Package 'paneltests'

November 7, 2026

Type Package

Title Panel Data Pre-Testing and Diagnostic Suite

Version 1.0.9

Date 2026-11-04

Description Pre-testing and diagnostic tools for panel data analysis. Researchers should run these tests before any panel regression to verify modelling assumptions. The package implements: (1) the Hsiao (2014, <ISBN:978-1-107-65763-2>) homogeneity F-tests (F1/F2/F3), Swamy (1970) <doi:10.2307/1913012> parameter heterogeneity test, and Pesaran (2004) <doi:10.2139/ssrn.572504> cross-sectional dependence test via xtpretest(); (2) missing-data detection, mechanism testing, and imputation for unbalanced panels via xtmispanel(); (3) quantile-regression cross-sectional dependence tests (T_tau and T-tilde_tau statistics) of Demetrescu, Hosseinkouchack and Rodrigues (2023) via xtcsdq(); and (4) the panel quantile-regression slope homogeneity S-hat and D-hat statistics of Galvao, Juhl, Montes-Rojas and Olmo (2017) <doi:10.1093/jjfinec/nbx016> via xtqsh(). Together these tests address three fundamental pre-testing questions: (i) are slopes homogeneous? (ii) is there cross-sectional dependence? and (iii) is the panel balanced and is missingness ignorable?

License GPL-3

Encoding UTF-8

Depends R (>= 3.5.0)

Imports stats, utils

Suggests testthat (>= 3.0.0), plm, zoo, quantreg

Config/testthat/edition 3

NeedsCompilation no

RoxygenNote 7.3.1

Author Muhammad Abdullah Alkhalaf [aut, cre, cph] (ORCID: <https://orcid.org/0009-0002-2677-9246>)

Maintainer Muhammad Abdullah Alkhalaf <muhammedalkhalaf@gmail.com>

Repository CRAN

Date/Publication 2026-11-07 00:30:02 UTC

Contents

print.xtcsdq 2
print.xtqsh 3
qsh_sample 3
summary.xtcsdq 4
summary.xtqsh 4
xtcsdq 5
xtkao 7
xtmispanel 9
xtpedroni 11
xtpretest 13
xtqsh 15

Index **17**

`print.xtcsdq` *Print method for xtcsdq objects*

Description

Print method for xtcsdq objects

Usage

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

Arguments

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

Value

Invisibly returns `x`.

print.xtqsh *Print Method for xtqsh Objects*

Description

Prints a formatted summary of an "xtqsh" test result.

Usage

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

Arguments

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

Value

Invisibly returns x.

qsh_sample *Sample Panel Dataset for Quantile Slope Homogeneity Testing*

Description

A simulated balanced panel dataset for demonstrating the quantile slope homogeneity test (xtqsh).

Usage

```
data(qsh_sample)
```

Format

A data frame with columns:

id Cross-sectional unit identifier.

time Time period identifier.

y Dependent variable.

x1 First explanatory variable.

x2 Second explanatory variable.

summary.xtcsdq *Summary method for xtcsdq objects*

Description

Summary method for xtcsdq objects

Usage

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

Arguments

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

Value

Invisibly returns object.

summary.xtqsh *Summary Method for xtqsh Objects*

Description

Prints a summary of an "xtqsh" test result.

Usage

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

Arguments

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

Value

Invisibly returns object.

`xtcsdq` *Tests of No Cross-Sectional Dependence in Panel Quantile Regressions*

Description

Tests the null hypothesis of no cross-sectional error dependence (CSD) in panel quantile regressions. Implements the T_tau and T-tilde_tau statistics of Demetrescu, Hosseinkouchack and Rodrigues (2023).

