

Package ‘phtt’

February 20, 2015

Type Package

Title Panel Data Analysis with Heterogeneous Time Trends

Version 3.1.2

Date 2014-03-22

Author Oualid Bada, Dominik Liebl

Maintainer Dominik Liebl <r.packages.mail@gmail.com>

Depends R(>= 2.10)

Imports pspline

Description The package provides estimation procedures for panel data with large dimensions n , T , and general forms of unobservable heterogeneous effects. Particularly, the estimation procedures are those of Bai (2009) and Kneip, Sickles, and Song (2012), which complement one another very well: both models assume the unobservable heterogeneous effects to have a factor structure. The method of Bai (2009) assumes that the factors are stationary, whereas the method of Kneip et al. (2012) allows the factors to be non-stationary. Additionally, the 'phtt' package provides a wide range of dimensionality criteria in order to estimate the number of the unobserved factors simultaneously with the remaining model parameters.

License GPL-2

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2014-08-15 16:49:41

R topics documented:

phtt-package	2
checkSpecif	2
Cigar	5
Eup	7
KSS	12
OptDim	16
Simulated Data for the KSS-Model: DGP1	20
Simulated Data for the KSS-Model: DGP2	22

Index**25**

phtt-package

*Panel Data Analysis with Heterogeneous Time Trends***Description**

The R-package phtt provides estimation procedures for panel data with large dimensions n , T , and general forms of unobservable heterogeneous effects. Particularly, the estimation procedures are those of Bai (2009) and Kneip, Sickles, and Song (2012), which complement one another very well: both models assume the unobservable heterogeneous effects to have a factor structure. The method of Bai (2009) assumes that the factors are stationary, whereas the method of Kneip et al. (2012) allows the factors to be non-stationary. Additionally, the phtt package provides a wide range of dimensionality criteria in order to estimate the number of the unobserved factors simultaneously with the remaining model parameters.

Details

Package: phtt
 Type: Package
 Version: 3.1
 Date: 2013-09-26
 License: GPL-2
 LazyLoad: yes

Author(s)

Oualid Bada, Dominik Liebl

References

- Kneip, A., Sickles, R. C., Song, W., 2012 “A new panel data treatment for heterogeneity in time trends”, *Econometric Theory*
- Bai, J., 2009 “Panel data models with interactive fixed effects”, *Econometrica*

checkSpecif

*Tests to check the model specifications***Description**

Tests to check the model specifications

Usage

```
checkSpecif(obj1, obj2, level = 0.05)
```

Arguments

- | | |
|-------|--|
| obj1 | <ul style="list-style-type: none"> • If obj2 is left unspecified the specification test proposed by Kneip, Sickles, and Song (2012) is computed. In this case obj1 can be an object of class 'KSS' or 'Eup'. The given KSS- or Eup-model needs unspecified factor dimensions (<code>factor.dim=NULL</code>). See also the Details. • If obj2 is specified by an object of class 'Eup' the Hausman-type test proposed by Bai (2009) is computed, which compares the model in obj1 with the model in obj2. The Hausman-type test of Bai applies to 'Eup'-objects only. See also the Details. |
| obj2 | An object of class 'Eup'. If obj2 is left unspecified the test proposed by Kneip, Sickles, and Song (2012) is computed. |
| level | The significance level. |

Details

This function is equipped with two types of specification-tests:

- The first specification-tests is the Hausman-type test proposed by Bai (2009), which is computed if obj1 as well as obj2 are specified by 'Eup'-objects. In this case the model given to the arguments obj1 and obj2 are compared with each other. Note that this test assumes that the (unobserved) true factor dimension is lower or equal to 2; see Section 9 in Bai (2009) for more details.

Given the assumption that there is only one unobserved common factor:

- Null-Hypothesis: 'The unobserved common factor is a classical individual or time effect'.
- Alternative-Hypothesis: 'The unobserved common factor is an arbitrary process'.

Given the assumption that there are two unobserved common factors:

- Null-Hypothesis: 'The two unobserved common factors are classical twoways effects'.
- Alternative-Hypothesis: 'The two unobserved common factors are arbitrary processes'.

- The second specification-test tests the existence of an additional factor structure beyond a classical additive effects model; as suggested in Kneip, Sickles, and Song (2012), which is applied if only obj1 is specified and obj2 is left unspecified. This test can be used for 'Eup'-objects as well as for 'KSS'-objects.

- Null-Hypothesis: 'There are no unobserved common factors beyond the classical individual, time, or twoways effects'.
- Alternative-Hypothesis: 'There are additional unobserved common factors'.

