

# Package ‘xtife’

July 22, 2026

**Type** Package

**Title** Interactive Fixed Effects Estimator for Panel Data

**Version** 0.1.4

**Date** 2026-07-22

**Description** Implements the interactive fixed effects (‘IFE’) panel estimator of Bai (2009) <[doi:10.3982/ECTA6135](https://doi.org/10.3982/ECTA6135)> for balanced and unbalanced panels, with optional additive unit and/or time fixed effects. Provides analytical standard errors (‘homoskedastic’, ‘HC1’ heteroskedasticity-robust, cluster-robust by unit, and heteroskedasticity- and autocorrelation-consistent), together with asymptotic incidental-parameter bias correction for large panels, including a dynamic extension for predetermined (lagged-dependent) regressors following Moon and Weidner (2017) <[doi:10.1017/S0266466615000328](https://doi.org/10.1017/S0266466615000328)>. The number of factors is chosen by information criteria (Bai and Ng 2002 <[doi:10.1111/1468-0262.00273](https://doi.org/10.1111/1468-0262.00273)>) or by singular value thresholding. Unbalanced panels are handled by an expectation-maximisation algorithm with nuclear-norm-regularised initialisation, with estimation, analytical inference, and bias correction following Su, Wang and Wang (2025) <[doi:10.2139/ssrn.5177283](https://doi.org/10.2139/ssrn.5177283)> and building on the matrix-completion and missing-data factor analysis of Bai and Ng (2021) <[doi:10.1080/01621459.2021.1967163](https://doi.org/10.1080/01621459.2021.1967163)>. All computations use base R only, with no external dependencies.

**License** GPL-2 | GPL-3

**Encoding** UTF-8

**Language** en-US

**LazyData** true

**RoxygenNote** 7.3.3

**Depends** R (>= 3.5.0)

**Imports** stats

**Suggests** testthat (>= 3.0.0), knitr, rmarkdown

**VignetteBuilder** knitr

**URL** <https://github.com/Rickchen0910/xtife>

**BugReports** <https://github.com/Rickchen0910/xtife/issues>  
**NeedsCompilation** no  
**Author** Binzhi Chen [aut, cre]  
**Maintainer** Binzhi Chen <Binzhi.Chen9@gmail.com>  
**Repository** CRAN  
**Date/Publication** 2026-07-22 18:40:02 UTC

Contents

|                            |           |
|----------------------------|-----------|
| cigar . . . . .            | 2         |
| ife . . . . .              | 3         |
| ife_select_r . . . . .     | 5         |
| ife_select_r_unb . . . . . | 6         |
| ife_unbalanced . . . . .   | 7         |
| print.ife . . . . .        | 10        |
| <b>Index</b>               | <b>12</b> |

---

|       |                                             |
|-------|---------------------------------------------|
| cigar | <i>Dataset on US Cigarette Demand Panel</i> |
|-------|---------------------------------------------|

---

Description

Balanced panel of cigarette sales and prices across 46 US states for 30 years (1963–1992). Originally used in Baltagi (1995) and widely used as a benchmark dataset for panel estimators.

Usage

cigar

Format

A data frame with 1,380 rows and 9 variables:

**state** US state identifier (integer, 1–46)  
**year** year (integer, 1963–1992)  
**price** cigarette price index  
**pop** state population  
**pop16** population aged 16 and over  
**cpi** consumer price index  
**ndi** per-capita disposable income  
**sales** per-capita cigarette sales (packs per person per year)  
**pimin** minimum cigarette price in adjoining states

## Source

Baltagi, B.H. (1995) *Econometric Analysis of Panel Data*. Wiley. Distributed with the **plm** R package (Croissant and Millo 2008).

## References

Baltagi, B.H. (1995). *Econometric Analysis of Panel Data*. Wiley.

Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: the plm package. *Journal of Statistical Software*, 27(2), 1–43. doi:[10.18637/jss.v027.i02](https://doi.org/10.18637/jss.v027.i02)

---

 ife

---

*Estimate Interactive Fixed Effects Model (Bai 2009)*


---

## Description

Fits the panel model

$$y_{it} = \alpha_i + \xi_t + X'_{it}\beta + \lambda'_i F_t + u_{it}$$

for balanced panel data with analytical standard errors.

