description
Vuong (1989)’s test for the comparison of non-nested factor copula models for mixed data. We compute the Vuong’s test between the factor copula model with BVN copulas (that is the standard factor model) and a competing factor copula model to reveal if the latter provides better fit than the standard factor model.
Usage

vuong.1f(cpar.bvn, cpar, copF1, continuous, ordinal, count, gl, param)
vuong.2f(cpar.bvn, cpar, copF1, copF2, continuous, ordinal, count, gl, param)

Arguments

cpar.bvn  copula parameters of the factor copula model with BVN copulas.
cpar      copula parameters of the competing factor copula model.
copF1     copula names for the first factor of the competing factor copula model.
copF2     copula names for the second factor of the competing factor copula model.
continuous matrix of continuous data.
ordinal    matrix of ordinal data.
count      matrix of count data.
gl         gauss-legendre quadrature points.
param      parameterization of estimated copula parameters. If FALSE, then cpar are the actual copula parameters without any transformation/reparameterization.

Value

A vector containing the following components:

z         the test statistic.
p.value   the p-value.
CI.left   lower/left endpoint of 95% confidence interval.
CI.right  upper/right endpoint of 95% confidence interval.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References


Examples

```r
#------------------------------------------------
# Setting quadreture points
nq <- 25
gl <- gauss.quad.prob(nq)
#------------------------------------------------
```
Vuong.StructuredFactor

Vuong's test for the comparison of bi-factor and second-order copula models

Description

The Vuong's test (Vuong, 1989) is the sample version of the difference in Kullback-Leibler divergence between two models and can be used to differentiate two parametric models which could be non-nested. For the Vuong's test we provide the 95% confidence interval of the Vuong’s test statistic (Joe, 2014, page 258). If the interval does not contain 0, then the best fitted model according to the AICs is better if the interval is completely above 0.

Usage

vuong_structured(models, cpar.m1, copnames.m1, cpar.m2, copnames.m2, y, grpsize)

Arguments

models

choose a number from (1,2,3,4). 1: Model1 is bifactor, Model2 is bifactor. 2: Model1 is second-order, Model2 is second-order. 3: Model1 is second-order, Model2 is bifactor. 4: Model1 is bifactor, Model2 is nested.
Vuong.StructuredFactor

**cpar.m1**
vector of copula parameters for model 1, starting with copula parameters that link the items with common factor (bifactor), or group factors with common factor (second-order).

**cpar.m2**
vector of copula parameters for model 2, starting with copula parameters that link the items with common factor (bifactor), or group factors with common factor (second-order).

**copnames.m1**
vector of names of copula families for model 1, starting with copulas that link the items with common factor (bifactor), or group factors with common factor (second-order).

**copnames.m2**
vector of names of copula families for model 2, starting with copulas that link the items with common factor (bifactor), or group factors with common factor (second-order).

**y**
matrix of ordinal data.

**grpsize**
vector indicating the size for each group, e.g., c(4,4,4) indicating four items in all three groups.

**Value**
A vector containing the following components:

- **z** the test statistic.
- **p.value** the \( p \)-value.
- **CI.left** lower/left endpoint of 95\% confidence interval.
- **CI.right** upper/right endpoint of 95\% confidence interval.

**Author(s)**
Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

**References**


**Examples**

```
#------------------------------------------------
# Setting quadreture points
nq <- 25
g1 <- gauss.quad.prob(nq)
#------------------------------------------------

# TAS Data
```
data(TAS)
grp1=c(1,3,6,7,9,13,14)
grp2=c(2,4,11,12,17)
grp3=c(5,8,10,15,16,18,19,20)
ydat=TAS[,c(grp1,grp2,grp3)]

d=ncol(ydat);d
n=nrow(ydat);n

#Rearrange items within testlets
g1=length(grp1)
g2=length(grp2)
g3=length(grp3)
grpsize=c(g1,g2,g3)#group size
#number of groups
ngrp=length(grpsize)