Author(s)

Oualid Bada, Dominik Liebl

References

- Bai, J., 2009 “Panel data models with interactive fixed effects”, *Econometrica*
- Kneip, A., Sickles, R. C., Song, W., 2012 “A New Panel Data Treatment for Heterogeneity in Time Trends”, *Econometric Theory*

See Also

KSS, Eup, OptDim

Examples

```
## See the example in 'help(Cigar)' in order to take a look at the
## data set 'Cigar'

#####
## DATA ##
#####

data(Cigar)
## Panel-Dimensions:
N <- 46
T <- 30
## Dependent variable:
## Cigarette-Sales per Capita
l.Consumption <- log(matrix(Cigar$sales, T,N))
d.l.Consumption <- diff(l.Consumption)
## Independent variables:
## Consumer Price Index
cpi <- matrix(Cigar$cpi, T,N)
## Real Price per Pack of Cigarettes
l.Price <- log(matrix(Cigar$price, T,N)/cpi)
d.l.Price <- diff(l.Price)
## Real Disposable Income per Capita
l.Income <- log(matrix(Cigar$ndi, T,N)/cpi)
d.l.Income <- diff(l.Income)

#####
## Testing the Sufficiency of a classical 'twoways' effects model: ##
## Hausman-type Test of Bai (2009) ##
#####

## Model under the null Hypothesis:
twoways.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
factor.dim = 0, additive.effects = "twoways")

## Model under the alternative Hypothesis:
not.twoways.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
factor.dim = 2, additive.effects = "none")

#####
## Test: ##
```

```
#####

## (This test returns an error message, since the (unobserved) true
## factor dimension is probably greater than 2.)
## Not run:
checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj, level = 0.01)

## End(Not run)

#####
## Testing the Existence of additional (unobserved) common Factors ##
## Specification Test of Kneip, Sickles, and Song (2012)          ##
#####

## For the model of Bai (2009):
Eup.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
  additive.effects = "twoways")

## Test:
checkSpecif(Eup.obj, level = 0.01)

## For the model of Kneip, Sickles, and Song (2012):
KSS.obj <- KSS(l.Consumption ~ -1 + l.Price + l.Income,
  additive.effects = "twoways")

## Test:
checkSpecif(KSS.obj, level = 0.01)
```

Cigar

Cigarette Consumption

Description

a panel of N=46 observations each with time-dimension T=30 from 1963 to 1992

total number of observations : 1380

observation : regional

country : United States

Usage

data(Cigar)

Format

A data frame containing :

state state abbreviation

year the year

price price per pack of cigarettes
pop population
pop16 population above the age of 16
cpi consumer price index (1983=100)
ndi per capita disposable income
sales cigarette sales in packs per capita
pimin minimum price in adjoining states per pack of cigarettes

Source

Online complements to Baltagi (2001). <http://www.wiley.com/legacy/wileychi/baltagi/>.

References

- Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.
- Baltagi, B.H. and D. Levin (1992) “Cigarette taxation: Raising revenues and reducing consumption”, *Structural Changes and Economic Dynamics*, **3**, 321–335.
- Baltagi, B.H., J.M. Griffin and W. Xiong (2000) “To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand”, *Review of Economics and Statistics*, **82**, 117–126.

Examples

```
data(Cigar)
## Panel-Dimensions:
N <- 46
T <- 30

## Dependent variable:

## Cigarette-Sales per Capita
l.Consumption <- log(matrix(Cigar$sales, T,N))

## Independent variables:

## Consumer Price Index
cpi <- matrix(Cigar$cpi, T,N)

## Real Price per Pack of Cigarettes
l.Price <- log(matrix(Cigar$price, T,N)/cpi)

## Real Disposable Income per Capita
l.Income <- log(matrix(Cigar$ndi, T,N)/cpi)

#####
## Plot the data ##
#####
```

```

par(mfrow=c(1,3))
## Dependent variable
matplot(l.Consumption, main="Log's of Cigarette -Sales\nper Capita",
        type="l", xlab="Time", ylab="")
## Independent variables
matplot(l.Price, main="Log's of Real-Prices of\nCigarettes per Pack",
        type="l", xlab="Time", ylab="")
matplot(l.Income, main="Log's of Real Disposable Income\nper Capita",
        type="l", xlab="Time", ylab="")
par(mfrow=c(1,1))

```

Eup

*Eup-Routine***Description**

Estimation of Panel Data Models with Interactive Fixed Effects.

Usage

```

Eup(formula,
     additive.effects = c("none", "individual", "time", "twoways"),
     dim.criterion    = c("PC1", "PC2", "PC3", "BIC3", "IC1", "IC2", "IC3",
                        "IPC1", "IPC2", "IPC3"),
     d.max            = NULL,
     sig2.hat         = NULL,
     factor.dim       = NULL,
     double.iteration = TRUE,
     start.beta       = NULL,
     max.iteration    = 500,
     convergence      = 1e-6,
     restrict.mode    = c("restrict.factors", "restrict.loadings"),
     ...)

