US Bank Integration 2.0

This is the restarting of the US bank integration project with better quality data compiled from WRDS.

Standard Industrial Classification Code

We include a short description of SIC codes and what they stand for. The following code scrapes the SIC code tables from the corresponding Wikipedia page. The full list can be found on the link: http://www.ehso.com/siccodes.php

```
url_SIC <- "https://en.wikipedia.org/wiki/Standard_Industrial_Classification"
webpage <- xml2::read_html(url_SIC)
SIC_tables <- rvest::html_nodes(webpage, "table")
table_1 <- rvest::html_table(SIC_tables[2], fill = F)
knitr::kable(table_1[[1]], caption = "SIC codes for industries.
             The acronym NEC stands for 'not elsewhere classified'")
```

Table 1: SIC codes for industries. The acronym NEC stands for 'not elsewhere classified'

Range of SIC Codes	Division
0100-0999	Agriculture, Forestry and Fishing
1000-1499	Mining
1500-1799	Construction
1800-1999	not used
2000-3999	Manufacturing
4000-4999	Transportation, Communications, Electric, Gas and Sanitary service
5000-5199	Wholesale Trade
5200-5999	Retail Trade
6000-6799	Finance, Insurance and Real Estate
7000-8999	Services
9100-9729	Public Administration
9900-9999	Nonclassifiable

As is clearly visible, the banking and finance industry is found between the SIC codes 6000 to 6799. Hence in order to collect information on banks, we need to focus on this particular subset. In the following set of codes, we display only the finance industry subgroups in the above range.

```
table_2 <- rvest::html_table(SIC_tables[3], fill = F)
sic_fin <- seq(6000, 6799)
table_2_fin <- table_2[[1]] %>% dplyr::filter(`SIC Code` %in% sic_fin)
knitr::kable(table_2_fin, caption = "SIC codes for finance and
             related industry groups")
```

Table 2: SIC codes for finance and related industry groups

SIC Code	Industry
6012	Pay Day Lenders
6021	National Commercial Banks
6022	State Commercial Banks
6029	Commercial Banks, NEC
6035	Savings Institution, Federally Chartered
6036	Savings Institutions, Not Federally Chartered
6099	Functions Related To Depository Banking, NEC
6111	Federal & Federally Sponsored Credit Agencies
6141	Personal Credit Institutions
6153	Short-Term Business Credit Institutions
6159	Miscellaneous Business Credit Institution
6162	Mortgage Bankers & Loan Correspondents
6163	Loan Brokers
6172	Finance Lessors
6189	Asset-Backed Securities
6199	Finance Services
6200	Security & Commodity Brokers, Dealers, Exchanges & Services
6211	Security Brokers, Dealers & Flotation Companies
6221	Commodity Contracts Brokers & Dealers
6282	Investment Advice
6311	Life Insurance
6321	Accident & Health Insurance
6324	Hospital & Medical Service Plans
6331	Fire, Marine & Casualty Insurance
6351	Surety Insurance
6361	Title Insurance
6399	Insurance Carriers, NEC
6411	Insurance Agents, Brokers & Service
6500	Real Estate
6510	Real Estate Operators (No Developers) & Lessors
6512	Operators of Nonresidential Buildings
6513	Operators of Apartment Buildings
6519	Lessors of Real Property, NEC
6531	Real Estate Agents & Managers (For Others)
6532	Real Estate Dealers (For Their Own Account)
6552	Land Subdividers & Developers (No Cemeteries)
6770	Blank Checks
6792	Oil Royalty Traders
6794	Patent Owners & Lessors
6795	Mineral Royalty Traders
6798	Real Estate Investment Trusts
6799	Investors, NEC

Further, we extract the relevant portion of the SIC table from the full list to focus on the

banking and finance industry:

