DATA UPDATE 1 FOR 2024: THE DATA SPEAKS, BUT WHAT IS IT SAYING?

Respect data, but don’t revere it!

My Data History

- In January 1993, I was valuing a retail company, and I found myself wondering what a reasonable margin was for a firm operating in the retail business.
  - In pursuit of an answer to that question, I used company-specific data from Value Line, one of the earliest entrants into the investment data business, to compute an industry average.
  - The numbers that I computed opened my eyes to how much perspective on the high, low, and typical values, i.e., the distribution of margins, helped in valuing the company, and how little information there was available, at least at that time, on this dimension.
- That year, I computed these industry-level statistics for five variables that I found myself using repeatedly in my valuations, and once I had them, I could not think of a good reason to keep them secret.
- After all, I had no plans on becoming a data service, and making them available to others cost me absolutely nothing. In fact, that year, my sharing was limited to the students in my classes, but in the years following, as the internet became an integral part of our lives, I extended that sharing to anyone who happened to stumble upon my website.

An Annual Data Ritual

- That process has become a start-of-the-year ritual, and as data has become more accessible and my data analysis tools more powerful, those five variables have expanded out to more than two hundred variables, and my reach has extended from the US stocks that Value Line followed to all publicly traded companies across the globe on much more wide-reaching databases.
- Along the way, more people than I ever imagined have found my data of use, and while I still have no desire to be a data service, I have an obligation to be transparent about my data analysis processes.
- I have also developed a practice in the last decade of spending much of January exploring what the data tells us, and does not tell us, about the investing, financing and dividend choices that companies made during the most recent year.
- In this, the first of the data posts for this year, I will describe my data, in terms of geographic spread and industrial breakdown, the variables that I estimate and report on, the choices I make when I analyze data, as well as caveats on best uses and biggest misuses of the data.

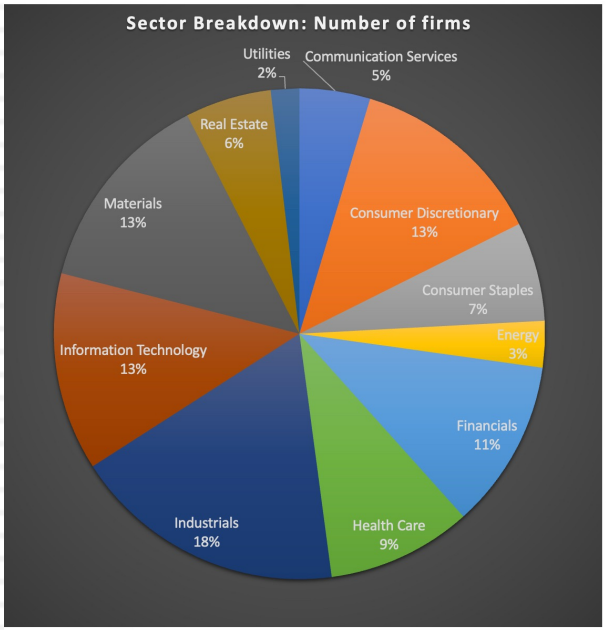
The Sample

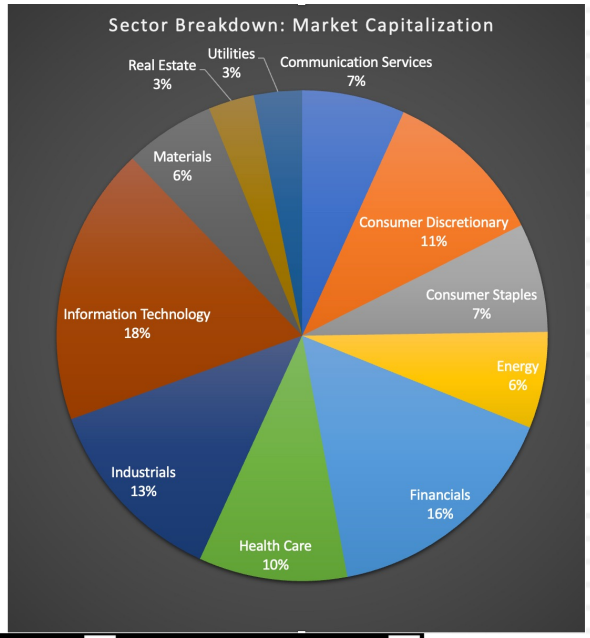
- While there are numerous services, including many free ones, that report data statistics, broken down by geography and industry, many look at only subsamples (companies in the most widely used indices, large market cap companies, only liquid markets), often with sensible rationale – that these companies carry the largest weight in markets or have the most reliable information on them.
- Early in my estimation life, I decided that while this rationale made sense, the sampling, no matter how well intentioned, created sampling bias.
- Thus, looking at only the companies in the S&P 500 may give you more reliable data, with fewer missing observations, but your results will reflect what large market cap companies in any sector or industry do, rather than what is typical for that industry.

