

READ ME FILE

Title: Designing an Efficient Reference Rate: Lessons from SOFIA

Authors: James Brugler, Calebe de Roure, Marta Khomyn, Max Prakoso and Talis Putniņš

Description

This 'read me' file details the replication files for RDP 2026-03. As per data provision agreements with the data providers, underlying data to this project cannot be disclosed. Nevertheless, this file discloses the code created for the RDP assuming access to the data.

All data used to plot graphs appearing in the RDP are available to the public and can be found in the spreadsheet 'rdp-2026-03-graph-data'.

This supplementary information is divided in three parts. First, it describes the files enclosed. Second, it describes which files are used to create each Figure and Table in the RDP. Third, it notes the Noise Analysis replication package contained in this zip file.

Files

The code files in this section were run using RStudio 2025.05.1 and Stata/MP 18.0.

1. 1.variables_time_series_analysis.R: creates the timeseriesdata.xlsx file that later will be used for the time series empirical analysis.

File dependencies:

- cashRateDaily.csv (provided)
- businessDaySchedule.csv (provided)

2. 2_timeseries_analysis.do: creates the time series empirical analysis.

File dependencies:

- timeseriesdata.xlsx (calculated using 1.variables_time_series_analysis.R)
- sign_noise_sofia_vwap_rbachck (calculated, see Noise Analysis replication package)
- bid-ask_3y.dta (not provided, access to Bloomberg data required)

3. 3.variables_paneldata_analysis.R: creates the paneldata.xlsx file that will later be used for the panel data empirical analysis.

File dependencies:

- cashRateDaily.csv (provided)
- businessDaySchedule.csv (provided)
- sign_noise_sofia_vwap_rbachck (calculated, see Noise Analysis replication package)

4. 4.paneldata_analysis.do: creates the panel data empirical analysis.

File dependencies:

- paneldata.xlsx (calculated using 3.variables_paneldata_analysis.R)

5. 5.Figures3-7.R: creates Figures 3 to 7.

File dependencies:

- cashRateDaily.csv (provided)
- businessDaySchedule.csv (provided)

6. 6.expert_judgement_analysis.do: creates data for Figure 9.

File dependencies:

- timeseriesdata.xlsx (calculated using 1.variables_time_series_analysis.R)

7. 7.SOFIA_minus_2stddev.R: creates data for Figure 8.

File dependencies:

- cashRateDaily.csv (provided)
- businessDaySchedule.csv (provided)

8. 8.expert_judgement_regression.do: creates empirical analysis.

File dependencies:

- timeseriesdata.xlsx (calculated using 1.variables_time_series_analysis.R)
- sign_noise_sofia_vwap_rbachck (calculated, see Noise Analysis replication package)

Outputs created

Table 1: Descriptive statistics of time series variables

Run 2_timeseries_analysis.do

File dependencies:

- 1.variables_time_series_analysis.R
- sofia_signal_noise_extract_final.ipynb (see Noise Analysis replication package)

Figure 1: Related-party transactions

Data in 1.variables_time_series_analysis.R

Figure 2: Market concentration and trading volume

Data in 1.variables_time_series_analysis.R

Table 2: Descriptive statistics for transaction-level variables

Run 4.paneldata_analysis.do

File dependencies:

- 3.variables_paneldata_analysis
- sofia_signal_noise_extract_final.ipynb (see Noise Analysis replication package)

Table 3: Time series regression results

Run 2_timeseries_analysis.do

File dependencies:

- 1.variables_time_series_analysis.R

Figure 3: Beta SOFIA excluding related-party transactions

Run 5.Figures3-7.R

Table 4: Panel data regression results

Run 4.paneldata_analysis.do

File dependencies:

- 3.variables_paneldata_analysis.R

Table 5: Noise share comparison between ASX's beta SOFIA and synthetic SOFIAs

Noise share output from sofia_vwap_rba_vs_vwap_alt_noise_shares_final.ipynb (see Noise Analysis replication package)

Figure 4: Beta SOFIA versus 25/95 trims

Run 5.Figures3-7.R

Figure 5: Beta SOFIA versus 25/95 trims, orders > \$2 million

Run 5.Figures3-7.R

Figure 6: Beta SOFIA versus 25/75 trims

Run 5.Figures3-7.R

Figure 7: Beta SOFIA versus 25/75 trims, orders > \$2 million

Run 5.Figures3-7.R

Figure 8: Beta SOFIA versus 25/mean + two standard deviations

Run 7.SOFIA_minus_2stddev.R

Figure 9: Binned scatter plot of data

- Panel 1: Transaction volume and noise
- Panel 2: Transaction volume and noise (bottom 5 percentile observations)
- Panel 3: Number of participants and noise
- Panel 4: Number of participants and noise (bottom 5 percentile observations)
- Panel 5: Number of trades and noise
- Panel 6: Number of trades and noise (bottom 5 percentile observations)

Run 6.expert_judgement_analysis.do

Table 6: Regression estimates of average noise in low liquidity days

Run 8.expert_judgement_regression.do

Figure A1: Binned scatter plot after new treatment of holidays

Run 6.expert_judgement_analysis.do

Noise Analysis

Note that there is a separate README.html file found in the 'Noise Analysis' folder that documents the replication package found within and how to use it.

The replication package consists of two Jupyter notebooks (which execute the analysis) and three Python modules (which contain model definitions and helper functions).

24 June 2026