

Package ‘Rcope’

March 3, 2026

Type Package

Title Tools to Cope with Endogeneity Problems

Version 1.0.0

Description Researchers across disciplines often face biased regression model estimates due to endogenous regressors correlated with the error term. Traditional solutions require instrumental variables (IVs), which are often difficult to find and validate. This package provides flexible, alternative IV-free methods using copulas, as described in the practical guide to endogeneity correction using copulas (Yi Qian, Tony Koschmann, and Hui Xie 2025) <[doi:10.1177/00222429251410844](https://doi.org/10.1177/00222429251410844)>. The current version implements the two-stage copula endogeneity correction (2sCOPE) method to fit models with continuous endogenous regressors and both continuous and discrete exogenous regressors, as described in Fan Yang, Yi Qian, and Hui Xie (2024) <[doi:10.1177/00222437241296453](https://doi.org/10.1177/00222437241296453)>. Using this method, users can address regressor endogeneity problems in nonexperimental data without requiring IVs.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports dplyr, Formula, car

RoxygenNote 7.3.3

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Config/Needs/quarto false

Depends R (>= 3.5)

NeedsCompilation no

Author Anthony Obrzut [aut, cre],

Yi Qian [aut],

Hui Xie [aut]

Maintainer Anthony Obrzut <anthony_obrzut@sfu.ca>

Repository CRAN

Date/Publication 2026-03-03 10:20:24 UTC

Contents

CCF	2
diapers	4
print.ccf	5
print.tscope	5
print.tscope.fit	6
relevance_test	6
tscope	7
tscope.fit	10

Index	12
--------------	-----------

CCF	<i>CCF: Copula Control Function</i>
-----	-------------------------------------

Description

CCF() computes copula control functions (CCFs) that can be added in the outcome model as control variables to correct for endogeneity. which returns P^* , W^* , and the first-stage residuals.

Usage

```
CCF(formula, data)
```

Arguments

formula	A formula describing the model to be fitted. The details of model specification are given under “Details”.
data	a data frame, list, or environment containing the variables in the model.

Details

The formula argument is either in the 1-bar form $Y \sim X \mid P$ or the 2-bar form $Y \sim X \mid P \mid W$, where X represents the explanatory variable(s) in the Y model, P represents the continuous endogenous regressors, and W represents the exogenous regressors. If X contains no exogenous regressors, then the 2sCOPE model reduces to the simpler model in Park and Gupta (2012) and returns P^* (the copula transformation of P) as CCF and W^* (the copula transformation of W) as null. When the structural outcome model includes an intercept, copula transformations of regressors in P and W use the optimized algorithm (Equation 9 in Qian, Koschmann, and Xie, 2025) to avoid estimation bias.

The function CCF() will compute copula control function for each endogenous regressor specified in P . Only first-order terms of endogenous regressors need to be included in P , even when the structural outcome model contains higher-order terms of endogenous regressors. This is because including copula control functions for the first-order endogenous regressors is sufficient to control for

endogeneity, while adding control functions for higher-order endogenous terms—such as interactions among endogenous regressors, interactions between endogenous and exogenous regressors, or squared endogenous regressors—is unnecessary and can substantially degrade the performance of copula correction (Qian, Koschmann, and Xie, 2025). This parsimonious treatment of higher-order endogenous regressors is a merit of copula correction.

Thus, if X contains no higher-order terms of endogenous regressors, the simpler 1-bar form $Y \sim X | P$ can be used, and `CCF()` treats all regressors in X except those in P as exogenous. When X includes higher-order endogenous terms, the 2-bar form $Y \sim X | P | W$ should be used to explicitly specify the exogenous regressors in W and ensure that the higher-order endogenous terms are not treated as exogenous variables.

