



Yanni Lan <lanyanni@bu.edu>

MSBA Capstone Cohort B - Team 1

Peter Wysocki <wysockip@bu.edu>

Fri, Feb 7, 2020 at 11:03 AM

To: Yanni Lan <lanyanni@bu.edu>

Cc: Kiki Hu <xuxiaoqi@bu.edu>, cniu@bu.edu, yuyangs@bu.edu, zhangsq@bu.edu

Dear MSBA Team 1 (Cohort B),

As follow-up to our meeting about your capstone project, I am listing below information about the two key data sources we discussed: (1) The Reuters-21578 news data set, and (2) Historical U.S. stock market data available at Questrom through WRDS. I provide more details below. I hope you find this information useful for your project.

(1) Reuters-21578 data set

- I have attached a PDF file with a recent academic article about using the text from daily Reuters news articles to predict overall stock market volatility (note that this is different than predicting future stock returns).

- You may also find the following web links useful to get a better understanding about how to access and process the Reuters-21578 data set:

<https://miguelmalvarez.com/2015/03/20/classifying-reuters-21578-collection-with-python-representing-the-data/>

<https://miguelmalvarez.com/2016/11/07/classifying-reuters-21578-collection-with-python/>

<https://martin-thoma.com/nlp-reuters/>

<https://www.kaggle.com/nltkdata/reuters>

<https://archive.ics.uci.edu/ml/datasets/reuters-21578+text+categorization+collection>

<http://www.daviddlewis.com/resources/testcollections/reuters21578/>

(2) Historical U.S. Stock market data available at Questrom through WRDS:

MSBA students can access financial data using the WRDS database at BU using the BA870 WRDS account (see below).

IMPORTANT: YOU DO NOT NEED TO REQUEST/CREATE A SEPARATE WRDS ACCOUNT.

What is WRDS?

<https://wrds-web.wharton.upenn.edu/wrds/about/WRDS%20FAQs.cfm>

IMPORTANT:

The WRDS services are for academic and non-commercial research purposes only. Users may not use data downloaded from the WRDS database for any non-academic or commercial endeavor.

How to access WRDS for BU MSBA Class BA870 (Spring 2020)?

Web Location: <https://wrds.wharton.upenn.edu>

Username (case sensitive): ba870bu

Password (case sensitive): BosCommAveMSBA2020

Important: This username and password is only for BA870 students. Do not share this username and password with anyone else.

When you log-in, you will be asked to "Review and Accept the WRDS Terms of Use".

How to find a time series of stock prices in WRDS?

The Quick Answer to "How to find a time series of stock prices in WRDS?":

Step 1: Log into WRDS (See Above)

Step 2: Navigate to Web Interface for CRSP Stock Market Information

After logging in (and accepting WRDS Terms of Service), you will see a horizontal list of options at the very top of the website (in dark Blue):

Get Data | Analytics | Classroom | Research | Support

Click on “**Get Data**”, then under the “**Vendors**” Heading, Click on “**CRSP**”

Under **Annual Updates** on left select **Stock / Security Files**

You then have the options to select

Monthly Stock File (with monthly reporting of stock prices and returns),

Daily Stock File (with monthly reporting of stock prices and returns),

Stock Market Indexes (Daily or Monthly information on Indices such as S&P500)

This example will show how to access Monthly Stock Prices

- Select **Monthly Stock File**

Step 3: Date Range

- Set **Beginning Date** month and year, e.g. Jan 1925
- Set **Ending Date** month and year, e.g. Dec 2008
- Data availability: CRSP monthly stock price data cover the period from December 1925 through the present. Daily data begin in July 1962. NYSE data are available throughout the entire period; AMEX data start in 1962, and NASDAQ data begin in 1972.

Step 4: Search

- Set code format to **CUSIP**
- Click **Code lookup** under **Manually enter company codes**; a new window will open
- Type company name (Type *Apple Inc*) into search box; click **Is Exactly**
- Copy and paste the company's CUSIP back into the **Company Codes** box on the main WRDS search page
- Enter multiple codes by leaving a single space between them
- Leave all other boxes in this step as is

Step 5: Variables

- In **Identifying Information** check **Company Name**
- In **Time Series Information** check **Price**
- Leave all other boxes in this step as is

Step 6: Output

- Choose **Tab-delimited text (*.txt)** format
- Leave all other boxes in this step as is
- Click **Submit Query**
- Data Request Summary window will open
- Right mouse click on the highlighted file name
- Choose **Save Target As** option
- Save as a .txt file to your computer
- Open file using Excel, Python, R or other text editor for viewing and processing

The Longer Answer to “How to find a time series of stock prices in WRDS?”:

Read the following information to best determine which data you need:

What is CRSP?

<https://wrds-www.wharton.upenn.edu/pages/support/data-overview/wrds-overview-crsp-us-stock-database/>

How to Get Financial Data using WRDS Web Queries?

<https://wrds-www.wharton.upenn.edu/pages/support/getting-started/wrds-web-queries/>

Finding Data by Concepts on WRDS (such as Stock Prices):

<https://vimeo.com/138246389>

Finding Stock Data from CRSP (Stock Price Data) on WRDS:

<https://vimeo.com/138985880>

Overview of CRSP Database:

(Note: The relevant Web Query information starts at 6:30 in the video)

<https://vimeo.com/137366884>

WRDS – Basic data retrieval:

<https://vimeo.com/221873921>

=====

Professor Peter D. Wysocki
Department of Accounting
Questrom School of Business
Boston University
[595 Commonwealth Avenue](#)
[Boston, MA 02215](#)
wysockip@bu.edu

[Quoted text hidden]



News and Stock Volatility.pdf

266K