

# Data Handbook for Daily Price, Volume and Earnings Call Transcripts of Russell 3000

SAIF-Tower Research Capital Live Learning Project

Muchen Hu\*

April 17, 2023

## Contents

<b>1</b>	<b>Introduction to Russell 3000</b>	<b>2</b>
<b>2</b>	<b>Daily Price and Volume</b>	<b>2</b>
2.1	Data Source . . . . .	2
2.2	Data Format . . . . .	2
<b>3</b>	<b>Earnings Call Transcripts</b>	<b>3</b>
3.1	Basics . . . . .	3
3.1.1	What is earnings call for? . . . . .	3
3.1.2	How is earnings call (transcript) structured? . . . . .	3
3.1.3	When does earnings call happen? . . . . .	4
3.1.4	How frequently does earnings call happen? . . . . .	4
3.1.5	How fast can we get the transcript? . . . . .	4
3.2	Data Source . . . . .	4
3.3	Dataset and Data Format . . . . .	4

---

\*For any further questions, please kindly contact me at [mchu.22@saif.sjtu.edu.cn](mailto:mchu.22@saif.sjtu.edu.cn).

# 1 Introduction to Russell 3000

The Russell 3000 Index is a market-capitalization-weighted equity index maintained by FTSE Russell that provides exposure to the entire U.S. stock market. The index tracks the performance of the 3,000 largest U.S.-traded stocks, which represent about 97% of all U.S.-incorporated equity securities.

Stocks in the Russell 3000 index are **reconstituted once a year**, on the last Friday in June. May is the ranking month, when all eligible U.S. companies are lined up to form the preliminary Russell Reconstitution portfolio. At this time, all eligible securities are ranked by their current market capitalization. This ensures that growing or shrinking companies are accurately represented in the overall index.

## 2 Daily Price and Volume

### 2.1 Data Source

We first obtained Russell 3000 constituents from 2010-01-01 to 2022-03-31 from Wind. Note that the constitution is time-varying. Then we retrieved daily trading information of Russell 3000 from CRSP Dataset<sup>1</sup>.

Each stock is uniquely identified with CUSIP instead of TICKER. A CUSIP is a nine-character numeric or alphanumeric code (e.g. 037833100 for Apple or 38259P508 for Google) that uniquely identifies a North American financial security for the purposes of facilitating clearing and settlement of trades. TICKER can be confusing, especially when M&A happen.

### 2.2 Data Format

Daily trading information of Russell 3000 can be found under *PriceVolumeDaily* folder in our shared drive. Pickle files can be read as dataframe using *pd.read\_pickle(your path)*.

Name	Content	Size	Kind
DlyAsk	Daily Ask	130.4MB	pickle file
DlyBid	Daily Bid	130.4MB	pickle file
DlyClose_raw	Daily Close	130.4MB	pickle file
DlyFacPrc	Daily Factor To Adjust Price	130.4MB	pickle file
DlyHigh_raw	Daily High	130.4MB	pickle file
DlyLow_raw	Daily Low	130.4MB	pickle file
DlyNumTrd	Daily Number Of Trades	130.4MB	pickle file
DlyOpen_raw	Daily Open	130.4MB	pickle file
DlyVol	Daily Volume	130.4MB	pickle file
universe	In Russell 3000 or not	16.4MB	pickle file

<sup>1</sup><https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/crsp/stocks-and-indices/crsp-stock-and-indices-version-2/>

## 3 Earnings Call Transcripts

### 3.1 Basics

#### 3.1.1 What is earnings call for?

An earnings call is a conference call between the management of a public company, analysts, investors, and the media to discuss the company's financial results during a given reporting period, such as a quarter or a fiscal year. An earnings call is usually preceded by an earnings report, which contains summary information on financial performance for the period.

During an earnings call, company management discusses the details of its SEC Form 10-Q (quarterly report) or 10-K (annual report). Federal securities laws mandate that publicly traded companies provide certain information in these forms, including detailed financial results, along with a more qualitative discussion.

Analysts use the information they learn from the earnings call in fundamental analysis of the company. Fundamental analysis begins with the company's financial statements. Analysts will comb through these statements in addition to listening in on verbal cues that company management gives during the earnings call. Analysts may ask questions during an earnings call related to main concepts or even details in the footnotes that focus on inventory and "less accumulated depreciation" lines.