Value

A list of class "ccf" containing the following components:

<code>ccf</code>	a matrix of the first-stage residuals as copula control functions.
<code>pstar</code>	a matrix representing P^*
<code>wstar</code>	a matrix representing W^*

References

- Qian, Y., Koschmann, A., & Xie, H. (2025). *EXPRESS: A Practical Guide to Endogeneity Correction Using Copulas*. *Journal of Marketing*. doi:10.1177/00222429251410844
- Park, S., & Gupta, S. (2012). *Handling endogenous regressors by joint estimation using copulas*. *Marketing Science*, 31(4), 567-586.
- Yang, F., Qian, Y., & Xie, H. (2025). *Addressing Endogeneity Using a Two-Stage Copula Generated Regressor Approach*. *Journal of Marketing Research*, 62(4), 601-623. doi:10.1177/00222437241296453

Examples

```
data("diapers") #load data

### Specify logPrice as endogenous using the 1-bar option
#run the copula control function
ccf_1bar <- CCF(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4|logPrice,data=diapers)
#print the first 5 elements of the first-stage residuals
head(ccf_1bar$ccf, 5)
head(ccf_1bar$pstar, 5) #print the first 5 elements of P*
head(ccf_1bar$wstar, 5) #print the first 5 elements of W*

### Specify logPrice as endogenous and the rest of the variables as exogenous
#using the 2-bar option, which will produce the same results,
ccf_2bar <- CCF(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4|logPrice|
  Fshare+week+Q2+Q3+Q4, data = diapers) #run the copula control function
head(ccf_2bar$ccf, 5) #print first 5 elements of the 1st-stage resid
head(ccf_2bar$pstar, 5) #print first 5 elements of P*
```

```

head(ccf_2bar$wstar, 5) #print first 5 elements of W*

### Run Park & Gupta (2012) by specifying logPrice as the only regressor,
### which is endogenous.

#run the copula control function
ccf_pg <- CCF(logVol ~ logPrice|logPrice, data = diapers)
head(ccf_pg$ccf, 5) #print first 5 elements of the 1st-stage resid
head(ccf_pg$pstar, 5) #print first 5 elements of P*
head(ccf_pg$wstar, 5) #print first 5 elements of W*
# notice how the 1st-stage residuals and P* are equivalent, and wstar is NULL

```

diapers

diapers

Description

This dataset is a modified dataset from Qian, Koschmann, Xie (2024). The purpose of this data is to evaluate the price endogeneity issue in diaper sales in Buffalo, NY from 2002-2006. Data was collected over 261 weeks. The data contains a response variable `logVol`, an endogenous explanatory variable `logPrice`, and exogenous explanatory variables `Fshare`, `week`, `Q2`, `Q3`, and `Q4`. Retail price, represented by `logPrice` in this data, is often considered endogenous in various marketing settings due to potential unmeasured product characteristics or demand shocks that can influence both consumers' and retailers' decisions. Further information on the dataset can be found Qian, Koschmann, Xie (2024).

Usage

diapers

Format

A data frame with 261 rows and 7 variables:

logVol numeric variable representing the log of total diapers sold in one week.

logPrice numeric variable representing the log of diaper retail price in American dollars.

Fshare numeric variable representing the category feature intensity.

week numeric variable representing the week number within the time-frame

Q2 binary variable representing the second quarter of the year

Q3 binary variable representing the third quarter of the year

Q4 binary variable representing the fourth quarter of the year

Source

Qian, Y., Koschmann, A., & Xie, H. (2025). *EXPRESS: A Practical Guide to Endogeneity Correction Using Copulas*. Journal of Marketing. doi:10.1177/00222429251410844

print.ccf	<i>Print method for CCF</i>
-----------	-----------------------------

Description

Print method for objects of class ccf

Usage

```
## S3 method for class 'ccf'  
print(x, ...)
```

Arguments

x	an object of class "ccf"
...	Additional arguments (currently ignored).

