MS&E 245A - Downloading Financial Data

Overview

There are numerous options for accessing financial data pertinent to MS&E 245A. Some of these are freely available to the public, such as Yahoo Finance and Google Finance. Others are geared towards business users and must be accessed on Stanford through special terminals, such as Bloomberg.[1] Others are geared towards academic users and can be accessed online by supplying one's Stanford ID and password, such as Wharton Research Data Services (WRDS).[2] This guide will focus on several ways of accessing data from Yahoo finance which should prove sufficient for most or all of your needs for MS&E 245A. Nevertheless, you are more than welcome to consult other data sources for additional analyses.

Important Data Notes

Regardless of how you get your data, there are several important things to be aware of as you do so:

Prices vs. Adjusted Prices

Frequently when you access price and returns data for a single stock, fund, etc. you will see two types of prices listed. One, a regular closing price (i.e. the stock price at the end of the trading day) and another, an "adjusted closing price." One factor that the adjusted price takes into account is stock splits. Thus, for example, suppose a stock is trading at $100 one day and the next day does a 2:1 split and thus is now trading at $50. The regular prices will show a drop of $100 to $50 but the adjusted price will show a constant $50. Furthermore, the adjusted price will multiply *all* prices prior to the stock split by 0.5 (in the case of a 2:1 split). This makes intuitive sense in some respects - nothing fundamental about the market valuation of a stock has changed when it splits, so we don't want that to affect the stock return calculations we make. By multiplying all prices prior to the split by a factor proportional to the split ratio, we ensure a steady series of prices that we can directly analyze.

Adjusted prices also account for dividends. The principle for accounting for these is just the same as it is for stock splits, though the practical details are a bit more complicated. Suppose a company's stock is selling at $50 per share on day 1. On day 2, the company issues a $2 per share dividend to all of its stock owners. The price of the stock will reliably drop by about $2 per share immediately afterwards.[3] Why? The shares represent a proportional claim on the total value of the company. If a company has 100 million shares, and each is selling for $50 a piece, then it means the market thinks the company is worth about $5 billion. If the company then gives away 4% of that value to shareholders in the form of a dividend, then that is no longer value that a shareholder can claim a fractional ownership of through owning the share. But, the investors won't be too sad - even though their share price drops by about $2, they will get $2 in cash from the dividend. Furthermore, if the investors want to, they can reinvest that cash right back into more stock, so that they will still have $50 worth of stock in the company.[4]

[1] The Engineering library on the second floor of Huang has two Bloomberg terminals, and the librarians there are trained to assist users in accessing the data.

[2] See `http://www.gsb.stanford.edu/library/articles/databases/links/wrds-doc` for details.

[3] Note: dividend payments are announced in advance. Thus, by the time that a dividend payment is actually made, there is essentially no uncertainty amongst investors as to how much it will be.

[4] Some companies even offer the option of having this done automatically to avoid transaction costs.

Dividends are important to consider when calculating returns on stocks and ETFs. The general assumption is that dividend payments are reinvested as soon as they are made. Fortunately, by using adjusted closing prices, dividend payments are automatically accounted for. Consider the following example:

On day 1, company A's stock is selling for $40 per share. 1 year later, the company's stock is selling for $50 per share. The company then pays a $2 dividend to shareholders, after which its stock drops to $48. One way to calculate the investors return would thus be look at (a) closing stock prices and (b) dividends. The guides below give information on how to pull both types of stock data. In this example then, we would see a closing stock price of $40 at the beginning of the year, a closing price of $48 at the end of the year, and a dividend of $2, also at the end of the year. We would calculate the investor's return as:

$$\text{return} = \frac{(48+2)-40}{40} = 25\%$$

Thus, if you are calculating returns using closing prices, it is very important to be mindful of dividends and when they occur, because otherwise, you will arrive at inaccurate calculations for stock returns.

Alternatively, if you use adjusted returns, this problem is solved for you. The formula for calculating adjusted returns in the presence of dividends is as follows:

1. Calculate the dividend adjustment factor as follows:

$$\text{adjustment factor} := \frac{\text{Price before dividend payment} - \text{Dividend amount}}{\text{Price before dividend payment}}$$

   In the example above, this would be:

$$\frac{50-2}{50} = 0.96$$

2. Multiply all stock prices prior to the dividend payment date by this adjustment factor.

Thus, in our example from above, if we looked at adjusted stock prices, instead of seeing the stock begin at $40 per share 1 year ago, we would instead see it begin at $\$40 * (0.96) = \$38.4$. Then, instead of seeing it grow to $48 plus a $2 dividend at the end of the year, we would instead just see it grow to $48. We would then calculate the return as:

$$\text{return} = \frac{48-38.4}{38.4} = 25\%$$

Thus, by using the adjusted returns, we automatically incorporate the dividend payments and save ourselves the trouble of needing to carefully keep track of when dividends were paid and in how much.

