Package 'fbi'

January 7, 2022

Type Package

Title Factor-Based Imputation and FRED-MD/QD Data Set

Version 0.5.0

Date 2022-01-07

Author Yankang (Bennie) Chen [aut, cre],
   Serena Ng [aut],
   Jushan Bai [aut]

Maintainer Yankang (Bennie) Chen <yankang.chen@yale.edu>

Description Factor-Based imputation of missing values in panel data and and
   manipulation of the FRED-MD/QD Data Set. It estimates the factor model in the
   panel data based on the methods in Bai and Ng (2002)
   <doi:10.1111/1468-0262.00273> and Bai and Ng (2019)
   <doi:10.1016/j.jeconom.2019.04.021>. It then computes the missing
   values using the Tall-Wide method (Bai and Ng (2021)
   <doi:10.1080/01621459.2021.1967163>) or the Tall-Project method (Cahan, Bai, and Ng
   (2021)<arXiv:2103.03045>). It also facilitates loading, preparing, and
   interpreting the FRED-
   MD/QD data set <https://research.stlouisfed.org/econ/mccracken/fred-databases>.

URL https://github.com/cykbennie/fbi

Depends R (>= 3.5.0)

Imports stats, readr, pracma

License GPL-3 + LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

1

fbi-package   *Factor-Based Imputation and FRED-MD/QD Data Set*

Description

The `fbi` package contains functions to estimate factor models and impute missing data based on factor models. It also includes functions to load and prepare the FRED-MD/QD data set.

Details

See `vignette("factor_fred",package = "fbi")` for an example using the FRED-MD dataset (<https://research.stlouisfed.org/econ/mccracken/fred-databases/>).

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models*. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00273>

Jushan Bai and Serena Ng (2019), *Rank regularized estimation of approximate factor models*. <https://doi.org/10.1016/j.jeconom.2019.04.021>

Jushan Bai and Serena Ng (2021), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data*. <https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1967163>

Ercument Cahan, Jushan Bai, and Serena Ng (2021), *Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions*. <https://arxiv.org/abs/2103.03045>

apc	*Factor Model of Balanced Panel Data*

Description

apc estimates the factor model of a given balanced panel data.

Usage

```
apc(X, kmax)
```

Arguments

X	a matrix of size T by N.
kmax	integer, indicating the maximum number of factors.

Value

a list of elements:

X	the original data
kmax	the maximum number of factors
Fhat	estimated F
Lamhat	estimated Lambda
Chat	equals Fhat x Lamhat'
Dhat	estimated diagonal matrix D, of dim kmax by kmax
d	first kmax elements of Dhat
d0	diagonal elements of Dhat
ehat	equals X - Chat

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models*.
https://doi.org/10.1111/1468-0262.00273

---

demeanXY   *Demean Panel Data*

Description

demeanXY demeans the panel data.

Usage

```
demeanXY(X, N, T, N0, T0)
```

Arguments

X	detaframe or matrix of the original panel data.
N	integer, total number of columns of the panel data.
T	integer, total number of rows of the panel data.
N0	integer, the number of columns in the panel data with full data availability.
T0	integer, the number of rows in the panel data with full data availability.

Value

a list of elements:

X1	demeaned data
FE	estimated fixed effects matrix

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

---

describe_md   *Describe selected variables in the FRED-MD Data Set*

Description

describe_md provides a description of the selected variables in the FRED-MD data set.

Usage

```
describe_md(varname, name.only = TRUE, verbose = FALSE)
```

Arguments

varname	string or a vector strings of the FRED variable name, such as GDPC1.
name.only	logical. If TRUE, return a dataframe with variable names and types of transformation only; if FALSE, return a dataframe with more details.
verbose	logical, indicating whether or not descriptions should be printed.

Value

a vector of variable names, or a data frame with detailed descriptions.