```

Arguments

formula An object of class 'formula' where the arguments are matrices. The number of rows has to be equal to the temporal dimension and the number of columns has to be equal to the number of individuals. The details of model specification are given under 'Details'.

additive.effects

Type of Data Transformations:

- "none": for no transformation
- "individual": for within transformation
- "time": for between transformation
- "twoways": for twoways transformation

<code>dim.criterion</code>	The dimensionality criterion to be used if <code>factor.dim</code> is left unspecified. The default criterion is "PC1".
<code>d.max</code>	Maximal dimension used in the dimensionality-criteria of Bai (2009). The default (<code>d.max=NULL</code>) yields to an internal selection of <code>d.max</code> .
<code>sig2.hat</code>	The squared standard deviation of the error-term required for the computation of some dimensionality criteria. The user can specify it instead of <code>d.max</code> . The default (<code>sig2.hat=NULL</code>) yields to an internal estimation.
<code>factor.dim</code>	Dimension of Factor-Structure, pre-specified by the user. The default (<code>factor.dim=NULL</code>) yields to an internal estimation.
<code>double.iteration</code>	logical. If FALSE the update of the factor dimension <code>d</code> will be done simultaneously with remaining model parameters without alternating between inner and outer iteration. This may speed up computations, but the convergence is less stable than in the default setting.
<code>start.beta</code>	allows the user to give a vector of starting values for the slope parameters.
<code>max.iteration</code>	controls the maximum number of iterations. The default is '500'.
<code>convergence</code>	Convergence condition of the estimators. The default is '1e-6'.
<code>restrict.mode</code>	Type of Restriction on the Factor-Structure: <ul style="list-style-type: none"> • "<code>restrict.factors</code>": Factors are restricted to have an euclidean norm of 1. • "<code>restrict.loadings</code>": Factor-Loadings are restricted to have an euclidean norm of 1.
<code>...</code>	Additional arguments to be passed to the low level functions.

Details

'Eup' is a function to estimate equidistant panel data models with unobserved multiple time varying individual effects. The considered model is given by

$$Y_{it} = \sum_{j=1}^P \beta_j X_{itj} + v_{it} + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

Where the individual time trends, v_{it} , are assumed to come from a finite dimensional factor model:

$$v_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}, \quad \lambda_{il}, f_{lt} \in R.$$

- `formula` Usual 'formula'-object. If you wish to estimate a model without an intercept use '-1' in the formula-specification. Each Variable has to be given as a TxN-matrix. Missing values are not allowed.
- `additive.effects`
 - "none": The data is not transformed, except for a subtraction of the overall mean, if the model is estimated with an intercept. The assumed model can be written as

$$Y_{it} = \mu + \sum_{j=1}^P \beta_j X_{itj} + v_{it} + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

- "individual": This is the "within"-model, which assumes that there are time-constant individual effects, α_i , besides the individual time trends v_{it} . The model can be written as

$$Y_{it} = \mu + \alpha_i + \sum_{j=1}^P \beta_j X_{itj} + v_{it} + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

- "time": This is the "between"-model, which assumes that there is a common time trend (for all individuals), θ_t . The model can be written as

$$Y_{it} = \mu + \theta_t + \sum_{j=1}^P \beta_j X_{itj} + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

- "twoways": This is the "twoways"-model ("within" & "between"), which assumes that there are time-constant individual effects, α_i , and a common time trend, θ_t . The model can be written as

$$Y_{it} = \mu + \alpha_i + \theta_t + \sum_{j=1}^P \beta_j X_{itj} + \tau_i + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

Inferences about the slope parameters can be obtained by using the method `summary()`. The type of correlation and heteroskedasticity in the idiosyncratic errors can be specified by choosing the corresponding number for the argument `error.type = c(1, 2, 3, 4, 5, 6, 7, 8)` in `summary()`, where

- 1: indicates the presence of i.i.d. errors,
- 2: indicates the presence of cross-section heteroskedasticity with

$$n/T \rightarrow 0$$

,

- 3: indicates the presence of cross-section correlation and heteroskedasticity with

$$n/T \rightarrow 0$$

,

- 4: indicates the presence of heteroskedasticity in the time dimension with

$$T/n \rightarrow 0$$

,

- 5: indicates the presence of correlation and heteroskedasticity in the time dimension with

$$T/n \rightarrow 0$$

,

- 6: indicates the presence of both time and cross-section dimensions with

$$T/n^2 \rightarrow$$

and

$$n/T^2 \rightarrow 0$$

,

- 7: indicates the presence of both time and cross-section dimensions with

$$n/T \rightarrow c > 0$$

, and

- 8: indicates the presence of correlation and heteroskedasticity in both time and cross-section dimensions with

$$n/T \rightarrow c > 0$$

.

The default is 1. In presence of serial correlations (cases 5 and 8), the kernel weights required for estimating the long-run covariance can be externally specified by given a vector of weights in the argument `kernel.weights`. By default, the function uses internally the linearly decreasing weights of Newey and West (1987) and a truncation at the lower integer part of

$$\min(\sqrt{n}, \sqrt{T})$$

. If case 7 or 8 are chosen, the method `summary()` calculates the realization of the bias corrected estimators and gives appropriate inference. The bias corrected coefficients can be called by using the method `coef()` to the object produced by `summary()`.

