Package ‘HighFreq’

May 18, 2018

Type Package

Title High Frequency Time Series Management

Version 0.1

Date 2017-10-26

Author Jerzy Pawlowski (algoquant)

Maintainer Jerzy Pawlowski <jp3900@nyu.edu>

Description Functions for chaining and joining time series, scrubbing bad data, managing time zones and alligning time indices, converting TAQ data to OHLC format, aggregating data to lower frequency, estimating volatility, skew, and higher moments.

License GPL (>= 2)

Depends rutils,
quantmod,

Imports TTR,
RcppRoll,
caTools,
lubridate,
Rcpp (>= 0.12.16),
RcppParallel

LinkingTo Rcpp, RcppArmadillo, RcppParallel

SystemRequirements GNU make, C++11

Remotes github::algoquant/rutils,

VignetteBuilder knitr

LazyData true

ByteCompile true

Repository GitHub

URL https://github.com/algoquant/HighFreq

RoxygenNote 6.0.1

R topics documented:

`agg_regate` *Calculate the aggregation (weighted average) of a statistical estimator over a* OHLC *time series.*

Description

Calculate the aggregation (weighted average) of a statistical estimator over a *OHLC* time series.

Usage

```
agg_regate(oh_lc, mo_ment = "run_variance", weight_ed = TRUE, ...)
```

Arguments

oh_lc	An *OHLC* time series of prices and trading volumes, in *xts* format.
mo_ment	A *character* string representing function for estimating the moment.
weight_ed	*Boolean* argument: should estimate be weighted by the trading volume? (default is TRUE)
...	additional parameters to the mo_ment function.

Details

The function agg_regate() calculates a single number representing the volume weighted average of an estimator over the *OHLC* time series of prices. By default the sum is trade volume weighted.

Value

A single *numeric* value equal to the volume weighted average of an estimator over the time series.

Examples

```
# calculate weighted average variance for SPY (single number)
vari_ance <- agg_regate(oh_lc=HighFreq::SPY, mo_ment="run_variance")
# calculate time series of daily skew estimates for SPY
skew_daily <- apply.daily(x=HighFreq::SPY, FUN=agg_regate, mo_ment="run_skew")
```

---

calc_eigen — *Calculate the eigen decomposition of the covariance matrix of returns using* RcppArmadillo.

Description

Calculate the eigen decomposition of the covariance matrix of returns using *RcppArmadillo*.

Usage

```
calc_eigen(mat_rix)
```

Arguments

mat_rix	A numeric *matrix* of returns data.

Details

The function calc_eigen() first calculates the covariance matrix of the mat_rix, and then calculates its eigen decomposition.

Value

A list with two elements: a numeric *vector* of eigenvalues (named "values"), and a numeric *matrix* of eigenvectors (named "vectors").

Examples

```
## Not run:
# Create random matrix
mat_rix <- matrix(rnorm(500), nc=5)
# Calculate eigen decomposition
ei_gen <- HighFreq::calc_eigen(scale(mat_rix, scale=FALSE))
# Calculate PCA
pc_a <- prcomp(mat_rix)
# Compare PCA with eigen decomposition
all.equal(pc_a$sdev^2, drop(ei_gen$values))
all.equal(abs(unname(pc_a$rotation)), abs(ei_gen$vectors))

## End(Not run)
```

---

calc_inv — *Calculate the regularized inverse of the covariance matrix of returns using* RcppArmadillo.

Description

Calculate the regularized inverse of the covariance matrix of returns using *RcppArmadillo*.

Usage

```
calc_inv(mat_rix, max_eigen)
```

Arguments

`mat_rix`	A numeric *matrix* of returns data.
`max_eigen`	An *integer* equal to the regularization intensity (the number of eigenvalues and eigenvectors used for calculating the regularized inverse).

Details

The function `calc_inv()` first calculates the covariance matrix of the `mat_rix`, and then it calculates the regularized inverse from the truncated eigen decomposition. It uses only the largest `max_eigen` eigenvalues and their corresponding eigenvectors.

Value

A numeric *matrix* equal to the regularized inverse.

Examples

```
## Not run:
# Create random matrix
mat_rix <- matrix(rnorm(500), nc=5)
max_eigen <- 3
# Calculate regularized inverse using RcppArmadillo
in_verse <- HighFreq::calc_inv(mat_rix, max_eigen)
# Calculate regularized inverse from eigen decomposition in R
ei_gen <- eigen(cov(mat_rix))
inverse_r <-  ei_gen$vectors[, 1:max_eigen] %*% (t(ei_gen$vectors[, 1:max_eigen]) / ei_gen$values[1:max_eigen]
```

```
# Compare RcppArmadillo with R
all.equal(in_verse, inverse_r)

## End(Not run)
```

calc_lm *Perform multivariate linear regression using RcppArmadillo.*

Description

Perform multivariate linear regression using RcppArmadillo.

Usage

```
calc_lm(res_ponse, de_sign)
```

Arguments

res_ponse	A numeric *vector* of response data.
de_sign	A numeric *matrix* of design (predictor i.e. explanatory) data.

Details

The function `calc_lm()` performs the same calculations as the function `lm()` from package *stats*. It uses *RcppArmadillo* and is about *10* times faster than `lm()`. The code was inspired by this article (but it's not identical to it): http://gallery.rcpp.org/articles/fast-linear-model-with-armadillo/

Value

A named list with three elements: a numeric *matrix* of coefficients (named *"coefficients"*), the *z-score* of the last residual (named *"z_score"*), and a numeric *vector* with the R-squared and F-statistic (named *"stats"*). The numeric *matrix* of coefficients named *"coefficients"* containes the alpha and beta coefficients, and their *t-values* and *p-values*.

Examples

```
## Not run:
# Define design matrix with explanatory variables
len_gth <- 100; n_var <- 5
de_sign <- matrix(rnorm(n_var*len_gth), nc=n_var)
# response equals linear form plus error terms
weight_s <- rnorm(n_var)
res_ponse <- -3 + de_sign %*% weight_s + rnorm(len_gth, sd=0.5)
# perform multivariate regression using lm()
reg_model <- lm(res_ponse ~ de_sign)
sum_mary <- summary(reg_model)
# perform multivariate regression using calc_lm()
reg_model_arma <- calc_lm(res_ponse=res_ponse, de_sign=de_sign)
reg_model_arma$coefficients
# compare the outputs of both functions
all.equal(reg_model_arma$coefficients[, "coeff"], unname(coef(reg_model)))
all.equal(unname(reg_model_arma$coefficients), unname(sum_mary$coefficients))
all.equal(drop(reg_model_arma$residuals), unname(reg_model$residuals))
```

```
all.equal(unname(reg_model_arma$stats), c(sum_mary$r.squared, unname(sum_mary$fstatistic[1])))

## End(Not run)
```

calc_scaled *Scale (standardize) the columns of a* matrix *of data using* RcppArmadillo.

Description

Scale (standardize) the columns of a *matrix* of data using *RcppArmadillo*.

Usage

```
calc_scaled(mat_rix, use_median = FALSE)
```

Arguments

mat_rix	A numeric *matrix* of data.
use_median	A *Boolean* argument: if TRUE then the centrality (central tendency) is calculated as the *median* and the dispersion is calculated as the *median absolute deviation* (*MAD*). If use_median is FALSE then the centrality is calculated as the *mean* and the dispersion is calculated as the *standard deviation*. (The default is FALSE)

Details

The function calc_scaled() scales (standardizes) the columns of the mat_rix argument using *RcppArmadillo*. If the argument use_median is FALSE (the default), then it performs a similar calculation as the standard *R* function scale(), and it calculates the centrality (central tendency) as the *mean* and the dispersion as the *standard deviation*. If the argument use_median is TRUE, then it calculates the centrality as the *median* and the dispersion as the *median absolute deviation* (*MAD*).

The function calc_scaled() uses *RcppArmadillo* and is about *5* times faster than function scale(), for a matrix with *1,000* rows and *20* columns.

Value

A numeric *matrix* with the same dimensions as the input argument mat_rix.

Examples

```
## Not run:
in_nov <- matrix(rnorm(2000), nc=2)
foo <- calc_scaled(mat_rix=in_nov, use_median=FALSE)
bar <- scale(in_nov)
all.equal(as.numeric(foo), as.numeric(bar))
library(microbenchmark)
summary(microbenchmark(
  pure_r=scale(in_nov),
  rcpp=calc_scaled(mat_rix=in_nov, use_median=FALSE),
  times=100))[, c(1, 4, 5)]  # end microbenchmark summary

## End(Not run)
```

calc_variance — *Calculate the variance of an* OHLC *time series, using different range estimators for variance.*

Description

Calculate the variance of an *OHLC* time series, using different range estimators for variance.

Usage

```
calc_variance(oh_lc, calc_method = "yang_zhang", sca_le = TRUE)
```

Arguments

`oh_lc` An *OHLC* time series of prices in *xts* format.

