Package 'hdcce'

September 15, 2026

Type Package

Title Estimation and Inference for High-Dimensional Panel Data Models with Interactive Fixed Effects

Version 0.1.0

Description Estimation and inference for panel data models with interactive fixed effects. The methods cover i) the linear specification of Ruecker, M., Vogt, M., Linton, O. and Walsh, C. (2025) ``Estimation and Inference in High-Dimensional Panel Data Models with Interactive Fixed Effects" <doi:10.3982/QE2308>, and ii) the dictionary design of Ruecker, M., Vogt, M. and Linton, O. (2026) ``High-Dimensional Panel Data Models with Interactive Fixed Effects: Beyond the Linear Case" <doi:10.48550/arXiv.2608.02055>.

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 3.5.0)

Imports stats, graphics, glmnet, mvtnorm

Suggests knitr, rmarkdown

VignetteBuilder knitr

LazyData true

LazyDataCompression xz

RoxygenNote 7.3.2

URL https://github.com/RueckerM/hdcce

NeedsCompilation no

Author Maximilian Ruecker [aut, cre], Michael Vogt [aut], Oliver Linton [aut], Christopher Walsh [aut]

Maintainer Maximilian Ruecker <maximilian.ruecker@uni-ulm.de>

Repository CRAN

Date/Publication 2026-09-15 14:20:02 UTC

Contents

`data_estimation` *Simulated panel data for estimation*

Description

A balanced panel simulated with `generate_data` from the design of the simulation study in Rücker et al. (2025), used to illustrate `hdcce_estimator`. The panel is sorted such that the first `obs_T = 20` rows belong to unit 1, the next 20 to unit 2, etc.

Usage

```
data_estimation
```

Format

A list with two components:

x numeric matrix of dimension 400 x 61 containing the regressors.

y numeric vector of length 400 containing the response.

Source

Simulated with `generate_data(obs_N = 20, obs_T = 20, p = 61, mu = rep(1, 63), RHO = 0.25)` after `set.seed(2024)`.

References

Rücker, M., Vogt, M., Linton, O. and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4), 1457–1509. doi:10.3982/QE2308.

Rücker, M., Vogt, M. and Linton, O. (2026) "High-Dimensional Panel Data Models with Interactive Fixed Effects: Beyond the Linear Case" https://arxiv.org/pdf/2608.02055.

data_inference *Simulated panel data for inference*

Description

As `data_estimation`, but with three response vectors differing in the coefficient of the first regressor; used to illustrate `hdcce_inference`.

Usage

```
data_inference
```

Format

A list with two components:

- **x** numeric matrix of dimension 400 x 61 containing the regressors.
- **y** numeric matrix of dimension 400 x 3; column k is generated with coefficient of the first regressor equal to c** = 0, 0.1 and 0.2.

Source

Simulated with `generate_data(obs_N = 20, obs_T = 20, p = 61, mu = rep(1, 63), RHO = 0.25)` after `set.seed(2024)`.

References

Rücker, M., Vogt, M., Linton, O. and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4), 1457–1509. doi:10.3982/QE2308

Rücker, M., Vogt, M. and Linton, O. (2026) "High-Dimensional Panel Data Models with Interactive Fixed Effects: Beyond the Linear Case" <https://arxiv.org/pdf/2608.02055>.

generate_data *Generate toy panel data*

Description

`generate_data` simulates panel data as in the simulation section of Rücker et al. (2025). It was used to create the data sets `data_estimation` and `data_inference` shipped with the package.

Usage

```
generate_data(obs_N, obs_T, p, mu, RHO)
```

Arguments

obs_N	Number of cross-section units.
obs_T	Number of time periods.
p	Number of regressors/covariates.
mu	Vector for individual loadings.
RHO	Pairwise correlation coefficient of the regressors.

Value

List containing the balanced panel data for estimation (data_estimation) and inference (data_inference). data_estimation$y contains the dependent variables and data_estimation$x the regressors. data_inference is in the same spirit with the minor difference that data_inference$y has three columns (one for each signal c** = 0, 0.1, 0.2). In both cases, the panels are sorted such that first T observations are those for unit 1, followed by the T observations for unit 2, etc.

References

Rücker, M., Vogt, M., Linton, O. and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4), 1457–1509. doi:10.3982/QE2308.

Rücker, M., Vogt, M. and Linton, O. (2026) "High-Dimensional Panel Data Models with Interactive Fixed Effects: Beyond the Linear Case" https://arxiv.org/pdf/2608.02055.