Inclusion Criteria

- Since I am lucky enough to have access to databases that carry data on all publicly traded stocks, I choose all publicly traded companies, with a market price that exceeds zero, as my universe, for computing all statistics. In January 2024, that universe had 47,698 companies.
- Geographically, these companies are incorporated in 134 countries, and while you can download the number of companies listed, by country, in a dataset at the end of this post, I break the companies down by region into six broad groupings – United States, Europe (including both EU and non-EU countries, but with a few East European countries excluded), Asia excluding Japan, Japan, Australia & Canada (as a combined group) and Emerging Markets (which include all countries not in the other groupings)/

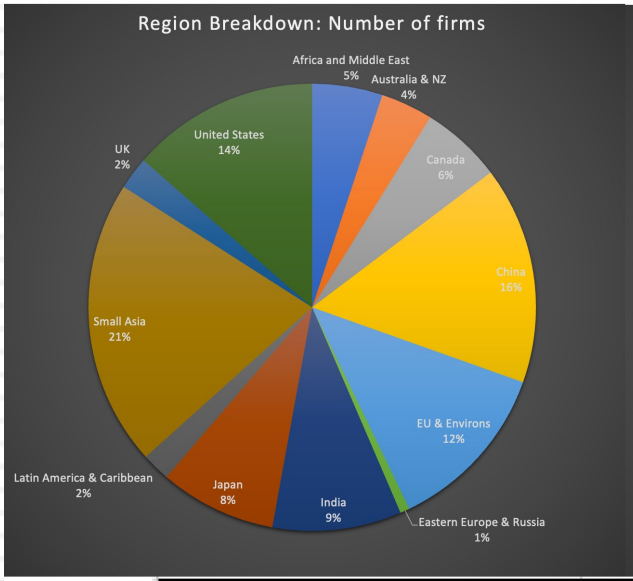
The Data Universe- By Sector

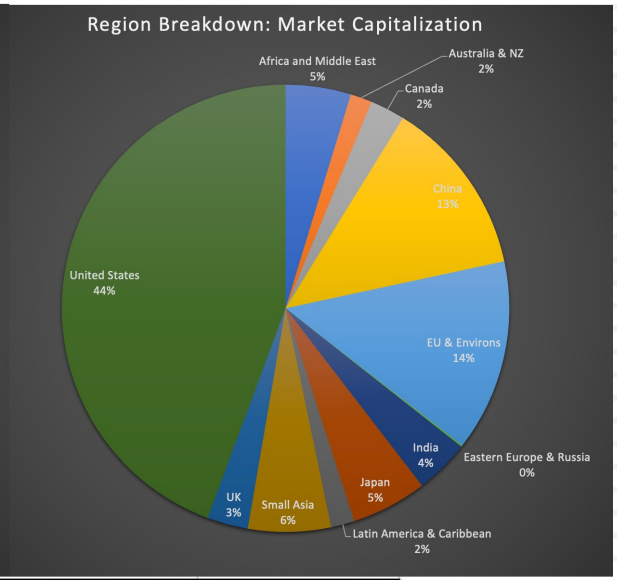




Sector	Number of firms	Market Capitalization
Communication Services	2,218	$7,518,062
Consumer Discretionary	6,185	$11,949,766
Consumer Staples	3,115	$7,975,760
Energy	1,459	$7,044,040
Financials	5,317	$17,673,838
Health Care	4,545	$10,842,708
Industrials	8,590	$14,020,183
Information Technology	6,222	$20,376,739
Materials	6,433	$6,606,719
Real Estate	2,704	$3,343,427
Utilities	893	$3,515,439