`calc_method` A *character* string representing the method for estimating variance. The methods include:

- "close" close to close,
- "garman_klass" Garman-Klass,
- "garman_klass_yz" Garman-Klass with account for close-to-open price jumps,
- "rogers_satchell" Rogers-Satchell,
- "yang_zhang" Yang-Zhang,

(default is `"yang_zhang"`)

`sca_le` *Boolean* argument: should the returns be divided by the number of seconds in each period? (default is `TRUE`)

Details

The function `calc_variance()` calculates the variance estimate from *OHLC* prices, using several different variance estimation methods based on the range of *OHLC* prices.

The methods `"close"`, `"garman_klass_yz"`, and `"yang_zhang"` do account for close-to-open price jumps, while the methods `"garman_klass"` and `"rogers_satchell"` do not account for close-to-open price jumps.

The default method is `"yang_zhang"`, which theoretically has the lowest standard error among unbiased estimators.

If `sca_le` is `TRUE` (the default), then the variance is divided by the squared differences of the time index (which scales the variance to units of variance per second squared.) This is useful for example, when calculating variance from minutely bar data, because dividing returns by the number of seconds decreases the effect of overnight price jumps.

If `sca_le` is `TRUE` (the default), then the variance is expressed in the scale of the time index of the *OHLC* time series. For example, if the time index is in seconds, then the variance is given in units of variance per second squared. If the time index is in days, then the variance is equal to the variance per day squared.

The function `calc_variance()` performs the same calculations as the function `run_variance()` and then calculates the average of the spot variance estimates.

Value

A single *numeric* value equal to the variance.

Examples

```
# create minutely OHLC time series of random prices
oh_lc <- HighFreq::random_ohlc()
# calculate variance of oh_lc
vari_ance <- HighFreq::calc_variance(oh_lc)
# calculate variance of SPY
vari_ance <- HighFreq::calc_variance(HighFreq::SPY, calc_method="yang_zhang")
# calculate variance of SPY without accounting for overnight jumps
vari_ance <- HighFreq::calc_variance(HighFreq::SPY, calc_method="rogers_satchell")
```

`calc_weights` *Calculate the scaled weights of the portfolio with the maximum Sharpe ratio.*

Description

Calculate the scaled weights of the portfolio with the maximum Sharpe ratio.

Usage

```
calc_weights(re_turns, max_eigen, al_pha)
```

Arguments

re_turns	A numeric *matrix* of excess returns data (the returns in excess of the risk-free rate).
max_eigen	An *integer* equal to the regularization intensity (the number of eigenvalues and eigenvectors used for calculating the regularized inverse).
al_pha	The shrinkage intensity.

Details

The function `calc_weights()` calculates the scaled weights of the portfolio with the maximum Sharpe ratio, using *RcppArmadillo*.

It first calculates the regularized inverse of the covariance matrix of returns using function `HighFreq::calc_inv()`. It then estimates the vector of mean returns and applies shrinkage to it, by shrinking it to its common mean value. The shrinkage intensity `al_pha` determines the amount of shrinkage that is applied, with `al_pha = 0` representing no shrinkage (with the estimator of mean returns equal to the means of the columns of `re_turns`), and `al_pha = 1` representing complete shrinkage (with the estimator of mean returns equal to the single mean of all the columns of `re_turns`)

The function `calc_weights()` calculates the weights by multiplying the inverse of the covariance matrix times the estimator of the mean returns. It finally scales the weights by their sum, so that the sum of the weights is equal to `1`.

Value

A numeric *vector* of the same length as the number of columns of `re_turns`.

Examples

```
## Not run:
# Calculate ETF prices
sym_bols <- colnames(rutils::env_etf$price_s)
sym_bols <- sym_bols[!(sym_bols=="VXX")]
price_s <- rutils::env_etf$price_s[, sym_bols]
# Carry forward non-NA prices
price_s <- zoo::na.locf(price_s)
price_s <- na.omit(price_s)
# Calculate simple ETF returns
re_turns <- rutils::diff_it(price_s)
# Calculate covariance matrix
ei_gen <- eigen(cov(re_turns))
# Calculate regularized inverse of covariance matrix
max_eigen <- 3
eigen_vec <- ei_gen$vectors[, 1:max_eigen]
eigen_val <- ei_gen$values[1:max_eigen]
in_verse <- eigen_vec %*% (t(eigen_vec) / eigen_val)
# Define shrinkage intensity and apply shrinkage to the mean returns
al_pha <- 0.5
col_means <- colMeans(re_turns)
col_means <- ((1-al_pha)*col_means + al_pha*mean(col_means))
# Calculate weights using R
weight_s <- in_verse %*% col_means
weights_r <- drop(weight_s/sum(weight_s))
# Calculate weights using RcppArmadillo
weight_s <- drop(HighFreq::calc_weights(re_turns, max_eigen, al_pha=al_pha))
all.equal(weight_s, weights_r)

## End(Not run)
```

---

hf_data *High frequency data sets*

Description

hf_data.RData is a file containing the datasets:

SPY an xts time series containing 1-minute OHLC bar data for the SPY etf, from 2008-01-02 to 2014-05-19. SPY contains 625,425 rows of data, each row contains a single minute bar.

TLT an xts time series containing 1-minute OHLC bar data for the TLT etf, up to 2014-05-19.

VXX an xts time series containing 1-minute OHLC bar data for the VXX etf, up to 2014-05-19.

Usage

```
data(hf_data)  # not required - data is lazy load
```

Format

Each xts time series contains OHLC data, with each row containing a single minute bar:

Open Open price in the bar

High High price in the bar

Low Low price in the bar

Close Close price in the bar

Volume trading volume in the bar

Source

https://wrds-web.wharton.upenn.edu/wrds/

References

Wharton Research Data Service (WRDS)

Examples

```
# data(hf_data)  # not required - data is lazy load
head(SPY)
chart_Series(x=SPY["2009"])
```

---

random_ohlc — *Calculate a random* OHLC *time series of prices and trading volumes, in* xts *format.*

Description

Calculate a random *OHLC* time series either by simulating random prices following geometric Brownian motion, or by randomly sampling from an input time series.

Usage

```
random_ohlc(oh_lc = NULL, re_duce = TRUE, vol_at = 6.5e-05, dri_ft = 0,
  in_dex = seq(from = as.POSIXct(paste(Sys.Date() - 3, "09:30:00")), to =
  as.POSIXct(paste(Sys.Date() - 1, "16:00:00")), by = "1 sec"), ...)
```

Arguments

`oh_lc`	An *OHLC* time series of prices and trading volumes, in *xts* format (default is *NULL*).
`re_duce`	*Boolean* argument: should `oh_lc` time series be transformed to reduced form? (default is `TRUE`)
`vol_at`	The volatility per period of the `in_dex` time index (default is `6.5e-05` per second, or about `0.01=1.0%` per day).
`dri_ft`	The drift per period of the `in_dex` time index (default is 0.0).
`in_dex`	The time index for the *OHLC* time series.

Details

If the input oh_lc time series is *NULL* (the default), then the function random_ohlc() simulates a minutely *OHLC* time series of random prices following geometric Brownian motion, over the two previous calendar days.

If the input oh_lc time series is not *NULL*, then the rows of oh_lc are randomly sampled, to produce a random time series.

If re_duce is TRUE (the default), then the oh_lc time series is first transformed to reduced form, then randomly sampled, and finally converted to standard form.

Note: randomly sampling from an intraday time series over multiple days will cause the overnight price jumps to be re-arranged into intraday price jumps. This will cause moment estimates to become inflated compared to the original time series.

Value

An *xts* time series with the same dimensions and the same time index as the input oh_lc time series.

Examples

```
# create minutely synthetic OHLC time series of random prices
oh_lc <- HighFreq::random_ohlc()
# create random time series from SPY by randomly sampling it
oh_lc <- HighFreq::random_ohlc(oh_lc=HighFreq::SPY["2012-02-13/2012-02-15"])
```

---

random_taq — *Calculate a random* TAQ *time series of prices and trading volumes, in* xts *format.*

Description

Calculate a *TAQ* time series of random prices following geometric Brownian motion, combined with random trading volumes.

Usage

```
random_taq(vol_at = 6.5e-05, dri_ft = 0, in_dex = seq(from =
  as.POSIXct(paste(Sys.Date() - 3, "09:30:00")), to =
  as.POSIXct(paste(Sys.Date() - 1, "16:00:00")), by = "1 sec"),
  bid_offer = 0.001, ...)
```

Arguments

vol_at	The volatility per period of the in_dex time index (default is 6.5e-05 per second, or about 0.01=1.0% per day).
dri_ft	The drift per period of the in_dex time index (default is 0.0).
in_dex	The time index for the *TAQ* time series.
bid_offer	The bid-offer spread expressed as a fraction of the prices (default is 0.001=10bps).

Details

The function `random_taq()` calculates an *xts* time series with four columns containing random prices following geometric Brownian motion: the bid, ask, and trade prices, combined with random trade volume data. If `in_dex` isn't supplied as an argument, then by default it's equal to the secondly index over the two previous calendar days.

Value

An *xts* time series, with time index equal to the input `in_dex` time index, and with four columns containing the bid, ask, and trade prices, and the trade volume.

