INSTRUCTIONS TO RUN CODE FOR THE EMPIRICAL AND NUMERICAL RESULTS PRESENTED ON "AN INFORMATION-BASED THEORY OF FINANCIAL INTERMEDIATION"

Overview

This readme file contains instructions on how to run the code in the replication package that generates Tables 1-8, and Figures 5, 7 and 8 for the empirical components in the paper and Figures 3, 4, and 6 for the numerical results.

Contact: bethune@virginia.edu, bruno@sultanum.com, trachter@gmail.com

Data Availability Statement

The paper uses confidential regulatory data from the Trade Information Warehouse (TIW) of the Depository Trust and Clearing House Corporation (DTCC) on credit-default-swap transactions. Board staff (which, for these purposes may also include Federal Reserve Bank personnel with a need to know) uses the Data in carrying out the Board's responsibilities in connection with prudential supervision, monetary policy, and systemic risk monitoring. The Board treats the Data as confidential and exempt from disclosure under 5 U.S.C. 552(b)(4) and 552(b)(8). All outward facing documents in this replication package, including the code, figures and tables, and this readme file went through the approval at The Board to verify compliance with the above statutes.

In carrying out the Board's responsibilities, staff may incorporate portions of the Data, either in unmodified ('raw') or summary ('aggregate') form, into Board works. These works include research papers (such as papers prepared for publication in academic journals), speeches, testimony, reports, official publications (such as the monetary policy report), and internal policy documents (such as the Tealbook and the financial stability report).
Reserve Bank staff requesting to use the data for 'research' purposes must submit justification specific to the Board's responsibilities via the National Access Management System (NAMS) request form.

We obtained access request approval through employment by the Federal Reserve Bank of Richmond. The raw data was last downloaded for this project in the fourth quarter of 2017.

Other academic papers using the same data set include: Du, Wenxin, Salil Gadgil, Michael B. Gordy, and Clara Vega. "Counterparty risk and counterparty choice in the credit default swap market." *Available at SSRN 2845567* (2019).

The 13-F data is available from the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) at the Securities and Exchange Commission (SEC) publicly available at https://www.sec.gov/edgar/searchedgar/companysearch.html.

Dataset list

Note: the dataset below is not provided in replication files due to confidentiality requirements (see the Data Availability Statement (DAS) above). We have provided the variable references used in the code that is included in the replication files for clarity.

- \complete data for paper\cleaned tiw data\tiw matched clean.dta
    - Each observation is a trade in this data, and it contains all trades in the indices used in the paper. Below we describe the variables.
        - refid: trade identifier.
        - date: trade date.
        - year: trade year.
        - quarter: trade quarter.
        - entity: underlying CDS index.
        - buyer: buyer name.
        - buyer_accountid: account identifier for the buyer.
        - namecik_buyer: matched EDGAR CIK name of the buyer.
        - cik_buyer: matched buyer CIK.
        - buyer_report_YEARqQUARTER_13F: dummy of whether the buyer in the trade file a 13 F in the respective YEAR and QUARTER. Where YEAR ranges from 2013 to 2017; and QUARTER from 1 to 4.
        - buyer_datefiled_YEARqQUARTER_13F: day that the buyer in the trade file a 13 F in the respective YEAR and QUARTER. Where YEAR ranges from 2013 to 2017; and QUARTER from 1 to 4. If applicable.
        - seller: seller name.
        - seller_accountid: account identifier for the seller.
        - namecik_seller: matched EDGAR CIK name of the seller.
        - cik_seller: matched seller CIK.
        - seller_report_YEARqQUARTER_13F: dummy of whether the seller in the trade file a 13 F in the respective YEAR and QUARTER. Where YEAR ranges from 2013 to 2017; and QUARTER from 1 to 4.
        - seller_datefiled_YEARqQUARTER_13F: day that the seller in the trade file a 13 F in the respective YEAR and QUARTER. Where YEAR ranges from 2013 to 2017; and QUARTER from 1 to 4. If applicable.
- \complete data for paper\raw EDGAR data\masterYEARqQUARTER.dta and \complete data for paper\raw EDGAR data\masterYEARqQUARTER.txt
    - Where YEAR ranges from 2013 to 2017, and QUARTER ranges from 1 to 4.
    - Each observation is a 13-F filed. The data is available from the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) at the Securities and Exchange Commission (SEC).