The Data Universe- By Region





Region	Number of firm	Market Cap
Africa and Middle East	2,423	$5,225,482
Australia & NZ	1,798	$1,738,485
Canada	2,791	$2,791,564
China	7,504	$14,220,441
EU & Environs	5,925	$15,385,625
Eastern Europe & Russia	325	$117,635
India	4,446	$4,387,483
Japan	4,020	$6,060,760
Latin America & Caribbean	984	$1,882,716
Small Asia	9,876	$6,620,480
UK	1,125	$3,279,251
United States	6,481	$49,157,037

Caveats on Grouping

- I am sure that there are countries or at least one country (your own) that I have miscategorized, I have three points to make, representing a combination of mea culpas and explanations.
  - First, these categorizations were created close to twenty years ago, when I first started looking a global data, and many countries that were emerging markets then have developed into more mature markets now.
  - Second, I use these groupings to compute industry averages, by grouping, as well as global averages, and nothing stops you from using the average of a different grouping in your valuation.
  - Third, the emerging market grouping is now a large and unwieldy one, including most of Asia (other than Japan), Africa, the Middle East, portions of Eastern Europe and Russia and Latin America. Consequently, I do report industry averages for the two fastest growing emerging markets in India and China.

The Variables

- This entire exercise of collecting and analyzing data is a selfish one, insofar as I compute the data variables that I find useful when doing corporate financial analysis, valuation, or investment analysis.
- I also have quirks in how I compute widely used statistics like accounting returns on capital or debt ratios, and I will stay with those quirks, no matter what the accounting rule writers say.
- Thus, I have treated leases as debt in computing debt ratios all through the decades that I have been computing this statistic, even though accounting rules did not do so until 2019, and capitalized R&D, even though accounting has not made that judgment yet.

Corporate Finance - Variables

Corporate Governance & Descriptive
1. Insider, CEO & Institutional holdings
2. Aggregate operating numbers
3. Employee Count & Compensation

Investing Principle		Financing Principle		Dividend Principle	
Hurdle Rate	**Project Returns**	**Financing Mix**	**Financing Type**	**Cash Return**	**Dividends/Buybacks**
1. Beta & Risk	1. Return on Equity	1. Debt Ratios & Fundamentals	1. Debt Details	1. Dividends and Potential Dividends (FCFE)	1. Divdends & Buybacks
2. Equity Risk Premiums	2. Return on (invested) capital	2. Ratings & Spreads	2. Lease Effect	2. Dividend yield & payout	
3. Default Spreads	3. Margins & ROC				
4. Costs of equity & capital	4. Excess Returns on investments				
	5. Market alpha				

Valuation - Variables

	Valuation		Pricing
Growth & Reinvestment	**Profitability**	**Risk**	**Multiples**
1. Historical Growth in Revenues & Earnings	1. Profit Margins	1. Costs of equity & capital	1. Earnings Multiples
2. Fundamental Growth in Equity Earnings	2. Return on Equity	2. Standard Deviation in Equity/Firm Value	2. Book Value Multiples
3. Fundamenal Growth in Operating Earnings	3. Return on Invested Capital		3. Revenue Multiples
4. Long term Reinvestment (Cap Ex & Acquisitons)			4. EBIT & EBITDA multiplse

Macro Data?

- My data is primarily micro-focused, since there are other services that are much better positioned to provide macro data (on inflation, interest rates, exchange rates etc.).
  - My favorite remains the Federal Reserve data site in St. Louis (know as FRED, and one of the great free data resources in the world), but there are a few macro data items that I estimate, primarily because they are not as easily available, or if available, are exposed to estimation choices.
  - I report annual historical returns on asset classes (stocks, bonds, real estate, gold) going back to 1928, mostly because data services seem to focus on individual asset classes and partly because I want to make sure that returns are computed the way I want them to be.
  - I also have implied equity risk premiums (forward-looking and dynamic estimate of what investors are pricing stocks to earn in the future) for the S&P 500 going back annually to 1960 and monthly to 2008, and equity risk premiums for countries.