Usage

```
xtcsdq(
  formula = NULL,
  data = NULL,
  index = NULL,
  quantiles,
  mode = c("pooled", "individual", "residuals"),
  residuals = NULL,
  bandwidth = NULL,
  correction = TRUE
)
```

Arguments

`formula`	A formula of the form `y ~ x1 + x2 + ...`. Required for `mode = "pooled"` (default) and `mode = "individual"`. Not used when `residuals` is provided.
`data`	A data frame containing the panel data in long format. Required unless `residuals` is provided.
`index`	A character vector of length 2: `c("id_var", "time_var")`. Required unless `residuals` is provided.
`quantiles`	A numeric vector of quantile levels, each strictly between 0 and 1.
`mode`	Estimation mode: `"pooled"` (default, pooled FE-QR), `"individual"` (per-unit QR), or `"residuals"` (provide pre-computed residuals via the `residuals` argument).
`residuals`	A list (or named list) of numeric vectors or a matrix with one column per quantile, containing pre-computed QR residuals. Only used when `mode = "residuals"`.
`bandwidth`	Numeric. KDE bandwidth for sparsity estimation. If `NULL` (default), uses $0.35(NT)^{-0.2}$.
`correction`	Logical. If `TRUE` (default), reports the bias-corrected T-tilde statistic in addition to T_tau.

Details

The T_tau statistic (Equation 3 in Demetrescu et al., 2023) tests for CSD by examining pairwise correlations of demeaned QR residuals across units. Under the null of no CSD, T_tau is asymptotically standard normal.

The bias-corrected version T-tilde_tau (Equation 5) subtracts two correction terms that account for the estimation uncertainty in the QR slope and the sparsity at the quantile. Reject H0 for large positive values.

The portmanteau statistic $M_K = K^{-1} \sum_{q=1}^{K} T_\tau^{(q)}$ aggregates across K quantile levels.

The KDE bandwidth defaults to $h = 0.35(NT)^{-0.2}$ as in the original paper.

Value

An object of class `"xtcsdq"` with components:

T_tau Numeric vector of T_tau statistics (one per quantile).

Ttilde_tau Numeric vector of bias-corrected T-tilde_tau statistics.

pval_T p-values for T_tau.

pval_Ttilde p-values for T-tilde_tau.

fhat KDE density estimates at zero (one per quantile).

M_K Portmanteau statistic (average of T_tau over quantiles).

Mtilde_K Bias-corrected portmanteau statistic.

pval_M p-value for M_K.

pval_Mc p-value for Mtilde_K.

quantiles Quantile levels used.

N Number of cross-sectional units.

TT Number of time periods.

bandwidth KDE bandwidth used.

References

Demetrescu, M., Hosseinkouchack, M. and Rodrigues, P.M.M. (2023). Tests of no cross-sectional error dependence in panel quantile regressions. *Ruhr Economic Papers*, No. 1041. doi:10.4419/96973210

Examples

```
set.seed(42)
n <- 8; tt <- 20
dat <- data.frame(
  id   = rep(1:n, each = tt),
  time = rep(1:tt, times = n),
  y    = rnorm(n * tt),
  x1   = rnorm(n * tt)
)
res <- xtcsdq(y ~ x1, data = dat, index = c("id", "time"),
```

```
              quantiles = c(0.25, 0.5, 0.75))
print(res)
summary(res)
```

---

xtkao *Kao (1999) Panel Cointegration Tests*

Description

Residual-based panel cointegration tests of Kao (1999) with a common cointegrating vector and panel fixed effects: the modified DF t, DF t and ADF t (which allow endogenous regressors through the long-run variance adjustment) and the unadjusted modified DF t and DF t (strict exogeneity, no serial correlation). The default `convention = "kao"` follows Kao (1999) and reproduces the finite-sample sizes of Kao (1999, Table 4); `convention = "stata"` follows the Stata command `xtcointtest kao` (see Details). The p-values are left-tail normal probabilities, the direction of rejection stated by Kao (1999).