Value

No return value, prints contents of the "ccf" object.

print.tscope	<i>Print method for tscope</i>
--------------	--------------------------------

Description

Print method for objects of class tscope

Usage

```
## S3 method for class 'tscope'  
print(x, ...)
```

Arguments

x	an object of class "tscope"
...	Additional arguments (currently ignored).

Value

No return value, prints contents of the "tscope" object.

```
print.tscope.fit      Print method for tscope.fit
```

Description

Print method for objects of class `tscope.fit`

Usage

```
## S3 method for class 'tscope.fit'
print(x, ...)
```

Arguments

`x` an object of class "tscope.fit"
`...` Additional arguments (currently ignored).

Value

No return value, prints contents of the "tscope.fit" object.

```
relevance_test      relevance_test: Test the relevance of the exogenous regressors.
```

Description

This test is needed only if the endogenous regressor P is close to be normally distributed. In this case, 2sCOPE can leverage correlated exogenous regressors to achieve model identification. This function conducts a test for the relevance of exogenous regressor(s), i.e. the effect of W^* on P^* . Test statistics greater than 10 are reported in a table. The formula argument must be in the form $Y \sim X | P$ or $Y \sim X | P | W$, where X represents the explanatory variable(s), P represents the endogenous explanatory variable(s), and W represents the exogenous explanatory variable(s).

Usage

```
relevance_test(ccf_obj)
```

Arguments

`ccf_obj` an object of class `ccf` returned from the function `CCF()` containing the model of interest for which the relevance test will be conducted for.

Details

This test is needed only if the endogenous regressor P is close to be normally distributed. If the endogenous regressor P is found to have insufficient nonnormality (the Kolmogorov-Smirnov (KS) normality test p -value > 0.05), then 2sCOPE can leverage correlated exogenous regressors to achieve model identification. To compensate for the lack of nonnormality of endogenous regressor P , at least one exogenous and continuous regressor W needs to satisfy the following two conditions: (1) sufficient nonnormality, and (2) sufficient association with the endogenous regressor P . A conservative rule of thumb for such a W is the p -value from the KS test on W being < 0.001 and a sufficient association with P (F statistic for the effect of W^* on P^* > 10 in the first-stage regression). This function `relevance_test()` checks the condition (2) above. When these conditions are met, 2sCOPE is expected to yield consistent estimates even if P is normally distributed. When these conditions are not met, Yang, Qian, and Xie (2025) suggest gauging potential bias of 2sCOPE for data at hand via a bootstrap procedure described there, and using 2sCOPE only if the potential bias is small. In order for this function to work as intended, the user must supply a `ccf` object as an argument to the function.

Value

No return value; prints out the results of the relevance test.

References

Qian, Y., Koschmann, A., & Xie, H. (2025). *EXPRESS: A Practical Guide to Endogeneity Correction Using Copulas*. Journal of Marketing. doi:10.1177/00222429251410844

Yang, F., Qian, Y., & Xie, H. (2025). *Addressing Endogeneity Using a Two-Stage Copula Generated Regressor Approach*. Journal of Marketing Research, 62(4), 601-623. doi:10.1177/00222437241296453

Examples

```
data("diapers") #load data

### Specify logPrice as endogenous using the 1-bar option,
#run the copula control function
cop_ctrl_fn <- CCF(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4|logPrice,
                  data = diapers)

relevance_test(cop_ctrl_fn) #run relevance test
```

Description

Fit the two-stage copula endogeneity (2sCOPE) control function regression for addressing regressor endogeneity.

Usage

```
tscope(formula, data, nboot = 500)
```

Arguments

formula	a formula describing the model to be fitted. The details of model specification are given under “Details”.
data	a data frame, list, or environment containing the variables in the model.
nboot	a numeric value representing the number of desired bootstrap samples taken to compute the standard errors of the 2sCOPE model estimates. nboot = 1 will not compute any standard errors, only parameter estimates.