#### 3.1.2 How is earnings call (transcript) structured?

Earnings calls often begin with the moderator issuing a safe harbor statement, which advises that the call may include forward-looking statements.

After the safe harbor statement has been made, the company's managers take over the call. Commonly, a company is represented by C-level executives. Depending on a company and its corporate hierarchy, the number of participating executives may vary. However, the two key executives that are always present in the earnings call are the chief executive officer (CEO) and chief financial officer (CFO).

The final portion of the earnings call is reserved for the Q&A session. During this session, investors, analysts, and other participants in the call have an opportunity to ask the company's management questions regarding the presented financial results. Note that the company's managers have the right to decline or defer their answers for certain questions.

**Usually, Q&A session contains more informative signals for the firm's financial performance, as speakers are less prepared for random questions raised by analysts.**

The transcript for an earnings call contains all the information mentioned above.

**[Company name + Report period + Release date]**

**Company participants / Executives**

[Name] – [Position]

**Conference call participants / Analysts**

[Name] – [Company]

**Executives Structure**

Who is(are) the main speaker(s)?

Who is most active in Q&A session?

**Opening**

Operators or one of company participants

Declaration and introduction to executives

**Company report / presentation**

CEO, President, CFO, CMO, VP, SVP.....(usually CEO and CFO present)

Overall performance, update on performance / development of different product lines, financials(usually CFO).....

**Facts**

**Sentiment**

From which section

From whose words

Relative to industry/history

**Q&A Session**

analyst by analyst

### 3.1.3 When does earnings call happen?

Normally earnings calls are scheduled before the market opens or after the market closes<sup>2</sup>. Before the earnings calls, earnings announcements are usually already released.

**In our case, we built the dataset in order to utilize the earnings call transcripts to predict the next open to close return rates. Therefore, if the earnings call happens before the market opens, its time stamp remains unchanged. Otherwise, its time stamp will be mapped to the next day.**

For example, if one earnings call was released at Aug.10, 2015 9:00 AM(before 9:30 AM), we aligned it with Aug.10, 2015; if it was released later, we aligned it with Aug.11, 2015(actually should be the next trading day).

Note that the release date(e.g. Aug.10,2015) is different from report date(e.g. Q2 2015).

### 3.1.4 How frequently does earnings call happen?

Public companies have an earnings call each quarter. These quarterly calls are where companies share financial information like revenue, profit, losses, and earnings per share (EPS). Quarterly calls align with a company’s fiscal year, **which can be different from the calendar year under GAAP(different from the practice in China).**

**Therefore, one can imagine that, unlike daily information, earnings call transcripts come less frequently and regularly.**

### 3.1.5 How fast can we get the transcript?

Currently, we have two data sources for earnings call transcripts: CapitalIQ and SeekingAlpha. CapitalIQ is a well-maintained database while SeekingAlpha is a website for data and news.

From CapitalIQ, there can be at most one day delay to get a neat, corrected and uniformly formatted earnings call transcript. From SeekingAlpha, a newly released transcript with errors can be accessed within a few hours.

Pros		Cons
CapitalIQ	no errors, nicely formatted, CUSIP identifier	longer delay
SeekingAlpha	shorter delay	need web scraper, TICKER identifier

Considering all pros and cons, we choose CapitalIQ as the main dataset and SeekingAlpha as the validation dataset to make sure data from different sources is consistent without conflicts.

## 3.2 Data Source

The transcripts are downloaded from CapitalIQ using CUSIP identifier from 2015-01-01 to 2022-03-31. Besides earnings call transcripts, CapitalIQ also provides other similar conference call transcripts, such as shareholders conference. We downloaded those transcripts together to enrich our dataset.

## 3.3 Dataset and Data Format

There are two datasets(folders) for transcripts in our shared drive.

<sup>2</sup><https://finance.yahoo.com/calendar/earnings/>

One is called *EarningsCallTextByCUISP*, which prioritizes CUISP first. This dataset is suitable for any case where users want to get all transcripts for one particular company since 2015.

The other one is called *TranscriptsAlongTime*, which prioritizes time first. This dataset is suitable for cross-sectional analysis, where users want to get all transcripts from different companies released on a particular trading day on a rolling basis.

Note that a folder for a particular company or trading day can be empty.

Universe(stock): EarningsCallTestByCUISP ~22% missing

CUISP1	CUISP2	CUISP3	CUISP4
cuisp1.txt	None	None	None

Daily Timestamp: TranscriptsAlongTime

~44 stock per day since 2015