Value

'Eup' returns an object of 'class' "'Eup'" containing the following components:

- `dat.matrix`: Whole data set stored within a $(N \times T) \times (p+1)$ -Matrix, where P is the number of independent variables without the intercept.
- `formula`: returns the used formula object.
- `dat.dim`: Vector of length 3: $c(T, N, p)$
- `slope.para`: Beta-parameters
- `names`: Names of the dependent and independent variables.
- `is.intercept`: logical. Used an intercept in the formula?: TRUE or FALSE
- `additive.effects`: Additive effect type. One of: "none", "individual", "time", "twoways".
- `Intercept`: Intercept-parameter. Tacks the value 0 if it is not specified in the model.
- `Add.Ind.Eff`: Estimated values of additive individual effects. If additive individual effects are not specified in the model, the function returns a vector of zeros.
- `Add.Tim.Eff`: Estimated values of additive time effects. If this effects are not specified in the model, the function returns a vector of zeros.
- `unob.factors`: $T \times d$ -matrix of estimated unobserved common factors, where 'd' is the number of used factors.

- ind.loadings: Nxd-matrix of loadings parameters.
- unob.fact.stru: TxN-matrix of the estimated factor structure. Each column represents an estimated individual unobserved time trend.
- used.dim: Used dimension 'd' to calculate the factor structure.
- proposed.dim: Indicates whether the user has specified the factor dimension or not.
- optimal.dim: The optimal dimension calculated internally.
- factor.dim: The user-specified factor dimension. Default is NULL
- d.max: The maximum number of factors used to estimate the optimal dimension.
- dim.criterion: The used dimensionality criterion.
- OvMeans: A vector that contains the overall means of the observed variables (Y and X).
- ColMean: A matrix that contains the column means of the observed variables (Y and X).
- RowMean: A matrix that contains the row means of the observed variables (Y and X).
- max.iteration: The maximum number of iterations. The default is '500'.
- convergence: The convergence condition. The default is '1e-6'.
- start.beta: A vector of user-specified starting values for the estimation of the beta-parameters. Default is NULL.
- Nbr.iteration: Number of iterations required for the computation.
- fitted.values: Fitted values.
- orig.Y: Original values of the dependent variable.
- residuals: Original values of the dependent variable.
- sig2.hat.dim: user-specified variance estimator of the errors. Default is NULL.
- sig2.hat: Estimated variance of the error term.
- degrees.of.freedom: Degrees of freedom of the residuals.
- call

Author(s)

Oualid Bada

References

- Bai, J., 2009 "Panel data models with interactive fixed effects", *Econometrica*
- Bada, O. and Kneip, A., 2014 "Parameter Cascading for Panel Models with Unknown Number of Unobserved Factors: An Application to the Credit Spread Puzzle", *Computational Statistics & Data Analysis (forthcoming)*

See Also

KSS, OptDim

Examples

```
## See the example in 'help(Cigar)' in order to take a look at the
## data set 'Cigar'

#####
## DATA ##
#####

data(Cigar)
## Panel-Dimensions:
N <- 46
T <- 30
## Dependent variable:
## Cigarette-Sales per Capita
d.l.Consumption <- diff(log(matrix(Cigar$sales, T,N)))
## Independent variables:
## Consumer Price Index
cpi <- matrix(Cigar$spi, T,N)
## Real Price per Pack of Cigarettes
d.l.Price <- diff(log(matrix(Cigar$price, T,N)/cpi))
## Real Disposable Income per Capita
d.l.Income <- diff(log(matrix(Cigar$ndi, T,N)/cpi))

## Estimation:
Eup.fit <- Eup(d.l.Consumption~d.l.Price+d.l.Income)
(Eup.fit.sum <- summary(Eup.fit))

## Plot the components of the estimated individual effects

plot(Eup.fit.sum)
```

KSS

KSS-Routine

Description

Estimation of Panel Data Models with Heterogeneous Time Trends

Usage

```
KSS(formula,
     additive.effects = c("none", "individual", "time", "twoways"),
     consult.dim.crit = FALSE,
     d.max            = NULL,
     sig2.hat        = NULL,
     factor.dim      = NULL,
     level           = 0.01,
     spar            = NULL,
     CV              = FALSE,
```

```

convergence      = 1e-6,
restrict.mode    = c("restrict.factors", "restrict.loadings"), ...)