Details

The formula argument is either in the 1-bar form $Y \sim X \mid P$ or the 2-bar form $Y \sim X \mid P \mid W$, where X represents the explanatory variable(s) in the Y model, P represents the continuous endogenous regressors, and W represents the exogenous regressors. If X contains no exogenous regressors, then the 2sCOPE model reduces to the simpler model in Park and Gupta (2012) and returns P^* (the copula transformation of P) as CCF and W^* (the copula transformation of W) as null. When the structural outcome model includes an intercept, copula transformations of regressors in P and W use the optimized algorithm (Equation 9 in Qian, Koschmann, and Xie, 2025) to avoid estimation bias.

The function will add copula control function for each endogenous regressor specified in P . Only first-order terms of endogenous regressors need to be included in P , even when the structural outcome model contains higher-order terms of endogenous regressors. This is because including copula control functions for the first-order endogenous regressors is sufficient to control for endogeneity, while adding control functions for higher-order endogenous terms—such as interactions among endogenous regressors, interactions between endogenous and exogenous regressors, or squared endogenous regressors—is unnecessary and can substantially degrade the performance of copula correction (Qian, Koschmann, and Xie, 2025). This parsimonious treatment of higher-order endogenous regressors is a merit of copula correction.

Thus, if X contains no higher-order terms of endogenous regressors, the simpler 1-bar form $Y \sim X \mid P$ can be used, and `tscope()` treats all regressors in X except those in P as exogenous. When X includes higher-order endogenous terms, the 2-bar form $Y \sim X \mid P \mid W$ should be used to explicitly specify the exogenous regressors in W and ensure that the higher-order endogenous terms are not treated as exogenous variables.

The extra generated regressors are denoted by `ccf`: followed by the associated endogenous regressor in the model output. The correlations between the endogenous regressors and the structural error of the model are denoted by `cor`: followed by the associated endogenous regressor.

Value

a data.frame of class "tscope" containing the following components:

Est	the coefficients and other contents of the 2sCOPE model. The first section contains the coefficient estimates of the original regressors. The second section contains the coefficient estimates of the generated regressors (also known as copula terms or copula control functions). The third section contains the correlation(s) between the endogenous regressor(s) and the structural error of the model, which represents the strength and size of the endogeneity of the model, as well as sigma representing the standard deviation of the structural error term.
boot.SE	standard errors for the coefficient estimates obtained from bootstrapping
z value	z score of the associated coefficient estimate
Pr(> z)	p-value of the associated coefficient estimate

References

- Qian, Y., Koschmann, A., & Xie, H. (2025). *EXPRESS: A Practical Guide to Endogeneity Correction Using Copulas*. *Journal of Marketing*. doi:10.1177/00222429251410844
- Park, S., & Gupta, S. (2012). *Handling endogenous regressors by joint estimation using copulas*. *Marketing Science*, 31(4), 567-586.
- Yang, F., Qian, Y., & Xie, H. (2025). *Addressing Endogeneity Using a Two-Stage Copula Generated Regressor Approach*. *Journal of Marketing Research*, 62(4), 601-623. doi:10.1177/00222437241296453

Examples

```
data("diapers") #load data

#run a OLS model to compare results to 2sCOPE
ols <- lm(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4, data = diapers)
summary(ols)

#run 2sCOPE with 1-bar option
tscope_model_1bar <- tscope(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4|logPrice,
  data = diapers, nboot = 300)
tscope_model_1bar

#run 2sCOPE with 2-bar option
tscope_model_2bar <- tscope(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4 |logPrice|
  Fshare+week+Q2+Q3+Q4,
  data = diapers, nboot = 300)
tscope_model_2bar

#notice how both the 1-bar and 2-bar options produce the same parameter estimates,
#and that the results differ from OLS after correcting for endogeneity.
#the standard errors are not the same because they are obtained from bootstrapping.

#run Park and Gupta (2012) model
pg <- tscope(logVol ~ logPrice|logPrice, data = diapers, nboot = 300)
pg
```

tscope.fit

tscope.fit: Fitter Function for 2sCOPE

Description

Basic computing engine called by `tscope()`

Usage

```
tscope.fit(formula, data)
```

Arguments

formula	a formula describing the model to be fitted. The details of model specification are given under “Details”.
data	a data frame, list, or environment containing the variables in the model.