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

References

Michael W. McCracken and Serena Ng (2015), *FRED-MD Updated Appendix.* [https://s3.](https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf) [amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf](https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf)

Examples

```
library(fbi)
varnames <- describe_md(c("RPI", "RETAILx"), name.only = TRUE, verbose = FALSE)
```

---

describe_qd *Describe selected variables in the FRED-QD Data Set*

Description

describe_qd provides a description of the selected variables in the FRED-QD data set.

Usage

```
describe_qd(varname, name.only = TRUE, verbose = FALSE)
```

Arguments

varname	string or a vector strings of the FRED variable name, such as GDPC1.
name.only	logical. If TRUE, return a dataframe with variable names and types of transformation only; if FALSE, return a dataframe with more details.
verbose	logical, indicating whether or not descriptions should be printed.

Value

a vector of variable names, or a data frame with detailed descriptions.

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

References

Michael W. McCracken and Serena Ng (2020), *FRED-QD Updated Appendix.* [https://s3.amazonaws.](https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QD_appendix.pdf) [com/files.fred.stlouisfed.org/fred-md/FRED-QD_appendix.pdf](https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QD_appendix.pdf)

Examples

```
library(fbi)
varnames <- describe_qd(c("GDPC1", "Y033RC1Q027SBEAx"), name.only = TRUE, verbose = FALSE)
```

---

fredmd *Loading FRED-MD Data Set*

Description

fredmd loads the official FRED-MD data set and provides a few tools to manipulate the data set.

Usage

```
fredmd(file = "", date_start = NULL, date_end = NULL, transform = TRUE)
```

Arguments

file	Either a path to a file, a connection, or literal data (either a single string or a raw vector).
date_start	Date or NULL, the start date (included) of the data selection. If NULL, select till the latest data available.
date_end	Date or NULL, the end date (included) of the data selection. If NULL, select up to the earliest data available.
transform	logical, indicating Whether or not the FRED-MD data set should be transformed according to the transformation code.

Value

a subset of the (transformed) FRED-MD data of class fredmd.

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

References

Michael W. McCracken and Serena Ng (2015), *FRED-MD and FRED-QD: Monthly and Quarterly Databases for Macroeconomic Research*. https://research.stlouisfed.org/econ/mccracken/fred-databases/

---

fredmd_description *FRED-MD Data Set Description*

Description

A description of the FRED-MD data set.

Usage

```
data(fredmd_description)
```

Format

A data frame with 135 rows and 9 variables. The variables are as follows:

id series ID number

tcode code of transformation

ttype type of transformation

fred variable name used in the FRED-MD data set

description description of the series

gsi variable name used in the Global Insights Basic Economics Database (GSI)

gsi:description description of the series in GSI

group group of the series

edited logical, indicating if the data has been editted

varname "X" + id

Source

The `fredmd_description` data were obtained from Michael W. McCracken and Serena Ng (2015), *FRED-MD Updated Appendix*. https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf

---

fredqd	*Loading FRED-QD Data Set*

Description

`fredqd` loads the official FRED-QD data set and provides a few tools to manipulate the data set.

Usage

```
fredqd(file = "", date_start = NULL, date_end = NULL, transform = TRUE)
```

Arguments

`file`	Either a path to a file, a connection, or literal data (either a single string or a raw vector).
`date_start`	Date or `NULL`, the start date (included) of the data selection. If `NULL`, select till the latest data available.
`date_end`	Date or `NULL`, the end date (included) of the data selection. If `NULL`, select up to the earliest data available.
`transform`	logical, indicating Whether or not the FRED-QD data set should be transformed according to the transformation code.

Value

a subset of the (transformed) FRED-QD data of class `fredqd`.

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

References

Michael W. McCracken and Serena Ng (2015), *FRED-MD and FRED-QD: Monthly and Quarterly Databases for Macroeconomic Research*. https://research.stlouisfed.org/econ/mccracken/fred-databases/

---

fredqd_description	*FRED-QD Data Set Description*

Description

A description of the FRED-QD data set.

