Package: xtfifevd (via r-universe)

November 1, 2026

Type Package

Title Panel Fixed Effects Filtered and Variance Decomposition Estimation

Version 1.1.0

Date 2026-09-30

Description Implements fixed effects estimators for time-invariant variables in panel data models. Provides three estimation methods: FEVD (Fixed Effects Vector Decomposition) from Plumper and Troeger (2007) <doi:10.1093/pan/mpm002>, and FEF (Fixed Effects Filtered) and FEF-IV (instrumental variables variant) from Pesaran and Zhou (2018) <doi:10.1080/07474938.2016.1222225>. All methods use the Pesaran and Zhou variance estimators, which account for generated regressor uncertainty, and report the full covariance matrix of the time-varying, time-invariant and intercept coefficients.

License GPL-3

URL https://github.com/muhammedalkhalaf/xtfifevd

BugReports https://github.com/muhammedalkhalaf/xtfifevd/issues

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Depends R (>= 4.1.0)

Imports stats, utils

Suggests plm, testthat (>= 3.0.0)

Config/testthat/edition 3

Repository https://muhammedalkhalaf.r-universe.dev

Date/Publication 2026-10-02 22:48:07 UTC

RemoteUrl https://github.com/muhammedalkhalaf/xtfifevd

RemoteRef HEAD

RemoteSha 728458ba0589cba43c0c853167d16b364574c1a9

Contents

bw_ratio 2
xtfifevd 3

Index **9**

`bw_ratio` *Between/Within SD Ratio for Time-Invariant Variables*

Description

Computes the between-panel and within-panel standard deviations for specified variables, along with their ratio. This diagnostic helps assess whether FEVD/FEF methods may improve upon standard FE estimation.

Usage

```
bw_ratio(data, variables, id)
```

Arguments

`data`	A data frame containing the panel data.
`variables`	A character vector of variable names to analyze.
`id`	Character string naming the panel identifier variable.

Details

For truly time-invariant variables, the within-panel SD should be zero (or near-zero due to numerical precision), giving an infinite ratio. Plumper and Troeger (2007, p. 134) define the b/w ratio as the between SD of a variable divided by its within SD, which is what this function reports.

Plumper and Troeger (2007, Section 6) compare the root mean squared error of the FE and FEVD estimators of a rarely changing variable in a Monte Carlo design with N = 30 units and T = 20 periods (their footnote 13 reports that the findings for rarely changing variables were confirmed for N in 15 to 100 and T in 20 to 100). The b/w ratio above which FEVD has the lower RMSE depends on the (unobservable) correlation between the rarely changing variable and the unit effects: about 0.2 when the correlation is 0, about 1.7 at a correlation of 0.3, about 2.8 at 0.5 and close to 3.8 at 0.8 (their Fig. 4, pp. 135-136). The often quoted threshold of 1.7 is therefore a result for corr(z, u) = 0.3 from their Fig. 4 (N = 30, T = 20), and Plumper and Troeger themselves write that they "cannot offer a simple rule of thumb", adding (p. 136) that "the odds are that at a b/w ratio of at least 2.8, the variable is better included into the stage 2 estimation". This function prints these thresholds for orientation only.

Two caveats from Plumper and Troeger apply. First, the FEVD (and FEF) coefficient on a time-invariant or rarely changing variable is biased whenever that variable is correlated with the unobserved unit effects (p. 129); the bias is the usual omitted variable bias and does not vanish with a high b/w ratio. Second, this correlation cannot be observed or tested, because the unit effects are unobservable (p. 135); the trade-off between the bias of FEVD and the inefficiency of FE for rarely changing variables therefore rests on an untestable assumption.