Funds vs. Indices

The first week's investment strategy writeup introduces the concepts of index funds that track certain indices such as the S&P 500 (the 500 largest public companies in the US). We will return to indices and funds of various types throughout the investment simulation for MS&E 245A. A key concept to understand is that an index is in some respects a theoretical construct. An index is defined by:

1. A set of stocks or other securities; and
2. A set of weights specifying how much of each security to hold as compared to the others.

Thus, the S&P 500 index specifies the set of the 500 largest public companies with weights for each company determined by the market cap of that company divided by the total market cap of all companies. You cannot, however, *directly* own an index or shares in an index. Instead, what you can do is to own shares in a *fund* that *tracks* an index by holding shares of stocks in an index in proportions specified by the weights that define an index.

The distinction between an index and a fund becomes particularly relevant when searching for historical data for at least two reasons:

1. **Differences in Time Coverage:** In the first week's investment strategy and throughout this data guide, we use the example of the ticker symbol VOO which tracks the S&P 500 index. If you go to look up financial data under the ticker symbol VOO, however, you will notice that none is available prior to 2010. Does this mean that the S&P 500 index didn't exist prior to 2010? Certainly not. But, it does mean that the VOO index fund was created then. Thus, if you are conducting research on an investing strategy that involves investing in the S&P 500 index, the VOO fund might be the fund in which you choose to invest in order to implement that strategy, but it might not be what you use to get data to analyze the implications of your strategy. Instead, if you want to consider a broader time frame, you might choose to seek out data on the underlying S&P 500 index itself.

2. **Differences in Return:** As discussed in the first week's strategy document, index funds are very cheap and efficient. But, this doesn't mean that they are completely free of costs from management fees, transaction expenses (for buying and selling stocks), and other "frictions." Similarly, due to issues of "tracking error," most funds returns will slightly deviate from those of the indexes they seek to track.[5] For these reasons and others[6], the actual returns that you experience from a strategy that invests in a given index fund that tracks an index will tend to deviate by a small amount from the returns of the index itself. For the purposes of 245A, we will generally not consider these issues in depth, but they are important to keep in mind when designing investment strategies in a real professional or personal setting.

Looking up index data on Yahoo Finance: Yahoo finance uses the ^ symbol before a set of letters to designate a "ticker" for an index. For example, ^GSPC is the symbol that Yahoo Finance uses to track the S&P 500 index. You'll notice that if you look for data under this, Yahoo will have information going back to 1950 - 60 years more data than it has on VOO! A partial list of these index tickers can be found here. Some googling around for the name of the index plus "index" and "yahoo finance," and/or searching in the search bar at the top of the Yahoo finance page, may be necessary to track down the proper symbols for other indices that you are interested in. Also, keep in mind that not every exchange traded fund has an accompanying established index for which you can find longer periods of financial data, as you can with the S&P 500. For instance, some ~~stoned~~ bored investment managers recently created an exchange traded fund that tracks a basket of stocks of companies who market their products and services primarily to millennials. You're obviously not going to find an actual index matching this fund with data going back to 1950!

Details and Complications in Historical Index Prices

A complication when dealing with historical information on the returns of various indices is that the companies within them do not stay the same over time. Consider an index that tracks some number of stocks based on a size criteria (e.g. 100 biggest, ranks 2000 to 3000 in size, etc.). As time passes and the size of companies changes, some companies will leave and re-enter the index. Beyond this, however, companies within the index may merge with other companies, split into separate companies, "go private" meaning that all of their stock is bought by private investors, go bankrupt, and many other such events. All of these will have impacts on the returns from investing in the index. But, the details of properly accounting for these events are complex and not always well handled by basic index returns data one finds from a place like Yahoo or even Bloomberg or WRDS. For the purposes of MS&E 245A you can ignore these issues. But, if you are in a real life situation with large amounts of money on the line, these issues are worth looking in to.

[5] Pro tip: In reality, many index funds don't actually hold all of the stocks in an index. Instead, in order to reduce transaction costs of buying and selling small amounts of large numbers of stocks, they will seek to choose a smaller number of stocks within that index, which, when combined in certain proportions, very closely match the performance of the index overall. The difference between the theoretical performance of the real index (which would come from holding all of the stocks in it, but generally will not account for transaction costs of buying and selling those stocks) and the actual performance of a fund that tracks that index, is known as the "tracking error." We will not directly get into discussions of tracking error in 245A, but you may well see it referenced when researching information on index funds.