Industry Groupings

Industry Group	*# firms*	*Market Cap*	*Industry Group*	*# firms*	*Market Cap*	*Industry Group*	*# firms*	*Market Cap*
Advertising	391	$224,201	Farming/Agriculture	430	$398,920	Real Estate (General/Diversified)	336	$349,936
Aerospace/Defense	289	$1,439,195	Financial Svcs. (Non-bank & Insurance)	1,258	$3,582,060	Real Estate (Operations & Services)	748	$605,711
Air Transport	155	$518,545	Food Processing	1,657	$2,011,577	Recreation	326	$273,780
Apparel	1,152	$1,478,223	Food Wholesalers	173	$113,695	Reinsurance	34	$185,259
Auto & Truck	165	$2,136,341	Furn/Home Furnishings	383	$313,506	Restaurant/Dining	394	$923,539
Auto Parts	761	$706,940	Green & Renewable Energy	253	$456,760	Retail (Automotive)	204	$271,424
Bank (Money Center)	607	$7,110,555	Healthcare Products	849	$1,987,975	Retail (Building Supply)	120	$579,260
Banks (Regional)	881	$883,969	Healthcare Support Services	463	$1,391,430	Retail (Distributors)	1,028	$992,388
Beverage (Alcoholic)	220	$1,275,022	Heathcare Information and Technology	443	$1,102,040	Retail (General)	256	$3,134,012
Beverage (Soft)	100	$825,654	Homebuilding	173	$318,721	Retail (Grocery and Food)	201	$583,957
Broadcasting	130	$101,911	Hospitals/Healthcare Facilities	232	$375,074	Retail (REITs)	123	$325,517
Brokerage & Investment Banking	590	$1,069,614	Hotel/Gaming	650	$1,028,281	Retail (Special Lines)	639	$991,985
Building Materials	457	$693,847	Household Products	178	$660,432	Rubber& Tires	89	$136,217
Business & Consumer Services	980	$1,234,236	Information Services	81	$107,794	Semiconductor	647	$4,611,471
Cable TV	48	$322,207	Insurance (General)	202	$837,614	Semiconductor Equip	367	$1,245,563
Chemical (Basic)	888	$954,815	Insurance (Life)	137	$1,087,860	Shipbuilding & Marine	348	$459,132
Chemical (Diversified)	64	$176,649	Insurance (Prop/Cas.)	228	$869,959	Shoe	85	$276,388
Chemical (Specialty)	949	$1,835,821	Investments & Asset Management	1,374	$1,109,737	Software (Entertainment)	317	$3,343,776
Coal & Related Energy	215	$394,489	Machinery	1,492	$1,671,106	Software (Internet)	151	$378,159
Computer Services	1,164	$1,501,051	Metals & Mining	1,815	$1,318,181	Software (System & Application)	1,616	$6,259,238
Computers/Peripherals	338	$4,018,549	Office Equipment & Services	137	$43,298	Steel	718	$846,514
Construction Supplies	453	$553,433	Oil/Gas (Integrated)	36	$4,024,439	Telecom (Wireless)	98	$985,632
Diversified	326	$2,331,777	Oil/Gas (Production and Exploration)	590	$1,086,852	Telecom. Equipment	453	$639,628
Drugs (Biotechnology)	1,259	$1,750,957	Oil/Gas Distribution	174	$501,659	Telecom. Services	288	$1,371,917
Drugs (Pharmaceutical)	1,299	$4,235,230	Oilfield Svcs/Equip.	444	$1,036,602	Tobacco	56	$461,261
Education	260	$104,653	Packaging & Container	164	$122,266	Transportation	814	$1,715,842
Electrical Equipment	1,072	$1,657,484	Paper/Forest Products	266	$177,819	Transportation (Railroads)	49	$670,805
Electronics (Consumer & Office)	127	$240,924	Power	488	$2,270,880	Trucking	106	$175,304
Electronics (General)	1,486	$1,723,080	Precious Metals	853	$452,241	Utility (General)	50	$651,500
Engineering/Construction	1,623	$1,752,426	Publishing & Newspapers	205	$106,764	Utility (Water)	102	$136,300
Entertainment	741	$1,061,653	R.E.I.T.	633	$1,515,764			
Environmental & Waste Services	383	$377,555	Real Estate (Development)	884	$512,953			