Computational Requirements

Below we list the software used:

- RStudio Version 1.2.1335
    - Required libraries: data.table (v1.14.0), stringdist (v0.9.7)
- Stata MP 16 for Windows (64-bit x86-64)
- C++ version 9.3.0 (compiled with g++)

    - Required libraries other than what is standard: gsl (v 2.7), CBLAS (v 3.10), OpenMP
- iPython 7.25.0 interactive python shell using Python 3.9.6 (64-bit)
    - Required libraries: numpy (v1.2.1), matplotlib (v.2.2.5), os (v3.9), sys (v3.9.6)

The total run time is from 2 to 5 hours. We recommend a computer with at least 64GB of memory. Total run time just to run the computational part of the code is approximately 2 hours.

Descriptions of Programs

The folders in this zip file are:

- folder "complete data for paper" – is a placeholder for the raw data set as well as all the intermediary and cleaned data created when running the code (note, these datasets are not provided due to data restrictions, see DAS above). The subfolders are:
    - "\complete data for paper\cleaned tiw data"
        - tiw matched clean.dta – this file is not provided but referenced in other code. It contains the confidential data containing CDS transactions.
    - "\complete data for paper\regressions" – this folder is used to store intermediate data files created to run regressions. Once you run the code, it will contain the following files:
        - regressions setup US.dta
        - regressions setup US, filers.dta
        - regressions setup US, filing buyers and sellers.dta
        - regressions setup US, central 40-60.dta
        - regressions setup US, central 20-80.dta
        - regressions setup US, central 40-60, filers.dta
        - regressions setup US, central 20-80, filers.dta
        - regressions setup all indexes, filers.dta
        - regressions setup delays, filers.dta
    - "\complete data for paper\empirical figures" – contains files used for Figures 5, 7, and 8.
        - hist_delay_all.png – figure 5
        - share_buying_no_weight_all.png – figure 7 left panel
        - dev_share_buying_scatter_log_centrality1.png – figure 7 right panel
        - centrality_vs_counter_w_dealers.png – figure 8 left panel
        - centrality_vs_counter_no_separate_dealer.png – figure 8 right
    - "\complete data for paper\raw EDGAR data" – contains raw data on 13-F filings
- folder "complete code for paper" – contains all the empirical code described in the instructions below.
    - regressions setup.do – prepares data for regressions

  The subfolders are:
    - "\complete code for paper\1. cleaning"
        - identifying the top 5 traders.do – identifies the top-5 traders
        - Subset and format tiw matched clean.do – re-organize data in different orientations
    - "\complete code for paper\2. setup balanced panel"
        - 13f regressions SETUP for US indexes.do – creates panel data for US
        - 13f regressions SETUP for all indexes.do – creates panel data for all indices
        - generate official report dates. do – create dataset with official report dates

        - 13f regressions SETUP for US indexes, delays.do – creates panel data for US with filling delay variable
    - "\complete code for paper\3. main paper
        - Table 1. and 8. trading statistics.do – generates stats for tables 1 and 8
        - Table 2 and 5. impact of 13f filing on trade, different datasets.do – generates regs for tables 2 and 5
        - Table 3. impact of 13f filing on trade, different week spans.do – generates regs for table 3
        - Table 4. impact of 13f filing by index types.do – generates regs for table 3
        - Figures 7-8. Centrality and buying share.do – generates figures 7 and 8
        - Figures 5. Delays.do – generates figure 5
    - "\complete code for paper\4. appendix"
        - Table 6 and 7. impact of 13f filing on trade, delay window.do – generates regs for tables 6 and 7
        - Table 9. impact of 13f filing on trade, buyside v sellside.do – generates regs for table 9
- folder "numerical code" – contains all the programs needed to produce the numerical results. The subfolders and files are:
    - fig3code - holds code to compute model for Figure 3
        - functions.C – module to compute pricing mechanism
        - globals.C – defines model and program parameters
        - II_model.C – module to compute model
        - main.C – main program file
        - main.exe – compiled executable
        - plots.py – program to produce figure
        - model_output – folder that holds model output
        - log.txt – log file
    - fig4code - holds files to compute model for Figure 4 (sub-file description identical to those in fig3code)
    - fig6code - holds files to compute model for Figure 6 (sub-file description identical to those in fig3code)
    - figures – holds final computational figures
- readme.pdf (this file)