Usage

```
data(fredqd_description)
```

Format

A data frame with 248 rows and 10 variables. The variables are as follows:

id series ID number

sw_id series ID number in SW (2012)

tcode code of transformation

ttype type of transformation

sw_factors logical, indicating whether a series was used in SW (2012) when constructing factors

fred_mnemonic mnemonic in FRED-QD

sw_mnemonic mnemonic used in SW (2012)

description a brief definition of the series

group group of the series

varname "X" + id

Source

The `fredqd_description` data were obtained from Michael W. McCracken and Serena Ng (2020), *FRED-QD Updated Appendix*. https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QD_appendix.pdf

removeFE   *Remove Fixed Effects from the Panel Data*

Description

removeFE removes fixed effects from the panel data.

Usage

```
removeFE(X, N, T, N0, T0)
```

Arguments

X	detaframe or matrix of the original panel data.
N	integer, total number of columns of the panel data.
T	integer, total number of rows of the panel data.
N0	integer, the number of columns in the panel data with full data availability.
T0	integer, the number of rows in the panel data with full data availability.

Value

a list of elements:

X1	demeaned data
FE	estimated fixed effects matrix

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

res_overlay.twtp   *Residual Overlay*

Description

res_overlay.twtp estimates the covariance and correlation matrix of the unbalanced panel data using the method of residual overlay.

Usage

```
res_overlay.twtp(object, method = 1, S = 500)
```

Arguments

object	an object of class 'tw' or 'tp', i.e. the output of tw_apc or tp_apc.
method	integer 1 to 4, indicating which residual overlay method to use. They correspond to the four methods described in the paper.
S	the number of iterations.

Value

a list of elements:

method	the method of residual overlay
S	the number of iterations
cov	estimated covariance matrix
cor	estimated correlation matrix

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Cahan, E., Bai, J. and Ng, S. 2019, Factor Based Imputation of Missing Data and Covariance Matrix Estimation. unpublished manuscript, Columbia University

---

rm_outliers.fredmd	*Remove outliers of the FRED-MD Data Set*

Description

rm_outliers.fredmd removes outliers of the FRED-MD data set produced by the fredmd function.

Usage

```
rm_outliers.fredmd(object)
```

Arguments

object	an object of class fredmd.

Value

FRED-MD data of class fredmd with outliers removed.

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

References

Michael W. McCracken and Serena Ng (2015), *FRED-MD and FRED-QD: Monthly and Quarterly Databases for Macroeconomic Research.* https://research.stlouisfed.org/econ/mccracken/fred-databases/

---

rpca	*Estimation of Approximate Factor Models*

Description

`rpca` estimates the approximate factor models of the given matrix.

Usage

```
rpca(X, kmax, standardize = FALSE, tau = 0)
```

Arguments

X	a matrix of size T by N.
kmax	integer, indicating the maximum number of factors.
standardize	logical, indicating Whether or not X should be centered and scaled.
tau	numeric, specifying the parameter in the rank-regularized estimation. If `tau = 0`, then rank regularization is not used.

Value

a list of elements:

X

kmax

standardize

tau

ic2

pc2k

pc20

Fhat

Lamhat

Chat

Sigma

IC2

PC2k

PC20

fhat

lamhat

d

d0

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models*. https://doi.org/10.1111/1468-0262.00273

Jushan Bai and Serena Ng (2019), *Rank regularized estimation of approximate factor models*. https://doi.org/10.1016/j.jeconom.2019.04.021

---

se.rpca	*Standard Error of C^hat*

Description

se.rpca produces the estimated standard error of C^hat produced by the rpca function.

Usage

```
se.rpca(object, xpoints, qq)
```

Arguments

object	an object of class rpca.
xpoints	placeholder.
qq	placeholder.

