Package 'PanJen'

November 1, 2017

Version 1.6

Date 2017-11-01

Title A Semi-Parametric Test for Specifying Functional Form

Author Toke Emil Panduro <tepp@ifro.ku.dk>, Cathrine Ulla Jensen <cuj@ifro.ku.dk>

Maintainer Cathrine Ulla Jensen <cuj@ifro.ku.dk>

Description A central decision in a parametric regression is how to specify the relation between an dependent variable and each explanatory variable. This package provides a semi-parametric tool for comparing different transformations of an explanatory variables in a parametric regression. The functions is relevant in a situation, where you would use a box-cox or Box-Tidwell transformations. In contrast to the classic power-transformations, the methods in this package allows for theoretical driven user input and the possibility to compare with a non-parametric transformation.

Depends R (>= 3.4), mgcv

License GPL (>= 2)

LazyLoad yes

LazyData yes

NeedsCompilation no

Repository CRAN

Date/Publication 2017-11-01 14:10:26 UTC

R topics documented:

---

choose.fform

Compare a number of user-specified transformations with a semiparametric smoothing and a model without the variable

---

Description

PanJen is built on the idea that the choice of a functional form can be extrapolated from model fit measures. The function provides a ranking of different transformations according to their Baysian Information Criterion (BIC). The BIC provides a relative goodness-of-fit measure while accounting for the complexity of the model. The function provides BIC for each transformation, a model without the variable in question and a socalled smoothing of the variable. The models are estimated a Generalized additive model (GAM).A GAM is a special case of the Generalized Linear Model (GLM), where it is possible to include one or more so called smoothing functions. A smoothing function is a non-parametric way to include a continuous independent variable in a parametric model and thus make it semi-parametric. Please see Wood (2006) for an elaboration. The printed output is a table with the transformations sorted according to their explanatory power measured by AIC. The table shows both AIC and BIC for each regression, where BIC penalize for the number of explanatory variables

Usage

```
choose.fform(data,base_form,variable,functionList, distribution=gaussian)
```

Arguments

data	A data.frame
variable	A character-string with the name of the variable to test
base_form	A formula-object with the regressions without the variable that is tested
functionList	A list of transformations. Please see the example for an elaboration
distribution	Assumed distribution, see mgcv-vignette for an elaboration. Default is gaussian

Value

rank.table	The printed table. The transformations are ranked according to their explanatory power measured by AIC. The table shows both the AIC and BIC value, where the BIC value penalizes for the number of variables
models	A list of estimated models
dataset	A dataframe with the dataset
fforms	The formula provided by the user

Author(s)

Toke Emil Panduro & Cathrine Ulla Jensen

References

Simon Wood, Generalized Additive Models: an introduction with R.,hapman and Hall/CRC, 2006

Examples

```
## Test a linear specification (x), a log (I(log(x))) and a square I(x^2))
library(PanJen)
data("hvidovre")
form<-formula(lprice ~brick+roof_tile+roof_cemen+rebuild90+rebuild00+y7+y8+y9)
fxlist= list(
  linear = function(x) x,
  sqr = function(x) x^2,
  log=function(x) log(x)
)

PanJenAreaC<-choose.fform(data=hvidovre,variable="area",base_form=form, functionList=fxlist)
```

---

fform	*Compare a number of transformations with a semiparametric smoothing and a model without the variable*

Description

PanJen is built on the idea that the choice of a functional form can be extrapolated from model fit measures. The function provides a ranking of different transformations according to their Baysian Information Criterion (BIC). The BIC provides a relative goodness-of-fit measure while accounting for the complexity of the model. The function provides BIC for a set of predefined transformations aswell as a model without the variable in question and a socalled smoothing of the variable. The models are estimated a Generalized additive model (GAM).A GAM is a special case of the Generalized Linear Model (GLM), where it is possible to include one or more so called smoothing functions. A smoothing function is a non-parametric way to include a continuous independent variable in a parametric model and thus make it semi-parametric. Please see Wood (2006) for an elaboration. The printed output is a table with the transformations sorted according to their explanatory power measured by AIC. The table shows both AIC and BIC for each regression, where BIC penalize for the number of explanatory variables

Usage

```
fform(data,variable,base_form, distribution=gaussian)
```

Arguments

data	A data.frame
variable	A character-string with the name of the variable to test
base_form	A formula-object with the regressions without the variable that is tested
distribution	Assumed distribution, see mgcv-vignette for an elaboration. Default is gaussian

Value

rank.table	The printed table. The transformations are ranked according to their explanatory power measured by AIC. The table shows both the AIC and BIC value, where the BIC value penalizes for the number of variables
models	A list of estimated models
dataset	A dataframe with the dataset

Author(s)

Toke Emil Panduro & Cathrine Ulla Jensen

References

Simon Wood, Generalized Additive Models: an introduction with R.,hapman and Hall/CRC, 2006

Examples

```
library(PanJen)
data("hvidovre")

form<-formula(lprice ~brick+roof_tile+roof_cemen+rebuild70+rebuild80+rebuild90+rebuild00+y7+y8+y9)
PanJenArea<-fform(data=hvidovre,variable="area",base_form=form)
```

---

hvidovre	*Houseprice data*

Description

A dataset with trade price and attributes for 900 homes in a Danish munnicipality

Usage

```
data("hvidovre")
```

Format

A data frame with 901 observations on the following 19 variables

lprice a numeric vector, price, log price in 1000 EUR

brick a numeric vector, dummy, wall made out of brick =1

roof_tile a numeric vector, roof_tiledummy, roof made out of tiles =1

roof_cemen a numeric vector, dummy, roof made out of cement

y7 a numeric vector, home sold in 2007

y8 a numeric vector, home sold in 2008

y9 a numeric vector, home sold in 2009

`rebuild70` a numeric vector, home rebuild in 1970's

`rebuild80` a numeric vector, home rebuild in 1980's

`rebuild90` a numeric vector, home rebuild in 1990's

`rebuild00` a numeric vector,home rebuild in 2000's

`area` a numeric vector,living area in square meters

`age` a numeric vector, build year

`bathrooms` a numeric vector, number of bathrooms

`highways` a numeric vector, distance to nearest highway in meters

`big_roads` a numeric vector, distance to nearest large road in meters

`railways` a numeric vector, railwaysdistance to nearest railway in meters

`nature_SLD` a numeric vector, nature_SLDdistance to nearest nature area in meters

`lake_SLD` a numeric vector,distance to nearest lake in meters

Source

Panduro et al (in review JAERE)

References

Panduro T.E., Jensen, C.U, Lundhede, T.H., von Graevenitz, K. and Thorsen, B.J., Estimating demand schedules in hedonic analysis: The case of urban parks, (in review JAERE)

Examples

```
data(hvidovre)
```

---

plotff	*This function plots objects generated by choose.fform or fform*

Description

plots the output from fform() and choose.fform()

Usage

```
plotff(input)
```

Arguments

`input` A PJ-objected produced by fform or choose.fform

Author(s)

Toke Emil Panduro and Cathrine Ulla Jensen

Examples

```
library(PanJen)
data("hvidovre")

form<-formula(lprice ~brick+roof_tile+roof_cemen+rebuild90+rebuild00+y7+y8+y9)
PanJenArea<-fform(data=hvidovre,variable="area",base_form=form)
plotff(PanJenArea)
```

Index