[6] e.g. tax issues for you, particularly in terms of treatment of dividends, short term and long term capital gains.

Downloading Data from the Web

Finding and downloading stock data from the web is intuitive for anyone experienced with a web browser. The only downside of it is that if you are downloading lots of stock data regularly, going through all of the steps to first find, then save the data, then import it into whatever program (Excel, Matlab, R, etc.) that you are using to do your analysis can get tedious. Nevertheless, web resources for stock data can still be useful, even if you ultimately plan to use a software interface such as one of those described below. This is because the web provides a convenient way to conduct initial screening and research to identify precisely which ticker symbols you want to download. To get historical stock data from Yahoo finance, follow these steps:

1. Go to Yahoo finance at `https://finance.yahoo.com/`
2. Enter the name of the ticker symbol that you are interested in the search bar at the top (e.g. "VOO")
3. Click on the "Historical Data" tab towards the middle of the page.
4. Click on the dates selected in the "Time Period" option to choose which dates to download data for.
    - Note, you can click on the box that says "Max" to download all available data. Even for a long period of time, the size of this file for one or even a few stocks will be quite small.
5. For the "Show" option, select either "Historical Prices" (the default) or "Dividends," depending on what you want.
6. Click the "Apply" button to enact the selections you have made.
7. Click the "Download Data" link to then save the data as a .csv file.

Downloading Data using R

Getting Access to R

R can be downloaded and installed by visiting the following website: `https://cran.r-project.org/`. Simply click on the options at the top to download for Linux, Mac OSX, or Windows, and install the program as you normally would.

Downloading Data Using R

There are numerous R packages that can be used to access financial data in R. For a partial list, see here. One such package that this guide will focus on is called `quantmod`. You can install this package in R (only needs to be done once per computer) using:

```
install.packages("quantmod", dependencies=TRUE, repos = 'http://cran.stat.ucla.edu/')
```

Then, each time you want to use it, just load it using:

```
library(quantmod)
```

Once loaded, you can fetch data for a given ticker symbol from Yahoo's financial data using:

```
VOO = getSymbols("VOO", auto.assign = FALSE)
```

The output will look like this:

```
> VOO
           VOO.Open VOO.High VOO.Low VOO.Close VOO.Volume VOO.Adjusted
2010-09-09 51.25 51.25 50.57 50.66 26500 90.02559
2010-09-10 50.84 50.93 50.65 50.89 8600 90.43432
2010-09-13 51.48 51.57 51.25 51.53 33700 91.57163
2010-09-14 51.42 51.74 51.19 51.52 59400 91.55386
2010-09-15 51.31 51.69 51.20 51.65 9200 91.78488
```

Notes

- Initially, the results of this will be an xts object, which is a type of time series object. If you are more familiar working with, for example, data.frame objects in R, you can convert it by using `data.frame(VOO)`
- If we don't specify any dates, we will get all available daily observations for the stock. We can also, specify certain dates, using:

```
SP500 = getSymbols(
  "^GSPC",
  from = "2011-01-01",
  to = "2012-01-01",
  auto.assign = FALSE)
```

A Note on Dates: The date specified in the `to =` argument of `getSymbols()` is the date such that all data fetched comes *before* that date. Thus, e.g. if you want data for a single month, you would specify `from = "2011-01-01"` and `to = "2012-02-01"` rather than using `to = "2012-01-31"`.

Dividends

To get dividends, we use, e.g.:

```
Divs = getDividends("VOO",
             from = "2011-01-01",
             to = Sys.Date(), ## current date
             src = "yahoo",
             auto.assign = FALSE)
```

Get Weekly Data

To get weekly data using R, just use the `to.weekly()` function:

```
GSPC = getSymbols(
  "^GSPC",
  from = "2011-01-01",
  to = "2014-01-01",
  auto.assign = FALSE)

GSPC = to.weekly(GSPC, indexAt = "startof")
```

Comments

- The process of considering weekly data is simply one of considering one price observation at regular intervals each week. E.g. at the close of the business day each Friday (or Thursday, if Friday is a holiday).
- The `indexAt` argument lets you specify exactly how to select which day of the week is regularly used. If you select the value `"startof"` then the first weekly observation will be the from the date of the first daily observation in your data. This is, for example, what you will want for Project 2.

Extensions

- You can read more about the options for the `getSymbols()` and `getDividends()` functions through the general help interface in R, e.g. `?getSymbols()`
- Full package documentation for quantmod is available here.