## Usage

```
ife(
  formula,
  data,
  index,
  r = 1L,
  force = "two-way",
  se = "standard",
  bias_corr = FALSE,
  method = "static",
  M1 = 1L,
  tol = 1e-09,
  max_iter = 10000L
)
```

## Arguments

|         |                                                               |
|---------|---------------------------------------------------------------|
| formula | R formula: outcome ~ covariate1 + covariate2 + ...            |
| data    | data.frame in long format (one row per unit-time observation) |
| index   | character(2): c("unit_id_column", "time_id_column")           |
| r       | integer >= 0, number of interactive factors (default 1)       |

|           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| force     | additive FE specification: "none"   "unit"   "time"   "two-way" (default "two-way"). Additive unit effects $\alpha_i$ and time effects $\xi_t$ are removed via the standard within transformation (iterative demeaning) before the SVD algorithm runs, following Bai (2009) Section 3. Bai (2009, p.1) shows that two-way additive effects are a special case of the interactive structure with $r = 2$ (setting $F_t = (1, \xi_t)'$ and $\lambda_i = (\alpha_i, 1)'$ ), so the IFE estimator remains consistent when additive effects are present regardless of the force choice, but pre-demeaning improves efficiency. |
| se        | SE type: "standard"   "robust"   "cluster" (default "standard"; "cluster" clusters by unit id)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| bias_corr | logical; if TRUE apply bias correction. For method = "static" uses the two-term Bai (2009) Sec. 7 correction ( $B/N + C/T$ ). For method = "dynamic" uses the three-term Moon and Weidner (2017) correction ( $B1/T + B2/N + B3/T$ ). Requires $r > 0$ and at least one covariate. (default FALSE)                                                                                                                                                                                                                                                                                                                        |
| method    | "static" (default) for Bai (2009) strictly-exogenous regressors; "dynamic" for Moon and Weidner (2017) predetermined regressors (e.g. lagged dependent variable). The dynamic estimator uses double projection $M_\Lambda M_F$ on $X$ in the SVD loop.                                                                                                                                                                                                                                                                                                                                                                    |
| M1        | integer; lag bandwidth for the B1 dynamic bias term (default 1L). Only used when method = "dynamic" and bias_corr = TRUE.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| tol       | convergence tolerance (default $1e-9$ )                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| max_iter  | maximum iterations (default 10000L)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

## Value

An S3 object of class "ife" with the following components:

- coef – named p-vector of estimated coefficients
- vcov –  $p \times p$  variance-covariance matrix
- se – named p-vector of standard errors
- tstat – named p-vector of t-statistics
- pval – named p-vector of two-sided p-values
- ci –  $p \times 2$  matrix of 95% confidence intervals (CI.lower, CI.upper)
- table – data.frame coefficient table (Estimate, Std.Error, t.value, Pr.t, CI.lower, CI.upper)
- F\_hat –  $T \times r$  estimated factor matrix
- Lambda\_hat –  $N \times r$  estimated loading matrix
- residuals –  $T \times N$  residual matrix (full model)
- sigma2 – estimated error variance
- df – residual degrees of freedom
- n\_iter – iterations to convergence
- converged – logical
- N, T, r, force, se\_type – model dimensions and options
- call – matched call

## References

- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279. doi:[10.3982/ECTA6135](https://doi.org/10.3982/ECTA6135)
- Moon, H.R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33, 158–195. doi:[10.1017/S0266466615000328](https://doi.org/10.1017/S0266466615000328)
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221. doi:[10.1111/14680262.00273](https://doi.org/10.1111/14680262.00273)

## Examples

```
data(cigar, package = "xtife")
fit <- ife(sales ~ price, data = cigar, index = c("state", "year"),
          r = 2, force = "two-way", se = "standard")
print(fit)
```

---

ife\_select\_r

---

Select the Number of Factors via Information Criteria

---

## Description

Fits the IFE model for  $r = 0, 1, \dots, r\_max$  and evaluates five information criteria at each value of  $r$ . Returns IC1, IC2, and IC3 from Bai and Ng (2002) Proposition 1, applied to IFE residuals per Bai (2009) Section 9.4, plus a BIC-style penalty (IC\_bic) and a small-sample-corrected prediction criterion (PC) from Bai (2009). The criterion-minimising  $r$  for each IC is flagged with "\*" in the printed table, and a data-driven recommendation (favouring IC\_bic when the Bai-Ng criteria decrease monotonically) is displayed.

