Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

The dataset used in this paper consists of high-frequency transaction prices of the most active contracts of the e-mini S&P 500 futures (e-mini for short), a stock market index futures contract traded on the CME Group's electronic trading platform Globex. The data cover the period from November 30, 2005, through November 30, 2020.

Availability

☐ Data **are** publicly available

☒ Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

Publicly available data

☐ Data are available online at:

☐ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Discussion of lack of publicly available data:

This is proprietary data set and cannot be made publicly available. A synthetic data set generated from the same procudure as that specified in Section A.3 of the Supplementary Appendix is provided. The simulated data set captures the main characteristic regarding the volatility diurnal pattern of the real data. However, results based on this pseudo-dataset may differ from the real data analyses.

Description

File format(s)

☒ CSV or other plain text:
☐ Software-specific binary format (.Rda, Python pickle, etc.):
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (described here):

Data dictionary

☐ Provided by the authors in the following file(s):
☒ Data file(s) is (are) self-describiing (e.g., netCDF files)
☐ Available at the following URL:

Additional information (optional)

The pseudo-dataset is a series of 6,112,471 simulated logarithmic prices at the 30-second frequency. This series is further split into two segments (files). The two segments (files) are series of 2,705,431 and 3,407,041 log prices with volatility diurnal patterns calibrated to that of the real data covering the periods 2005-2010 and 2015-2020, respectively.

Part 2: Code

Abstract

This code mainly consists of two parts. The first part implements the calendar effect estimation procedure together with 95% confidence intervals. The second part implements the test procedure.

Description

Code format(s)

☒ Script files
- ☒ R ☐ Python ☐ Matlab
- ☐ Other:

☐ Package
- ☐ R ☐ Python ☐ MATLAB toolbox
- ☐ Other:

☐ Reproducible report
- ☐ R Markdown ☐ Jupyter notebook
- ☐ Other:

☐ Shell script

☐ Other (described here):

Supporting software requirements

Version of primary software used

R version 3.5.1.

Libraries and dependencies used by the code

None.

Supporting system/hardware requirements (optional)

Parallelization used

☒ No parallel code used
☐ Multi-core parallelization on a single machine/node
Number of cores used:
☐ Multi-machine/multi-node parallelization
Number of nodes and cores used:

License

☒ MIT License (default)
☐ BSD
☐ GPL v3.0
☐ Creative Commons
☐ Other (described here):

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper

☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

☐ All tables and figures in the paper

☐ Selected tables and figures in the paper, as explained and justified here:

Workflow details

Location

The workflow is available:

☒ As part of the paper's supplementary material

☐ In this Git repository:

☐ Other:

Format(s)

☐ Single master code file

☐ Wrapper (shell) script(s)

☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

☐ Text file (e.g., a readme-style file) that documents workflow

☐ Makefile

☒ Other (more detail in 'Instructions' below)

Instructions

The code in the format of a single R script file is self-describing. It illustrates how the key numerical results (estimation and test) of the paper are obtained as long as one or two series of trade prices observed at an appropriate frequency is provided as inputs.

After loading a series of logarithmic prices at the thirty-second frequency (the default setting, but easily adjustable), the first part of the code produces estimate of calendar effect function together with its 95% confidence intervals (both pointwise and uniform) for the period over which the pseudo-prices are observed. The plot produced by this part of the code resembles that of Figure 4 in the Appendix Section B.2 as well as the bottom plots of Figure 1 in Section 6.2 of the manuscript.

Meanwhile, after loading two series of logarithmic prices observed over two time periods, the second part of the code produces the p-value of the test for the equality of calendar effects of the two samples associated with this pair of time periods. The outputs of this part of the code resemble that of Table 1 in Section 6.3 of the manuscript.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

☐ <1 minute
☒ 1-10 minutes
☐ 10-60 minutes
☐ 1-8 hours
☐ >8 hours
☐ Not feasible to run on a desktop machine, as described here:

Additional documentation (optional)

Notes (optional)