Usage

```
xtkao(
  data,
  formula,
  index,
  lags = 1L,
  lag_select = c("none", "aic", "bic"),
  kernel_lags = NULL,
  convention = c("kao", "stata")
)
```

Arguments

`data`	A balanced panel `data.frame` in long format.
`formula`	`y ~ x1 + x2 + ...`; terms are evaluated with `model.frame` and `model.matrix`, so `log(x)` or `I(x^2)` are allowed.
`index`	Character vector `c("panel_id", "time_id")`.
`lags`	Integer. Number of lagged differences in the pooled ADF regression (default 1); with `lag_select` it is the maximum and the lag is chosen by AIC or BIC over 1, ..., `lags` on the common sample of the largest regression.
`lag_select`	`"none"` (default), `"aic"` or `"bic"`.
`kernel_lags`	Integer bandwidth of the Bartlett kernel for the long-run covariance matrices, applied to every panel, or `NULL` (default) for `floor(4 (T/100)^(2/9))` with $T = T_0 - 1$. Stata selects a bandwidth per panel with the Newey-West automatic rule and reports the average; that rule is not reproduced here.
`convention`	`"kao"` (default) or `"stata"`: the sample and innovation conventions described in Details.

Details

With T_0 periods per panel, the two conventions are:

`convention = "kao"`**.** The fixed-effects regression of y on x (Kao 1999, eq. 4) uses all T_0 periods. The innovations entering the long-run covariance matrices are the first differences $\hat{w}_{it} = (\Delta y_{it}, \Delta x'_{it})'$, not demeaned (Kao 1999, eqs. 10-12). The pooled DF regression $\hat{e}_{it} = \rho \hat{e}_{i,t-1} + \nu_{it}$ (eqs. 7-8) and the pooled ADF regression with p common lagged differences (eq. 9) use the residual variances s_e^2 and s_ν^2 computed as the residual sum of squares divided by $N(T_0 - 1)$, the number of observations of the DF regression, and the modified DF statistics use $\sqrt{N}T_0(\hat{\rho} - 1)$. With these choices the five statistics reproduce the sizes of Kao (1999, Table 4) in a replication of its design (two ADF lags, Bartlett window 5, $T = 10, 25, 50$, $N = 2$ to 50, 5000 replications per cell): mean absolute differences of 0.002 to 0.009 across the five columns (`inst/validation/kao_table4.R`).

Size in short panels. Both conventions use the asymptotic N(0, 1) critical values and over-reject in short panels, as Kao (1999, Table 4) reports. In a simulation with $N = 6$, $T_0 = 17$, one regressor, no cointegration, lags chosen by AIC up to 2 and the default bandwidth (1000 replications), the rejection rates at the nominal 5 percent level were 0.13 to 0.19 with `"kao"` and 0.05 to 0.17 with `"stata"` across the five statistics, and 0.35 to 0.50 and 0.21 to 0.43 with four regressors. Rejections in such samples are weak evidence of cointegration.

`convention = "stata"`**.** As in `xtcointtest kao`, the sample first loses one period to the pooled AR(1) regressions (without constant) of each variable on its own lag; their residuals, demeaned within panel, are the innovations. The fixed-effects regression uses the remaining $T_0 - 1$ periods, the DF regression $T_0 - 2$, and the residual variances and the modified DF statistics use $T = T_0 - 2$. This is the computation of versions 1.0.8 and earlier. The unadjusted statistics reproduce `xtcointtest kao` to at least 1e-6; the modified DF t and DF t do so when the same kernel bandwidth is used for all panels (verified on a six-country panel, four specifications, bandwidths 0, 1 and 2). Stata's ADF t differs by up to 0.19 (standardized units) on the reference data; its internal convention could not be identified from the documentation. The AR(1) regressions of this convention have no constant, so its statistics change when a panel-specific constant is added to a variable; the statistics of `convention = "kao"` do not.

Input checks. The function stops when a (panel, time) pair is duplicated, when a numeric or `Date` time index is not equally spaced within a panel, when the response has no within-panel variation, when a regressor is constant within panels or collinear with the others after removing the fixed effects, when `kernel_lags` is not a non-negative integer smaller than the number of innovations per panel, and when `lags` leaves too few observations for the pooled ADF regression or makes it fit exactly.