Value

A data frame with columns:

variable Variable name

sd_between Between-panel standard deviation

sd_within Within-panel standard deviation

bw_ratio Ratio of between to within SD

References

Plumper, T. and Troeger, V. E. (2007). Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis*, 15(2), 124-139. doi:10.1093/pan/mpm002

Examples

```
# Create example data
set.seed(42)
N <- 50
T <- 5
id <- rep(1:N, each = T)
z_invariant <- rep(rnorm(N), each = T)  # Truly time-invariant
z_slow <- rep(rnorm(N), each = T) + rnorm(N * T, sd = 0.1)  # Slowly varying
x_varying <- rnorm(N * T)  # Time-varying

data <- data.frame(id = id, z_inv = z_invariant,
                   z_slow = z_slow, x = x_varying)

bw_ratio(data, c("z_inv", "z_slow", "x"), id = "id")
```

---

xtfifevd *Panel Fixed Effects Estimation for Time-Invariant Variables*

Description

Estimates panel models with time-invariant regressors using FEVD, FEF, or FEF-IV methods. Standard fixed effects estimation cannot identify coefficients on time-invariant variables; these methods decompose or filter the unit effects to recover these coefficients.

Usage

```
xtfifevd(
  formula,
  data,
  id,
  time,
```

```
  method = c("fevd", "fef", "fef_iv"),
  instruments = NULL,
  vcov_beta = c("robust", "classical"),
  na.action = na.omit
)

fevd(
  formula,
  data,
  id,
  time,
  vcov_beta = c("robust", "classical"),
  na.action = na.omit
)

fef(
  formula,
  data,
  id,
  time,
  vcov_beta = c("robust", "classical"),
  na.action = na.omit
)

fef_iv(
  formula,
  data,
  id,
  time,
  instruments,
  vcov_beta = c("robust", "classical"),
  na.action = na.omit
)
```

Arguments

`formula`	A formula of the form `y ~ x1 + x2 | z1 + z2` where terms before `|` are time-varying and terms after `|` are time-invariant. Both parts are processed with `stats::model.matrix()`, so transformations (`log(x)`, `I(x^2)`, `x:w`, `log(y)` on the left-hand side) and factors (with their contrasts) are handled as in `stats::lm()`. Intercept columns are removed; the model intercept is estimated in stage 2.
`data`	A data frame containing the variables.
`id`	Character string naming the panel (individual) identifier variable.
`time`	Character string naming the time identifier variable.
`method`	Estimation method: `"fevd"` (default), `"fef"`, or `"fef_iv"`.
`instruments`	For `method = "fef_iv"`, a one-sided formula specifying instrumental variables, e.g., `~ iv1 + iv2`.

`vcov_beta`	Covariance estimator for the stage 1 (within) coefficients `beta`: `"robust"` (default) is the Arellano-type panel-robust matrix of Pesaran and Zhou (2018, eq. 18), valid under heteroskedasticity and serial correlation within panels; `"classical"` is the homoskedastic $\hat{\sigma}_e^2 (X'MX)^{-1}$. The chosen matrix is also used inside the Pesaran and Zhou variance of `gamma` and of the intercept.
`na.action`	How to handle missing values. Default is `na.omit`.

Details

Model:

The panel model is:

$$y_{it} = \alpha + \alpha_i + x_{it}'\beta + z_i'\gamma + \varepsilon_{it}$$

where x_{it} are time-varying regressors, z_i are time-invariant regressors, and α_i are individual effects that may be correlated with x_{it}.

Stage 1 (all methods):

Within (fixed effects) regression of y_{it} on x_{it} yields $\hat{\beta}$ and the time-averaged FE residuals $\bar{u}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$ (Pesaran and Zhou 2018, eq. 3).

Stage 2:

- **FEF**: unit-level OLS of $\bar{u}_i$ on an intercept and z_i (Pesaran and Zhou eq. 4 and 5).
- **FEF-IV**: unit-level 2SLS using instruments r_i (Pesaran and Zhou eq. 48).
- **FEVD**: as FEF; the residuals h_i are the unexplained part of the unit effect (Plumper and Troeger 2007, eq. 6).

