Package: hatemicoint (via r-universe)

November 1, 2026

Type Package

Title Hatemi-J Cointegration Test with Two Unknown Regime Shifts

Version 1.1.0

Description Implements the Hatemi-J (2008) cointegration test which allows for two unknown structural breaks (regime shifts) in the cointegrating relationship. The test provides three test statistics: ADF* (Augmented Dickey-Fuller), Zt* (Phillips-Perron Z_t), and Za* (Phillips-Perron Z_alpha), along with endogenously determined break dates. Critical values are based on simulations from Hatemi-J (2008) <doi:10.1007/s00181-007-0175-9>. The long-run variance in the Phillips statistics is estimated by default with a prewhitened quadratic spectral kernel and the automatic bandwidth of Andrews (1991) <doi:10.2307/2938229>, following Andrews and Monahan (1992) <doi:10.2307/2951574>.

License GPL-3

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

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

Date 2026-09-30

Encoding UTF-8

Depends R (>= 3.5.0)

Imports stats

Suggests testthat (>= 3.0.0)

RoxygenNote 7.3.3

Config/testthat/edition 3

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

Date/Publication 2026-10-02 22:47:47 UTC

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

RemoteRef HEAD

RemoteSha 38179abcd0f34a35d73ea25be5edba664089db8e

Contents

hatemicoint-package 2
hatemicoint 3
print.hatemicoint 7
summary.hatemicoint 7

Index **8**

`hatemicoint-package` *hatemicoint: Hatemi-J Cointegration Test with Two Unknown Regime Shifts*

Description

Implements the Hatemi-J (2008) cointegration test which allows for two unknown structural breaks (regime shifts) in the cointegrating relationship. The test provides three test statistics: ADF* (Augmented Dickey-Fuller), Zt* (Phillips-Perron Z_t), and Za* (Phillips-Perron Z_alpha), along with endogenously determined break dates.

Details

The main function in this package is `hatemicoint`, which performs the cointegration test with two structural breaks.

The test is particularly useful when:

- Standard cointegration tests fail to reject the null of no cointegration
- There is reason to believe the relationship has changed over time
- Two structural breaks are suspected in the data

Test Statistics

ADF* Augmented Dickey-Fuller test on residuals with optimal lag selection

Zt* Phillips-Perron Z_t test with kernel-based long-run variance

Za* Phillips-Perron Z_alpha test with kernel-based long-run variance

Critical Values

Critical values depend on the number of regressors (k = 1, 2, 3, or 4) and are taken from Table 1 of Hatemi-J (2008). The null hypothesis of no cointegration is rejected when the test statistic is smaller (more negative) than the critical value.

Author(s)

Maintainer: Muhammad Alkhalaf <muhammedalkhalaf@gmail.com> (ORCID) [copyright holder]

Authors:

- Muhammad Alkhalaf <muhammedalkhalaf@gmail.com> (ORCID) [copyright holder]

References

Hatemi-J, A. (2008). Tests for cointegration with two unknown regime shifts with an application to financial market integration. *Empirical Economics*, 35, 497-505. doi:10.1007/s0018100701759

Gregory, A.W. and Hansen, B.E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics*, 70(1), 99-126. doi:10.1016/03044076(69)416857

See Also

Useful links:

- https://github.com/muhammedalkhalaf/hatemicoint
- Report bugs at https://github.com/muhammedalkhalaf/hatemicoint/issues

---

`hatemicoint` *Hatemi-J Cointegration Test with Two Unknown Regime Shifts*

Description

Performs the Hatemi-J (2008) cointegration test which allows for two unknown structural breaks (regime shifts) in the cointegrating relationship. The test searches over all admissible break date pairs and returns the minimum test statistics along with the endogenously determined break dates.

Usage

```
hatemicoint(
  y,
  x,
  maxlags = NULL,
  lag_selection = c("tstat", "aic", "sic", "fixed"),
  kernel = c("qs", "bartlett", "iid"),
  bwl = NULL,
  prewhite = TRUE,
  trimming = 0.15
)
```

Arguments

`y`	Numeric vector. The dependent variable (must be I(1)).
`x`	Numeric matrix or vector. The independent variable(s) (must be I(1)). Maximum of 4 regressors allowed (k <= 4).
`maxlags`	Integer. Maximum number of lags for the ADF regression (the lag itself when `lag_selection = "fixed"`). Default `NULL` means `floor(4 * (n/100)^(1/4))`.
`lag_selection`	Character. Lag selection rule: `"tstat"` (default), `"aic"`, `"sic"`, or `"fixed"`.
`kernel`	Character. Kernel for the long-run variance in the Phillips statistics: `"qs"` (quadratic spectral, default), `"bartlett"`, or `"iid"` (no correction).

`bwl`	Numeric. Bandwidth for the kernel. If `NULL` (default), the Andrews (1991) automatic bandwidth is used for `"qs"` and `round(4 * (n/100)^(2/9))` for `"bartlett"`. Ignored for `"iid"`.
`prewhite`	Logical. AR(1) prewhitening of the kernel estimator (Andrews and Monahan 1992). Default `TRUE`. Ignored for `"iid"`.
`trimming`	Numeric. Trimming parameter for the break point search. Must be between 0 and 0.5 (exclusive). Default is 0.15.