In both conventions the long-run covariance matrices are computed per panel with a Bartlett kernel and averaged over panels, and the conditional variances $\hat{\sigma}_\nu^2 = \hat{\sigma}_u^2 - \hat{\Sigma}_{u\epsilon}\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Sigma}_{\epsilon u}$ and $\hat{\omega}_\nu^2$ (same with the long-run matrix) are taken from the averaged matrices. The ADF t statistic is $t_{ADF} = (\hat{\rho} - 1)[\sum_i \hat{e}_i' Q_i \hat{e}_i]^{1/2} / s_\nu$ (Kao 1999, p. 9).

All five statistics are standard normal under the null and diverge to minus infinity under the alternative (Kao 1999, Theorem 2, Corollary 2 and Remarks 6 and 8), so the p-values are $\Phi(z)$, z being the standardized statistic. Kao (1999) tabulates finite-sample sizes for one regressor only (Table 4); the function warns when there are more than seven regressors (no finite-sample evidence at all), when $N < 2$ (the standardization is asymptotic in N) or when the panel has fewer than 10 periods.

Value

An object of class "xtkao": a list with stats, p (one-sided left-tail p-values pnorm(stat)), rho, t_rho, t_adf, adf_lags, kernel_lags, sigma2_nu, omega2_nu, N, T (periods of the DF regression), T0 (periods per panel in the data), convention and call.

References

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, 90(1), 1-44. doi:10.1016/S03044076(98)000232

Examples

```
set.seed(2)
N <- 6; TT <- 20
d <- do.call(rbind, lapply(1:N, function(i) {
  x <- cumsum(rnorm(TT)); y <- i + 0.5 * x + rnorm(TT)
  data.frame(id = i, t = 1:TT, y = y, x = x) }))
xtkao(d, y ~ x, index = c("id", "t"), kernel_lags = 1)
```

---

xtmispanel *Missing Data Detection and Imputation for Panel Data*

Description

Detects, diagnoses, and imputes missing values in panel (longitudinal) data sets. The function can produce summary tables (Module 1), test the missingness mechanism (Module 2), impute a target variable (Module 3), and run a cross-method sensitivity analysis (Module 4).

Usage

```
xtmispanel(
  data,
  vars = NULL,
  index,
  detect = TRUE,
  test = FALSE,
  impute = NULL,
  target = NULL,
  new_var = NULL,
  sensitivity = FALSE,
  knn_k = 5L
)
```

Arguments

data	A data.frame in long format.
vars	Character vector of variable names to analyse. If NULL (default), all numeric columns except the index are used.
index	Character vector of length 2: c("panel_id", "time_id").
detect	Logical. Run Module 1 (detection tables, default TRUE).
test	Logical. Run Module 2 (MCAR/MAR mechanism tests, default FALSE).
impute	Character or NULL. If a method name is given, run Module 3 (imputation). Supported methods: "mean", "median", "locf", "nocb", "linear", "spline", "pmm", "hotdeck", "knn", "rf", "em".
target	Character. Name of the variable to impute (required when impute is not NULL).
new_var	Character. Name of the output imputed variable (default "\{target\}_imp").
sensitivity	Logical. Run Module 4 (sensitivity analysis across all imputation methods, default FALSE).
knn_k	Integer. Number of neighbours for KNN imputation (default 5).

Value

A list (invisibly) with components:

detect Summary statistics per variable/panel/period.

test MCAR and MAR test results.

imputed The data frame augmented with the imputed column (when imputation is requested).

impute_stats Summary comparing original vs imputed.

sensitivity Sensitivity analysis results.

