DATA UPDATE 1 FOR 2026: THE PUSH AND PULL OF NUMBERS!

The Good and Bad Sides of Data!

Numbers, numbers everywhere..

- In my musings on valuation, I have long described myself as more of a number cruncher than a storyteller, but it is because I love numbers for their own sake, rather than a fondness for abstract mathematics.
- It is that love for numbers that has led me at the beginning of each year since the 1990s to take publicly available data on individual companies, both from their financial statements and from the markets that they are listed and traded on, and try to make sense of that data for a variety of reasons - to gain perspective, to use in my corporate financial analysis and valuations and to separate information from disinformation .
- As my access to data has improved, what started as a handful of datasets in my first data update in 1994 has expanded to cover a much wider array of statistics than I had initially envisioned, and my 2026 data updates are now ready. If you are interested in what they contain, please read on.

The (Golden?) Age of Data

- It is the golden age of data, if you look at the amount of data that we can access and the ease with which we can get to that data, but this may be just me, but I feel less at ease than I did forty years ago.
- After a year during which we heard more talk about data and data centers than ever before in history, usually in the context of how AI will change our lives, it is worth considering the draw that data has aways had on not just businesses but on individuals, as well as the dangers with the proliferation of data and the trust we put on that data.

Data's Good Side

1. Signal in the noise: Anyone who has to price/value a stock or assess a project at a firm has to make estimates in the face of contradictions, both in viewpoints and numbers. The entire point of good data analysis is **to find the signals in the noise**, allowing for reasoned judgments, albeit with the recognition that you will make mistakes.
2. Coping mechanism for uncertainty: Investors and businesses, when faced with uncertainty, often respond in unhealthy ways, with denial and paralysis as common responses. Data can help in two ways, first by helping you **picture the range of possible outcomes** and second by bringing **in tools (simulations, data visualizations) for incorporating uncertainty** into your decision-making.
3. Prescription against tunnel vision: It is **easy to get bogged down in details**, when faced with having to make investment decisions, and lose perspective. One of the advantages of looking at data differences over time and across firms is that it can help you elevate and regain perspective.
4. Shield from disinformation: I find that people make up stuff and present it as fact, and it is getting worse over time. While it is easy to blame social media, which has provided a megaphone for these fabulists, I read and hear statements in the media, ostensibly from experts, politicians and regulators, that cause me to do double takes since they are not just wrong, but easily provable as wrong, with the data.

Data's Bad Side

1. False precision: It is undeniable that attaching a number to something that worries you, whether it be your health or your finances, can provide a sense of comfort, but there is the danger with treating estimates as facts.
2. The Role of Bias: I have long argued that we are all biased, albeit in varying degrees and in different directions, and that bias will find its way into the choices we make. With data, this can play out consciously, where we use data estimates that feed into our biases and avoid estimates that work in the opposite direction, but more dangerously, they can also play out subconsciously, in the choices we make.
3. Lazy mean reversion: In a series of posts that I wrote about value investing, at least as practiced by many of its old-time practitioners, I argued that it was built around mean reversion, the assumption that the world (and markets) will revert to historic norms.
4. The data did it: As we put data on a pedestal, treating the numbers from emerge from it as the truth, there is also the danger that some analysts who use it view themselves as purely data engineers. While they make recommendations based upon the data, they also refuse to take ownership for their own prescriptions, arguing that it is the data that is responsible.

The Data Explosion: The Net Effet

- As the data that we collect and have access to gets richer and deeper, and the tools that we have to analyze that data become more powerful, there are some who see a utopian world where this data access and analysis leads to better decisions and policy as output.
- Many analysts now complain that they have too much data, not too little, and struggle with data overload. At the same time, a version of Gresham's law seems to be kicking in, where bad data (or misinformation) often drives out good data, leading to worse decisions and policy choices.
- My advice, gingerly offered, is that as you access data, it is caveat emptor, and that you should do the following with any data (including my own):
  1. Consider the biases and priors of the data provider.
  2. Not use data that comes from black boxes, where providers refuse to detail how they arrived at numbers.
  3. Crosscheck with alternate data providers, for consistency.

Data Coverage: Raw Data

- In 1993, when I did my first estimates, I subscribed to Value Line and **bought their company-specific data**, which about 2000 US companies and included a subset of items on financial statements, on a compact disc.
- In 2025, **my access to data has widened**, especially because my NYU affiliation gives me access S&P Capital IQ and a Bloomberg terminal, which I supplement with subscriptions (mostly free) to online data.
- It is worth noting that these **almost all the data from these providers is in the public domain**, either in the form of company filings for disclosure or in government macroeconomic data, and the primary benefit (and it is a big one) is easy access.

