Package 'xtdhcoint'

October 31, 2026

Type Package

Title Durbin-Hausman Panel Cointegration Tests

Version 1.0.5

Date 2026-10-30

Description Implements the Durbin-Hausman panel cointegration tests of Westerlund (2008) <doi:10.1002/jae.967>. The tests are robust to cross-sectional dependence through common factor extraction using principal components. Provides both group-mean (DHg) and panel (DHp) test statistics with automatic factor number selection via information criteria.

License GPL-3

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

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

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports stats

Suggests testthat (>= 3.0.0), knitr, rmarkdown

RoxygenNote 7.3.3

NeedsCompilation no

Author Muhammad Alkhalaf [aut, cre, cph] (ORCID: <https://orcid.org/0009-0002-2677-9246>), Joakim Westerlund [ctb] (Original GAUSS code author)

Maintainer Muhammad Alkhalaf <muhammedalkhalaf@gmail.com>

Repository CRAN

Date/Publication 2026-10-31 06:50:02 UTC

Contents

coef.xtdhcoint 2
fisher_panel 2
print.xtdhcoint 3
summary.xtdhcoint 4
xtdhcoint 4

Index **8**

`coef.xtdhcoint` *Coef Method for xtdhcoint Objects*

Description

Extracts the test statistics from the Durbin-Hausman test.

Usage

```
## S3 method for class 'xtdhcoint'
coef(object, ...)
```

Arguments

`object`	An object of class `"xtdhcoint"`.
`...`	Additional arguments (currently ignored).

Value

Named numeric vector of z-statistics and p-values.

`fisher_panel` *Fisher Effect Panel Dataset*

Description

Simulated panel dataset for testing the Fisher effect (cointegration between inflation and nominal interest rates). The data are generated with cointegration and common factors, suitable for demonstrating the Durbin-Hausman panel cointegration tests.

Usage

```
fisher_panel
```

Format

A data frame with 1000 observations and 4 variables:

country Character. Country identifier (Country01 through Country20).

year Integer. Year (1970-2019).

inflation Numeric. Annual inflation rate (percent).

interest Numeric. Nominal interest rate (percent).

Details

The data are simulated according to:

$$inflation_{it} = \alpha_i + \beta_i \cdot interest_{it} + F_t' \lambda_i + e_{it}$$

where F_t are common factors, λ_i are factor loadings, and e_{it} is a stationary cointegration error (under cointegration).

The cointegrating coefficient β_i is approximately 1, consistent with the Fisher hypothesis.

Source

Simulated data.

Examples

```
data(fisher_panel)
head(fisher_panel)
table(fisher_panel$country)
```

---

print.xtdhcoint *Print Method for xtdhcoint Objects*

Description

Prints the results of the Durbin-Hausman panel cointegration test.

Usage

```
## S3 method for class 'xtdhcoint'
print(x, ...)
```

Arguments

x	An object of class "xtdhcoint".
...	Additional arguments (currently ignored).

Value

Invisibly returns the input object.

summary.xtdhcoint *Summary Method for xtdhcoint Objects*

Description

Provides a detailed summary of Durbin-Hausman test results with decision tables at various significance levels.

Usage

```
## S3 method for class 'xtdhcoint'
summary(object, level = 0.95, ...)
```

Arguments

object	An object of class "xtdhcoint".
level	Numeric. Confidence level for hypothesis testing. Default is 0.95.
...	Additional arguments (currently ignored).

Value

Invisibly returns a list with summary statistics.

xtdhcoint *Durbin-Hausman Panel Cointegration Tests*

Description

Performs the Durbin-Hausman panel cointegration tests of Westerlund (2008). The tests are robust to cross-sectional dependence through common factor extraction using principal components.