Instructions to Replicators for Empirical Results

This section includes detailed instructions on how to run the programs to produce Tables 1-8, and Figures 5, 7 and 8. The programs i) setup a balanced panel and ii) analyze the various samples and print output in tables and figures. Note: the instructions below state which folders the data sets have to be downloaded into in order for the code to run.

1. Setup Balanced Panel for Regressions

Subset final data set

This step takes “\complete data for paper\cleaned tiw data\tiw matched clean.dta" and creates two different dataset orientations, long and short. The "long" data set eliminates the distinction between buyer and seller and treats each participant as a "trader" or “accountid”. This file also creates datasets that are US-index specific.

- Run file: “\complete code for paper\1. cleaning\Subset and format tiw matched clean.do”
- Run file “\complete code for paper\1. cleaning\identifying the top 5 traders.do"
  - Finds the top 5 traders. The account id of these traders were manually entered in the other files. Due to the confidentiality of the data, in the current codes we replaced the IDs with “TTTTTTT”. It has to be updated for the code to run correctly.

Construct panel data

Below we describe the steps to construct the panel data used in the regressions. For reference, here is a glossary of the variable names used in these steps:

- N_: denotes the number of trades; e.g. N_top5 is the number of trades with a top 5 trader.
- D_: is a dummy (0,1) for whether or not a trade happened; e.g. D_top5 = 1 means that there was 1 or more trades that occurred with a top 5 trader.
- R_: is a dummy (0,1) for whether or not a trader reported.
- s: the prefix s means that the variable is standardized by frequency.
- _b: the suffix _b means that the variable refers to buyers. Likewise, the suffix _s means that the variable is refers to sellers.

To Run:

- Run file “\complete code for paper\2. setup balanced panel\13f regressions SETUP for US indexes.do" that sets up a balanced panel data set where each observation is a unique “index x institution x week”.
  - Here are the steps the file completes:
    1. Calculate the number of trades with top 5 per trader. Save these .dta files separately. They will be merged into the main file.
    2. Create the “skeleton" data set that has one observation per index x institution x week (ijt).
    3. Calculate the number of trades per index x institution x week (this time we don't care who the counterparty is).
    4. Merge the skeleton with the variables in Steps 1 and 3.
    5. Generate the dummy variables for whether a trade happened for a particular index x institution x week.
    6. Set the data up as panel data. Generate the dummy variables Rjt to indicate whether or not a report was filed in that week for that trader and generate the lagged report variables, ie. whether or not the trader reported last week, two weeks ago, etc.
    7. Subset and format the files necessary for our regressions.
  - The output is saved as:
    - “\complete data for paper\regressions\regressions setup US.dta”
    - “\complete data for paper\regressions\regressions setup US, filers.dta”
    - “\complete data for paper\regressions\regressions setup US, filing buyers and sellers.dta”
    - “\complete data for paper\regressions\regressions setup US, central 40-60.dta”
    - “\complete data for paper\regressions\regressions setup US, central 20-80.dta”
    - “\complete data for paper\regressions\regressions setup US, central 40-60, filers.dta”

        - "\complete data for paper\regressions\regressions setup US, central 20-80, filers.dta"
- Run file "\complete code for paper\2. setup balanced panel\13f regressions SETUP for all indexes.do" that sets up the regressions for all entities, only keeping traders who filed.
    - The steps are the same as "13f regressions SETUP for US indexes.do".
    - The output is saved as: "\complete data for paper\regressions\regressions setup all indexes, filers.dta"
- Run file "\complete code for paper\2. setup balanced panel\13f regressions SETUP for US indexes, delays.do"
    - The file follows the same steps as "13f regressions SETUP for US indexes.do" but adds the delay variable. The results for this regression end up in the appendix.
    - The output is saved as: "\complete data for paper\regressions\regressions setup delays, filers.dta"

2. Create Output for Tables and Figures

This code should not need any updates if more recent data is introduced, aside from having to change the data sets that are imported. The data in Tables 1 through 9 is printed in the Stata terminal.