Examples

```
# create secondly TAQ time series of random prices
ta_q <- HighFreq::random_taq()
# create random TAQ time series from SPY index
ta_q <- HighFreq::random_taq(in_dex=index(HighFreq::SPY["2012-02-13/2012-02-15"]))
```

---

remove_jumps — *Remove overnight close-to-open price jumps from an* OHLC *time series, by adding adjustment terms to its prices.*

Description

Remove overnight close-to-open price jumps from an *OHLC* time series, by adding adjustment terms to its prices.

Usage

```
remove_jumps(oh_lc)
```

Arguments

`oh_lc`	An *OHLC* time series of prices and trading volumes, in *xts* format.

Details

The function `remove_jumps()` removes the overnight close-to-open price jumps from an *OHLC* time series, by adjusting its prices so that the first *Open* price of the day is equal to the last *Close* price of the previous day.

The function `remove_jumps()` adds adjustment terms to all the *OHLC* prices, so that intra-day returns and volatilities are not affected.

The function `remove_jumps()` identifies overnight periods as those that are greater than 60 seconds. This assumes that intra-day periods between neighboring rows of data are 60 seconds or less.

The time index of the `oh_lc` time series is assumed to be in *POSIXct* format, so that its internal value is equal to the number of seconds that have elapsed since the *epoch*.

Value

An *OHLC* time series with the same dimensions and the same time index as the input `oh_lc` time series.

Examples

```
# remove overnight close-to-open price jumps from SPY data
oh_lc <- remove_jumps(HighFreq::SPY)
```

`roll_apply` *Apply an aggregation function over a rolling look-back interval and the end points of an* OHLC *time series.*

Description

Apply an aggregation function over a rolling look-back interval and the end points of an *OHLC* time series.

Usage

```
roll_apply(x_ts, agg_fun, look_back = 2, end_points = seq_along(x_ts),
  by_columns = FALSE, out_xts = TRUE, ...)
```

Arguments

`x_ts`	An *OHLC* time series of prices and trading volumes, in *xts* format.
`agg_fun`	The name of the aggregation function to be applied over a rolling look-back interval.
`look_back`	The number of end points in the look-back interval used for applying the aggregation function (including the current row).
`end_points`	An integer vector of end points.
`by_columns`	*Boolean* argument: should the function `agg_fun()` be applied column-wise (individually), or should it be applied to all the columns combined? (default is `FALSE`)
`out_xts`	*Boolean* argument: should the output be coerced into an *xts* series? (default is `TRUE`)
`...`	additional parameters to the agg_fun function.

Details

The function `roll_apply()` applies an aggregation function over a rolling look-back interval attached at the end points of an *OHLC* time series.

`HighFreq::roll_apply()` performs similar operations to the functions `rollapply()` and `period.apply()` from package xts, and also the function `apply.rolling()` from package PerformanceAnalytics. (The function `rollapply()` isn't exported from the package xts.)

But `HighFreq::roll_apply()` is faster because it performs less type-checking and skips other overhead. Unlike the other functions, `roll_apply()` doesn't produce any leading *NA* values.

The function `roll_apply()` can be called in two different ways, depending on the argument `end_points`. If the argument `end_points` isn't explicitly passed to `roll_apply()`, then the default value is used, and `roll_apply()` performs aggregations over overlapping intervals at each point in time.

If the argument `end_points` is explicitly passed to `roll_apply()`, then `roll_apply()` performs aggregations over intervals attached at the end_points. If look_back=2 then the aggregations are performed over non-overlapping intervals, otherwise they are performed over overlapping intervals.

If the argument out_xts is TRUE (the default) then the output is coerced into an *xts* series, with the number of rows equal to the length of argument end_points. Otherwise a list is returned, with the length equal to the length of argument end_points.

If out_xts is TRUE and the aggregation function agg_fun() returns a single value, then roll_apply() returns an *xts* time series with a single column. If out_xts is TRUE and if agg_fun() returns a vector of values, then roll_apply() returns an *xts* time series with multiple columns, equal to the length of the vector returned by the aggregation function agg_fun().

Value

Either an *xts* time series with the number of rows equal to the length of argument end_points, or a list the length of argument end_points.

Examples

```
# extract a single day of SPY data
oh_lc <- HighFreq::SPY["2012-02-13"]
inter_val <- 11  # number of data points between end points
look_back <- 4  # number of end points in look-back interval
# calculate the rolling sums of oh_lc columns over a rolling look-back interval
agg_regations <- roll_apply(oh_lc, agg_fun=sum, look_back=look_back, by_columns=TRUE)
# apply a vector-valued aggregation function over a rolling look-back interval
agg_function <- function(oh_lc)  c(max(oh_lc[, 2]), min(oh_lc[, 3]))
agg_regations <- roll_apply(oh_lc, agg_fun=agg_function, look_back=look_back)
# define end points at 11-minute intervals (HighFreq::SPY is minutely bars)
end_points <- rutils::end_points(oh_lc, inter_val=inter_val)
# calculate the sums of oh_lc columns over end_points using non-overlapping intervals
agg_regations <- roll_apply(oh_lc, agg_fun=sum, end_points=end_points, by_columns=TRUE)
# apply a vector-valued aggregation function over the end_points of oh_lc
# using overlapping intervals
agg_regations <- roll_apply(oh_lc, agg_fun=agg_function,
                            look_back=5, end_points=end_points)
```

---

roll_backtest *Perform a backtest simulation of a trading strategy (model) over a vector of end points along a time series of prices.*

Description

Perform a backtest simulation of a trading strategy (model) over a vector of end points along a time series of prices.

Usage

```
roll_backtest(x_ts, train_func, trade_func, look_back = look_forward,
  look_forward, end_points = rutils::calc_endpoints(x_ts, look_forward), ...)
```

Arguments

x_ts	A time series of prices, asset returns, trading volumes, and other data, in *xts* format.
train_func	The name of the function for training (calibrating) a forecasting model, to be applied over a rolling look-back interval.

`trade_func`	The name of the trading model function, to be applied over a rolling look-forward interval.
`look_back`	The size of the look-back interval, equal to the number of rows of data used for training the forecasting model.
`look_forward`	The size of the look-forward interval, equal to the number of rows of data used for trading the strategy.
`end_points`	A vector of end points along the rows of the `x_ts` time series, given as either integers or dates.
`...`	additional parameters to the functions `train_func()` and `trade_func()`.

Details

The function `roll_backtest()` performs a rolling backtest simulation of a trading strategy over a vector of end points. At each end point, it trains (calibrates) a forecasting model using past data taken from the `x_ts` time series over the look-back interval, and applies the forecasts to the `trade_func()` trading model, using out-of-sample future data from the look-forward interval.

The function `trade_func()` should simulate the trading model, and it should return a named list with at least two elements: a named vector of performance statistics, and an *xts* time series of out-of-sample returns. The list returned by `trade_func()` can also have additional elements, like the in-sample calibrated model statistics, etc.

The function `roll_backtest()` returns a named list containing the lists returned by function `trade_func()`. The list names are equal to the *end_points* dates. The number of list elements is equal to the number of *end_points* minus two (because the first and last end points can't be included in the backtest).

Value

An *xts* time series with the number of rows equal to the number of end points minus two.

Examples

```
## Not run:
# combine two time series of prices
price_s <- cbind(rutils::env_etf$XLU, rutils::env_etf$XLP)
look_back <- 252
look_forward <- 22
# define end points
end_points <- rutils::calc_endpoints(price_s, look_forward)
# perform back-test
back_test <- roll_backtest(end_points=end_points,
    look_forward=look_forward,
    look_back=look_back,
    train_func = train_model,
    trade_func = trade_model,
    model_params = model_params,
    trading_params = trading_params,
    x_ts=price_s)

## End(Not run)
```

`roll_hurst` *Calculate a time series of* Hurst *exponents over a rolling look-back interval.*

Description

Calculate a time series of *Hurst* exponents over a rolling look-back interval.

Usage

```
roll_hurst(oh_lc, look_back = 11)
```

Arguments

`oh_lc`	An *OHLC* time series of prices in *xts* format.
`look_back`	The size of the look-back interval, equal to the number of rows of data used for aggregating the *OHLC* prices.

Details

The function `roll_hurst()` calculates a time series of *Hurst* exponents from *OHLC* prices, over a rolling look-back interval.

The *Hurst* exponent is defined as the logarithm of the ratio of the price range, divided by the standard deviation of returns, and divided by the logarithm of the interval length.

The function `roll_hurst()` doesn't use the same definition as the rescaled range definition of the *Hurst* exponent. First, because the price range is calculated using *High* and *Low* prices, which produces bigger range values, and higher *Hurst* exponent estimates. Second, because the *Hurst* exponent is estimated using a single aggregation interval, instead of multiple intervals in the rescaled range definition.

The rationale for using a different definition of the *Hurst* exponent is that it's designed to be a technical indicator for use as input into trading models, rather than an estimator for statistical analysis.

Value

An *xts* time series with a single column and the same number of rows as the argument `oh_lc`.