The Data 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	3. Tax rates			
4. Costs of equity & capital	4. Excess Returns on investments				
	5. Market alpha				

	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
5. R&D			
6. Working capital needs			

Consistency and Transparency

- In computing these variables, I have tried to stay consistent with my own thinking and teaching and transparent about my usage.
  - As an illustration for consistency, I have argued for three decades that **lease commitments should be treated as debt** and that **R&D expenditures are capital, not operating, expenses**, and my calculations have always reflected those views, even if they were at odds with the accounting rules.
  - In 2019, the **accounting rules caught up with my views on lease debt**, and while the numbers that I report on debt ratios and invested capital are now closer to the accounting numbers, I continue to do my own computations of lease debt and report on divergences with accounting estimates.
  - **With R&D, I remain at odds with accountants**, and I report on the affected numbers (like margins and accounting return) with and without my adjustments.
- On the transparency front, you can find the details of how I computed each variable at this link, and it is entirely possible that you may not agree with my computation, it is in the open.

Estimation Details

- With the micro data, I report on industry values rather than on individual companies, for two reasons. The first is that my **raw data providers are understandably protective of their company-level data** and will have a dim view of my entry into that space. The second is that if you want company-level data for an individual company or even a subset, **that data is, for the most part, already available in the financial filings of the company**.
- For global statistics, **where companies in different countries are included within each industry, and report their financials in different currencies**, I download the **data converted into US dollars**. Thus, numbers that are in absolute value (like total market capitalization) are in US dollars, but most of the statistics that I report are **ratios or fractions**, where currency is not an issue, at least for measurement.
- While computing industry averages may seem like a trivial computational challenge, there are two problems you face in large datasets of diverse companies.
  - The first is that there will be individual companies where the data is missing or not available, as is the case with PE ratios for companies with negative earnings.
  - The second is that the companies within a group can vary in size with very small and large companies in the mix.
  - My solution, and you may not like it, it to compute aggregated values of variable, and use these aggregated values to compute the representative statistics.

My geographical breakdown

Grouping	*Includes*	*Number of firms*
Broad Groupings		
United States	Publicly traded companies that are incorporated and headquartered in the United States, though that list includes many companies that do some or even much of their business elsewhere in the world.	5994
Global	Publicly traded companies that are incorporated and headquartered anywhere in the world.	48156
Europe	Publicly traded companies that are incorporated and headquartered in the EU, the UK, Switzerland and Norway.	
Japan	Publicly traded companies that are incorporated and headquartered in Japan.	3965
Emerging Markets	Publicly traded companies in Asia, Latin America, Africa, portions of Europe and Russia.	27360
Rest	Publicly traded companies that are incorporated and headquartered in Australia, New Zealand and Canada.	4458
Emerging market subgroups		
India	Publicly traded companies that are incorporated and headquartered in India	5170
China	Publicly traded companies that are incorporated and headquartered in China and Hong Kong.	7775

My Data Universe

- In the first part of this post, I noted how bias can skew data analysis, and **one of the biggest sources of bias is sampling**, where you pick a subset of companies and draw the wrong conclusions about companies.
- To reduce this sampling bias, I **include all publicly traded companies that have a market capitalization that exceeds zero in my sample**, yielding a total sample size of 48,156 companies in my data universe.
- Note that there will be some sampling bias still left insofar as **unlisted and privately owned businesses are not included**, but since disclosure requirements for these businesses are much spottier, it is unlikely that we will have datasets that include these ignored companies in the sample in the near future.

The World in 2025, by Sub-Region…

Sub Region	Number of firms	Market Cap (12/31/25	% of Global	Market Cap (12/31/24)	% of Global	Change in Market Cap in 2025	% Change in 2025
Africa and Middle East	2555	$5,518,209	3.70%	$5,229,701	4.26%	$288,508	5.52%
Australia & NZ	1702	$1,976,650	1.33%	$1,674,438	1.36%	$302,212	18.05%
Canada	2576	$3,853,912	2.58%	$2,867,174	2.34%	$986,739	34.42%
China	7775	$20,574,298	13.80%	$15,532,353	12.65%	$5,041,945	32.46%
Eastern Europe & Russia	439	$157,119	0.11%	$110,329	0.09%	$46,791	42.41%
EU & Environs	5649	$19,081,368	12.79%	$14,833,102	12.08%	$4,248,266	28.64%
India	5170	$5,334,871	3.58%	$5,164,032	4.21%	$170,839	3.31%
Japan	3965	$7,406,131	4.97%	$6,113,618	4.98%	$1,292,513	21.14%
Latin America & Caribbean	980	$2,162,227	1.45%	$1,583,781	1.29%	$578,446	36.52%
Small Asia	10367	$9,273,403	6.22%	$6,719,876	5.47%	$2,553,527	38.00%
UK	984	$4,027,733	2.70%	$3,209,469	2.61%	$818,264	25.50%
United States	5994	$69,772,850	46.78%	$59,752,305	48.66%	$10,020,545	16.77%
Grand Total	**48156**	**$149,138,771**	**100.00%**	**$122,790,179**	**100.00%**	**$26,348,592**	**21.46%**