SIC Code	Industry
60	DEPOSITORY INSTITUTIONS
601	CENTRAL RESERVE DEPOSITORY INSTITUTIONS
6011	FEDERAL RESERVE BANKS
6019	CENTRAL RESERVE DEPOSITORY INSTITUTIONS, NEC
602	COMMERCIAL BANKS
6021	NATIONAL COMMERCIAL BANKS
6022	STATE COMMERCIAL BANKS
6029	COMMERCIAL BANKS, NEC
603	SAVINGS INSTITUTIONS
6035	SAVINGS INSTITUTIONS, FEDERALLY CHARTERED
6036	SAVINGS INSTITUTIONS, NOT FEDERALLY CHARTERED
606	CREDIT UNIONS
6061	CREDIT UNIONS, FEDERALLY CHARTERED
6062	CREDIT UNIONS, NOT FEDERALLY CHARTERED
608	FOREIGN BANKING AND BRANCHES AND AGENCIES OF FOREIGN BANKS
6081	BRANCHES AND AGENCIES OF FOREIGN BANKS
6082	FOREIGN TRADE AND INTERNATIONAL BANKING INSTITUTIONS
609	FUNCTIONS RELATED TO DEPOSITORY BANKING
6091	NONDEPOSIT TRUST FACILITIES
6099	FUNCTIONS RELATED TO DEPOSITORY BANKING, NEC
61	NONDEPOSITORY CREDIT INSTITUTIONS
611	FEDERAL AND FEDERALLY-SPONSORED CREDIT AGENCIES
6111	FEDERAL AND FEDERALLY-SPONSORED CREDIT AGENCIES
614	PERSONAL CREDIT INSTITUTIONS
6141	PERSONAL CREDIT INSTITUTIONS
615	BUSINESS CREDIT INSTITUTIONS
6153	SHORT-TERM BUSINESS CREDIT INSTITUTIONS, EXCEPT AGRICULTURAL
6159	MISCELLANEOUS BUSINESS CREDIT INSTITUTIONS
616	MORTGAGE BANKERS AND BROKERS
6162	MORTGAGE BANKERS AND LOAN CORRESPONDENTS
6163	LOAN BROKERS
67	HOLDING AND OTHER INVESTMENT OFFICES
671	HOLDING OFFICES
6712	OFFICES OF BANK HOLDING COMPANIES
6719	OFFICES OF HOLDING COMPANIES, NEC

We follow Fahlenbrach, Prilmeier, and Stulz (2016) in constructing our sample of US banks. We quote them (p. 6):

> We construct our sample as follows. We search the CRSP database for all firms that have an SIC code between 6020 and 6079 (Commercial Banks, Savings Institutions, and Credit Unions) or from 6710 through 6712 (Offices of Bank Holding Companies) at some point in the firm's history.

Additionally we also seek guidance from the sample construction in Adrian and Brunnermeier

(2016):

> Banks correspond to SIC codes 60, 61, and 6712; insurance companies correspond to SIC codes 63–64, real estate companies correspond to SIC codes 65–6, and broker-dealers are SIC code 67 (except for the bank holding companies, 6712).

Since our sample and definition of the US banking system follows the discussion in Fahlenbrach, Prilmeier, and Stulz (2016) more closely, we try to construct our sample as close to theirs as we can.

Directory Management

First we need to be able to navigate to the directory where data are stored.

```
data_folder_path <- "../Data_Bank_Int/"
file_path <- paste0(data_folder_path, "SICCD_6020-6079_6710-6712_20171105.dta")
```

The relevant file here is a Stata file with extension `.dta`. There is a special reason for choosing the Stata file format instead of the usual Excel or `.csv` format preferred for doing data analysis in `R`. This is so since the file containing daily stock prices of all banks in the US is very large and its number of rows exceeds the maximum row limit in Excel, which is around 1 million. Hence if one were to download and read the data via Excel, only the first one million or so rows will be displayed and stored.[1]

Tidy Reading

After navigating to the relevant folder, one needs to be able to read the data file, stored in this case, with a `.dta` extension.

```
data <- haven::read_dta(file_path)

bank_name <- data %>% dplyr::distinct(comnam)
```

References

Adrian, Tobias, and Markus K Brunnermeier. 2016. "CoVaR." *American Economic Review* 106 (7): 1705–41.

Fahlenbrach, Rüdiger, Robert Prilmeier, and Renè M Stulz. 2016. "Why Does Fast Loan Growth Predict Poor Performance for Banks?"

[1] A previous difficulty encountered in this project was that the list of banks corresponding to SIC codes between 6000-6799 did not include the G-SIBs. However, the reason for that mystery was the lack of rows from 1-2 million being displayed/stored in Excel due to its size limitation.