Examples

```
# calculate rolling Hurst for SPY in March 2009
hurst_rolling <- roll_hurst(oh_lc=HighFreq::SPY["2009-03"], look_back=11)
chart_Series(hurst_rolling["2009-03-10/2009-03-12"], name="SPY hurst_rolling")
```

`roll_moment` *Calculate a vector of statistics over an* OHLC *time series, and calculate a rolling mean over the statistics.*

Description

Calculate a vector of statistics over an *OHLC* time series, and calculate a rolling mean over the statistics.

Usage

```
roll_moment(oh_lc, mo_ment = "run_variance", look_back = 11,
  weight_ed = TRUE, ...)
```

Arguments

`oh_lc`	An *OHLC* time series of prices and trading volumes, in *xts* format.
`mo_ment`	The name of the function for estimating statistics of a single row of *OHLC* data, such as volatility, skew, and higher moments.
`look_back`	The size of the look-back interval, equal to the number of rows of data used for calculating the rolling mean.
`weight_ed`	*Boolean* argument: should statistic be weighted by trade volume? (default `TRUE`)
`...`	additional parameters to the mo_ment function.

Details

The function `roll_moment()` calculates a vector of statistics over an *OHLC* time series, such as volatility, skew, and higher moments. The statistics could also be any other aggregation of a single row of *OHLC* data, for example the *High* price minus the *Low* price squared. The length of the vector of statistics is equal to the number of rows of the argument `oh_lc`. Then it calculates a trade volume weighted rolling mean over the vector of statistics over and calculate statistics.

Value

An *xts* time series with a single column and the same number of rows as the argument `oh_lc`.

Examples

```
# calculate time series of rolling variance and skew estimates
var_rolling <- roll_moment(oh_lc=HighFreq::SPY, look_back=21)
skew_rolling <- roll_moment(oh_lc=HighFreq::SPY, mo_ment="run_skew", look_back=21)
skew_rolling <- skew_rolling/(var_rolling)^(1.5)
skew_rolling[1, ] <- 0
skew_rolling <- rutils::na_locf(skew_rolling)
```

roll_portf *Simulate (backtest) a rolling portfolio optimization strategy, using* RcppArmadillo.

Description

Simulate (backtest) a rolling portfolio optimization strategy, using *RcppArmadillo*.

Usage

```
roll_portf(ex_cess, re_turns, start_points, end_points, max_eigen, al_pha)
```

Arguments

re_turns	A numeric *matrix* of excess returns data (the returns in excess of the risk-free rate).
max_eigen	An *integer* equal to the regularization intensity (the number of eigenvalues and eigenvectors used for calculating the regularized inverse).
al_pha	The shrinkage intensity.

Details

The function `roll_portf()` calculates the scaled weights of the portfolio with the maximum Sharpe ratio, using *RcppArmadillo*.

It first calculates the regularized inverse of the covariance matrix of returns using function `HighFreq::calc_weights()`. It then estimates the vector of mean returns and applies shrinkage to it, by shrinking it to its common mean value. The shrinkage intensity `al_pha` determines the amount of shrinkage that is applied, with `al_pha = 0` representing no shrinkage (with the estimator of mean returns equal to the means of the columns of `re_turns`), and `al_pha = 1` representing complete shrinkage (with the estimator of mean returns equal to the single mean of all the columns of `re_turns`)

The function `roll_portf()` calculates the weights by multiplying the inverse of the covariance matrix times the estimator of the mean returns. It finally scales the weights by their sum, so that the sum of the weights is equal to `1`.

The function `roll_portf()` calculates a *vector* of rolling variance estimates, from over a *vector* of returns, using *Rcpp*.

Value

A numeric *vector* of the same length as the number of rows of `re_turns`.

Examples

```
## Not run:
# Calculate ETF prices
sym_bols <- colnames(rutils::env_etf$price_s)
sym_bols <- sym_bols[!(sym_bols=="VXX")]
price_s <- rutils::env_etf$price_s[, sym_bols]
# Carry forward non-NA prices
price_s <- zoo::na.locf(price_s)
price_s <- na.omit(price_s)
# Calculate simple ETF returns
```

```
re_turns <- rutils::diff_it(price_s)
# Calculate the daily excess returns
# risk_free is the daily risk-free rate
risk_free <- 0.03/260
ex_cess <- re_turns - risk_free
# Define monthly end_points without initial warmpup period
end_points <- rutils::calc_endpoints(re_turns, inter_val="months")
end_points <- end_points[end_points>50]
len_gth <- NROW(end_points)
# Define 12-month look_back interval and start_points over sliding window
look_back <- 12
start_points <- c(rep_len(1, look_back-1), end_points[1:(len_gth-look_back+1)])
# Define shrinkage and regularization intensities
al_pha <- 0.5
max_eigen <- 3
# Simulate a monthly rolling portfolio optimization strategy
roll_rets <- HighFreq::roll_portf(ex_cess, re_turns,
                                  start_points-1, end_points-1,
                                  max_eigen, al_pha)
roll_rets <- xts(roll_rets, index(re_turns))
colnames(roll_rets) <- "roll_rets"
# Plot dygraph of strategy
dygraphs::dygraph(cumsum(strat_rets),
  main="Cumulative Returns of Max Sharpe Portfolio Strategy")

## End(Not run)
```

`roll_scale` *Perform a rolling scaling (standardization) of the columns of a* matrix *of data using* RcppArmadillo.

Description

Perform a rolling scaling (standardization) of the columns of a *matrix* of data using *RcppArmadillo*.

Usage

```
roll_scale(mat_rix, look_back, use_median = FALSE)
```

Arguments

`mat_rix`	A numeric *matrix* of data.
`look_back`	The length of the look-back interval, equal to the number of rows of data used in the scaling.
`use_median`	A *Boolean* argument: if `TRUE` then the centrality (central tendency) is calculated as the *median* and the dispersion is calculated as the *median absolute deviation* (*MAD*). If `use_median` is `FALSE` then the centrality is calculated as the *mean* and the dispersion is calculated as the *standard deviation*. (The default is `FALSE`)

Details

The function `roll_scale()` performs a rolling scaling (standardization) of the columns of the `mat_rix` argument using *RcppArmadillo*. If the argument `use_median` is `FALSE` (the default), then it performs a similar calculation as the function `roll::roll_scale()`. If the argument `use_median` is `TRUE`, then it calculates the centrality as the *median* and the dispersion as the *median absolute deviation* (*MAD*).

Value

A numeric *matrix* with the same dimensions as the input argument `mat_rix`.

Examples

```
## Not run:
mat_rix <- matrix(rnorm(20000), nc=2)
look_back <- 11
rolled_scaled <- roll::roll_scale(data=mat_rix, width=look_back, min_obs=1)
rolled_scaled2 <- roll_scale(mat_rix=mat_rix, look_back=look_back, use_median=FALSE)
all.equal(rolled_scaled[-1, ], rolled_scaled2[-1, ])

## End(Not run)
```

`roll_sharpe` *Calculate a time series of Sharpe ratios over a rolling look-back interval for an* OHLC *time series.*

Description

Calculate a time series of Sharpe ratios over a rolling look-back interval for an *OHLC* time series.

Usage

```
roll_sharpe(oh_lc, look_back = 11)
```

Arguments

`oh_lc`	An *OHLC* time series of prices in *xts* format.
`look_back`	The size of the look-back interval, equal to the number of rows of data used for aggregating the *OHLC* prices.

Details

The function `roll_sharpe()` calculates the rolling Sharpe ratio defined as the ratio of percentage returns over the look-back interval, divided by the average volatility of percentage returns.

Value

An *xts* time series with a single column and the same number of rows as the argument `oh_lc`.

Examples

```
# calculate rolling Sharpe ratio over SPY
sharpe_rolling <- roll_sharpe(oh_lc=HighFreq::SPY, look_back=11)
```

roll_sum — *Calculate the rolling sum over a vector using* Rcpp.

Description

Calculate the rolling sum over a vector using *Rcpp*.

Usage

```
roll_sum(vec_tor, look_back)
```

Arguments

vec_tor	A numeric *vector* of data.
look_back	The length of the look-back interval, equal to the number of elements of data used for calculating the sum.

Details

The function `roll_sum()` calculates a *vector* of rolling sums, over a *vector* of data, using *Rcpp*. The function `roll_sum()` is over *6* times faster than `rutils::roll_sum()` which uses vectorized *R*.

Value

A numeric *vector* of the same length as the argument `vec_tor`.

Examples

```
## Not run:
# calculate rolling sums over 11-period intervals
sum_rolling <- HighFreq::roll_sum(rnorm(1000), look_back=11)

## End(Not run)
```

roll_var — *Calculate a time series of variance estimates over a rolling look-back interval for an* OHLC *time series of prices, using different range estimators for variance.*

Description

Currently only works for vectors

Usage

```
roll_var(vec_tor, look_back)
```

Arguments

- `look_back` The size of the look-back interval, equal to the number of rows of data used for calculating the variance.
- `oh_lc` An *OHLC* time series of prices in *xts* format.
- `calc_method` *character* string representing method for estimating variance. The methods include:
  - "close" close to close,
  - "garman_klass" Garman-Klass,
  - "garman_klass_yz" Garman-Klass with account for close-to-open price jumps,
  - "rogers_satchell" Rogers-Satchell,
  - "yang_zhang" Yang-Zhang,

  (The default is `"yang_zhang"`)
- `sca_le` *Boolean* argument: should the returns be divided by the number of seconds in each period? (The default is `TRUE`)

Details

The function `roll_var()` calculates a time series of rolling variance estimates of percentage returns, from over a *vector* of returns, using several different variance estimation methods based on the range of *OHLC* prices.