Stage 3 (FEVD only):

Pooled OLS of y_{it} on an intercept, x_{it}, z_i and h_i (Plumper and Troeger 2007, eq. 7). Plumper and Troeger's stage 2 equation (5) is printed without an intercept, but their $\hat{u}_i$ (eq. 4) contains the model constant and their stage 3 equation (7) has an intercept; the package therefore includes an intercept in stage 2, which makes the stage 3 estimates identical to FEF and the coefficient δ on h_i identically equal to 1 (Pesaran and Zhou 2018, Proposition 3). The identity holds in balanced and unbalanced panels: with $(a, \hat{\beta}, \hat{\gamma}, 1)$ the stage 3 residuals are the within residuals, which sum to zero within every unit and are orthogonal to x_{it}, so they satisfy the stage 3 normal equations exactly. Without the stage 2 intercept the FEVD estimator is in general biased (Pesaran and Zhou 2018, Section 3.4). Plumper and Troeger are silent on unbalanced panels; the package runs stage 2 at the unit level without weights, one observation per panel unit (Pesaran and Zhou's FEF), rather than at the observation level, where units would be weighted by T_i.

The naive stage 3 OLS standard errors of Plumper and Troeger are too small for the time-invariant coefficients because they ignore that h_i is a generated regressor (Breusch, Ward, Nguyen and Kompas 2010, Theorem 3; Greene 2011; Pesaran and Zhou 2018). In a Monte Carlo check by the package author (400 replications, N = 200, T = 8, two time-varying and two time-invariant regressors, AR(1) errors with coefficient 0.8 and heteroskedastic across units, x correlated with the unit effects) their 95 percent coverage for γ was 35 to 40 percent, against 92 to 95 percent for the Pesaran and Zhou standard errors. They are returned in `stage3$se_naive` for reference only and are never used for inference.

Rarely changing variables:

If a variable after | varies within panels, a warning is issued and its unit (panel) mean is used as z_i in all stages. This is the package's choice; Plumper and Troeger (2007) do not specify how rarely changing variables enter stage 2.

Variance estimation:

The `gamma` block of `vcov` is Pesaran and Zhou (2018) equation 17 (FEF and FEVD) or equation 51 (FEF-IV), which account for the estimation uncertainty of $\hat{\beta}$ through the matrix `vcov_beta`. The `beta` block is `vcov_beta` itself (robust eq. 18 by default). The remaining blocks follow from Pesaran and Zhou eq. (A.11),

$$\hat{\gamma} - \gamma = Q_{zz}^{-1} \left[N^{-1} \sum_i (z_i - \bar{z}) v_i - Q_{z\bar{x}} (\hat{\beta} - \beta) \right],$$

so that $Cov(\hat{\gamma}, \hat{\beta}) = -Q_{zz}^{-1} Q_{z\bar{x}} Var(\hat{\beta})$, and from eq. (5), $\hat{\alpha} = \bar{u} - \bar{z}'\hat{\gamma}$, by the delta method with $c = \bar{x} - Q'_{z\bar{x}} Q_{zz}^{-1} \bar{z}$: $Var(\hat{\alpha}) = N^{-2} \sum_i w_i^2 \hat{v}_i^2 + c' Var(\hat{\beta}) c$, $w_i = 1 - \bar{z}' Q_{zz}^{-1} (z_i - \bar{z})$, with the corresponding covariances. For FEF-IV the same derivation is used with $Q_{zz}^{-1}(z_i - \bar{z})$ replaced by $H_{zr}(r_i - \bar{r})$ and $Q_{z\bar{x}}$ by $Q_{r\bar{x}}$. As in Pesaran and Zhou's eq. (17), the cross term between the unit-level scores and $\hat{\beta}$ (the term defined in their eq. 15, negligible under their condition 16) is dropped in $Var(\hat{\alpha})$ and in $Cov(\hat{\gamma}, \hat{\beta})$.