```

Arguments

<code>formula</code>	An object of class 'formula'.
<code>additive.effects</code>	Type of Data Transformations: <ul style="list-style-type: none"> • "none": for no transformation • "individual": for within transformation • "time": for between transformation • "twoways": for twoways transformation
<code>consult.dim.crit</code>	logical. <ul style="list-style-type: none"> • If <code>consult.dim.crit</code> is FALSE (default) and <code>factor.dim</code> is NULL: Only the dimensionality criterion of Kneip, Sickles & Song 2012 is used. • If <code>consult.dim.crit</code> is TRUE and <code>factor.dim</code> is NULL: All implemented dimensionality criteria as implemented in the function <code>OptDim()</code> are computed and the user has to select one proposed dimension via a GUI.
<code>d.max</code>	A maximal dimension needed for some dimensionality-criteria that are implemented in the function <code>OptDim()</code> . The default (<code>d.max=NULL</code>) yields to an internal selection of <code>d.max</code> .
<code>sig2.hat</code>	Standard deviation of the error-term. The default (<code>sig2.hat=NULL</code>) yields to an internal estimation of <code>sig2.hat</code> .
<code>factor.dim</code>	Dimension of Factor-Structure. The default (<code>factor.dim=NULL</code>) yields to an internal estimation of <code>factor.dim</code> .
<code>level</code>	Significance-level for Dimensionality-Criterion of Kneip, Sickles & Song 2012.
<code>spar</code>	Smoothing parameter for spline smoothing of the residuals. If (<code>spar=NULL</code>) (default) and <code>CV=FALSE</code> <code>spar</code> is determined via generalized cross validation (GCV).
<code>CV</code>	logical. Selects the procedure for the determination of the smoothing parameter <code>spar</code> . <ul style="list-style-type: none"> • If <code>CV=FALSE</code> (default) and <code>spar=NULL</code>: The smoothing parameter <code>spar</code> is determined by GCV. • If <code>CV=TRUE</code> and <code>spar=NULL</code>: The smoothing parameter <code>spar</code> is determined by Leave-one-out cross validation (CV).
<code>convergence</code>	Convergence criterion for the CV-optimization of the smoothing parameter <code>spar</code> . Default is <code>convergence=1e-6</code> .
<code>restrict.mode</code>	Type of Restriction on the Factor-Structure: <ul style="list-style-type: none"> • "restrict.factors": Factors are restricted to have an euclidean norm of 1. • "restrict.loadings": Factor-Loadings are restricted to have an euclidean norm of 1.
<code>...</code>	Additional arguments to be passed to the low level functions.

Details

'KSS' is a function to estimate panel data models with unobserved heterogeneous time trends $v_i(t)$. The considered model in Kneip, Sickles & Song (2012) is given by

$$Y_{it} = \theta_t + \sum_{j=1}^P \beta_j X_{itj} + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

Where the individual time trends, $v_i(t)$, are assumed to come from a finite dimensional factor model

$$v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t), \quad \lambda_{il} \in R, \quad f_l \in L^2[0, T].$$

The unobserved functions $v_i(t)$ can be interpreted as smooth functions of a continuous argument t , as well as stochastic processes for discrete argument t .

- formula Usual 'formula'-object. If you wish to estimate a model without an intercept use '-1' in the formula-specification. Each Variable has to be given as a TxN-matrix. Missing values are not allowed.
- additive.effects
 - "none": The data is not transformed, except for an eventually subtraction of the overall mean; if the model is estimated with an intercept. The assumed model can be written as

$$Y_{it} = \mu + \sum_{j=1}^P \beta_j X_{itj} + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

- "individual": This is the "within"-model, which assumes that there are time-constant individual effects, α_i , besides the individual time trends $v_i(t)$. The model can be written as

$$Y_{it} = \mu + \sum_{j=1}^P \beta_j X_{itj} + v_i(t) + \alpha_i + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

- "time": This is the "between"-model, which assumes that there is a common (for all individuals) time trend, θ_t . The model can be written as

$$Y_{it} = \mu + \theta_t + \sum_{j=1}^P \beta_j X_{itj} + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

- "twoways": This is the "twoways"-model ("within" & "between"), which assumes that there are time-constant individual effects, α_i , and a common time trend, θ_t . The model can be written as

$$Y_{it} = \mu + \theta_t + \sum_{j=1}^P \beta_j X_{itj} + \alpha_i + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; t = 1, \dots, T.$$

The parameter 'mu' is set to zero if '-1' is used in formula.

Value

'KSS' returns an object of 'class' "'KSS"'.

An object of class "'KSS'" is a list containing at least the following components:

- `dat.matrix`: Whole data set stored within a $(N \times T) \times (p+1)$ -Matrix, where P is the number of independent variables without the intercept.
- `dat.dim`: Vector of length 3: $c(T, N, p)$
- `slope.para`: Beta-parameters
- `beta.V`: Covariance matrix of the beta-parameters.
- `names`: Names of the dependent and independent variables.
- `is.intercept`: Used an intercept in the formula?: TRUE or FALSE
- `additive.effects`: Additive effect type. One of: "none", "individual", "time", "twoways".
- `Intercept`: Intercept-parameter
- `Add.Ind.Eff`: Estimated values of additive individual effects.
- `Add.Tim.Eff`: Estimated values of additive time effects.
- `unob.factors`: $T \times d$ -matrix of estimated unobserved common factors, where 'd' is the number of used factors.
- `ind.loadings`: $N \times d$ -matrix of loadings parameters.
- `unob.fact.stru`: $T \times N$ -matrix of the estimated factor structure. Each column represents an estimated individual unobserved time trend.
- `used.dim`: Used dimensionality of the factor structure.
- `optimal.dim`: List of proposed dimensionalities.
- `fitted.values`: Fitted values.
- `orig.Y`: Original values of the dependent variable.
- `residuals`: Residuals
- `sig2.hat`: Estimated variance of the error term.
- `degrees.of.freedom`: Degrees of freedom of the residuals.
- `call`