If `sca_le` is `TRUE` (the default), then the variance is divided by the squared differences of the time index (which scales the variance to units of variance per second squared.) This is useful for example, when calculating intra-day variance from minutely bar data, because dividing returns by the number of seconds decreases the effect of overnight price jumps.

If `sca_le` is `TRUE` (the default), then the variance is expressed in the scale of the time index of the *OHLC* time series. For example, if the time index is in seconds, then the variance is given in units of variance per second squared. If the time index is in days, then the variance is equal to the variance per day squared.

The time index of the `oh_lc` time series is assumed to be in *POSIXct* format, so that its internal value is equal to the number of seconds that have elapsed since the *epoch*.

The methods `"close"`, `"garman_klass_yz"`, and `"yang_zhang"` do account for close-to-open price jumps, while the methods `"garman_klass"` and `"rogers_satchell"` do not account for close-to-open price jumps.

The default method is `"yang_zhang"`, which theoretically has the lowest standard error among unbiased estimators.

The function `roll_var()` performs the same calculations as the function `volatility()` from package TTR, but it's a little faster because it uses function RcppRoll::roll_sd(), and it performs less data validation.

Value

An *xts* time series with a single column and the same number of rows as the argument `oh_lc`.

Examples

```
## Not run:
# create minutely OHLC time series of random prices
oh_lc <- HighFreq::random_ohlc()
# calculate variance estimates for oh_lc over a 21 period interval
```

```
var_rolling <- HighFreq::roll_var(oh_lc, look_back=21)
# calculate variance estimates for SPY
var_rolling <- HighFreq::roll_var(HighFreq::SPY, calc_method="yang_zhang")
# calculate SPY variance without accounting for overnight jumps
var_rolling <- HighFreq::roll_var(HighFreq::SPY, calc_method="rogers_satchell")

## End(Not run)
```

`roll_variance` *Calculate a time series of variance estimates over a rolling look-back interval for an* OHLC *time series of prices, using different range estimators for variance.*

Description

Calculate a time series of variance estimates over a rolling look-back interval for an *OHLC* time series of prices, using different range estimators for variance.

Usage

```
roll_variance(oh_lc, look_back = 11, calc_method = "yang_zhang",
  sca_le = TRUE)
```

Arguments

`oh_lc` An *OHLC* time series of prices in *xts* format.

`look_back` The size of the look-back interval, equal to the number of rows of data used for calculating the variance.

`calc_method` *character* string representing method for estimating variance. The methods include:

- "close" close to close,
- "garman_klass" Garman-Klass,
- "garman_klass_yz" Garman-Klass with account for close-to-open price jumps,
- "rogers_satchell" Rogers-Satchell,
- "yang_zhang" Yang-Zhang,

(default is `"yang_zhang"`)

`sca_le` *Boolean* argument: should the returns be divided by the number of seconds in each period? (default is `TRUE`)

Details

The function `roll_variance()` calculates a time series of variance estimates of percentage returns, from *OHLC* prices, using several different variance estimation methods based on the range of *OHLC* prices.

If `sca_le` is `TRUE` (the default), then the variance is divided by the squared differences of the time index (which scales the variance to units of variance per second squared.) This is useful for example, when calculating intra-day variance from minutely bar data, because dividing returns by the number of seconds decreases the effect of overnight price jumps.

If sca_le is TRUE (the default), then the variance is expressed in the scale of the time index of the *OHLC* time series. For example, if the time index is in seconds, then the variance is given in units of variance per second squared. If the time index is in days, then the variance is equal to the variance per day squared.

The time index of the oh_lc time series is assumed to be in *POSIXct* format, so that its internal value is equal to the number of seconds that have elapsed since the *epoch*.

The methods "close", "garman_klass_yz", and "yang_zhang" do account for close-to-open price jumps, while the methods "garman_klass" and "rogers_satchell" do not account for close-to-open price jumps.

The default method is "yang_zhang", which theoretically has the lowest standard error among unbiased estimators.

The function roll_variance() performs the same calculations as the function volatility() from package TTR, but it's a little faster because it uses function RcppRoll::roll_sd(), and it performs less data validation.

Value

An *xts* time series with a single column and the same number of rows as the argument oh_lc.

Examples

```
# create minutely OHLC time series of random prices
oh_lc <- HighFreq::random_ohlc()
# calculate variance estimates for oh_lc over a 21 period interval
var_rolling <- HighFreq::roll_variance(oh_lc, look_back=21)
# calculate variance estimates for SPY
var_rolling <- HighFreq::roll_variance(HighFreq::SPY, calc_method="yang_zhang")
# calculate SPY variance without accounting for overnight jumps
var_rolling <- HighFreq::roll_variance(HighFreq::SPY, calc_method="rogers_satchell")
```

---

roll_vwap — *Calculate the volume-weighted average price of an* OHLC *time series over a rolling look-back interval.*

Description

Performs the same operation as function VWAP() from package VWAP, but using vectorized functions, so it's a little faster.

Usage

```
roll_vwap(oh_lc, x_ts = oh_lc[, 4], look_back)
```

Arguments

oh_lc	An *OHLC* time series of prices in *xts* format.
x_ts	A single-column *xts* time series.
look_back	The size of the look-back interval, equal to the number of rows of data used for calculating the average price.

Details

The function `roll_vwap()` calculates the volume-weighted average closing price, defined as the sum of the prices multiplied by trading volumes in the look-back interval, divided by the sum of trading volumes in the interval. If the argument `x_ts` is passed in explicitly, then its volume-weighted average value over time is calculated.

Value

An *xts* time series with a single column and the same number of rows as the argument `oh_lc`.

Examples

```
# calculate and plot rolling volume-weighted average closing prices (VWAP)
prices_rolling <- roll_vwap(oh_lc=HighFreq::SPY["2013-11"], look_back=11)
chart_Series(HighFreq::SPY["2013-11-12"], name="SPY prices")
add_TA(prices_rolling["2013-11-12"], on=1, col="red", lwd=2)
legend("top", legend=c("SPY prices", "VWAP prices"),
bg="white", lty=c(1, 1), lwd=c(2, 2),
col=c("black", "red"), bty="n")
# calculate running returns
returns_running <- run_returns(x_ts=HighFreq::SPY)
# calculate the rolling volume-weighted average returns
roll_vwap(oh_lc=HighFreq::SPY, x_ts=returns_running, look_back=11)
```

`roll_wsum` *Calculate the rolling weighted sum over a vector of data using* RcppArmadillo.

Description

Calculate the rolling weighted sum over a vector of data using *RcppArmadillo*.

Usage

```
roll_wsum(vec_tor, wei_ghts)
```

Arguments

`vec_tor`	A numeric *vector* of data.
`wei_ghts`	A numeric *vector* of weights.

Details

The function `roll_wsum()` calculates the rolling weighted sum of a vector over its past values (a convolution with the *vector* of weights), using *RcppArmadillo*. It performs a similar calculation as the standard *R* function `filter(x=vec_tor, filter=wei_ghts, method="convolution", sides=1)`, but it's about *6* times faster, and it doesn't produce any *NA* values.

Value

A numeric *vector* of the same length as the argument `vec_tor`.

Examples

```
## Not run:
# First example
# create vector from historical prices
vec_tor <- as.numeric(rutils::env_etf$VTI[, 6])
# create simple weights
wei_ghts <- c(1, rep(0, 10))
# calculate rolling weighted sum
weight_ed <- HighFreq::roll_wsum(vec_tor=vec_tor, wei_ghts=rev(wei_ghts))
# compare with original
all.equal(vec_tor, as.numeric(weight_ed))
# Second example
# create exponentially decaying weights
wei_ghts <- exp(-0.2*1:11)
wei_ghts <- wei_ghts/sum(wei_ghts)
# calculate rolling weighted sum
weight_ed <- HighFreq::roll_wsum(vec_tor=vec_tor, wei_ghts=rev(wei_ghts))
# calculate rolling weighted sum using filter()
filter_ed <- filter(x=vec_tor, filter=wei_ghts, method="convolution", sides=1)
# compare both methods
all.equal(as.numeric(filter_ed[-(1:11)]), as.numeric(weight_ed[-(1:11)]))

## End(Not run)
```

`roll_zscores` *Perform rolling regressions over the rows of the design matrix, and calculate a* vector *of z-scores of the residuals.*

Description

Perform rolling regressions over the rows of the design matrix, and calculate a *vector* of z-scores of the residuals.

Usage

```
roll_zscores(res_ponse, de_sign, look_back)
```

Arguments

`res_ponse`	A numeric *vector* of response data.
`de_sign`	A numeric *matrix* of design (predictor i.e. explanatory) data.
`look_back`	The length of the look-back interval, equal to the number of elements of data used for calculating the regressions.