## Usage

```
ife_select_r(
  formula,
  data,
  index,
  r_max = NULL,
  force = "two-way",
  verbose = TRUE,
  tol = 1e-09,
  max_iter = 10000L
)
```

## Arguments

|         |                                                            |
|---------|------------------------------------------------------------|
| formula | R formula passed to ife()                                  |
| data    | long-format data.frame                                     |
| index   | character(2): c("unit_id", "time_id")                      |
| r_max   | maximum r to consider (default: min(8, floor(min(N,T)/2))) |

|          |                                                                                                                |
|----------|----------------------------------------------------------------------------------------------------------------|
| force    | additive FE type (default "two-way")                                                                           |
| verbose  | logical; if TRUE (default) print progress and results table to the console. Set to FALSE for silent operation. |
| tol      | convergence tolerance (default 1e-9)                                                                           |
| max_iter | maximum iterations (default 10000L)                                                                            |

Value

(invisibly) a data.frame with columns r, V\_r, IC1, IC2, IC3, IC\_bic, PC, converged, and attribute "suggested" (named integer vector giving the IC-minimising r for each criterion).

References

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279. doi:10.3982/ECTA6135

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221. doi:10.1111/14680262.00273

Examples

```
data(cigar, package = "xtife")
sel <- ife_select_r(sales ~ price, data = cigar,
                   index = c("state", "year"), r_max = 4)
```

---

|                  |                                                                 |
|------------------|-----------------------------------------------------------------|
| ife_select_r_unb | <i>Factor Number Selection for Unbalanced Panel IFE via SVT</i> |
|------------------|-----------------------------------------------------------------|

---

Description

Estimates the number of interactive factors in an unbalanced panel by the singular value thresholding (SVT) rule of Su, Wang and Wang (2025), applied to the nuclear-norm-regularised (soft-imputed) matrix — a missing-data counterpart of the information-criterion rules of Bai and Ng (2002).

Usage

```
ife_select_r_unb(formula, data, index, c_f = 0.6, nu_NT = NULL, verbose = TRUE)
```

Arguments

|         |                                                                                                                                  |
|---------|----------------------------------------------------------------------------------------------------------------------------------|
| formula | R formula: outcome ~ covariate1 + covariate2 + ...                                                                               |
| data    | Data frame in long format.                                                                                                       |
| index   | Character vector of length 2: c("unit_col", "time_col").                                                                         |
| c_f     | SVT threshold constant (default 0.6).                                                                                            |
| nu_NT   | Optional scalar or vector of NNR penalty values. If NULL (default), cross-validates over c(0.01, 0.1, 1, 10) * sqrt(max(N, TT)). |
| verbose | Logical; print result table. Default TRUE.                                                                                       |

**Value**

Invisibly returns a list with components `r_hat`, `sv` (normalised singular values), `threshold`, `c_f`, `c_NT`, and `nu_used`.

**References**

- Su, L., Wang, F. and Wang, Y. (2025). Estimation and inference for interactive fixed effects panel data models with unbalanced panels. SSRN Working Paper No. 5177283. doi:10.2139/ssrn.5177283
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221. doi:10.1111/14680262.00273
- Bai, J. and Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 116(536), 1746–1763. doi:10.1080/01621459.2021.1967163

**Examples**

```
data(cigar, package = "xtife")
set.seed(42)
cigar_unb <- cigar[sample(nrow(cigar), 1200L), ]
ife_select_r_unb(sales ~ price, data = cigar_unb,
                 index = c("state", "year"))
```

ife\_unbalanced

*Unbalanced Panel Interactive Fixed Effects Estimator***Description**

Fits the interactive fixed effects model

$$Y_{it} = \alpha_i + \xi_t + X'_{it}\beta + \lambda'_i F_t + u_{it}$$

for unbalanced panels (units observed at different sets of time periods), where the additive unit effects  $\alpha_i$  and time effects  $\xi_t$  are controlled by force. The estimation and inference theory follows Su, Wang and Wang (2025). Estimation uses an alternating outer loop that updates  $\hat{\beta}$  and the structure  $(\hat{\alpha}, \hat{\xi}, \hat{\lambda}, \hat{F})$ , with an expectation-maximisation (EM) inner loop — in the spirit of Bai (2009, Appendix B) and the missing-data factor / matrix-completion framework of Bai and Ng (2021) — that imputes the unobserved cells from the current structure and re-estimates the additive and interactive components on the completed panel. An optional nuclear-norm-regularised (soft-impute) warm start (Mazumder, Hastie and Tibshirani 2010) is available via `init = "nnr"`.