References

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198-1202. doi:10.1080/01621459.1988.10478722

Examples

```
set.seed(1)
df <- data.frame(
  id   = rep(1:4, each = 8),
  time = rep(1:8, times = 4),
  y    = c(rnorm(32))
)
# introduce some NAs
df$y[c(3, 11, 20)] <- NA
res <- xtmispanel(df, vars = "y", index = c("id", "time"), detect = TRUE)
```

`xtpedroni` *Pedroni (1999, 2004) Panel Cointegration Tests*

Description

Residual-based panel cointegration tests of Pedroni (1999, 2004) with panel-specific cointegrating vectors: four "panel" (same-AR) statistics (modified variance ratio, modified Phillips-Perron t, Phillips-Perron t, ADF t) and three "group" (panel-specific-AR) statistics (modified Phillips-Perron t, Phillips-Perron t, ADF t). The statistics are those of Pedroni (1999, Table 1); the sample conventions are those of the Stata command `xtcointtest pedroni`, which the statistics reproduce numerically (see Details). The p-values follow the tails stated by Pedroni (1999, p. 668).

Usage

```
xtpedroni(
  data,
  formula,
  index,
  trend = FALSE,
  lags = 1L,
  lag_select = c("none", "aic", "bic"),
  kernel_lags = NULL
)
```

Arguments

`data`	A balanced panel `data.frame` in long format.
`formula`	`y ~ x1 + x2 + ...`; the cointegrating regression is estimated panel by panel with a constant (and a trend if `trend = TRUE`). Terms are evaluated with `model.frame` and `model.matrix`, so transformations such as `log(x)` or `I(x^2)` are allowed; each column of the model matrix counts as one regressor.
`index`	Character vector `c("panel_id", "time_id")`.
`trend`	Logical. Include panel-specific linear trends (default `FALSE`).
`lags`	Integer. Number of lagged differences in the ADF regressions (default 1). With `lag_select = "aic"` or `"bic"` it is the maximum, and the common lag (shared by all panels) is chosen on the pooled ADF regression over 1, ..., `lags`, as Stata does.
`lag_select`	`"none"` (default), `"aic"` or `"bic"`.
`kernel_lags`	Integer bandwidth of the Bartlett kernel used for the long-run variances, or `NULL` (default) for the rule `floor(4 (T/100)^(2/9))` where T is the number of periods after differencing. Stata reports the bandwidth it used ("Lags:" in the header); pass that value to reproduce Stata output exactly.

Details

Let $\hat{e}_{it}$ be the residuals of the panel-specific level regression on all T_0 periods. All sums run over the $T = T_0 - 1$ periods for which first differences exist. For each panel the DF regression $\Delta\hat{e}_{it} = (\rho_i - 1)\hat{e}_{i,t-1} + u_{it}$ (no constant) is estimated; its residuals are demeaned and give $\hat{s}_i^2 = \sum u_{it}^2/T$, the Bartlett long-run variance $\hat{\sigma}_i^2$ of u_{it} (autocovariances divided by T) and $\hat{\lambda}_i = (\hat{\sigma}_i^2 - \hat{s}_i^2)/2$. $\hat{L}_{11i}^2$ is the conditional long-run variance of Δy given Δx, computed from the Bartlett long-run covariance matrix of the demeaned first differences. The ADF regression (panel-specific ρ_i, no constant) on the $T - p$ periods with `lags` $= p$ lagged differences available gives, from its demeaned residuals, $\hat{s}_i^{*2}$; the ADF statistics use the sums $\sum \hat{e}_{i,t-1}\Delta\hat{e}_{it}$ and $\sum \hat{e}_{i,t-1}^2$ over that reduced sample.

Standardization uses $(\chi - \mu\sqrt{N})/\sqrt{\nu}$. For $k = 1$ regressor the moments are those of Pedroni (2004, Corollary 1): case 2 (estimated fixed effects) when `trend = FALSE` and case 3 (estimated fixed effects and trends) when `trend = TRUE`. For $k = 2, \ldots, 7$ they are those of Pedroni (1999, Table 2) with $m = k$, rows "heterogeneous intercepts included" and "heterogeneous deterministic trends and intercepts included". More than seven regressors are refused because no moments are tabulated. The moments are large-T approximations; the function warns when $N < 2$ (the standardization is asymptotic in N) or when the panel has fewer than 10 periods.