Author(s)

Dominik Liebl

References

- Kneip, A., Sickles, R. C., Song, W., 2012 "A New Panel Data Treatment for Heterogeneity in Time Trends", *Econometric Theory*

See Also

Eup

Examples

```
## See the example in 'help(Cigar)' in order to take a look at the
## data set Cigar

#####
## DATA ##
#####

data(Cigar)
## Panel-Dimensions:
N <- 46
T <- 30
## Dependent variable:
## Cigarette-Sales per Capita
l.Consumption <- log(matrix(Cigar$sales, T,N))
## Independent variables:
## Consumer Price Index
cpi <- matrix(Cigar$spi, T,N)
## Real Price per Pack of Cigarettes
l.Price <- log(matrix(Cigar$price, T,N)/cpi)
## Real Disposable Income per Capita
l.Income <- log(matrix(Cigar$ndi, T,N)/cpi)

## Estimation:
KSS.fit <- KSS(l.Consumption~l.Price+l.Income, CV=TRUE)
(KSS.fit.sum <- summary(KSS.fit))
plot(KSS.fit.sum)
```

OptDim

Estimation of the Factor Dimension

Description

Functions for the Estimation of the Factor Dimension

Usage

```
OptDim(Obj,
       criteria = c("PC1", "PC2", "PC3", "BIC3",
                    "IC1", "IC2", "IC3",
                    "IPC1", "IPC2", "IPC3",
                    "ABC.IC1", "ABC.IC2",
                    "KSS.C",
                    "ED", "ER", "GR"),
       standardize = FALSE,
       d.max,
       sig2.hat,
       spar,
       level = 0.01,
```



```
c.grid = seq(0, 5, length.out = 128),
T.seq, n.seq)
```

Arguments

Obj	The function requires either a Txn matrix or an object with class "Eup" or "KSS".
criteria	A character vector that contains the desired criteria to be used. If it is left unspecified, the function returns the result of all 16 criteria.
standardize	logical. If TRUE the input variable will be standardized. Default is FALSE.
d.max	Maximal dimension used in the dimensionality-criteria of Bai (2009). The default (d.max=NULL) yields to an internal selection of d.max.
sig2.hat	The squared standard deviation of the error-term required for the computation of some dimensionality criteria. The user can specify it instead of d.max. The default (sig2.hat=NULL) yields to an internal estimation.
spar	Smoothing parameter used to calculate the criterion of Kneip, Sickles, and Song (2012). The default is NULL, which leads to internal computation.
level	The significance level used for the criterion of Kneip, Sickles, and Song (2012). The default is 0.01.
c.grid	Required only for computing "ABC.IC1" and "ABC.IC2". It specifies the grid interval in which the scaling parameter of the penalty terms in "ABC.IC1" and "ABC.IC2" are calibrated. Default is <code>c.grid = seq(0, 5, length.out = 128)</code> .
T.seq	Required only for computing "ABC.IC1" and "ABC.IC2". It can be a vector containing different dimensions for T or an integer indicating the length of the sequence to be considered in calibrating "ABC.IC1" and "ABC.IC2". If it is left unspecified, the function determines internally a sequence of the form <code>seq((T-C), T)</code> , where C is the square root of $\min\{T, 900\}$.
n.seq	Required only for computing "ABC.IC1" and "ABC.IC2". It can be a vector containing different dimensions for n or an integer indicating the length of the sequence to be considered in calibrating "ABC.IC1" and "ABC.IC2". If it is left unspecified, the function determines internally a sequence of the form <code>seq((n-D), n)</code> , where D is the square root of $\min\{n, 900\}$.

Details

The function 'OptDim' allows for a comparison of the optimal factor dimensions obtained from different panel criteria (in total 13). This criteria are adjusted for panel data with diverging T and N.

Value

'OptDim' returns an object of 'class' "'OptDim'" containing a list with the following components:

criteria:	The name of the criteria specified by the user.
PC1:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .

PC2:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .
PC3:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .
IC1:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> .
IC2:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> .
IC3:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> .
IPC1:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .
IPC2:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .
IPC3:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .
KSS.C:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> and/ or <code>sig2.hat</code> .
ED:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> .
ER:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> .
GR:	If specified in <code>criteria</code> a table is returned with the optimal dimension, the empirical standard deviation of the residuals, and some other informations required internally by the criterion, such as <code>d.max</code> .
summary:	A table (in a matrix form) containing all the estimated dimensions obtained by the specified criteria.
BaiNgC:	A logical vector required for further internal computations.
BaiC:	A logical vector required for further internal computations.
KSSC:	A logical vector required for further internal computations.
OnatC:	A logical vector required for further internal computations.
RHC:	A logical vector required for further internal computations.
obj:	The argument 'Obj' given to the function <code>OptDim()</code> .
cl:	Object of mode "call".