Usage

```
xtdhcoint(
  formula,
  data,
  id,
  time,
  kmax = 5,
  criterion = c("ic", "pc", "aic", "bic"),
  penalty = NULL,
  bandwidth = -1,
  predet = 0,
  nfactors = NULL
)
```

Arguments

`formula`	A formula specifying the cointegrating relationship (e.g., `y ~ x1 + x2`). The intercept is never estimated: the unit-specific constant of the levels model vanishes when the data are first-differenced (Westerlund, 2008, eq. 12).
`data`	A data frame containing panel data. Missing values in the response, the regressors, `id` or `time` are not allowed; the function stops and reports their count and row positions.
`id`	Character string naming the cross-sectional unit identifier.
`time`	Character string naming the time period identifier. Units are aligned on this variable (see Details, "Unbalanced panels").
`kmax`	Integer. Maximum number of common factors to consider in the search $k = 0, 1, \ldots, kmax$. Default is 5. Values of at least $\min(N, T-1)$ are reduced to $\min(N, T-1) - 1$ with a message.
`criterion`	Character string specifying the information criterion for factor number selection. One of `"ic"` (default, the IC criteria of Bai and Ng, 2002, eq. 9), `"pc"` (their PC criteria), `"aic"` (their AIC3) or `"bic"` (their BIC3).
`penalty`	Integer 1, 2 or 3 selecting the penalty function $g(N, T)$ of the `"ic"` and `"pc"` criteria (the default, `NULL`, means 1). Bai and Ng (2002) define a single AIC3 and a single BIC3, so `penalty` must be `NULL` when `criterion` is `"aic"` or `"bic"`; any other combination is an error.
`bandwidth`	Integer. Bartlett kernel bandwidth M for the long-run variance estimator. If -1 (default) the rule of Westerlund (2008, Section 3.4) is used: $M = \lfloor 4(T/100)^{2/9} \rfloor$, the same for every unit, where T is the number of time periods in the analysed window. This is a fixed function of T, not the data-dependent Newey and West (1994) procedure. A value of 0 uses the contemporaneous variance only. Values other than -1 or a non-negative integer smaller than $T - 1$ are an error.
`predet`	Numeric. If non-zero, the slope(s) of the first-difference regression are not estimated but set to the given value(s) (Westerlund, 2008, Remark 4). Either a single number applied to every regressor or a vector of length K. Default is 0 (estimate the slopes by OLS).
`nfactors`	Integer or `NULL`. If supplied, the number of common factors is fixed at this value (0 means no defactoring) and the information criterion is not used. Useful when `N` is small (see Details).

Details

The Durbin-Hausman tests examine the null hypothesis of no cointegration against the alternative of cointegration. The tests are based on comparing OLS and instrumental variable estimators of the autoregressive parameter in the cumulated residuals.

Computation. Following Westerlund (2008, Section 3.2), for each unit the first differences of the response are regressed on the first differences of the regressors over $t = 2, \ldots, T$ without an intercept (eq. 12). Common factors are extracted from the $(T-1) \times N$ matrix of residuals by principal components, the defactored differences are recumulated from zero, $\tilde{e}_{it} = \sum_{j=2}^{t} \Delta\tilde{e}_{ij}$ with $\tilde{e}_{i1} = 0$, and the autoregression (13) is fitted over $t = 2, \ldots, T$. The statistics are therefore invariant to adding a constant to the response or to any regressor.

DHg (Group-mean statistic): Tests the null of no cointegration against the heterogeneous alternative that at least some units are cointegrated. Under the null, the standardized statistic is asymptotically N(0,1).

DHp (Panel statistic): Tests the null of no cointegration against the homogeneous alternative that all units are cointegrated with a common autoregressive parameter. Under the null, the standardized statistic is asymptotically N(0,1).

Both statistics are computed using right-tail critical values. Large positive values indicate evidence of cointegration.

Number of factors. The number of common factors is selected by minimising the chosen criterion over $k = 0, 1, \ldots, kmax$ (Westerlund, 2008, Remark 5; Bai and Ng, 2002, Theorem 2). $k = 0$ means that no factor is removed. For the `"pc"`, `"aic"` and `"bic"` criteria the scale $\hat{\sigma}^2$ is $V(kmax)$, as allowed by Bai and Ng (2002, p. 201).