Direction of rejection and p-values. The raw statistics are Pedroni's: under the alternative the panel variance ratio diverges to plus infinity and the other six statistics diverge to minus infinity (Pedroni 1999, p. 668). The p-values are therefore $1 - \Phi(z)$ for the panel variance ratio and $\Phi(z)$ for the other six. Versions up to 1.0.7 returned $1 - \Phi(|z|)$ for all seven statistics, which rejects a true null at twice the nominal rate and gives a small p-value to a statistic of the wrong sign.

Agreement with Stata. With the same kernel bandwidth and ADF lag, all seven standardized statistics reproduce `xtcointtest pedroni` to at least 1e-6 (verified on a six-country panel, four specifications with four to six regressors, bandwidths 0, 1 and 2). Two conventions of that command are kept because they are needed for this agreement and are stated here as the package's choice: the residuals of the DF and ADF regressions are demeaned before the variances are computed, and the panel ADF t uses $\tilde{s}^{*2} = N^{-1}\sum_i \hat{L}_{11i}^{-2}\hat{s}_i^{*2}$ (weighted), whereas Pedroni (1999, Table 1) writes the unweighted average. The p-values are not part of the comparison; they follow the paper as described above. Stata's default bandwidth is chosen per panel by the Newey and West (1994) algorithm and the header reports the average; that algorithm is not reproduced here, so the default rule of this function (or an integer `kernel_lags`) will in general differ from a Stata run with `kernel(bartlett nwest)` unless all panels received the same bandwidth (as when the reported average is an integer such as 0.00).

Input checks. The function stops when a (panel, time) pair is duplicated, when a numeric or `Date` time index is not equally spaced within a panel, when the response is constant in a panel, when a regressor is constant or collinear with the others and the deterministic terms in a panel, when `kernel_lags` is not a non-negative integer smaller than the number of differenced observations per panel, and when `lags` leaves no more observations than coefficients in the panel-specific ADF regressions or makes one of them fit exactly.

Value

An object of class `"xtpedroni"`: a list with `stats` (standardized statistics), `p` (one-sided p-values in the tail stated by Pedroni (1999, p. 668): `1 - pnorm(z)` for the panel variance ratio, which diverges to plus infinity under the alternative, and `pnorm(z)` for the other six statistics, which diverge to minus infinity), `tail` (the tail used for each statistic), `raw` (the unstandardized statistics), `moments` (means and variances used), `N`, `T`, `k`, `adf_lags`, `kernel_lags` and `call`.

References

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61(S1), 653-670. doi:10.1111/1468-0084.61.s1.14

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory*, 20(3), 597-625. doi:10.1017/S0266466604203073

Examples

```
set.seed(1)
N <- 6; TT <- 20
d <- do.call(rbind, lapply(1:N, function(i) {
  x <- cumsum(rnorm(TT)); y <- 1 + 0.5 * x + rnorm(TT)
  data.frame(id = i, t = 1:TT, y = y, x = x) }))
r <- xtpedroni(d, y ~ x, index = c("id", "t"), kernel_lags = 2)
r$stats
```

`xtpretest` *Comprehensive Panel Data Pre-Testing Suite*

Description

Performs a full battery of panel data pre-tests: Hsiao (2014) homogeneity F-tests, robust (HC1) versions, Swamy (1970) parameter heterogeneity test, cross-sectional dependence (Pesaran 2004), and panel summary statistics.