Value

standard error of C^hat

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models*. https://doi.org/10.1111/1468-0262.00273

Jushan Bai and Serena Ng (2019), *Rank regularized estimation of approximate factor models*. https://doi.org/10.1016/j.jeconom.2019.04.021

Jushan Bai and Serena Ng (2021), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data*. https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1967163

Ercument Cahan, Jushan Bai, and Serena Ng (2021), *Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions*. https://arxiv.org/abs/2103.03045

---

se.tp *Standard Error of Selected Points (TP)*

Description

se.tp produces the estimated standard error of C^hat produced by the [tp_apc](tp_apc) function.

Usage

```
se.tp(object, npoints, tpoints, qq, re_estimate = TRUE)
```

Arguments

object	an object of class 'tp'.
npoints	integer or vector of integers, indicating i of the (i,t) pair(s) of interest.
tpoints	integer or vector of integers, indicating t of the (i,t) pair(s) of interest.
qq	placeholder.
re_estimate	logical. If 'FALSE', use first pass estimation (Lemma 2 of Cahan, Bai, and Ng (2021)). If 'TRUE', use re-estimation (Proposition 1).

Value

a list of elements:

tpoints	t's of the (i,t) pair(s) of interest
npoints	i's of the (i,t) pair(s) of interest
re_estimate	logical. If 'FALSE', use first pass estimation; if 'TRUE', use re-estimation.
Fhat	estimated F
Lamhat	estimated Lambda
Chat	euqals Fhat x Lamhat'
SigmaC	estimated variance of C
SigmaF	estimated variance of F
SigmaL	estimated variance of L

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models*. https://doi.org/10.1111/1468-0262.00273

Jushan Bai and Serena Ng (2019), *Rank regularized estimation of approximate factor models*. https://doi.org/10.1016/j.jeconom.2019.04.021

Jushan Bai and Serena Ng (2021), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data*. https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1967163

Ercument Cahan, Jushan Bai, and Serena Ng (2021), *Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions*. https://arxiv.org/abs/2103.03045

---

se.tw	*Standard Error of Selected Points (TW)*

Description

se.tw produces the estimated standard error of C^hat produced by the tw_apc function.

Usage

```
se.tw(object, npoints, tpoints, qq, re_estimate)
```

Arguments

object	an object of class 'tw'.
npoints	integer or vector of integers, indicating i of the (i,t) pair(s) of interest.
tpoints	integer or vector of integers, indicating t of the (i,t) pair(s) of interest.
qq	placeholder.
re_estimate	logical. If 'FALSE', use first pass estimation. If 'TRUE', use re-estimation.

Value

a list of elements:

tpoints	t's of the (i,t) pair(s) of interest
npoints	i's of the (i,t) pair(s) of interest
re_estimate	logical. If 'FALSE', use first pass estimation; if 'TRUE', use re-estimation.
Fhat	estimated F
Lamhat	estimated Lambda
Chat	euqals Fhat x Lamhat'
SigmaC	estimated variance of C
SigmaF	estimated variance of F
SigmaL	estimated variance of L

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models.* https://doi.org/10.1111/1468-0262.00273

Jushan Bai and Serena Ng (2019), *Rank regularized estimation of approximate factor models.* https://doi.org/10.1016/j.jeconom.2019.04.021

Jushan Bai and Serena Ng (2021), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data.* https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1967163

Ercument Cahan, Jushan Bai, and Serena Ng (2021), *Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions.* https://arxiv.org/abs/2103.03045

---

se0	*Standard Error of Selected Points (Baseline)*

Description

se0 produces the estimated standard error of C^hat produced by the `tw_apc` or `tp_apc` function.

Usage

```
se0(object, npoints, tpoints, qq)
```

Arguments

object	an object of class 'tw' or 'tp'.
npoints	integer or vector of integers, indicating i of the (i,t) pair(s) of interest.
tpoints	integer or vector of integers, indicating t of the (i,t) pair(s) of interest.
qq	placeholder.