Details

The function `roll_zscores()` performs rolling regressions along the rows of the design matrix `de_sign`, using function `calc_lm()`. For each regression it calculates the *z-score* of the last residual value. It returns a numeric *vector* of the *z-scores*.

Value

A numeric *vector* of the same length as the number of rows of `de_sign`.

Examples

```
## Not run:
# calculate Z-scores from rolling time series regression using RcppArmadillo
look_back <- 11
clo_se <- as.numeric(Cl(rutils::env_etf$VTI))
date_s <- xts::.index(rutils::env_etf$VTI)
z_scores <- HighFreq::roll_zscores(res_ponse=clo_se,
 de_sign=matrix(as.numeric(date_s), nc=1),
 look_back=look_back)
# Define design matrix with explanatory variables
len_gth <- 100; n_var <- 5
de_sign <- matrix(rnorm(n_var*len_gth), nc=n_var)
# response equals linear form plus error terms
weight_s <- rnorm(n_var)
res_ponse <- -3 + de_sign %*% weight_s + rnorm(len_gth, sd=0.5)
# calculate Z-scores from rolling multivariate regression using RcppArmadillo
look_back <- 11
z_scores <- HighFreq::roll_zscores(res_ponse=res_ponse, de_sign=de_sign, look_back=look_back)
# calculate z-scores in R from rolling multivariate regression using lm()
z_scores_r <- sapply(1:NROW(de_sign), function(ro_w) {
  if (ro_w==1) return(0)
  start_point <- max(1, ro_w-look_back+1)
  sub_response <- res_ponse[start_point:ro_w]
  sub_design <- de_sign[start_point:ro_w, ]
  reg_model <- lm(sub_response ~ sub_design)
  resid_uals <- reg_model$residuals
  resid_uals[NROW(resid_uals)]/sd(resid_uals)
})  # end sapply
# compare the outputs of both functions
all.equal(unname(z_scores[-(1:look_back)]),
  unname(z_scores_r[-(1:look_back)]))

## End(Not run)
```

---

run_returns — *Calculate single period percentage returns from either* TAQ *or* OHLC *prices.*

Description

Calculate single period percentage returns from either *TAQ* or *OHLC* prices.

Usage

```
run_returns(x_ts, lagg = 1, col_umn = 4, sca_le = TRUE)
```

Arguments

x_ts	An *xts* time series of either *TAQ* or *OHLC* data.
lagg	An integer equal to the number of time periods of lag. (default is 1)
col_umn	The column number to extract from the *OHLC* data. (default is 4, or the *Close* prices column)
sca_le	*Boolean* argument: should the returns be divided by the number of seconds in each period? (default is TRUE)

Details

The function run_returns() calculates the percentage returns for either *TAQ* or *OHLC* data, defined as the difference of log prices. Multi-period returns can be calculated by setting the lag parameter to values greater than 1 (the default).

If sca_le is TRUE (the default), then the returns are divided by the differences of the time index (which scales the returns to units of returns per second.)

The time index of the x_ts time series is assumed to be in *POSIXct* format, so that its internal value is equal to the number of seconds that have elapsed since the *epoch*.

If sca_le is TRUE (the default), then the returns are expressed in the scale of the time index of the x_ts time series. For example, if the time index is in seconds, then the returns are given in units of returns per second. If the time index is in days, then the returns are equal to the returns per day.

The function run_returns() identifies the x_ts time series as *TAQ* data when it has six columns, otherwise assumes it's *OHLC* data. By default, for *OHLC* data, it differences the *Close* prices, but can also difference other prices depending on the value of col_umn.

Value

A single-column *xts* time series of returns.

Examples

```
# calculate secondly returns from TAQ data
re_turns <- HighFreq::run_returns(x_ts=HighFreq::SPY_TAQ)
# calculate close to close returns
re_turns <- HighFreq::run_returns(x_ts=HighFreq::SPY)
# calculate open to open returns
re_turns <- HighFreq::run_returns(x_ts=HighFreq::SPY, col_umn=1)
```

---

run_sharpe — *Calculate time series of Sharpe-like statistics for each row of a* OHLC *time series.*

Description

Calculate time series of Sharpe-like statistics for each row of a *OHLC* time series.

Usage

```
run_sharpe(oh_lc, calc_method = "close")
```

Arguments

oh_lc	An *OHLC* time series of prices in *xts* format.
calc_method	A *character* string representing method for estimating the Sharpe-like exponent.

Details

The function run_sharpe() calculates Sharpe-like statistics for each row of a *OHLC* time series. The Sharpe-like statistic is defined as the ratio of the difference between *Close* minus *Open* prices divided by the difference between *High* minus *Low* prices. This statistic may also be interpreted as something like a Hurst exponent for a single row of data. The motivation for the Sharpe-like statistic is the notion that if prices are trending in the same direction inside a given time bar of data, then this statistic is close to either 1 or -1.

Value

An *xts* time series with the same number of rows as the argument oh_lc.

Examples

```
# calculate time series of running Sharpe ratios for SPY
sharpe_running <- run_sharpe(HighFreq::SPY)
```

---

run_skew — *Calculate time series of skew estimates from a* OHLC *time series, assuming zero drift.*

Description

Calculate time series of skew estimates from a *OHLC* time series, assuming zero drift.

Usage

```
run_skew(oh_lc, calc_method = "rogers_satchell")
```

Arguments

oh_lc	An *OHLC* time series of prices in *xts* format.
calc_method	A *character* string representing method for estimating skew.

Details

The function run_skew() calculates a time series of skew estimates from *OHLC* prices, one for each row of *OHLC* data. The skew estimates are expressed in the time scale of the index of the *OHLC* time series. For example, if the time index is in seconds, then the skew is given in units of skew per second. If the time index is in days, then the skew is equal to the skew per day.

Currently only the "close" skew estimation method is correct (assuming zero drift), while the "rogers_satchell" method produces a skew-like indicator, proportional to the skew. The default method is "rogers_satchell".

Value

A time series of skew estimates.

Examples

```
# calculate time series of skew estimates for SPY
sk_ew <- HighFreq::run_skew(HighFreq::SPY)
```

`run_variance` *Calculate a time series of point estimates of variance for an* OHLC *time series, using different range estimators for variance.*

Description

Calculates the point variance estimates from individual rows of *OHLC* prices (rows of data), using the squared differences of *OHLC* prices at each point in time, without averaging them over time.

Usage

```
run_variance(oh_lc, calc_method = "yang_zhang", sca_le = TRUE)
```

Arguments

`oh_lc`	An *OHLC* time series of prices in *xts* format.
`calc_method`	A *character* string representing the method for estimating variance. The methods include: • "close" close to close, • "garman_klass" Garman-Klass, • "garman_klass_yz" Garman-Klass with account for close-to-open price jumps, • "rogers_satchell" Rogers-Satchell, • "yang_zhang" Yang-Zhang, (default is `"yang_zhang"`)
`sca_le`	*Boolean* argument: should the returns be divided by the number of seconds in each period? (default is `TRUE`)

Details

The function `run_variance()` calculates a time series of point variance estimates of percentage returns, from *OHLC* prices, without averaging them over time. For example, the method `"close"` simply calculates the squares of the differences of the log *Close* prices.

The other methods calculate the squares of other possible differences of the log *OHLC* prices. This way the point variance estimates only depend on the price differences within individual rows of data (and possibly from the neighboring rows.) All the methods are implemented assuming zero drift, since the calculations are performed only for a single row of data, at a single point in time.

The user can choose from several different variance estimation methods. The methods `"close"`, `"garman_klass_yz"`, and `"yang_zhang"` do account for close-to-open price jumps, while the methods `"garman_klass"` and `"rogers_satchell"` do not account for close-to-open price jumps. The default method is `"yang_zhang"`, which theoretically has the lowest standard error among unbiased estimators.

The point variance estimates can be passed into function `roll_vwap()` to perform averaging, to calculate rolling variance estimates. This is appropriate only for the methods `"garman_klass"` and `"rogers_satchell"`, since they don't require subtracting the rolling mean from the point variance estimates.

The point variance estimates can also be considered to be technical indicators, and can be used as inputs into trading models.

If sca_le is TRUE (the default), then the variance is divided by the squared differences of the time index (which scales the variance to units of variance per second squared.) This is useful for example, when calculating intra-day variance from minutely bar data, because dividing returns by the number of seconds decreases the effect of overnight price jumps.

If sca_le is TRUE (the default), then the variance is expressed in the scale of the time index of the *OHLC* time series. For example, if the time index is in seconds, then the variance is given in units of variance per second squared. If the time index is in days, then the variance is equal to the variance per day squared.

The time index of the oh_lc time series is assumed to be in *POSIXct* format, so that its internal value is equal to the number of seconds that have elapsed since the *epoch*.

The function run_variance() performs similar calculations to the function volatility() from package TTR, but it assumes zero drift, and doesn't calculate a running sum using runSum(). It's also a little faster because it performs less data validation.

Value

An *xts* time series with a single column and the same number of rows as the argument oh_lc.

