Package 'TAQMNGR'

May 20, 2018

Version 2018.5-1

Date 2018-05-12

Title Manage Tick-by-Tick Transaction Data

Author Francesco Calvori, Fabrizio Cipollini, Giampiero M. Gallo and 'gzstream' authors.

Maintainer Fabrizio Cipollini <cipollini.fabrizio@gmail.com>

Depends R (>= 3.0-2)

LinkingTo Rcpp (>= 0.11.0)

Imports

Suggests

SystemRequirements zlib headers and library

Description Manager of tick-by-tick transaction data that performs 'cleaning', 'aggregation' and 'import' in an efficient and fast way. The package engine, written in C++, exploits the 'zlib' and 'gzstream' libraries to handle gzipped data without need to uncompress them. 'Cleaning' and 'aggregation' are performed according to Brownlees and Gallo (2006) <DOI:10.1016/j.csda.2006.09.030>. Currently, TAQMNGR processes raw data from WRDS (Wharton Research Data Service, <https://wrds-web.wharton.upenn.edu/wrds/>).

License GPL (>= 2)

URL https://cran.r-project.org/package=TAQMNGR

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-05-20 06:37:40 UTC

R topics documented:

TAQ *TAQ Manager*

Description

Manage tick-by-tick transaction data

Usage

```
TAQ.CleanTickByTick(dirInput, dirOutput,

  window = 80, deltaTrimmed = 0.10, granularity = 0.04, useCleaned = TRUE)

TAQ.Aggregate(dirInput, symbol, bin, useAggregated = TRUE)

TAQ.Report(dirInput, symbol)

TAQ.Read(dirInput, symbol, import = NULL, startDate, endDate, bin)
```

Arguments

dirInput	A *character* scalar: the input directory. For the function CleanTickByTick(), dirInput is the name of the folder containing the raw data files. In this case it is important that dirInput includes only .gz files to be cleaned. For the remaining functions, dirInput is the name of the folder including the previously cleaned/aggregated data (appearing as dirOutput in the function CleanTickByTick()).
dirOutput	A *character* scalar: the output directory. It must be different from dirInput.
window	A *numeric* integer scalar: the window size for the trimming procedure of data clean (see Details).
deltaTrimmed	A *numeric* scalar into $(0,1)$: the trimming proportion (see Details).
granularity	A *numeric* positive scalar: the granularity parameter (see Details).
useCleaned	A *logical* scalar: if TRUE, previously cleaned files (if any) are not cleaned again.
useAggregated	A *logical* scalar: if TRUE, previously aggregated data (if any) are not aggregated again.
symbol	A *character* (vector in TAQ.Aggregate(); scalar in TAQ.Report() and TAQ.Read()): the ticker symbols of interest.
startDate	A *numeric* integer scalar: the start date in the *yyyymmdd* format.
endDate	A *numeric* integer scalar: the end date in the *yyyymmdd* format.
bin	A *numeric* integer scalar: the bin size (in seconds) for aggregating data.

import — A *character*: the list of fields to be imported. One or more among:
"FIRST": First price in the bin.
"MIN": Min price in the bin.
"MAX": Max price in the bin.
"LAST": Last price in the bin.
"SIZE": First price in the bin.
"#TRADES": Number of trades in the bin.
"VWAP": Volume Weighted Average Price in the bin.
If NULL, all fields are imported.

Details

The meaning of the arguments window, deltaTrimmed, and granularity is detailed in the reference below.

References

Brownlees, C. T., and Gallo, G. M. (2006). Financial Econometric Analysis at Ultra–High Frequency: Data Handling Concerns, *Computational Statistics and Data Analysis* **51**, 2232–2245.

Examples

```
#### A fake dataset for running the example can be downloaded at
#### 'http://local.disia.unifi.it/cipollini/webpage-new/data/data_sample.txt.gz'
## Input
# dirInput  <- "path of the input folder"
# dirOutput <- "path of the output folder" ## Must be different from 'dirInput'
## Clean
# TAQ.CleanTickByTick(dirInput = dirInput, dirOutput = dirInput)
## Make the report (1 at a time)
# TAQ.Report(dirInput = dirOutput, symbol = c("DOG")) ## A scalar symbol
# TAQ.Report(dirInput = dirOutput, symbol = c("GNU")) ## A scalar symbol
## Aggregate
# TAQ.Aggregate(dirInput = dirOutput, symbol = c("DOG", "GNU"), bin = 300,
#   useAggregated = TRUE)
## Import data
# dog <- TAQ.Read(dirInput = dirOutput, symbol = "DOG",
#   startDate = 00010101, endDate = 20141231, bin = 300)
```

TAQMNGR *TAQ Manager*

Description

The package manages tick-by-tick transaction data, performing *cleaning*, *aggregation* and *import*.

Details

The package manages tick-by-tick transaction data, performing *cleaning*, *aggregation* and *import* in an efficient and fast way (the package engine is developed in C++).
Cleaning and Aggregation are performed according to Brownlees and Gallo (2006).

Package:	TAQMNGR
Type:	Package
Version:	2015.2-1
Date:	2015-02-21
License:	GPL (>= 2)

Note

Currently, the package processes raw data from WRDS (Wharton Research Data Service).
They have to satisfy the following requirements:

- all fields have to be included (select the 'Check All' button at the WRDS downloading page);
- select the fixed-width text and 'G zip' as output format and compression type, respectively, at the WRDS downloading page.

An example with fake raw data can be downloaded at http://local.disia.unifi.it/cipollini/webpage-new/data/data_sample.txt.gz.
The package uses the following libraries: 'Gzstream' (available at 'http://www.cs.unc.edu/Research/compgeom/gzstream/' under LGPL license), and 'zlib' (freely available at 'http://www.zlib.net/').

Author(s)

Francesco Calvori <francesco.calvori@gmail.com>,

Fabrizio Cipollini <cipollini.fabrizio@gmail.com>,

Giampiero M. Gallo <giampiero.gallo@gmail.com>.

Maintainer: <fabrizio.cipollini@gmail.com>

References

Brownlees, C. T., and Gallo, G. M. (2006). Financial Econometric Analysis at Ultra–High Frequency: Data Handling Concerns, *Computational Statistics and Data Analysis* **51**, 2232–2245.

Index