Enhanced TRACE Database

1 Overview

This Python repository creates the TRACE bond database as used in Dickerson, Mueller and Robotti (2023, DMR). It first aims to clean various data issues in the daily Trace Enhanced File. The cleaning procedure is largely based on the discussions presented in the works of Dick-Nielsen (2009) and (2014), as well as Bai, Bali, and Wen (2019). The script effectively handles Cancellation, Correction, Reversal, and Double entries in the daily data. Moreover, it is designed to cater to both pre and post the 2012/02/06 change in the Trace System. Thereafter, several scripts are available to create the monthly bond-level panel which variables included such as bond returns, credit spreads and others.
The data file, `trace_2002_2021.h5`, contains the actual data we use for the main results of DMR.

1.1 A note on the TRACE database

Creating the TRACE database from scratch (from the intraday to the final monthly panel) is subject to various coding design choices and possible heterogeneity across programming languages. We have processed the data in several ways, with different choices of functions and consistently arrive at a similar sized dataset with a set of results that are always closely aligned. We believe our choices (throughout the code base) are justified.
We strongly urge researchers to use the publicly available WRDS database going forward to completely eliminate any possible heterogeneity from having to 'self-process' the TRACE data from scratch.
If you believe the TRACE data preparation codebase can be improved, please contact: `a.dickerson@warwick.ac.uk`.

2 Requirements

- Python 3.9.13
- pandas
- numpy
- quantLib 1.29
- joblib 1.1.1
- wrds 3.1.2 (and access to the WRDS database and cloud)

3 Usage

1. Ensure you have Python installed on your system. If not, you can download it from the official Python website: `https://www.python.org/downloads/`
2. Update and install the required packages.
3. Run these scripts sequentially to produce the TRACE bond-level panel:
    - Run `MakeIntra_Daily.py`. This script outputs the daily bond-level panel with clean prices and volumes. Run on the WRDS cloud. Approximate runtime is ∼4.50 hours.

- Run `MakeBondDailyMetrics.py`. This script outputs the daily bond accrued interest, dirty prices, duration, convexity, and yields. With 14 cores, approximate runtime is ∼20 minutes.
- Run `MakeMonthlyMetrics.py`. This script outputs the monthly bond returns, excess returns, bond yields, duration, and convexity. Negligible runtime (under 1 minute).
- Run `MakeCreditSpreads.py`. This script estimates monthly bond credit spreads. With 14 cores, approximate runtime is ∼20 minutes.
- Run `MakeIlliquidity.py`. This script estimates monthly bond illiquidity following Bao et al. (2011). Negligible runtime (under 5 minutes).
- Run `MakeRatings.py` and `MakeAmountOutstanding.py` in any order. This script downloads bond ratings and bond amount outstanding. Negligible runtime (under 1 minute).
- Run `MakeDataBaseTRACE.py`. This script downloads the data processed in the prior scripts and generates the final database. Negligible runtime (under 1 minute).

4. Total runtime to generate the full Enhanced TRACE monthly bond-level panel from scratch is ∼5.00 hours. Please reach out to `a.dickerson@warwick.ac.uk` if you believe the runtime can be optimized. You will be given full credit.

4 Acknowledgements

1. Matthias Buchner (Trafigura) – code optimization.
2. Francis Cong (Morgan Stanley) – code related to QuantLib.
3. Mihai Mihut (Tilburg) – code optimization.
4. Freda Song Drechsler (WRDS) – providing SAS code to process the daily TRACE data.