

For Good Measure Betting on Growth versus Magnitude

DAN GEER



Dan Geer is the CISO for In-Q-Tel and a security researcher with a quantitative bent. He has a long history with the USENIX Association, including officer positions, program committees, etc. dan@geer.org

Whether you are Werner Heisenberg or Janet Yellen, your field of study includes measurements of position and velocity and how they interact. Poor measurements may be unable to nail down either the one or the other (much less both), but even if only the one is measured, there is likely to be some prediction that you will be able to make. In hard-to-measure situations, consistency of error distribution can be your friend—consistent errors help you to find the message in the body of noisy data. The reader probably knows all that.

Let's look at some data where position and velocity have been charted for enough years to get a feel for what is going on, and then we'll discuss what use we can make of it.

The (US) Federal Trade Commission has a program known as Sentinel [1]. Quoting its introduction,

Sentinel is the unique investigative cyber tool that provides members of the Consumer Sentinel Network with access to millions of consumer complaints. Consumer Sentinel includes complaints about:

- Identity Theft
- Do-Not-Call Registry violations
- Computers, the Internet, and Online Auctions
- Telemarketing Scams
- Advance-fee Loans and Credit Scams
- Immigration Services
- Sweepstakes, Lotteries, and Prizes
- Business Opportunities and Work-at-Home Schemes
- Health and Weight Loss Products
- Debt Collection, Credit Reports, and Financial Matters

Consumer Sentinel is based on the premise that sharing information can make law enforcement even more effective. To that end, the Consumer Sentinel Network provides law enforcement members with access to complaints provided directly to the Federal Trade Commission by consumers, as well as providing members with access to complaints shared by data contributors.

What Sentinel then produces are interpretive text reports backed by spreadsheets, all freely available for aggregate data. Querying the backing database for individual complaints is limited to law enforcement with a need to know.

For Good Measure: Betting on Growth versus Magnitude

Because the Sentinel data files are categorized enumerations of consumer complaints, there will be some errors. Let us assume that the errors are relatively constant over time, which is to say that trendlines and rank orderings are going to be instructive even if, say, there is persistent undercounting. The categories of consumer complaints for which we have at least six years of data are

- ◆ Advance Payments for Credit Services
- ◆ Auto-Related Complaints
- ◆ Banks & Lenders
- ◆ Business & Job Opportunities
- ◆ Buyers' Clubs
- ◆ Charitable Solicitations
- ◆ Computer Equipment & Software
- ◆ Credit Bureaus, Information Furnishers & Report Users
- ◆ Credit Cards
- ◆ Debt Collection
- ◆ Education
- ◆ Foreign Money Offers & Counterfeit Check Scams
- ◆ Grants
- ◆ Health Care
- ◆ Home Repair, Improvement & Products
- ◆ Identity Theft
- ◆ Impostor Scams
- ◆ Internet Auction
- ◆ Internet Services
- ◆ Investment-Related Complaints
- ◆ Magazines & Books
- ◆ Mortgage Foreclosure Relief & Debt Management
- ◆ Office Supplies & Services
- ◆ Prizes, Sweepstakes & Lotteries
- ◆ Real Estate
- ◆ Shop-at-Home & Catalog Sales
- ◆ Tax Preparers
- ◆ Telephone & Mobile Services
- ◆ Television & Electronic Media
- ◆ Travel, Vacations & Timeshare Plans

As of the time of writing, the most recent Sentinel data set available is for 2014. Let's think of the number of complaints as "position" and the rate of growth as "velocity." Converting both position and velocity into rank order and graphing them in the typical high/low quadrant style, we see considerable spread in Figure 1.

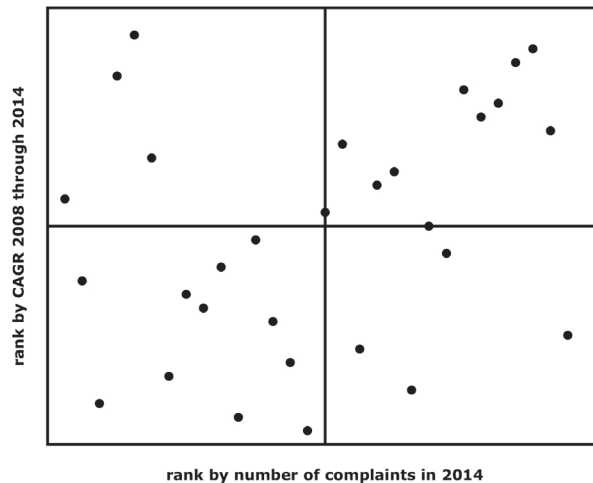


Figure 1: Size versus trajectory spread using the data from Sentinel categories with at least six years of data

The smallest and slowest growing are in the lower left quadrant, the biggest and fastest growing are in the upper right quadrant, etc.; you've seen this kind of graph before. The dot closest to the lower left corner is that for Real Estate, ranked as third smallest both in numbers of complaints (4,952) and in compound annual growth rate (or CAGR, which is actually declining at -5%). The dot closest to the upper right corner is that for Impostor Scams, ranked third highest in numbers of complaints (276,622) and first highest in growth rate of complaints (CAGR of 182%). If you were the person in charge, it is pretty clear that you'd put more manpower into impostor scams than into real estate fraud. That's an easy call.