Details

The Hatemi-J (2008) test extends the Gregory and Hansen (1996) cointegration test by allowing for two structural breaks instead of one. The test is based on the residuals from the cointegrating regression with regime shift dummies:

$$y_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \beta_0' x_t + \beta_1' D_{1t} x_t + \beta_2' D_{2t} x_t + u_t$$

where $D_{1t} = 1$ if $t > tb_1$ and $D_{2t} = 1$ if $t > tb_2$.

Three test statistics are computed for every break pair and the smallest value over the grid is reported (equations 7 to 9 of the paper):

- **ADF***: t-ratio of δ in $\Delta\hat{u}_t = \delta\hat{u}_{t-1} + \sum_{i=1}^{p} \phi_i \Delta\hat{u}_{t-i} + e_t$ (no intercept).
- **Zt***: Phillips Z_t statistic, equation (6) (see below for the square root).
- **Za***: Phillips Z_α statistic, equation (5).

Phillips statistics (equations 3 to 6). With $\hat{\rho}$ the OLS coefficient of $\hat{u}_{t-1}$ on $\hat{u}_t$ (no intercept), $v_t = \hat{u}_t - \hat{\rho}\hat{u}_{t-1}$, autocovariances $\hat{\gamma}(j) = n^{-1}\sum_t v_t v_{t-j}$ (equation 4), and $S = \sum_{t=1}^{n-1} \hat{u}_t^2$, the long-run variance is $\hat{\sigma}^2 = \hat{\gamma}(0) + 2\sum_{j\geq 1} w(j/B)\hat{\gamma}(j)$, the one-sided correction is $\hat{\lambda} = (\hat{\sigma}^2 - \hat{\gamma}(0))/2$, $\hat{\rho}^* = (\sum_{t=1}^{n-1} \hat{u}_t\hat{u}_{t+1} - n\hat{\lambda})/S$, $Z_\alpha = n(\hat{\rho}^* - 1)$ and $Z_t = (\hat{\rho}^* - 1)/\sqrt{\hat{\sigma}^2/S}$. The factor n in front of $\hat{\lambda}$ compensates the $1/n$ in equation (4), as in Phillips (1987); equation (3) of the paper prints the difference without it. The Z_t formula is the one of Gregory and Hansen (1996, p. 105) and Phillips (1987); equation (6) of the paper as printed omits the square root in the denominator (that form would grow with n). All sums use the normalisation of the paper (n, not $n-1$).

Long-run variance. The paper (footnote 3) estimates $\hat{\sigma}^2$ with a prewhitened quadratic spectral (QS) kernel with first-order autoregressive prewhitening and the automatic bandwidth of Andrews (1991), following Andrews and Monahan (1992). This is the default since version 1.1.0 (`kernel = "qs"`, `prewhite = TRUE`, `bwl = NULL`) and is implemented as follows for every break pair: an AR(1) without intercept is fitted to v_t, its coefficient a_1 is capped at 0.97 in absolute value, the AR(1) residuals e_t are used to compute the Andrews bandwidth $B = 1.3221(\hat{\alpha}(2)T)^{1/5}$ with $\hat{\alpha}(2) = 4\hat{\phi}^2/(1-\hat{\phi})^4$, where $\hat{\phi}$ is the AR(1) coefficient of e_t, also capped at 0.97 in absolute value (a package choice, not part of Andrews 1991), and T is the length of e_t; if the bandwidth is zero ($\hat{\phi} = 0$) the kernel estimate reduces to $\hat{\gamma}(0)$ of e_t. The QS kernel $k(x) = 3/z^2(\sin z/z - \cos z)$, $z = 6\pi x/5$, is applied to all lags $j = 1, \ldots, T-1$ of e_t (the QS kernel is not truncated), and the result is recoloured by dividing by $(1-a_1)^2$. With `kernel = "bartlett"` the weights are $1 - j/B$ for $j < B$ and zero beyond, that is the kernel $k(x) = 1 - |x|$ evaluated at $x = j/B$ as in the paper's $w(j/B)$ notation; versions before 1.1.0 used the Newey-West form $1 - j/(B+1)$. This is the package's choice, not a correction from the paper. With integer B only lags $1, \ldots, B-1$ enter, so `bwl = 1` gives no correction. If `bwl` is `NULL` the Bartlett bandwidth is `round(4 (n/100)^(2/9))`

(the rule used by earlier versions). With `kernel = "iid"` no serial-correlation correction is applied and Z_t equals the Dickey-Fuller t-ratio up to the residual-variance divisor. Versions before 1.1.0 used `kernel = "iid"` by default.