And by Sector…

Sub Region	*Number of firms*	*Market Cap (12/31/25*	*% of Market*	*Market Cap (12/31/24)*	*% of Global*	*Change in Market Cap in 2025*	*% Change in 2025*
Communication Services	2193	$11,882,165.61	7.97%	$9,264,536	7.55%	$2,617,630	28.25%
Consumer Discretionary	6358	$14,969,165.69	10.04%	$13,606,918	11.08%	$1,362,248	10.01%
Consumer Staples	3202	$8,572,309.48	5.75%	$7,903,381	6.44%	$668,928	8.46%
Energy	1402	$6,980,639.24	4.68%	$6,647,400	5.41%	$333,240	5.01%
Financials	5278	$26,118,764.85	17.51%	$20,975,019	17.08%	$5,143,746	24.52%
Health Care	4452	$12,408,326.67	8.32%	$10,680,074	8.70%	$1,728,252	16.18%
Industrials	9007	$19,031,983.73	12.76%	$15,225,027	12.40%	$3,806,957	25.00%
Information Technology	6234	$32,800,233.00	21.99%	$25,454,059	20.73%	$7,346,174	28.86%
Materials	6457	$8,307,099.64	5.57%	$6,032,005	4.91%	$2,275,095	37.72%
Real Estate	2671	$3,671,712.23	2.46%	$3,288,702	2.68%	$383,010	11.65%
Utilities	902	$4,396,371.32	2.95%	$3,713,059	3.02%	$683,312	18.40%
Grand Total	**48156**	**$149,138,771**	**100.00%**	**$122,790,179**	**100.00%**	**$26,348,592**	**21.46%**

Data Usage

1. For practitioners, not academic researchers: The data that I report is **for practitioners in corporate finance, investing and valuation, rather than academic researchers**. Thus, all of the data is on the current data link is data as of the start of January 2026 and can be used in assessments and analysis today. If you are doctoral student or researcher, you will be better served going to the raw data or having access to a full data service, but if you lack that access, and want to download and use my industry averages over time, you can use the archived data that I have.
2. Starting point, not ending point: If you do decide to use any of my data, please do recognize that it is the **starting point for your analysis, not a magic bullet**. Thus, if you are pricing a steel company in Thailand, you can start with the EV/EBITDA multiple that I report for emerging market steel companies, but you should adjust that multiple for the characteristics of the company being analyzed.
3. Take ownership: If you do use my data, whether it be on equity risk premiums or pricing ratios, you should understand how I compute these numbers (from my classes or writing) and **take ownership of the resulting analysis**.

The Bots are coming…

- In 2024, I talked about **the Damodaran Bot**, an AI entity that had read or watched everything that I have put online (classes, books, writing, spreadsheets) and talked about what I could do to stay ahead of its reach.
- I argued that AI bots will not only match, but be better than I am, at **mechanical and rule-based tasks**, and that my best pathways to creating a differential advantage was in finding aspects of my work that required multi-disciplinary (numbers plus narrative) and generalist thinking, with intuition and imagination playing a key role.
- As I looked at the process that I went through to put my datasets together, I realized that there was no aspect of it that a bot cannot do, and I plan to work on involving my bot more in my data update next year, with the end game of having it take over almost the entire process.

A More General Message…

- I do think that there is a message here for businesses that are built around collecting and processing data, and charging high prices for that service.
- Unless they can find other differentials, they are exposed to disruption, with AI doing much of what they do.
- More generally, to the extent that a great deal of quant investing has been built around smart numbers people working with large datasets to eke out excess returns, it will become more challenging, not less so, with AI in the mix.

Data Links

- Webpage: Damodaran.com
- Current data: https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datacurrent.html
- Archived data: https://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html
- Data variable definitions: https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/variable.htm