On the other hand, the dot farthest to the right in the lower right quadrant is Identity Theft, number one in total complaints (332,646 in 2014 meaning one every 20 seconds) but with the eighth-slowest rate of growth (CAGR of not quite 1%). Similarly, the dot closest to the boundary of the upper left quadrant is Tax Preparers, fifth smallest in total complaints (6,418) but with the highest growth rate of all (CAGR of 292%). For this pair of Identity Theft versus Tax Preparers, which one is more deserving of investment?

Some readers will look at that graph and ask, "Is there any correlation here?" No, there isn't—Spearman's $\rho = 0.280$, meaning there is nothing worth talking about, as you can see: one-third of all categories are well off the diagonal, i.e., categories come and go, which is no surprise when you have sentient opponents. Again, where would you put your money when you are in charge?

The three most common complaints are Identity Theft (17.7%), Debt Collection (15.0%), and Impostor Scams (14.8%), which together comprise almost half of all complaints. The three fastest growing are Tax Preparers (291% CAGR), Impostor Scams

For Good Measure: Betting on Growth versus Magnitude

(182% CAGR), and Telephone and Mobile Services (47% CAGR). Is it position or velocity—count or CAGR—that tells you where to put your money?

Suppose the current observed compound annual growth rates are sustained until 2020. By then, the three most common complaints would be Impostor Scams #1, Telephone and Mobile Services #2, and Debt Collection #3. Identity Theft would have fallen to #6 (behind Banks and Lenders #4 and Auto-Related Complaints #5). If you imagine that any serious countermeasure takes, say, five years to actually work, then should we be spending our money now on the predicted top three in 2020? Or does the sentient opponent make a five-year plan an exercise as useless as the Five Year Plans that many national bureaucracies so love?

We are not alone in facing this kind of problem, namely, do you spend your money (or other scarce resource) to solve problems you actually have now or to stave off problems you are going to have later? As the old saying goes, a stitch in time saves nine, so prevention is likely the better buy, but it is never the easy sell... When there is not enough vaccine to go around, do you vaccinate those most likely to soon sicken and die (children, perhaps) or those most likely to soon become transmission vectors (clinic workers, perhaps)? In the Middle East, Western governments are invited to choose between stability and justice. If you choose stability, then you must reinforce dictators' grip on power, regardless of how they treat their people. This was the West's policy during the Cold War—and it is Vladimir Putin's policy today. If, however, you choose justice, you must side with the crowds trying to throw off their rulers, even if this triggers the collapse of order [2].

I've written over and over on this problem from every different angle, including the disastrous practice of vendors abandoning code bases they don't want to support yet simultaneously refusing to open source the code they are abandoning. Or how many platforms in common use are provisioned with software that its maker can no longer build? Or the longer a deployed device stays deployed, the more likely it is that it cannot be found and the more certain it is that it cannot be updated if found; should we be putting money into having a, say, 20-year guarantee for updatability of autonomous devices with network connectivity, or is the embedding of sensors in damned near everything already past the point where such decisions are even relevant?

Nassim Taleb (in *The Black Swan*, for example) argues that when a distribution is fat-tailed, estimations of parameters based on historical experience will inevitably mislead, which means

[we are] undergoing a switch between [continuous low grade volatility] to...the process moving by jumps, with less and less variations outside of jumps. [3]

As I ponder that, I am more inclined to put my money on identifying, as best I can, problems that will grow than on problems that have grown. Easy for me to say, but killing dragons in their cribs beats dealing with them later on and, by and large, I can avoid the dragons that are already full size by just not doing the things that make me look like lunch. If you study the full Sentinel reports, you'll see what the demographics that spell "lunch" look like, such as the order of magnitude greater rate of identity theft in Miami (340.4 per 100,000 population per year) than in Bismarck, ND (27.9 per 100,000 population per year), or how demographics predict whether it is one's government benefits or one's credit cards that you are most likely to lose to an identity thief.

References

- [1] <https://www.ftc.gov/enforcement/consumer-sentinel-network>; <https://www.ftc.gov/enforcement/consumer-sentinel-network/reports>.
- [2] <http://www.telegraph.co.uk/news/worldnews/middleeast/12046082/Tony-Blair-has-learnt-important-lessons-from-Iraq-Its-a-shame-no-one-wants-to-listen.html>.
- [3] N. N. Taleb, "On the Super-Additivity and Estimation Biases of Quantile Contributions": www.fooledbyrandomness.com/longpeace.pdf.



Become a USENIX Supporter and Reach Your Target Audience

The USENIX Association welcomes industrial sponsorship and offers custom packages to help you promote your organization, programs, and products to our membership and conference attendees.

Whether you are interested in sales, recruiting top talent, or branding to a highly targeted audience, we offer key outreach for our sponsors. To learn more about becoming a USENIX Supporter, as well as our multiple conference sponsorship packages, please contact sponsorship@usenix.org.

Your support of the USENIX Association furthers our goal of fostering technical excellence and innovation in neutral forums. Sponsorship of USENIX keeps our conferences affordable for all and supports scholarships for students, equal representation of women and minorities in the computing research community, and the development of open source technology.

Learn more at:
www.usenix.org/supporter