Examples

```
# Simulate an example data set with
# N = 20, T = 20, RHO = 0.25, p = 61
#-------------------------------------------------------------------------------

# Set the parameters
#-------------------------------------------------------------------------------
# Set the number of cross-sections, time periods
obs_N  <- 20
obs_T  <- 20
p <- 61
# Specify the pairwise correlation among the regressors
RHO <- 0.25
# Specify the mean for the factor loadings
mu <- c(1, 1, 1, rep(1, (p-1)))

# Simulate the data
set.seed(2025)
sim <- generate_data(obs_N = obs_N, obs_T = obs_T, p,  mu = mu,
                             RHO = RHO)
dim(sim$data_estimation$x)
dim(sim$data_inference$y)     # one column per signal c** = 0, 0.1, 0.2
```

hdcce_estimator *Estimate HD panels with IFE*

Description

hdcce_estimator fits (i) the linear HD-CCE estimator of Rücker et al. (2025), or (ii) the HD-CCE estimator for a parametric additive (dictionary) design as in Rücker et al. (2026), when dictionaries is supplied.

Usage

```
hdcce_estimator(
  data,
  obs_N,
  obs_T,
  dictionaries = NULL,
  TRUNC = 0.01,
  NFACTORS = NULL,
  variant = c("Lasso", "LS"),
  lambda = NULL,
  NFOLDS = 10,
  foldid = NULL,
  scree_plot = FALSE,
  standardize = TRUE
)
```

Arguments

data	List containing the balanced panel data with data$y the dependent variable and data$x the p regressors. Both are sorted such that the first obs_T observations are those for unit 1, followed by the obs_T observations for unit 2, etc.
obs_N, obs_T	The number of cross-section units (obs_N) and the time series length (obs_T).
dictionaries	Either NULL (linear model, the default), an obs_N * obs_T x d matrix of dictionary transformations of data$x, or a list with components Phi (that matrix) and, optionally, group, a d-vector with entries in 1, . . . , p indicating which covariate each column transforms. When a dictionary is supplied, the design and the response are empirically centred across units before projection.
TRUNC	The truncation parameter tau used to estimate the number of factors. Default is 0.01.
NFACTORS	Allows to set the number of factors used in estimation. Default is NULL, so that the data driven choice with TRUNC is used.
variant	Either "Lasso" (default) or "LS" for the least squares variant. "LS" requires the design to have full column rank.
lambda	User specified lambda grid, on the scale of the penalised criterion in the paper (see Details).

NFOLDS	The number of folds (partitioned along the cross-section) used for cross-validation. Default is 10. Fold size can vary by one if obs_N is not divisible by NFOLDS.
foldid	Integer vector (obs_N * obs_T-dimensional) containing a fold label for each observation.
scree_plot	Logical; whether a scree plot of the eigendecomposition of Σ should be shown. The default is FALSE.
standardize	Logical; whether glmnet is called with the standardized projected data. The default is TRUE.

Details

The criterion in the paper is $(nT)^{-1}\sum_i \|\tilde{Y}_i - \tilde{X}_i b\|^2 + \lambda\|b\|_1$, whereas glmnet minimises $(2nT)^{-1}\sum_i \|\tilde{Y}_i - \tilde{X}_i b\|^2 + \lambda_g\|b\|_1$. The two penalties are related by $\lambda = 2\lambda_g$. Both the lambda argument and the returned Lambda are on the scale of the paper.

Value

A list with

coefs The coefficient estimates. A numeric vector, except when a lambda grid is supplied, in which case a matrix with one column per penalty.

K_hat The estimated number of factors, or NFACTORS if set.

Lambda The penalty selected by cross-validation, on the scale of the paper; the supplied grid if lambda was given; NULL for variant = "LS".

eigenvalues The normalised eigenvalues used to determine K_hat.

Projection_Matrix The computed projection matrix Pi_hat.

type Either "linear" or "dictionary".

call The matched call.

References

Rücker, M., Vogt, M., Linton, O. and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4), 1457–1509. doi:10.3982/QE2308

Rücker, M., Vogt, M. and Linton, O. (2026). High-dimensional panel data models with interactive fixed effects: beyond the linear case. https://arxiv.org/abs/2608.02055

Examples

```
data("data_estimation")
obs_N <- 20
obs_T <- 20
fit <- hdcce_estimator(data = data_estimation, obs_N = obs_N, obs_T = obs_T)
fit$K_hat
head(fit$coefs, 10)
```

hdcce_inference *Inference for HD panels with IFE*

Description

hdcce_inference performs high-dimensional inference with either (i) the desparsified HD-CCE estimator of Rücker et al. (2025), giving confidence intervals for individual coefficients, or (ii) the max-type significance test of Rücker et al. (2026) for the null that a covariate has no effect in the parametric additive model, when a dictionary is supplied.