**Usage**

```
ife_unbalanced(
  formula,
  data,
  index,
```

```

r = 1L,
force = "none",
se = "standard",
init = "ols",
bias_corr = FALSE,
exog = "strict",
L_T = NULL,
c_f = 0.6,
nu_NT = NULL,
tol = 1e-09,
max_iter = 10000L,
tol_em = 1e-07,
max_iter_em = 500L
)

```

### Arguments

|           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| formula   | R formula: outcome ~ covariate1 + covariate2 + ...                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| data      | Data frame in long format (one row per observed unit-time pair).                                                                                                                                                                                                                                                                                                                                                                                                                          |
| index     | Character vector of length 2: c("unit_col", "time_col").                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| r         | Positive integer. Number of interactive factors (default 1).                                                                                                                                                                                                                                                                                                                                                                                                                              |
| force     | Additive fixed effects to remove jointly with the factors: "none" (default; intercept-free interactive model), "unit", "time", or "two-way". Additive FE are estimated jointly with the factors via EM on the imputed (completed) panel, which is robust to informative (factor-correlated) missingness; use force = "two-way" for data with level/trend structure (matching the balanced ife() default). Note: with strong, near-collinear common trends, "two-way" can converge slowly. |
| se        | SE type: "standard" (homoskedastic), "robust" (HC1), "cluster" (cluster-robust by unit), or "hac" (HAC with a Bartlett kernel, for serially correlated errors). Default "standard".                                                                                                                                                                                                                                                                                                       |
| init      | Initialisation method: "ols" (default, grand-mean OLS) or "nnr" (nuclear-norm regularisation / soft-impute).                                                                                                                                                                                                                                                                                                                                                                              |
| bias_corr | Logical. Apply the analytical incidental-parameter bias correction. Supports both strictly and weakly exogenous regressors (controlled by exog). Default FALSE.                                                                                                                                                                                                                                                                                                                           |
| exog      | Exogeneity assumption: "strict" (default, regressors uncorrelated with past and future errors) or "weak" (weakly exogenous, e.g., lagged dependent variable $x_{it} = y_{i,t-1}$ ). When "weak" and bias_corr = TRUE, an additional dynamic bias term $\hat{b}_2$ is included.                                                                                                                                                                                                            |
| L_T       | Bartlett kernel bandwidth for HAC standard errors (se = "hac") and the dynamic bias term $\hat{b}_2$ (exog = "weak", bias_corr = TRUE). If NULL (default), set to $\lfloor 2T^{1/5} \rfloor$ after the panel dimensions are known.                                                                                                                                                                                                                                                        |
| c_f       | Singular-value-thresholding constant (default 0.6) used for factor-number selection. Used only when init = "nnr".                                                                                                                                                                                                                                                                                                                                                                         |
| nu_NT     | NNR penalty grid. If NULL (default), cross-validates over $c * \sqrt{\max(N, TT)}$ for c in c(0.01, 0.1, 1, 10).                                                                                                                                                                                                                                                                                                                                                                          |



|                          |                                                                                                    |
|--------------------------|----------------------------------------------------------------------------------------------------|
| <code>tol</code>         | Outer-loop convergence tolerance on $\max  \hat{\beta}^{new} - \hat{\beta}^{old} $ . Default 1e-9. |
| <code>max_iter</code>    | Maximum outer-loop iterations. Default 10000L.                                                     |
| <code>tol_em</code>      | Inner EM convergence tolerance. Default 1e-7.                                                      |
| <code>max_iter_em</code> | Maximum inner EM iterations per outer step. Default 500L.                                          |

## Details

**Inference.** Standard errors use a sandwich estimator on factor-projected regressors (the unbalanced analogue of the balanced formula of Bai 2009), following Su, Wang and Wang (2025), with heteroskedasticity-robust, cluster-robust (Arellano 1987; Cameron, Gelbach and Miller 2011) and HAC (Newey and West 1987) variants. The optional analytical bias correction of Su, Wang and Wang (2025) removes the leading incidental-parameter bias, extending the corrections of Bai (2009) and Moon and Weidner (2017) to the unbalanced and predetermined-regressor case.