Usage

```
xtpretest(data, formula, index, tests = "ALL", level = 0.05)
```

Arguments

`data`	A `data.frame` in long format containing all variables.
`formula`	A two-sided formula of the form `y ~ x1 + x2 + ...` specifying the dependent and independent variables. Terms are evaluated with `model.frame` and `model.matrix`, so `log(x)`, `I(x^2)` and interactions are allowed; every column of the model matrix (without intercept) counts as one regressor in K.
`index`	Character vector of length 2: `c("panel_id", "time_id")`.
`tests`	Character vector. Which modules to run. Possible values: `"summary"`, `"hsiao"`, `"robust"`, `"heterogeneity"`, `"csd"`. Default `"ALL"` runs everything.
`level`	Numeric. Significance level for decisions (default 0.05).

Details

Input checks. Missing values in the response or the regressors are refused with an error that reports their number and row positions; the tests have no defined treatment of missing cells. Fewer than two panels, or any panel with fewer than $K + 2$ observations (so that the individual regression of Hsiao's S_1 has no residual degrees of freedom), are refused with an error naming the panels. Exactly collinear regressor columns are detected from the rank of the QR decomposition of the pooled design and removed with a warning naming them; K is the rank, so the degrees of freedom refer to the regressors that remain. A regressor that is constant within one panel is not collinear in the pooled design; the HSIAO and HETEROGENEITY modules then stop with a message naming the panel (recorded in `failed`) because the individual regressions are not estimable, while the ROBUST module drops the aliased interaction columns and reports them.

Hsiao's F tests and their labels. The module HSIAO reports three F statistics labelled `F1`, `F2` and `F3` in the order overall homogeneity, slope homogeneity and intercept homogeneity. In Hsiao (2014, section 2.2.2) these are, respectively, eq. (2.2.14) (there called F_3), eq. (2.2.15) (there called F_1) and eq. (2.2.16) (there called F_4). The formulas and degrees of freedom are those of the book; only the labels differ, and they are kept for backward compatibility of the returned list.

Swamy's test. The module HETEROGENEITY computes Swamy (1970, eq. 6.2): $H = \sum_i (b_i - \hat{\beta})' X_i' X_i (b_i - \hat{\beta}) / s_{ii}$, where b_i are the slopes of the individual OLS regressions, X_i the within-panel demeaned regressors, $s_{ii} = RSS_i/(T_i - K - 1)$, and $\hat{\beta} = [\sum_i X_i' X_i / s_{ii}]^{-1} \sum_i (X_i' X_i / s_{ii}) b_i$ is the weighted pooled estimator of the paper. Versions up to 1.0.7 centred the statistic on the unweighted fixed-effects estimator, which over-rejects under cross-panel heteroskedasticity. H is referred to a chi-square with $K(N-1)$ degrees of freedom.

Cross-sectional dependence. The module CSD computes Pesaran's CD statistic $\sqrt{2/(N(N-1))} \sum_{i<j} \sqrt{T_{ij}} \hat{\rho}_{ij}$ from the residuals of the fixed-effects (within) regression with common slopes, using the pairwise-complete T_{ij} overlapping periods for each pair, with a two-sided normal p-value. The residuals of N individual OLS regressions would give a different value; the fixed-effects residuals are the package's choice.

ROBUST module and its size. The `"robust"` module is an extension of this package and is not part of Hsiao (2014), who only notes that the F-tests assume i.i.d. errors. It fits the fully interacted regression $y = a + Xb + \sum_{i>1} D_i g_i + \sum_{i>1} (D_i X) h_i + u$, with the first panel as reference, and computes HC1 Wald statistics (divided by the number of restrictions and referred to an F distribution) for slope homogeneity ($h_i = 0$, $(N-1)K$ restrictions) and for intercept homogeneity given heterogeneous slopes ($g_i = 0$, $N - 1$ restrictions). Columns aliased in the interacted regression (for example a regressor that is constant within a panel) are dropped and reported. A failure in any module no longer aborts the other modules.