Tables 1 and 8

Run file "\complete code for paper\3. Main paper\Table 1. and 8. trading statistics.do"

- can be customized by substituting different data sets for the ones included, which only involve US indexes 1. This file produces the statistics for Table 1. It uses the data files
    - \complete data for paper\cleaned tiw data\tiw matched clean US.dta"
    - \complete data for paper\cleaned tiw data\tiw matched clean longformat.dta"
    - \complete data for paper\cleaned tiw data\tiw matched clean longformat US.dta"

Tables 2 and 5

Run file "\complete code for paper\3. Main paper\Table 2 and 5. impact of 13f filing on trade, different datasets.do"
which calls the do-file **"\complete code for paper\regressions setup.do"**

- produces regressions for Table 1 and 5.

Table 3

Run file "\complete code for paper\3. Main paper\Table 3. impact of 13f filing on trade, different week spans.do"
which calls the do-file **"\complete code for paper\regressions setup.do"**

- produces regressions for Table 3.

Table 4

Run file "\complete code for paper\3. Main paper\Table 4. impact of 13f filing by index types.do"
which calls the do-file **"\complete code for paper\regressions setup.do"**

- produces regressions for Table 4.

Table 6 and 7

Run file "\complete code for paper\4. Appendix\Table 6 and 7. impact of 13f filing on trade, delay window.do"
which calls the do-file **"\complete code for paper\regressions setup.do"**

- produces regressions for Tables 6 and 7.

Table 9

Run file "\complete code for paper\4. Appendix\Table 9. impact of 13f filing on trade, buyside v sellside.do"
which calls the do-file **"\complete code for paper\regressions setup buyer and seller variables.do"**

- produces regressions for Table 9.

Figure 5

Run file "\complete code for paper\3. Main paper\ Figures 5. Delays.do"

- Produces figures 5.

Figures 7 and 8

Run file "\complete code for paper\3. Main paper\ Figures 7-8. Centrality and buying share.do"

- Produces figures 7a, 7b, 8a and 8b.

Instructions to Replicators for Numerical Results

This section includes detailed instructions on how to run the programs to produce Figures 3, 4, and 6.

To Run:

1. Open a terminal window
2. Change working directory to folder "computational code"
3. run "g++-9 –I **"include_path"** –L **"library_path"** ./fig3code/main.C –o ./fig3code/main –lgsl –lgslcblas –lm -fopenmp"
   - This step computes the model output for Figure 3
4. run "iPython ./fig3code/plots.py"
   - This step produces Figure 3 and puts output in folder "./computational code/figures
5. run "g++-9 –I **"include_path"** –L **"library_path"** ./fig4code/main.C –o ./fig4code/main –lgsl –lgslcblas –lm -fopenmp"
   - This step computes the model output for Figure 4
6. run "iPython ./fig4code/plots.py"
   - This step produces Figure 4 and puts output in folder "./computational code/figures
7. run "g++-9 –I **"include_path"** –L **"library_path"** ./fig6code/main.C –o ./fig6code/main –lgsl –lgslcblas –lm -fopenmp"
   - This step computes the model output for Figure 6

8. run “iPython ./fig6code/plots.py”
    - This step produces Figure 6 and puts output in folder “./computational code/figures

***Note: user needs to define their own “include_path” (e.g. /usr/local/include) and **“library_path”** (e.g. /usr/local/lib) in steps 3, 6, and 8 above.*

Data Citations

1. Depository Trust & Clearing Corporation, Trade Information Warehouse Data (Confidential Data), http://www.dtcc.com/derivatives-services/trade-information-warehouse.
2. “Form 13F.” Electronic Data Gathering, Analysis, and Retrieval system (EDGAR). Securities and Exchange Commission (SEC).

END OF README