**Additive fixed effects.** With `force = "none"` (default) the model is intercept-free and all heterogeneity is carried by the interactive factors. Setting `force = "unit"`, `"time"` or `"two-way"` estimates the additive effects jointly with the factors by demeaning the *imputed* (completed) panel inside the EM loop, which is robust to informative missingness; the degrees-of-freedom adjustment is propagated to the standard errors.

## Value

An S3 object of class `"ife_unb"` with components:

- coef** Named p-vector of estimated coefficients  $\hat{\beta}$  (bias-corrected when `bias_corr = TRUE`).
- coef\_raw** Named p-vector of uncorrected coefficients (only when `bias_corr = TRUE`).
- vcov** p x p variance-covariance matrix.
- se** Named p-vector of standard errors.
- tstat** Named p-vector of t-statistics.
- pval** Named p-vector of two-sided p-values.
- ci** p x 2 matrix of 95 percent confidence intervals.
- table** Data frame coefficient table.
- F\_hat** TT x r estimated factor matrix (normalised  $F'F/TT = I_r$ ).
- Lambda\_hat** N x r estimated loading matrix.
- residuals** n\_obs numeric vector of full-model residuals at observed cells.
- sigma2** Estimated error variance ( $\text{sum}(u^2)/df$ ).
- df** Residual degrees of freedom.
- n\_obs** Number of observed unit-time cells.
- n\_iter** Outer-loop iterations to convergence.
- converged** Logical.
- N, TT, r, se\_type** Model dimensions and options.
- init, bias\_corr, exog, L\_T** Options used.
- b\_hat, b2, b3, b4, b5, b6** Bias components (only when `bias_corr = TRUE`). b2 is a zero vector when `exog = "strict"`.

**y\_name, x\_names, id\_col, time\_col** Variable names.  
**unit\_vals, time\_vals** Unique unit and time identifiers.  
**unit\_idx, time\_idx** Integer index vectors for residuals.  
**call** Matched call.

## References

- Su, L., Wang, F. and Wang, Y. (2025). Estimation and inference for interactive fixed effects panel data models with unbalanced panels. SSRN Working Paper No. 5177283. [doi:10.2139/ssrn.5177283](#)
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279. [doi:10.3982/ECTA6135](#)
- Bai, J. and Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 116(536), 1746–1763. [doi:10.1080/01621459.2021.1967163](#)
- Mazumder, R., Hastie, T. and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11, 2287–2322.
- Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33, 158–195. [doi:10.1017/S0266466615000328](#)

## Examples

```
data(cigar, package = "xtife")
# Drop ~10 % of rows to create an unbalanced panel
set.seed(1)
cigar_unb <- cigar[sample(nrow(cigar), 1200L), ]
fit <- ife_unbalanced(sales ~ price, data = cigar_unb,
                     index = c("state", "year"), r = 2L)
print(fit)
```

---

print.ife

---

*Print an IFE Model Summary*


---

## Description

Prints a formatted summary of an object of class "ife", including panel dimensions, number of factors, additive fixed effect specification, SE type, and a coefficient table with standard errors, t-statistics, p-values, and 95% confidence intervals. If bias correction was applied, bias terms are also reported. Information criteria are printed when the object contains them (i.e., when called from `ife_select_r()`).

## Usage

```
## S3 method for class 'ife'
print(x, digits = 4, ...)
```

**Arguments**

|                     |                                          |
|---------------------|------------------------------------------|
| <code>x</code>      | an object of class "ife"                 |
| <code>digits</code> | number of significant digits (default 4) |
| <code>...</code>    | unused                                   |

**Value**

`x` invisibly.

**Examples**

```
data(cigar, package = "xtife")
fit <- ife(sales ~ price, data = cigar, index = c("state", "year"),
          r = 2, force = "two-way", se = "standard")
print(fit)
```

# Index

## \* datasets

cigar, [2](#)

cigar, [2](#)

ife, [3](#)

ife\_select\_r, [5](#)

ife\_select\_r\_unb, [6](#)

ife\_unbalanced, [7](#)

print.ife, [10](#)