Value

a list of elements:

tpoints	t's of the (i,t) pair(s) of interest
npoints	i's of the (i,t) pair(s) of interest
Fhat	estimated F
Lamhat	estimated Lambda
Chat	euqals Fhat x Lamhat'
SigmaC	estimated variance of C
SigmaF	estimated variance of F
SigmaL	estimated variance of L

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2002), *Determining the number of factors in approximate factor models*. https://doi.org/10.1111/1468-0262.00273

Jushan Bai and Serena Ng (2019), *Rank regularized estimation of approximate factor models*. https://doi.org/10.1016/j.jeconom.2019.04.021

Jushan Bai and Serena Ng (2021), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data*. https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1967163

Ercument Cahan, Jushan Bai, and Serena Ng (2021), *Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions*. https://arxiv.org/abs/2103.03045

---

tnt	*Estimate Treatment Effect*

Description

`tnt` estimates the treatment effect.

Usage

```
tnt(data, param)
```

Arguments

`data`	list containing x1, x2, y0, y1, N0, N1, T0, and T1.
`param`	list containing K, r, do_FE, do_IFE, and maxit1.

Value

a list of elements:

`est`

`SE`

`V`

`it1`

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2019), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data*. https://arxiv.org/abs/1910.06677

| tp_apc | *Tall-Project Imputation of Missing Value in Panel Data* |

Description

`tp_apc` imputes the missing values in a given panel data using the method of "Tall-Project".

Usage

```
tp_apc(X, kmax, center = FALSE, standardize = FALSE, re_estimate = TRUE)
```

Arguments

X	a matrix of size T by N with missing values.
kmax	integer, indicating the maximum number of factors.
center	logical, indicating whether or not X should be demeaned
standardize	logical, indicating whether or not X should be scaled.
re_estimate	logical, indicating whether or not output factors, 'Fhat', 'Lamhat', 'Dhat', and 'Chat', should be re-estimated from the imputed data.

Value

a list of elements:

Fhat	estimated F
Lamhat	estimated Lambda
Dhat	estimated D
Chat	euqals Fhat x Lamhat'
ehat	equals Xhat - Chat
data	X with missing data imputed
X	the original data with missing values
kmax	the maximum number of factors
center	logical, indicating whether or not X was demeaned in the algorithm
standardize	logical, indicating whether or not X was scaled in the algorithm
re_estimate	logical, indicating whether or not output factors, 'Fhat', 'Lamhat', 'Dhat', and 'Chat', were re-estimated from the imputed data

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Ercument Cahan, Jushan Bai, and Serena Ng (2021), *Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions*. https://arxiv.org/abs/2103.03045

---

tw_apc   *Tall-Wide Imputation of Missing Value in Panel Data*

Description

`tw_apc` imputes the missing values in a given panel data using the method of "Tall-Wide".

Usage

```
tw_apc(X, kmax, center = FALSE, standardize = FALSE, re_estimate = TRUE)
```

Arguments

X	a matrix of size T by N with missing values.
kmax	integer, indicating the maximum number of factors.
center	logical, indicating whether or not X should be demeaned
standardize	logical, indicating whether or not X should be scaled.
re_estimate	logical, indicating whether or not output factors, 'Fhat', 'Lamhat', 'Dhat', and 'Chat', should be re-estimated from the imputed data.

Value

a list of elements:

Fhat	estimated F
Lamhat	estimated Lambda
Dhat	estimated D
Chat	euqals Fhat x Lamhat'
ehat	equals Xhat - Chat
data	X with missing data imputed
X	the original data with missing values
kmax	the maximum number of factors
center	logical, indicating whether or not X was demeaned in the algorithm
standardize	logical, indicating whether or not X was scaled in the algorithm
re_estimate	logical, indicating whether or not output factors, 'Fhat', 'Lamhat', 'Dhat', and 'Chat', were re-estimated from the imputed data

Author(s)

Yankang (Bennie) Chen <yankang.chen@yale.edu>

Serena Ng <serena.ng@columbia.edu>

Jushan Bai <jushan.bai@columbia.edu>

References

Jushan Bai and Serena Ng (2021), *Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data*. https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1967163

Index