Examples

```
# create minutely OHLC time series of random prices
oh_lc <- HighFreq::random_ohlc()
# calculate variance estimates for oh_lc
var_running <- HighFreq::run_variance(oh_lc)
# calculate variance estimates for SPY
var_running <- HighFreq::run_variance(HighFreq::SPY, calc_method="yang_zhang")
# calculate SPY variance without overnight jumps
var_running <- HighFreq::run_variance(HighFreq::SPY, calc_method="rogers_satchell")
```

---

save_rets — *Load, scrub, aggregate, and rbind multiple days of* TAQ *data for a single symbol. Calculate returns and save them to a single* '*.RData' *file.*

Description

Load, scrub, aggregate, and rbind multiple days of *TAQ* data for a single symbol. Calculate returns and save them to a single '*.RData' file.

Usage

```
save_rets(sym_bol, data_dir = "E:/mktdata/sec/",
  output_dir = "E:/output/data/", look_back = 51, vol_mult = 2,
  period = "minutes", tzone = "America/New_York")
```

Arguments

sym_bol	A *character* string representing symbol or ticker.
data_dir	A *character* string representing directory containing input '*.RData' files.
output_dir	A *character* string representing directory containing output '*.RData' files.

`look_back`	The number of data points in rolling look-back interval for estimating rolling quantile.
`vol_mult`	The quantile multiplier.
`period`	The aggregation period.
`tzone`	The timezone to convert.

Details

The function `save_rets` loads multiple days of *TAQ* data, then scrubs, aggregates, and rbinds them into a *OHLC* time series. It then calculates returns using function `run_returns()`, and stores them in a variable named '`symbol.rets`', and saves them to a file called '`symbol.rets.RData`'. The *TAQ* data files are assumed to be stored in separate directories for each '`symbol`'. Each '`symbol`' has its own directory (named '`symbol`') in the '`data_dir`' directory. Each '`symbol`' directory contains multiple daily '`*.RData`' files, each file containing one day of *TAQ* data.

Value

A time series of returns and volume in *xts* format.

Examples

```
## Not run:
save_rets("SPY")

## End(Not run)
```

`save_rets_ohlc` *Load* OHLC *time series data for a single symbol, calculate its returns, and save them to a single '`*.RData`' file, without aggregation.*

Description

Load *OHLC* time series data for a single symbol, calculate its returns, and save them to a single '`*.RData`' file, without aggregation.

Usage

```
save_rets_ohlc(sym_bol, data_dir = "E:/output/data/",
  output_dir = "E:/output/data/")
```

Arguments

`sym_bol`	A *character* string representing symbol or ticker.
`data_dir`	A *character* string representing directory containing input '`*.RData`' files.
`output_dir`	A *character* string representing directory containing output '`*.RData`' files.

Details

The function `save_rets_ohlc()` loads *OHLC* time series data from a single file. It then calculates returns using function `run_returns()`, and stores them in a variable named '`symbol.rets`', and saves them to a file called '`symbol.rets.RData`'.

Value

A time series of returns and volume in *xts* format.

Examples

```
## Not run:
save_rets_ohlc("SPY")

## End(Not run)
```

`save_scrub_agg` *Load, scrub, aggregate, and rbind multiple days of* TAQ *data for a single symbol, and save the* OHLC *time series to a single '*`*.RData`*' file.*

Description

Load, scrub, aggregate, and rbind multiple days of *TAQ* data for a single symbol, and save the *OHLC* time series to a single '`*.RData`' file.

Usage

```
save_scrub_agg(sym_bol, data_dir = "E:/mktdata/sec/",
  output_dir = "E:/output/data/", look_back = 51, vol_mult = 2,
  period = "minutes", tzone = "America/New_York")
```

Arguments

`sym_bol`	A *character* string representing symbol or ticker.
`data_dir`	A *character* string representing directory containing input '`*.RData`' files.
`output_dir`	A *character* string representing directory containing output '`*.RData`' files.
`look_back`	The number of data points in rolling look-back interval for estimating rolling quantile.
`vol_mult`	The quantile multiplier.
`period`	The aggregation period.
`tzone`	The timezone to convert.

Details

The function `save_scrub_agg()` loads multiple days of *TAQ* data, then scrubs, aggregates, and rbinds them into a *OHLC* time series, and finally saves it to a single '`*.RData`' file. The *OHLC* time series is stored in a variable named '`symbol`', and then it's saved to a file named '`symbol.RData`' in the '`output_dir`' directory. The *TAQ* data files are assumed to be stored in separate directories for each '`symbol`'. Each '`symbol`' has its own directory (named '`symbol`') in the '`data_dir`' directory. Each '`symbol`' directory contains multiple daily '`*.RData`' files, each file containing one day of *TAQ* data.

Value

An *OHLC* time series in *xts* format.

Examples

```
## Not run:
# set data directories
data_dir <- "C:/Develop/data/hfreq/src/"
output_dir <- "C:/Develop/data/hfreq/scrub/"
sym_bol <- "SPY"
# aggregate SPY TAQ data to 15-min OHLC bar data, and save the data to a file
save_scrub_agg(sym_bol=sym_bol, data_dir=data_dir, output_dir=output_dir, period="15 min")

## End(Not run)
```

---

save_taq — *Load and scrub multiple days of* TAQ *data for a single symbol, and save it to multiple '`*.RData`' files.*

Description

Load and scrub multiple days of *TAQ* data for a single symbol, and save it to multiple '`*.RData`' files.

Usage

```
save_taq(sym_bol, data_dir = "E:/mktdata/sec/",
  output_dir = "E:/output/data/", look_back = 51, vol_mult = 2,
  tzone = "America/New_York")
```

Arguments

`sym_bol`	A *character* string representing symbol or ticker.
`data_dir`	A *character* string representing directory containing input '`*.RData`' files.
`output_dir`	A *character* string representing directory containing output '`*.RData`' files.
`look_back`	The number of data points in rolling look-back interval for estimating rolling quantile.
`vol_mult`	The quantile multiplier.
`tzone`	The timezone to convert.

Details

The function `save_taq()` loads multiple days of *TAQ* data, scrubs it, and saves the scrubbed TAQ data to individual '`*.RData`' files. It uses the same file names for output as the input file names. The *TAQ* data files are assumed to be stored in separate directories for each '`symbol`'. Each '`symbol`' has its own directory (named '`symbol`') in the '`data_dir`' directory. Each '`symbol`' directory contains multiple daily '`*.RData`' files, each file containing one day of *TAQ* data.

Value

a *TAQ* time series in *xts* format.

Examples

```
## Not run:
save_taq("SPY")

## End(Not run)
```

`scrub_agg` *Scrub a single day of* TAQ *data, aggregate it, and convert to* OHLC *format.*

Description

Scrub a single day of *TAQ* data, aggregate it, and convert to *OHLC* format.

Usage

```
scrub_agg(ta_q, look_back = 51, vol_mult = 2, period = "minutes",
  tzone = "America/New_York")
```

Arguments

`ta_q`	*TAQ* A time series in *xts* format.
`look_back`	The number of data points in rolling look-back interval for estimating rolling quantile.
`vol_mult`	The quantile multiplier.
`period`	The aggregation period.
`tzone`	The timezone to convert.

Details

The function `scrub_agg()` performs:

- index timezone conversion,
- data subset to trading hours,
- removal of duplicate time stamps,
- scrubbing of quotes with suspect bid-offer spreads,
- scrubbing of quotes with suspect price jumps,
- cbinding of mid prices with volume data,
- aggregation to OHLC using function `to.period()` from package *xts*,

Valid 'period' character strings include: "minutes", "3 min", "5 min", "10 min", "15 min", "30 min", and "hours". The time index of the output time series is rounded up to the next integer multiple of 'period'.

Value

A *OHLC* time series in *xts* format.

Examples

```
# create random TAQ prices
ta_q <- HighFreq::random_taq()
# aggregate to ten minutes OHLC data
oh_lc <- HighFreq::scrub_agg(ta_q, period="10 min")
chart_Series(oh_lc, name="random prices")
# scrub and aggregate a single day of SPY TAQ data to OHLC
oh_lc <- HighFreq::scrub_agg(ta_q=HighFreq::SPY_TAQ)
chart_Series(oh_lc, name=sym_bol)
```

---

scrub_taq — *Scrub a single day of* TAQ *data in* xts *format, without aggregation.*

Description

Scrub a single day of *TAQ* data in *xts* format, without aggregation.

Usage

```
scrub_taq(ta_q, look_back = 51, vol_mult = 2, tzone = "America/New_York")
```

Arguments

`ta_q`	*TAQ* A time series in *xts* format.
`look_back`	The number of data points in rolling look-back interval for estimating rolling quantile.
`vol_mult`	The quantile multiplier.
`tzone`	The timezone to convert.

Details

The function `scrub_taq()` performs the same scrubbing operations as `scrub_agg`, except it doesn't aggregate, and returns the *TAQ* data in *xts* format.

Value

A *TAQ* time series in *xts* format.

Examples

```
ta_q <- HighFreq::scrub_taq(ta_q=HighFreq::SPY_TAQ, look_back=11, vol_mult=1)
# create random TAQ prices and scrub them
ta_q <- HighFreq::random_taq()
ta_q <- HighFreq::scrub_taq(ta_q=ta_q)
ta_q <- HighFreq::scrub_taq(ta_q=ta_q, look_back=11, vol_mult=1)
```

`season_ality` *Perform seasonality aggregations over a single-column* xts *time series.*

Description

Perform seasonality aggregations over a single-column *xts* time series.