Details

The formula argument is either in the 1-bar form $Y \sim X \mid P$ or the 2-bar form $Y \sim X \mid P \mid W$, where X represents the explanatory variable(s) in the Y model, P represents the continuous endogenous regressors, and W represents the exogenous regressors. If X contains no exogenous regressors, then the 2sCOPE model reduces to the simpler model in Park and Gupta (2012) and returns P^* (the copula transformation of P) as CCF and W^* (the copula transformation of W) as null. When the structural outcome model includes an intercept, copula transformations of regressors in P and W use the optimized algorithm (Equation 9 in Qian, Koschmann, and Xie, 2025) to avoid estimation bias.

The function will add copula control function for each endogenous regressor specified in P . Only first-order terms of endogenous regressors need to be included in P , even when the structural outcome model contains higher-order terms of endogenous regressors. This is because including copula control functions for the first-order endogenous regressors is sufficient to control for endogeneity, while adding control functions for higher-order endogenous terms—such as interactions among endogenous regressors, interactions between endogenous and exogenous regressors, or squared endogenous regressors—is unnecessary and can substantially degrade the performance of copula correction (Qian, Koschmann, and Xie, 2025). This parsimonious treatment of higher-order endogenous regressors is a merit of copula correction.

Thus, if X contains no higher-order terms of endogenous regressors, the simpler 1-bar form $Y \sim X \mid P$ can be used, and `tscope()` treats all regressors in X except those in P as exogenous. When X includes higher-order endogenous terms, the 2-bar form $Y \sim X \mid P \mid W$ should be used to explicitly specify the exogenous regressors in W and ensure that the higher-order endogenous terms are not treated as exogenous variables.

Value

A numeric vector containing the coefficients of the original and generated regressors, including any high-order or interaction terms if present.

References

- Qian, Y., Koschmann, A., & Xie, H. (2025). *EXPRESS: A Practical Guide to Endogeneity Correction Using Copulas*. *Journal of Marketing*. doi:10.1177/00222429251410844
- Park, S., & Gupta, S. (2012). *Handling endogenous regressors by joint estimation using copulas*. *Marketing Science*, 31(4), 567-586.
- Yang, F., Qian, Y., & Xie, H. (2025). *Addressing Endogeneity Using a Two-Stage Copula Generated Regressor Approach*. *Journal of Marketing Research*, 62(4), 601-623. doi:10.1177/00222437241296453

Examples

```
data("diapers") #load data

# run a OLS model to compare results to 2sCOPE
ols <- lm(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4, data = diapers)
coef(ols)

tscope_model_1bar <- tscope.fit(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4|
logPrice, data = diapers) # run 2sCOPE with 1-bar option
tscope_model_1bar

tscope_model_2bar <- tscope.fit(logVol ~ logPrice+Fshare+week+Q2+Q3+Q4|
logPrice|Fshare+week+Q2+Q3+Q4, data = diapers) # run 2sCOPE with 2-bar option
tscope_model_2bar

# notice how both the 1-bar and 2-bar options produce the same parameter
# estimates, and that the results differ from OLS after correcting for endogeneity.

#run Park and Gupta (2012) model
pg <- tscope.fit(logVol ~ logPrice|logPrice, data = diapers)
pg
```

Index

* **datasets**

diapers, [4](#)

CCF, [2](#)

diapers, [4](#)

print.ccf, [5](#)

print.tscope, [5](#)

print.tscope.fit, [6](#)

relevance_test, [6](#)

tscope, [7](#)

tscope.fit, [10](#)