Number of regressors. The limiting distributions in Westerlund (2008, Theorem 1 and Remark 3) do not depend on the regressors, so the simulated moments of Remark 2 (5.5464, 36.7673, 0.5005, 0.3348) apply for any number of regressors K; there is a single set of critical values. What the theory does require is that the regressors are not cointegrated among themselves (Assumption 1(d)). With several regressors the user should check this condition; the function only reports K. With `predet` and $K > 1$, a single value is applied to every regressor; supply a vector of length K for different slopes.

Unbalanced panels. The test is defined for a balanced panel: the common factors F_t are extracted from a $(T-1) \times N$ matrix whose rows are calendar periods shared by all units (Westerlund, 2008, Section 3.1 and p. 201), and the criterion of Remark 5 sums over $t = 2, \ldots, T$ for every unit. When units are observed over different periods the function keeps the common time window, the set of values of `time` at which every unit is observed, drops the observations outside it with a message, and requires that window to consist of consecutive periods of the data (no unit may have a gap inside it). If the common window has gaps, or fewer than 10 periods, the function stops with an error. Duplicate values of `time` within a unit are an error.

Small N and N larger than T. The limiting distributions of Theorem 1 require both N and T to grow, with N/T going to zero. In a Monte Carlo study of this version (Westerlund's Section 3.4 design, nominal 5 percent level, one factor fixed with `nfactors = 1`, 1000 replications for each of two seed sets) the null rejection rate was 38 to 41 percent for DHp with $N = 2$ ($T = 50$ and 100), 23 to 24 percent for DHp with $N = 5$ ($T = 50$ and 100), and 57 to 60 percent for DHg with $N = 25$ and $T = 20$. The function therefore warns when $N < 5$ or $N > T$. The Bai and Ng (2002) criteria are also asymptotic in both N and T: "for very small N or T, the proposed criteria are inadequate" (p. 203). In their Table I ($r = 1$, $k_{max} = 8$) with $N = 10$, every criterion except BIC3 averages the upper bound 8, whereas BIC3 averages 7.18 ($T = 50$) and 5.88 ($T = 100$). The function warns when $N < 10$ and the number of factors is selected automatically; fix the number of factors with `nfactors` or report results for several values of `kmax`.

Value

An object of class `"xtdhcoint"` containing:

dhg Raw DHg (group-mean) statistic.

dhp Raw DHp (panel) statistic.

dhg_z Standardized DHg z-statistic.

dhp_z Standardized DHp z-statistic.

dhg_pval P-value for DHg (right-tail).

dhp_pval P-value for DHp (right-tail).

nf Number of common factors used (estimated, or fixed by `nfactors`). 0 means that no factor was removed.

nf_at_kmax Logical. `TRUE` when the information criterion selected the upper bound `kmax`.

nfactors The value of the `nfactors` argument.

N Number of cross-sectional units.

TT Number of time periods in the analysed window.

time_window The first and last value of `time` in the analysed window.

n_dropped Number of observations outside the common time window that were dropped.

K Number of regressors.

bandwidth Bandwidth used.

kmax Upper bound of the factor search actually used.

criterion Information criterion used.

penalty Penalty function used (`NA` for `"aic"` and `"bic"`).

call The matched call.

References

Westerlund, J. (2008). Panel cointegration tests of the Fisher effect. *Journal of Applied Econometrics*, 23(2), 193–233. doi:10.1002/jae.967

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

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4), 631–653. doi:10.2307/2297912

Examples

```
# Load example data
data(fisher_panel)


# Basic test with default settings
result <- xtdhcoint(inflation ~ interest, data = fisher_panel,
                    id = "country", time = "year")
print(result)
summary(result)
```

Index

∗ **datasets**
    fisher_panel, 2

coef.xtdhcoint, 2

fisher_panel, 2

print.xtdhcoint, 3

summary.xtdhcoint, 4

xtdhcoint, 4