Data Timing

- In computing the statistics for each of the variables, I have one overriding objective, which is to make sure that they reflect the most updated data that I have at the time that I compute them, which is usually the first week of January.
- That does lead to what some of you may view as timing contradictions, since any statistic based upon market data (costs of equity and capital, equity risk premiums, risk free rates) is updated to the date that I do the analysis (usually the values at the close of the last trading day of the prior year – Dec 31, 2023, for 2024 numbers), but any statistic that uses accounting numbers (revenues, earnings etc.) will reflect the most recent quarterly accounting filing.
- Thus, when computing my accounting return on equity in January 2024, I will be dividing the earnings from the four quarters ending in September 2023 (trailing twelve month) by the book value of equity at the end of September 2022. Since this is reflecting of what investors in the market have access to at the start of 2024, it fulfils my objective of being the most updated data, notwithstanding the timing mismatch.

Currency and Accounting Issues?

- There are two perils with computing statistics across companies in different markets.
  - The first is differences in accounting standards, and there is little that I can do about that other than point out that these differences have narrowed over time.
  - The other is the presence of multiple currencies, with companies in different countries reporting their financials in different currencies.
- The global database that I use for my raw data, S&P Capital IQ, gives me the option of getting all of the data in US dollars, and that allows for aggregation across global companies. In addition, most of the statistics I report are ratios rather than absolute values, and are thus amenable to averaging across multiple countries.

Statistical Choices

- Missing Data: there are data items where the reporting standards either don't require disclosure in some parts of the world (stock-based compensation) or disclosure is voluntary (employee numbers). When confronted with missing data, I do not throw the entire company out of my sample, but I report the statistics only across companies that report that data.
- Averages, Medians and Aggregates: The simple average, which is the workhorse statistic that most services report for variables, is often a poor measure of what is typical in an industry, either because the variable cannot be computed for many of the companies in the industry, or because, even when computed, it can take on outlier values. Having toyed with alternative approaches, the one that I find offers the best balance is the aggregated ratio. The resulting value uses all of the companies in the sample, reducing sampling bias, and is closer to a weighted average, alleviating the outlier effect. For a few variables, I do report the conventional average and median, just for comparison.

Using the Data

- As I noted earlier, the datasets that I report are designed for my use, in corporate financial analysis and valuation that I do in real time. Thus, I plan to use the 2024 data that you see, when I value companies or do corporate financial analysis during the year, and if you are a practitioner doing something similar, it should work for you. You can find this current data at this link, organized to reflect the categories.
- That said, there are some of you who are not doing your analysis in real time, either because you are in the appraisal business and must value your company as of the start of 2020 or 2021, or a researcher looking at changes over time. I do maintain the archived versions of my datasets for prior years on my webpage, and if you click on the relevant data, you can get the throwback data from prior years.
- There are two uses that my data is put to where you are on your own.
  - The first is in **legal disputes**, where one or both sides of the dispute seem to latch on to data on my site to make their (opposing) cases. While I clearly cannot stop that from happening, please keep me out of those fights, since there is a reason I don't do expert witness of legal appraisal work; courts are the graveyards for good sense in valuation.
  - The other is in **advocacy work**, where data from my site is often selectively used to advance a political or business argument. My dataset on what companies pay as tax rates seems to be a favored destination, and I have seen statistics from it used to advance arguments that US companies pay too much or too little in taxes.
- My datasets do not carry company-specific data, since my raw data providers (fairly) constrain me from sharing that data. Thus, if you want to find the cost of capital for Unilever or a return on capital for Apple, you will not find it on my site, but that data is available online already, or can be computed from the financial releases from these companies.

A Sharing Request

- If you do use the data, you don't have to thank me, or even acknowledge my contribution.
- Use it sensibly, take ownership of your analysis (don't blame my data for your value being too high or low) and pass on knowledge. It is one of the few things that you can share freely and become richer as you share more.
- Also, as with any large data exercise, I am sure that there are mistakes that have found their way into the data, and if you find them, let me know, and I will fix them as quickly as I can!