# M1 bifactor - M2 bifactor
cpar.m1 = rep(0.6,d*2)
copnames.m1 = rep("bvn",d*2)
cpar.m2 = rep(1.6,d*2)
copnames.m2 = rep("rgum",d*2)
vuong.bifactor = vuong_structured(models=1, cpar.m1, copnames.m1,
cpar.m2, copnames.m2,
y=ydat, grpsize)

# M1 second-order - M2 second-order
cpar.m1 = rep(0.6,d+ngrp)
copnames.m1 = rep("bvn",d+ngrp)
cpar.m2 = rep(1.6,d+ngrp)
copnames.m2 = rep("rgum",d+ngrp)
vuong.second_order = vuong_structured(models=2, cpar.m1,
copnames.m1, cpar.m2, copnames.m2, y=ydat, grpsize)

# M1 second-order - M2 bifactor
cpar.m1 = rep(0.6,d+ngrp)
copnames.m1 = rep("bvn",d+ngrp)
cpar.m2 = rep(1.6,d*2)
copnames.m2 = rep("rgum",d*2)
vuong.2ndO_bif = vuong_structured(models=3, cpar.m1, copnames.m1,
cpar.m2, copnames.m2,
y=ydat, grpsize)

# M1 bifactor - M2 second-order
cpar.m1 = rep(0.6,d*2)
copnames.m1 = rep("bvn",d*2)
cpar.m2 = rep(1.6,d+ngrp)
copnames.m2 = rep("rgum",d+ngrp)
vuong.bif_2ndO = vuong_structured(models=4, cpar.m1, copnames.m1,
cpar.m2, copnames.m2,
y=ydat, grpsize)
* correlation
  discrepancy, 3
  semicorr, 26
* datagen
  rFactor, 18
  rStructuredFactor, 20
* datasets
  GSS, 5
  PE, 17
  TAS, 28
* maximum likelihood
  Vuong.Factor, 29
  Vuong.StructuredFactor, 31
* models
  mle.Factor, 12
  mle.StructuredFactor, 14
* multivariate
  M2.Factor, 6
  M2.StructuredFactor, 8
  mle.Factor, 12
  mle.StructuredFactor, 14
  Select.Factor, 22
  Select.StructuredFactor, 24
* package
  FactorCopula-package, 2
* parameters
  mapping, 10
* univar
  transformation, 29
  continuous2ordinal (transformation), 29
  count2ordinal (transformation), 29
  discrepancy, 3
  FactorCopula-package, 2
  GSS, 5
  M2.1F (M2.Factor), 6
  M2.2F (M2.Factor), 6
  M2.Factor, 6
  M2.StructuredFactor, 8
  M2Bifactor (M2.StructuredFactor), 8
  M2Second_order (M2.StructuredFactor), 8
  mapping, 10
  mle.Factor, 12
  mle.StructuredFactor, 14
  mle1factor (mle.Factor), 12
  mle2factor (mle.Factor), 12
  mleBifactor (mle.StructuredFactor), 14
  mleSecond_order (mle.StructuredFactor), 14
  par2tau (mapping), 10
  PE, 17
  r1factor (rFactor), 18
  r2Factor (rFactor), 18
  rBifactor (rStructuredFactor), 20
  rFactor, 18
  rSecond_order (rStructuredFactor), 20
  rStructuredFactor, 20
  Select.Factor, 22
  Select.StructuredFactor, 24
  select1F (Select.Factor), 22
  select2F (Select.Factor), 22
  selectBifactor
    (Select.StructuredFactor), 24
  selectSecond_order
    (Select.StructuredFactor), 24
  semicorr, 26
  TAS, 28
  tau2par (mapping), 10
  transformation, 29
  vuong.1f (Vuong.Factor), 29
  vuong.2f (Vuong.Factor), 29
  Vuong.Factor, 29
  Vuong.StructuredFactor, 31
vuong_structured
    (Vuong.StructuredFactor), 31