The HC1 Wald statistic has $(N-1)K$ restrictions and $(K+1)N$ parameters, and it over-rejects a true null at small T. In the simulations of the 1.0.8 validation (i.i.d. normal errors, heterogeneous intercepts, homogeneous slopes, $K = 2$, nominal 5 percent) the rejection rates of the robust slope test were 33.4 percent ($N = 10$, $T = 20$; 8000 replications, Monte Carlo standard error 0.5 percentage points), 64.8 percent ($N = 25$, $T = 20$), 15.4 percent ($N = 10$, $T = 50$) and 11.8 percent ($N = 25$, $T = 100$); the last three come from 500 replications (standard error about 2 percentage points). The classical F test in the same cells rejects at 3.4 to 6.4 percent. The printed output repeats this note. Use the classical F test (with caution under heteroskedasticity, where it reached 8.4 percent) or a bootstrap of the Wald statistic; the estimator itself is unchanged.

Value

A list (invisibly) with components:

`summary` Panel summary statistics.

`hsiao` Hsiao homogeneity F-test results.

`robust` Robust HC1 Wald versions of the Hsiao F2 and F3 tests (see Details), with the names of any collinear columns dropped.

`swamy` Swamy heterogeneity test results.

`csd` Cross-sectional dependence test results.

`recommendation` Character. Suggested estimator.

`failed` Named list of error messages for modules that failed and were skipped (absent when all modules ran).

`dropped` Character vector of regressor columns removed because they were exactly collinear with the others (absent when none).

References

Hsiao, C. (2014). *Analysis of Panel Data* (3rd ed.). Cambridge University Press. doi:10.1017/CBO9781139839327

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2), 311-323. doi:10.2307/1913012

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. *Cambridge Working Paper in Economics*, No. 0435. doi:10.2139/ssrn.572504

Examples

```
set.seed(10)
n <- 5; t <- 10
df <- data.frame(
  id   = rep(1:n, each = t),
  time = rep(1:t, times = n),
  y    = rnorm(n * t),
  x1   = rnorm(n * t)
)
res <- xtpretest(df, y ~ x1, index = c("id", "time"),
                 tests = c("hsiao", "csd"))
```

`xtqsh` *Quantile Regression Slope Homogeneity Test for Panel Data*

Description

Tests the null hypothesis of slope homogeneity in panel quantile regressions. Implements the S-hat and D-hat statistics of Galvao et al. (2017).

Usage

```
xtqsh(formula, data, index, tau, bw = "hallsheather", marginal = FALSE)
```

Arguments

formula	A formula of the form y ~ x1 + x2 + ...
data	A data frame containing the panel data in long format.
index	Character vector of length 2: c("id_var", "time_var").
tau	Numeric vector of quantile levels, each strictly between 0 and 1.
bw	Bandwidth method: "hallsheather" (default) or "bofinger".
marginal	Logical. If TRUE, compute per-variable marginal tests. Default FALSE.

Value

An object of class "xtqsh" containing test statistics and p-values.

References

Galvao, A.F., Juhl, T., Montes-Rojas, G. and Olmo, J. (2017). Testing Slope Homogeneity in Quantile Regression Panel Data. *Journal of Financial Econometrics*, 16(2), 211-243.

Examples

```
set.seed(42)
n <- 10; tt <- 20
dat <- data.frame(
  id   = rep(1:n, each = tt),
  time = rep(1:tt, times = n),
  y    = rnorm(n * tt),
  x1   = rnorm(n * tt)
)
res <- xtqsh(y ~ x1, data = dat, index = c("id", "time"), tau = 0.5)
print(res)
```

Index

∗ **datasets**
qsh_sample, 3
∗ **htest**
xtqsh, 15
∗ **print**
print.xtqsh, 3
summary.xtqsh, 4

print.xtcsdq, 2
print.xtqsh, 3

qsh_sample, 3

summary.xtcsdq, 4
summary.xtqsh, 4

xtcsdq, 5
xtkao, 7
xtmispanel, 9
xtpedroni, 11
xtpretest, 13
xtqsh, *3*, 15