THE PROMISE AND PERILS OF DATA: MY JANUARY 2017 DATA UPDATE

Aswath Damodaran

Moneyball, my version!

- In Moneyball, Michael Lewis celebrated the success of Billy Beane, the general manager of the Oakland As, who managed to keep his team in contention for the World Series with a barebones budget and a reliance on numbers.
- At the start of every year, for the last 25, I have played my version of Moneyball with financial data and have shared that data on my website.
- For the last three years, I have also written about what I find in that data that may be of use to me in corporate finance and valuation.

Data Delusions

1. Numbers are precise: Numbers are only as precise as the process that delivers them and in business, that makes them imprecise.
2. Numbers are objective: One of the resentments that number crunchers have about story tellers is that the latter indulge in flights of fancy and are unashamed about bringing their biases into their stories and through them into pricing and investing. Numbers can be just as biased as stories, with the caveat that it is easier to hide biases with numbers.
3. Numbers put you in control: It is human nature to try to be in control and numbers serve us well, in that pursuit.

A Skeptic on Big Data

- Data is not information: Not all data is created equal. Data that is based on what you do is worth a lot more than what you say will do. In addition, as we collect and store more data, it is worth noting that data is not information.
- If everyone has it (data), no one has it: Drawing on a theme that "if everyone has it, no one has it", for data to have value, you have to some degree of exclusivity in access to that data a proprietary edge on processing that data.
- Not all data is actionable: , To convert that data to profits, you need to be able to find a way to monetize whatever data edge you have acquired. For companies that offer products and services, this will take the form of modifying existing products/services or coming up with new products/services to what you have learned from the data.

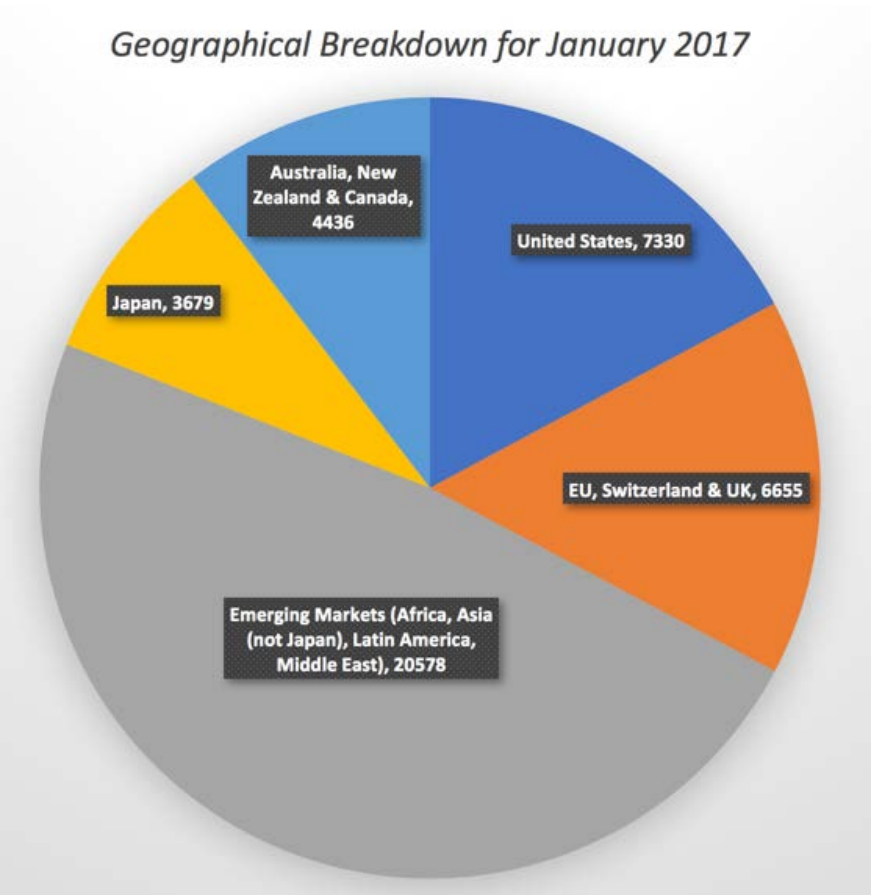
Process Details

1. Data Gathering: Raw data sources
2. Data Classification: Breaking down the data into groupings
3. Data Analysis: Metrics that are measured
4. Data Computation: Computational details can make a difference
5. Data Reporting: Numbers, Picture and Words
6. Data Access: Downloading and Using

1. Raw Data

- Sources: I use S&P Capital IQ, Bloomberg and a host of specialized services (Moody's, PRS etc.).
  - For company-specific data, the only criteria that I use for including a company is that it has to have a non-zero market capitalization, yielding a total of 42678 firms in my sample on January 1, 2017.
- Dating the Data: The data collected is as of January 1, 2017, but there are differences in updating with market and accounting data:
  - With market data (stock prices, market capitalization and interest rates) the numbers are as of the close of trading on December 31, 2016.
  - With accounting data, the data reflects the most recent twelve months (which would be through September 30, 2016 for calendar year companies).

2. Classification



- I break the data down into 94 industry groupings.
- The industry groupings are my own but are based upon S&P Capital IQ and Value Line classifications.

3. Key Numbers

- Reflecting my corporate finance bias, I break the company data down into:
  - Profitability and return measures:
  - Financial Leverage measures:
  - Dividend policy measures:
- Given my valuation focus, I also look at data on
  - Risk premiums (for equity and debt)
  - Risk parameters
  - Cash flow breakdowns (working capital, capital expenditures)
  - Growth rates (historical and sustainable)
- On pricing, I examine earnings, book value and revenue multiples, by industry group and across geographies.

4. Computational Details

- One of the lessons that I have learned from wrestling with the data is that computing even simple statistics requires making choices, which, in turn, can be affected by your biases.
- There are two key statistical issues
  - Sampling biases: As you compute values for some metrics, you will start losing some of your sample and if you lose enough, you have to worry about biased numbers.
  - Outliers: As you compute ratios, you will always get outliers, which, in turn, can skew your averages.
- An example with PE ratios
  - If you compute the average PE ratio across all companies with PE ratios, you lose about half your sample and you will be bedeviled with outliers even with the rest.
  - My solution is to use aggregated values and my estimate of PE ratio for an industry group is
    PE = Aggregate market cap of all companies in group/ Aggregate Net Income of all companies in group

5. Reporting

- Company-level versus Industry grouping: I don't make the company-level data accessible, sensitive to my raw data providers.
- Current data: You will see the data classified into risk, profitability, capital structure and dividend policy measures, reflecting my corporate finance focus, and then into pricing groups (earnings multiples, book value multiples and revenue multiples).
- Archived data: I also keep archived data from prior years (going back to 1999) at this link. Unfortunately, since I have had to switch raw data providers multiple times in the last 20 years, the data is not perfectly comparable over time, as both industry groupings and data measures change over time.

6. Data Access

- See it online: You can see the data online, by clicking on the appropriate link, but only for US data.
- Download excel file: You can download the data by clicking on the downloadable links. You can get the data at the global levels or for any of the five regions (US, Emerging Markets, Developed Europe, Japan and the Rest). You can also get the data for India and China.