Usage

```
season_ality(x_ts, in_dex = format(zoo::index(x_ts), "%H:%M"))
```

Arguments

`x_ts`	A single-column *xts* time series.
`in_dex`	A vector of *character* strings representing points in time, of the same length as the argument `x_ts`.

Details

The function `season_ality()` calculates the mean of values observed at the same points in time specified by the argument `in_dex`. An example of a daily seasonality aggregation is the average price of a stock between 9:30AM and 10:00AM every day, over many days. The argument `in_dex` is passed into function `tapply()`, and must be the same length as the argument `x_ts`.

Value

An *xts* time series with mean aggregations over the seasonality interval.

Examples

```
# calculate running variance of each minutely OHLC bar of data
x_ts <- run_variance(HighFreq::SPY)
# remove overnight variance spikes at "09:31"
in_dex <- format(index(x_ts), "%H:%M")
x_ts <- x_ts[!in_dex=="09:31", ]
# calculate daily seasonality of variance
var_seasonal <- season_ality(x_ts=x_ts)
chart_Series(x=var_seasonal, name=paste(colnames(var_seasonal),
  "daily seasonality of variance"))
```

`sim_arima` *Recursively filter a vector of innovations through a vector of* ARIMA *coefficients.*

Description

Recursively filter a vector of innovations through a vector of *ARIMA* coefficients.

Usage

```
sim_arima(in_nov, co_eff)
```

Arguments

in_nov	A numeric *vector* of innovations (random numbers).
co_eff	A numeric *vector* of *ARIMA* coefficients.

Details

The function `sim_arima()` recursively filters a vector of innovations through a vector of *ARIMA* coefficients, using *RcppArmadillo*. It performs the same calculation as the standard *R* function `filter(x=in_nov, filter=co_eff, method="recursive")`, but it's about *6* times faster.

Value

A numeric *vector* of the same length as the argument `in_nov`.

Examples

```
## Not run:
# create vector of innovations
in_nov <- rnorm(100)
# create ARIMA coefficients
co_eff <- c(-0.8, 0.2)
# calculate recursive filter using filter()
filter_ed <- filter(in_nov, filter=co_eff, method="recursive")
# calculate recursive filter using RcppArmadillo
ari_ma <- HighFreq::sim_arima(in_nov, rev(co_eff))
# compare the two methods
all.equal(as.numeric(ari_ma), as.numeric(filter_ed))

## End(Not run)
```

---

sim_garch — *Simulate a* GARCH *process using* Rcpp.

Description

Simulate a *GARCH* process using *Rcpp*.

Usage

```
sim_garch(om_ega, al_pha, be_ta, in_nov)
```

Arguments

om_ega	Parameter proportional to the long-term average level of variance.
al_pha	The weight associated with recent realized variance updates.
be_ta	The weight associated with the past variance estimates.
in_nov	A numeric *vector* of innovations (random numbers).

Details

The function `sim_garch()` simulates a *GARCH* process using *Rcpp*.

Value

A numeric *matrix* with two columns: the simulated returns and variance, and with the same number of rows as the length of the argument `in_nov`.

Examples

```
## Not run:
# Define the GARCH model parameters
om_ega <- 0.01
al_pha <- 0.5
be_ta <- 0.2
# Simulate the GARCH process using Rcpp
garch_rcpp <- sim_garch(om_ega=om_ega, al_pha=al_pha, be_ta=be_ta, in_nov=rnorm(10000))

## End(Not run)
```

`sim_ou` *Simulate an* Ornstein-Uhlenbeck *process using* Rcpp.

Description

Simulate an *Ornstein-Uhlenbeck* process using *Rcpp*.

Usage

```
sim_ou(eq_price, vol_at, the_ta, in_nov)
```

Arguments

`eq_price`	The equilibrium price.
`vol_at`	The volatility of returns.
`the_ta`	The strength of mean reversion.
`in_nov`	A numeric *vector* of innovations (random numbers).

Details

The function `sim_ou()` simulates an *Ornstein-Uhlenbeck* process using *Rcpp*, and returns a *vector* representing the time series of prices.

Value

A numeric *vector* representing the time series of prices, with the same length as the argument `in_nov`.

Examples

```
## Not run:
# Define the Ornstein-Uhlenbeck model parameters
eq_price <- 5.0
vol_at <- 0.01
the_ta <- 0.01
# Simulate Ornstein-Uhlenbeck process using Rcpp
price_s <- HighFreq::sim_ou_rcpp(eq_price=eq_price, vol_at=vol_at, the_ta=the_ta, in_nov=rnorm(1000))

## End(Not run)
```

`vari_ance` *Calculate the variance of a vector using* Rcpp.

Description

Calculate the variance of a vector using *Rcpp*.

Usage

```
vari_ance(vec_tor)
```

Arguments

`vec_tor`	A numeric *vector* of data.

Details

The function `vari_ance()` calculates the variance of a vector using *Rcpp*. The function `vari_ance()` is slightly faster than the *R* function `var()`.

Value

A single *numeric* value.

Examples

```
## Not run:
# calculate variance
HighFreq::vari_ance(rnorm(1000))

## End(Not run)
```

`which_extreme` *Calculate a* Boolean *vector that identifies extreme tail values in a single-column* xts *time series or vector, over a rolling look-back interval.*

Description

Calculate a *Boolean* vector that identifies extreme tail values in a single-column *xts* time series or vector, over a rolling look-back interval.

Usage

```
which_extreme(x_ts, look_back = 51, vol_mult = 2)
```

Arguments

`x_ts`	A single-column *xts* time series, or a *numeric* or *Boolean* vector.
`look_back`	The number of data points in rolling look-back interval for estimating rolling quantile.
`vol_mult`	The quantile multiplier.

Details

The function `which_extreme()` calculates a *Boolean* vector, with `TRUE` for values that belong to the extreme tails of the distribution of values.

The function `which_extreme()` applies a version of the Hampel median filter to identify extreme values, but instead of using the median absolute deviation (MAD), it uses the `0.9` quantile values calculated over a rolling look-back interval.

Extreme values are defined as those that exceed the product of the multiplier times the rolling quantile. Extreme values belong to the fat tails of the recent (trailing) distribution of values, so they are present only when the trailing distribution of values has fat tails. If the trailing distribution of values is closer to normal (without fat tails), then there are no extreme values.

The quantile multiplier `vol_mult` controls the threshold at which values are identified as extreme. Smaller quantile multiplier values will cause more values to be identified as extreme.

Value

A *Boolean* vector with the same number of rows as the input time series or vector.

Examples

```
# create local copy of SPY TAQ data
ta_q <- HighFreq::SPY_TAQ
# scrub quotes with suspect bid-offer spreads
bid_offer <- ta_q[, "Ask.Price"] - ta_q[, "Bid.Price"]
sus_pect <- which_extreme(bid_offer, look_back=51, vol_mult=3)
# remove suspect values
ta_q <- ta_q[!sus_pect]
```

which_jumps — *Calculate a* Boolean *vector that identifies isolated jumps (spikes) in a single-column* xts *time series or vector, over a rolling interval.*

Description

Calculate a *Boolean* vector that identifies isolated jumps (spikes) in a single-column *xts* time series or vector, over a rolling interval.

Usage

```
which_jumps(x_ts, look_back = 51, vol_mult = 2)
```

Arguments

x_ts	A single-column *xts* time series, or a *numeric* or *Boolean* vector.
look_back	The number of data points in rolling look-back interval for estimating rolling quantile.
vol_mult	The quantile multiplier.

Details

The function `which_jumps()` calculates a *Boolean* vector, with `TRUE` for values that are isolated jumps (spikes).

The function `which_jumps()` applies a version of the Hampel median filter to identify jumps, but instead of using the median absolute deviation (MAD), it uses the `0.9` quantile of returns calculated over a rolling interval. This is in contrast to function `which_extreme()`, which applies a Hampel filter to the values themselves, instead of the returns. Returns are defined as simple differences between neighboring values.

Jumps (or spikes), are defined as isolated values that are very different from the neighboring values, either before or after. Jumps create pairs of large neighboring returns of opposite sign.

Jumps (spikes) must satisfy two conditions:

1. Neighboring returns both exceed a multiple of the rolling quantile,
2. The sum of neighboring returns doesn't exceed that multiple.

The quantile multiplier `vol_mult` controls the threshold at which values are identified as jumps. Smaller quantile multiplier values will cause more values to be identified as jumps.

Value

A *Boolean* vector with the same number of rows as the input time series or vector.

Examples

```
# create local copy of SPY TAQ data
ta_q <- SPY_TAQ
# calculate mid prices
mid_prices <- 0.5 * (ta_q[, "Bid.Price"] + ta_q[, "Ask.Price"])
# replace whole rows containing suspect price jumps with NA, and perform locf()
ta_q[which_jumps(mid_prices, look_back=31, vol_mult=1.0), ] <- NA
ta_q <- zoo::na.locf(ta_q)
```

Index