



Alternative Data: Don't bark up the wrong tree (too early)!

By Ganesh Mani, Ph.D.

On a Monday in late April 2019, a boutique research firm—using publicly available corporate aircraft and flight data—reported that a Gulfstream V jet belonging to Occidental Petroleum had been sighted the day before in Omaha, Nebraska, the hometown of famed investor Warren Buffett. A day later, Buffett's Berkshire Hathaway announced its intention to invest \$10 billion in a preferred stake in Houston-based Occidental, a move that ultimately proved critical in helping Occidental trump Chevron in a takeover battle for Anadarko Petroleum. More immediately, however, Berkshire's official announcement of its involvement amplified price movements and trading volume in Occidental and Chevron shares relative to the prior day—when information about the jet data was known to just a handful of eagle-eyed investment sleuths.

Aircraft and flight logs are examples of "alternative data"—non-reportable information that some investors increasingly are embracing to gain an edge. Other examples include private opinion polls before a key political referendum; satellite and drone imagery of ports and mall parking lots; credit-card transactions; job postings; patent and trademark filings; and social-media sentiment, including ratings and reviews of products and services. Unlike financial and other data reportable to market regulators, which is readily available and widely disseminated, alternative data is often proprietary, affording its collectors or buyers information that other investors do not have.

Yet, the leap from today's alternative data to a future investment consensus is neither immediate nor guaranteed. Data vintage, authority, and provenance matter. Any changes in the data generation or collection process also can cause problems for those relying on the information to make investment decisions.

Just ask Bloomberg, which constructed and published a Tesla Model 3 production volume estimator, based on Vehicle Identification Numbers (VINs) from official U.S. government sources, social media reports, and input from owners of Tesla's electric cars. The Model 3 Tracker launched in February 2018. For most of 2018, as the market focused on Tesla's production ramp-up for the Model 3, Bloomberg's alternative-data-based estimate proved a good predictor of the company's quarterly output. Model 3 production numbers were a key driver of Tesla's stock price in late 2018 as well as in 2019's first quarter.

In early 2019, however, a rise in Tesla exports led to changes in how many extra VINs the company registered

pre-production to keep on hand a buffer of the numbers. These data - generation and collection - changes caused the Bloomberg tracker to veer off course, overestimating production numbers for the first quarter of 2019 by 26%. Tesla stock sold off early April, when the actual production and delivery numbers were announced, catching many investors relying on the Bloomberg tracker off guard.

To cite another example, the 2016 US Presidential Election was a significant black eye for election forecasters. Relying on alternative data, generated via opinion polls, almost all the pundits had predicted Hillary Clinton to be the victor. Similarly, the Brexit vote during June of 2016 and the relatively recent election of Scott Morrison as the Australian Prime Minister are additional instances where alternative data-informed methods led analysts astray. Investors across many asset classes - ranging from currencies to equities - were stunned by the volatility resulting from the surprise election results.

Artificial intelligence has recently emerged as a premium refiner of alternative data, especially to deal with surprises arising from missing, incomplete, and uncertain values; and, to normalize for transient effects, such as traffic-flow numbers affected by inclement weather or construction.

However, maximizing the utility of alternative data likely requires a hybrid analysis, combining humans and machines. Computers are able to process large amounts of data rapidly, while humans are able to apply judgment and empathy in ambiguous situations, often guided by past experiences, some from unrelated domains. Just as a pilot must interpret conflicting readings or flight-deck alerts in the cockpit, an analyst should use alternative data in concert with other information to make better choices or mitigate risk.

There are two significant caveats to keep in mind while employing alternative data. The first is that the data needs to imply a relatively direct connection to asset prices - a potential takeover usually implies a higher price for the company being acquired, an uptick in implied product demand points to increasing revenues for the manufacturing company in the near term; and, a candidate's improved election forecast may point to imminent implementation of previously-telegraphed policy changes. The second caveat is that often there is a significant delay between what is hinted by the data and a majority of the market participants acting on or realizing it. Consensus view or stock sentiment may not change overnight, and an asset can continue to remain mispriced. If the adjustment to pricing of a security takes a long time to materialize, it erodes

the internal rate of return (IRR), turning a potentially good investment decision into a mediocre one. Caveat (alternative data) emporium!



**FDP Advisory Board
Ganesh Mani, Ph.D.
Adjunct Faculty
Carnegie Mellon**

Ganesh Mani is on the adjunct Faculty of Carnegie Mellon University and is considered a thought leader in the areas of

investment management and AI / FinTech. He has been a pioneer in applying innovative techniques to multiple asset classes and has worked with large asset management firms incl. hedge funds, after having sold one of the earliest AI/ML-based investment management boutiques into SSgA (in the late nineties), nucleating the Advanced Research Center there.

Ganesh has been featured on ABC Nightline and in a Barron's cover story titled "New Brains on the Block". Mr. Mani has an MBA in Finance and a PhD in Artificial Intelligence from the University of Wisconsin-Madison, as well as an undergraduate degree in Computer Science from the Indian Institute of Technology, Bombay. Ganesh is a charter member of TIE (www.TIE.org), was an early member of the Chicago Quantitative alliance and is on the advisory board of the Journal of Financial Data Science.

The Financial Data Professional Institute (FDPI) was established by the CAIA Association to address the growing need in finance for a workforce that has the skills to perform in a digitized world where an increasing number of decisions will be data and analytics driven. The FDP curriculum introduces candidates to central concepts of machine learning and big data, including ethical and privacy issues, and their roles in various segments of the financial industry to boost and integrate quant knowledge into analytics' skills.