Usage

```
hdcce_inference(
  data,
  obs_N,
  obs_T,
  COEF_INDEX_VEC,
  dictionaries = NULL,
  TRUNC = 0.01,
  NFACTORS = NULL,
  NFOLDS = 10,
  foldid = NULL,
  alpha = c(0.01, 0.05, 0.1),
  HAC = 2,
  standardize = TRUE,
  s_rule = "lambda.1se",
  kernel = "epanechnikov",
  C = NULL,
  QUANT = 0.9,
  h = NULL,
  B = 5000
)
```

Arguments

data	List containing the balanced panel data with data$y the dependent variable and data$x the regressors. Both are sorted such that the first obs_T observations are those for unit 1, followed by the obs_T observations for unit 2, etc.
obs_N, obs_T	The number of cross-section units (obs_N) and the time series length (obs_T).
COEF_INDEX_VEC	Indices of the regressors to be tested.
dictionaries	Either NULL (linear model, the default) or a list with components Phi, an obs_N * obs_T x d matrix of dictionary transformations of data$x, and group, a d-vector with entries in 1,...,p indicating which covariate each column transforms. Optionally Phi_nw and group_nw give a separate dictionary for the nodewise regression; they default to Phi and group.

TRUNC	The truncation parameter tau used to estimate the number of factors. Default is 0.01.
NFACTORS	Allows to set the number of factors used in estimation. Default is NULL, so that the data driven choice with TRUNC is used.
NFOLDS	Number of folds for cross-validation in `cv.glmnet`.
foldid	Integer vector (`obs_N * obs_T`-dimensional) with a fold label for each observation.
alpha	Vector of significance levels.
HAC	Integer 1, 2 or 3, selecting the homoscedastic, heteroscedasticity robust, or HAC variance estimator. Only used in the linear case.
standardize	Logical; whether `glmnet` is called with standardized projected data.
s_rule	Either `"lambda.1se"` (default) or `"lambda.min"`, the rule used to select the penalties from cross-validation.
kernel	Kernel with support `[-1,1]` used for the test, either `"epanechnikov"`, `"biweight"`, or a function.
C	Half-width of the region of locations scanned by the test, in the units of the nodewise residuals: the statistic is maximised over $w \in [-C, C]$. The default is NULL, in which case C is set to the QUANT quantile of the absolute nodewise residuals, so that the interval $[-C, C]$ contains that share of them.
QUANT	Quantile used for the data-driven choice of C; ignored when C is supplied. The default is 0.9. Larger values reach further into the tails at the cost of thinner outer bumps.
h	Bandwidth of the kernel weights, in the same units as C. The default is NULL, in which case h is chosen by Silverman's rule of thumb applied to the nodewise residuals, $0.9\hat{\sigma}_u(nT)^{-1/5}$.
B	Number of Monte Carlo draws for the Gaussian coupling.

Details

The grid of locations is $\mathcal{W} = \{w \in [-C, C] : w = -C + (2\ell - 1)h,\ \ell = 1, 2, \ldots\}$, so that the supports of the kernel weights are disjoint and the number of locations is $L = \lfloor C/h + 1/2 \rfloor$

Value

A list with components `results` (a list with one entry per element of `COEF_INDEX_VEC`, named by the index), `alpha`, `obs_N`, `obs_T`, `type` and `call`. Each entry of `results` is a list with

coef_despar Desparsified estimate; NULL in the dictionary case.

se Estimated standard error; NULL in the dictionary case.

statistic z-statistic for H0: beta_j = 0, or the max-type statistic in the dictionary case..

critical_values Critical value for each level in `alpha`.

p_value Two-sided normal p-value, or Monte Carlo p-value in the dictionary case..

reject Logical, one entry per level in `alpha`.

confidence_band Confidence interval per level; NULL in the dictionary case.

profile Data frame with the grid of locations, the corresponding Psi_w and the number of residuals in each bump; `NULL` in the linear case.

K_hat, sigma_hat Estimated number of factors and error standard deviation.

References

Rücker, M., Vogt, M., Linton, O. and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4), 1457–1509. doi:10.3982/QE2308

Rücker, M., Vogt, M. and Linton, O. (2026). High-dimensional panel data models with interactive fixed effects: beyond the linear case. https://arxiv.org/abs/2608.02055

Examples

```
data("data_inference")
obs_N <- 20
obs_T <- 20
dat <- list(x = data_inference$x, y = data_inference$y[, 1])
fit <- hdcce_inference(dat, obs_N, obs_T, COEF_INDEX_VEC = 1)
fit$results[["1"]]$confidence_band
```

Index