Author(s)

Oualid Bada

References

- Ahn, S. C., Horenstein, A. R. 2013 “Eigenvalue ratio test for the number of factors”, *Econometrica*
- Bai, J., 2009 “Panel data models with interactive fixed effects”, *Econometrica*
- Bai, J. 2004 “Estimating cross-section common stochastic trends in nonstationary data”, *Journal of Econometrics*
- Bai, J., Ng, S. 2009 “Determining the number of factors in approximated factor models”, *Econometrica*
- Kneip, A., Sickles, R. C., Song, W., 2012 “A New Panel Data Treatment for Heterogeneity in Time Trends”, *Econometric Theory*
- Onatski, A. 2010 “Determining the number of factors from empirical distribution of eigenvalues”, *The Review of Economics and Statistics*

Examples

```
## See the example in 'help(Cigar)' in order to take a look at the
## data set 'Cigar'

#####
## DATA ##
#####

data(Cigar)
N <- 46
T <- 30

## Data: Cigarette-Sales per Capita
l.Consumption <- log(matrix(Cigar$sales, T,N))

## Calculation is based on the covariance matrix of l.Consumption

OptDim(l.Consumption)

## Calculation is based on the correlation matrix of l.Consumption

OptDim(l.Consumption, standardize = TRUE)

## Display the magnitude of the eigenvalues in percentage of the total variance

plot(OptDim(l.Consumption))
```

Simulated Data for the KSS-Model: DGPI

Simulated Panel-Data Set with Polynomial Factor Structure and exogenous regressors.

Description

A Panel-Data Sets with:

time-index : $t=1, \dots, T=30$

individual-index : $i=1, \dots, N=60$

This panel-data set has a polynomial factor structure (3 common factors) and *exogenous* regressors.

Usage

`data(DGP1)`

Format

A list containing :

Y dependent variable as N*T-vector

X1 first regressor as N*T-vector

X2 second regressor as N*T-vector

CF.1 first (unobserved) common factor:

$$CF.1(t) = 1$$

CF.2 second (unobserved) common factor:

$$CF.2(t) = \frac{t}{T}$$

CF.3 thrid (unobserved) common factor:

$$CF.3(t) = \left(\frac{t}{T}\right)^2$$

Remark: The time-index t is running faster than the individual-index i such that e.g. Y_{it} is ordered as:

$$Y_{11}, Y_{12}, \dots, Y_{1T}, Y_{21}, Y_{22}, \dots$$

Details

The panel-data set DPG1 is simulated according to the simulation-study in Kneip, Sickles & Song (2012):

$$Y_{it} = \beta_1 X_{it1} + \beta_2 X_{it2} + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; \quad t = 1, \dots, T$$

-Slope parameters:

$$\beta_1 = \beta_2 = 0.5$$

-Time varying individual effects being second order polynomials:

$$v_i(t) = \theta_{i0} + \theta_{i1} \frac{t}{T} + \theta_{i2} \left(\frac{t}{T} \right)^2$$

Where θ_{i0} , θ_{i1} , and θ_{i2} are iid as $N(0,4)$

The Regressors $X_{it}=(X_{it1}, X_{it2})'$ are simulated from a bivariate VAR model:

$$X_{it} = RX_{i,t-1} + \eta_{it} \quad \text{with} \quad R = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix} \quad \text{and} \quad \eta_{it} \sim N(0, I_2)$$

After this simulation, the N regressor-series

$$(X_{1i1}, X_{2i1})', \dots, (X_{1iT}, X_{2iT})'$$

are additionally shifted such that there are three different mean-value-clusters. Such that every third of the N regressor-series fluctuates around on of the following mean-values

$$\mu_1 = (5, 5)', \quad \mu_2 = (7.5, 7.5)', \quad \text{and} \quad \mu_3 = (10, 10)'$$

In this Panel-Data Set the regressors are exogenous. See Kneip, Sickles & Song (2012) for more details.

Author(s)

Dominik Liebl

References

- Kneip, A., Sickles, R. C., Song, W., 2012 “A New Panel Data Treatment for Heterogeneity in Time Trends”, *Econometric Theory*

Examples

```
data(DGP1)

## Dimensions
N <- 60
T <- 30

## Observed Variables
Y <- matrix(DGP1$Y, nrow=T, ncol=N)
X1 <- matrix(DGP1$X1, nrow=T, ncol=N)
```

```

X2  <- matrix(DGP1$X2, nrow=T,ncol=N)

## Unobserved common factors
CF.1 <- DGP1$CF.1[1:T]
CF.2 <- DGP1$CF.2[1:T]
CF.3 <- DGP1$CF.3[1:T]

## Take a look at the simulated data set DGP1:
par(mfrow=c(2,2))
matplot(Y, type="l", xlab="Time", ylab="", main="Depend Variable")
matplot(X1, type="l", xlab="Time", ylab="", main="First Regressor")
matplot(X2, type="l", xlab="Time", ylab="", main="Second Regressor")
## Usually unobserved common factors:
matplot(matrix(c(CF.1,
                 CF.2,
                 CF.3), nrow=T,ncol=3),
         type="l", xlab="Time", ylab="", main="Unobserved Common Factors")
par(mfrow=c(1,1))

## Estimation:
KSS.fit <- KSS(Y~-1+X1+X2)
(KSS.fit.sum <- summary(KSS.fit))

plot(KSS.fit.sum)

```

Simulated Data for the KSS-Model: DGP2

Simulated Panel-Data Set with Polynomial Factor Structure and endogenous regressors.