Value

An object of class `"xtfifevd"` containing:

coefficients Named vector of all coefficients (beta, gamma, `_cons`). For `method = "fevd"` these are the stage 3 pooled OLS estimates, which equal the FEF estimates (see Details).

vcov Full variance-covariance matrix of `coefficients` from the Pesaran and Zhou (2018) derivation, including the covariances between `beta`, `gamma` and the intercept (see Details).

beta Coefficients on time-varying variables

gamma Coefficients on time-invariant variables

intercept Overall intercept

delta (FEVD only) Stage 3 coefficient on the unexplained unit effect h_i; equal to 1 by construction.

stage3 (FEVD only) List with the full stage 3 pooled OLS coefficient vector (`coefficients`, including `h`), the naive OLS covariance matrix `vcov_naive` and standard errors `se_naive`. These naive standard errors are known to be too small for the time-invariant coefficients and are returned for reference only.

fef (FEVD only) The stage 1 and 2 (FEF) coefficients.

residuals Idiosyncratic (within) residuals from stage 1

fitted.values Fitted values $x'_{it}\hat{\beta} + \bar{z}'_i\hat{\gamma} + \hat{\alpha}$ (unit effects excluded)

sigma2_e Variance of the idiosyncratic error

sigma2_u Variance of the unexplained unit effect: the stage 2 residual variance minus $\hat{\sigma}_e^2 \,\mathrm{mean}(1/T_i)$, truncated at zero

vcov_beta The `vcov_beta` choice used

V_beta_robust, V_beta_classical Both covariance matrices of `beta`

V_gamma_pz The gamma block of vcov (Pesaran and Zhou eq. 17 or 51)

stage2_residuals Unit-level stage 2 residuals

N Total number of observations

N_g Number of groups (panels)

T_bar Average time periods per panel

balanced Logical, whether the panel is balanced

method Estimation method used

call The matched call

Functions

- `fevd()`: FEVD estimation (3-stage, Plumper and Troeger 2007, with Pesaran and Zhou 2018 standard errors)
- `fef()`: FEF estimation (2-stage, Pesaran and Zhou 2018)
- `fef_iv()`: FEF-IV estimation with instruments

References

Breusch, T., Ward, M. B., Nguyen, H. and Kompas, T. (2010). On the fixed-effects vector decomposition. MPRA Paper No. 21452. https://mpra.ub.uni-muenchen.de/21452/

Greene, W. H. (2011). Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models? *Political Analysis*, 19(2), 135-146. doi:10.1093/pan/mpq034

Plumper, T. and Troeger, V. E. (2007). Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis*, 15(2), 124-139. doi:10.1093/pan/mpm002

Pesaran, M. H. and Zhou, Q. (2018). Estimation of time-invariant effects in static panel data models. *Econometric Reviews*, 37(10), 1137-1171. doi:10.1080/07474938.2016.1222225

See Also

`fevd()`, `fef()`, `fef_iv()`, `bw_ratio()`

Examples

```
# Simulate panel data
set.seed(123)
N <- 100  # panels
T <- 10   # time periods
n <- N * T

# Generate data
id <- rep(1:N, each = T)
time <- rep(1:T, N)
alpha_i <- rep(rnorm(N), each = T)  # Fixed effects
z <- rep(rnorm(N), each = T)        # Time-invariant
x <- rnorm(n)                        # Time-varying
```

```
y <- 1 + 2 * x + 0.5 * z + alpha_i + rnorm(n, sd = 0.5)

data <- data.frame(id = id, time = time, y = y, x = x, z = z)

# Estimate with different methods
fit_fevd <- xtfifevd(y ~ x | z, data = data, id = "id", time = "time")
summary(fit_fevd)
fit_fevd$delta   # equals 1 by construction

fit_fef <- xtfifevd(y ~ x | z, data = data, id = "id", time = "time",
                    method = "fef")
summary(fit_fef)

# Transformations in the formula are allowed
fit_log <- fef(y ~ x + I(x^2) | z, data = data, id = "id", time = "time")
coef(fit_log)
```

Index

bw_ratio, 2
bw_ratio(), *7*

fef *(*xtfifevd*)*, 3
fef(), *7*
fef_iv *(*xtfifevd*)*, 3
fef_iv(), *7*
fevd *(*xtfifevd*)*, 3
fevd(), *7*

stats::lm(), *4*
stats::model.matrix(), *4*

xtfifevd, 3