ADF lag selection. The paper does not state a lag rule. The package compares the candidate lags $0, \ldots, maxlags$ on a common sample (the first `maxlags` usable observations are dropped for every candidate) and computes the final statistic for the chosen lag on the largest sample available for that lag. `"tstat"` is the general-to-specific rule (largest lag whose last coefficient has $|t| > 1.645$), `"aic"` and `"sic"` minimise $\log(RSS/T) + c(p+1)/T$ with $c = 2$ and $c = \log T$, and `"fixed"` uses `maxlags` lags. The lag is re-selected for every break pair, as in Gregory and Hansen (1996) style infimum tests. The default `maxlags = floor(4 (n/100)^(1/4))` is the Schwert rule and is the package's choice, not the paper's.

Break grid. With trimming τ (default 0.15), tb_1 runs from $\lceil \tau n \rceil$ to $\lfloor (1-2\tau)n \rfloor$ and tb_2 from $tb_1 + \lceil \tau n \rceil$ to $\lfloor (1-\tau)n \rfloor$, so that $tb_1 \geq \tau n$, $tb_2 \leq (1-\tau)n$ and $tb_2 - tb_1 \geq \tau n$ for every n. Break pairs for which a statistic cannot be computed (singular regressor matrix or non-positive long-run variance) are skipped with a warning and never replaced by a substitute value.

The null hypothesis is no cointegration. Rejection occurs when the test statistic is smaller (more negative) than the critical value. The critical values of Table 1 are asymptotic (response-surface intercepts). In a Monte Carlo with the data generating process of Section 4 of the paper (n = 100, m = 1, 400 replications, package defaults) the empirical 5 percent null quantiles were about -6.6 for ADF*, -6.5 for Zt* and -62 for Za*, against the asymptotic values -6.015, -6.015 and -76.003, so at n = 100 ADF* and Zt* reject more often than the nominal level and Za* much less often. With the asymptotic Table 1 critical values the package therefore does not reproduce Table 2 of the paper at n = 100: a 200 replication run with the defaults gave rejection frequencies of 0.175 (ADF*), 0.140 (Zt*) and 0.000 (Za*) under the null against 0.032, 0.069 and 0.069 in Table 2, and the ADF* size depends strongly on the lag rule (0.005 with a fixed lag of 4, 0.105 with lag 0, 0.175 with the default "tstat" rule). The package does not adjust for this.

Value

An object of class `"hatemicoint"` containing:

adf_min Minimum ADF* test statistic

tb1_adf First break location (observation number) for ADF*

tb2_adf Second break location (observation number) for ADF*

lag_adf ADF lag used at the ADF* break pair

zt_min Minimum Zt* test statistic

tb1_zt First break location for Zt*

tb2_zt Second break location for Zt*

bwl_zt Bandwidth used at the Zt* break pair (`NA` for "iid")

za_min Minimum Za* test statistic

tb1_za First break location for Za*

tb2_za Second break location for Za*

bwl_za Bandwidth used at the Za* break pair (`NA` for "iid")

cv_adfzt Critical values for ADF* and Zt* tests (1%, 5%, 10%)

cv_za Critical values for Za* test (1%, 5%, 10%)

nobs Number of observations

k Number of regressors

maxlags Maximum lags used

lag_selection Lag selection method used

kernel Kernel used

bwl Bandwidth argument (`NA` when chosen automatically)

prewhite Whether prewhitening was used

trimming Trimming parameter

n_skipped Number of break pairs skipped because a statistic could not be computed

References

Hatemi-J, A. (2008). Tests for cointegration with two unknown regime shifts with an application to financial market integration. *Empirical Economics*, 35, 497-505. doi:10.1007/s0018100701759

Gregory, A.W. and Hansen, B.E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics*, 70(1), 99-126. doi:10.1016/03044076(69)416857

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817-858. doi:10.2307/2938229

Andrews, D.W.K. and Monahan, J.C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4), 953-966. doi:10.2307/2951574

Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica*, 55(2), 277-301. doi:10.2307/1913237

Examples

```
# Generate example data with structural breaks
set.seed(123)
n <- 200
x <- cumsum(rnorm(n))

# Create cointegrated series with two breaks
y <- numeric(n)
y[1:70] <- 1 + 0.8 * x[1:70] + rnorm(70, sd = 0.5)
y[71:140] <- 3 + 1.2 * x[71:140] + rnorm(70, sd = 0.5)
y[141:200] <- 2 + 0.6 * x[141:200] + rnorm(60, sd = 0.5)

# Run the test
result <- hatemicoint(y, x)
print(result)
summary(result)
```

print.hatemicoint *Print Method for hatemicoint Objects*

Description

Print Method for hatemicoint Objects

Usage

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

Arguments

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

Value

Invisibly returns the input object.

summary.hatemicoint *Summary Method for hatemicoint Objects*

Description

Summary Method for hatemicoint Objects

Usage

```
## S3 method for class 'hatemicoint'
summary(object, ...)
```

Arguments

object	An object of class "hatemicoint".
...	Additional arguments (ignored).

Value

Invisibly returns a summary list with inference at various significance levels.

Index

∗ **package**
hatemicoint-package, 2

hatemicoint, 2, 3
hatemicoint-package, 2

print.hatemicoint, 7

summary.hatemicoint, 7