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.

```
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")
```

Table 1: SIC codes for industries

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

We follow Fahlenbrach, Prilmeier, and Stulz (2016) in constructing our sample of US banks.

We quote them:

> 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.

Directory Management

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

```
data_folder_path <- "../Data_Bank_Int/"
file_path <- paste0(data_folder_path, "SICCD_6000_6999_20171031.csv")
```

Tidy Reading

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

```
data <- readr::read_csv(file_path,
                        na = c("", "NA", "NaN", "C", "B"),
                        col_types =
                          cols(`Cumulative Factor to Adjust Shares/Vol` = col_double(),
                               `Cumulative Factor to Adjust Prices` = col_double(),
                               `Factor to Adjust Price` = col_double(),
                               `Factor to Adjust Shares` = col_double(),
                               `CUSIP Header` = col_character(),
                               CUSIP = col_character()
                               )
                        )

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

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