Description

A Panel-Data Sets with:

time-index : $t=1, \dots, T=30$

individual-index : $i=1, \dots, N=60$

This panel-data set has a polynomial factor structure (3 common factors) and *endogenous* regressors.

Usage

```
data(DGP2)
```

Format

A data frame containing :

Y dependent variable as $N \times T$ -vector

X1 first regressor as $N \times T$ -vector

X2 second regressor as N*T-vector

CF.1 first (unobserved) common factor:

$$CF.1(t) = 1$$

CF.2 second (unobserved) common factor:

$$CF.2(t) = \frac{t}{T}$$

CF.3 third (unobserved) common factor:

$$CF.3(t) = \left(\frac{t}{T}\right)^2$$

Remark: The time-index t is running "faster" than the individual-index i such that e.g. Y_{it} is ordered as:

$$Y_{11}, Y_{12}, \dots, Y_{1T}, Y_{21}, Y_{22}, \dots$$

Details

The panel-data set DPG2 is simulated according to the simulation-study in Kneip, Sickles & Song (2012):

$$Y_{it} = \beta_1 X_{it1} + \beta_2 X_{it2} + v_i(t) + \epsilon_{it} \quad i = 1, \dots, n; \quad t = 1, \dots, T$$

-Slope parameters:

$$\beta_1 = \beta_2 = 0.5$$

-Time varying individual effects being second order polynomials:

$$v_i(t) = \theta_{i0} + \theta_{i1} \frac{t}{T} + \theta_{i2} \left(\frac{t}{T}\right)^2$$

Where θ_{i0} , θ_{i1} , and θ_{i2} are iid as $N(0,4)$

The Regressors $X_{it} = (X_{it1}, X_{it2})'$ are simulated from a bivariate VAR model:

$$X_{it} = R X_{i,t-1} + \eta_{it} \quad \text{with} \quad R = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix} \quad \text{and} \quad \eta_{it} \sim N(0, I_2)$$

After this simulation, the N regressor-series

$$(X_{1i1}, X_{2i1})', \dots, (X_{1iT}, X_{2iT})'$$

are additionally shifted such that there are three different mean-value-clusters. Such that every third of the N regressor-series fluctuates around one of the following mean-values

$$\mu_1 = (5, 5)', \quad \mu_2 = (7.5, 7.5)', \quad \text{and} \quad \mu_3 = (10, 10)'$$

In this Panel-Data Set the regressor X_{it2} is made endogenous by the re-definition:

$$X_{it2} := X_{it2} + 0.5v_i(t)$$

See Kneip, Sickles & Song (2012) for more details.

Author(s)

Dominik Liebl

References

- Kneip, A., Sickles, R. C., Song, W., 2012 “A New Panel Data Treatment for Heterogeneity in Time Trends”, *Econometric Theory*

Examples

```

data(DGP2)

## Dimensions
N   <- 60
T   <- 30

## Observed Variables
Y   <- matrix(DGP2$Y, nrow=T,ncol=N)
X1  <- matrix(DGP2$X1, nrow=T,ncol=N)
X2  <- matrix(DGP2$X2, nrow=T,ncol=N)

## Unobserved common factors
CF.1 <- DGP2$CF.1[1:T]
CF.2 <- DGP2$CF.2[1:T]
CF.3 <- DGP2$CF.3[1:T]

## Take a look at the simulated data set DGP2:
par(mfrow=c(2,2))
matplot(Y, type="l", xlab="Time", ylab="", main="Depend Variable")
matplot(X1, type="l", xlab="Time", ylab="", main="First Regressor")
matplot(X2, type="l", xlab="Time", ylab="", main="Second Regressor")
## Usually unobserved common factors:
matplot(matrix(c(CF.1,
                 CF.2,
                 CF.3), nrow=T,ncol=3),
         type="l", xlab="Time", ylab="", main="Unobserved Common Factors")
par(mfrow=c(1,1))

## Estimation
KSS.fit <- KSS(Y~-1+X1+X2)
(KSS.fit.sum <- summary(KSS.fit))

plot(KSS.fit.sum)

```


Index

*Topic **datasets**

Simulated Data for the KSS-Model:

DGP1, [20](#)

Simulated Data for the KSS-Model:

DGP2, [22](#)

checkSpecif, [2](#)

checkSpecification (checkSpecif), [2](#)

Cigar, [5](#)

DGP1 (Simulated Data for the
KSS-Model: DGP1), [20](#)

DGP2 (Simulated Data for the
KSS-Model: DGP2), [22](#)

Eup, [7](#)

KSS, [12](#)

OptDim, [16](#)

phtt (phtt-package), [2](#)

phtt-package, [2](#)

Simulated Data for the KSS-Model:
DGP1, [20](#)

Simulated Data for the KSS-Model:
DGP2, [22](#)