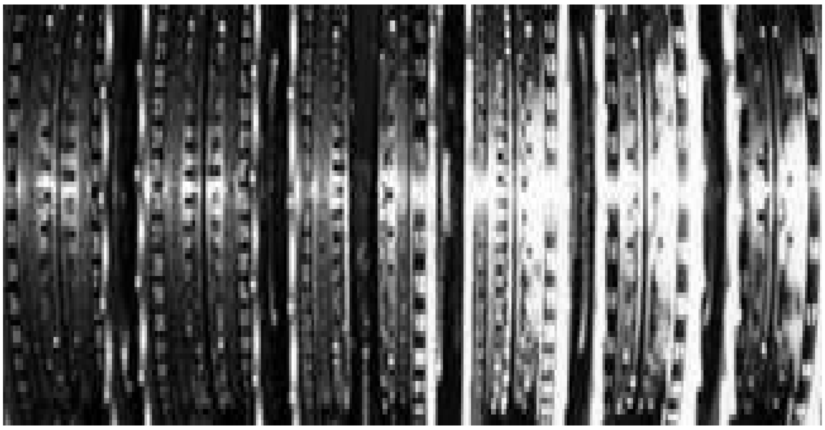
Big Data Brother: Power of Trading in the Hands of a Few

Tommi A. Vuorenmaa, Ph.D.

September 7, 2015



High speed automated news-based trading is popular while its more profound relation to society is not. Like in cryptography, structure and data context are essential to making correct decisions. The 23 paragraphs of this article each hold about the same amount of information to reflect this basic premise; to be exact, 111 words, with on average of 686 characters and 5488 bits, best illustrated by a crypto-device with 23 wheels and 111 positions (both prime numbers) in each of the wheels.

But artificial intelligence required to interpret news data need to be a whole lot smarter than in cryptography. As data turn bigger and less structured, and as trading machines become faster and smarter, financial markets approach the ideological goal of extreme efficiency where expectations of future are openly and precisely reflected in the current prices – and yet this same democracy allows power to concentrate in the hands of a few: the control of information in a free capitalistic society gradually slips into the hands of a Big Data Brother who has the capacity to exert control over financial markets and other dimensions of society in the fashion of the Orwellian dystopia. Its implications are dark: instead of extremely efficient markets, the markets are only falsely efficient. Capitalism does not guarantee democracy of a high degree and may fall in the trap of communism.

Article also available at http://versustakes.co (direct link). Keywords: automated trading, big data, bug data, capitalism, communism, entropy, hack data, high-frequency trading, news-based trading.

[main.txt]

1. “All animals are equal, but some animals are more equal than others,” wrote George Orwell in the 1940s when the grim Second World War handicapped Europe. In the 1990s, when I was an economics master's student, I joked after one finance conference dinner in Barcelona that in a stark contrast to a capitalistic society, in a communistic society stock market prices would be preset to fixed flawless trajectories: no news of any significance would ever appear for any company. Stock markets would be perfectly – albeit falsely – efficient and always reflect the “true” company value. There would exist an omnipotent Big Brother controlling news of any kind and their intrinsic value.
[words 111; characters 681; bits 5448]

2. Orwell's “animals” are like news pieces. In communism, all news – if they would exist – would be of the same exact value in economic terms: thus the quote “All animals are equal.” In that case, news would only serve to solidify the structure of that society, which is, metaphorically, as strict as the structure of this article. In capitalism, on the other hand, animals may come in different sizes and follow only a general statistical distribution of arrival times. Such animals have the potential to affect company valuations and economy unexpectedly. Their effect is highly context dependent. This fundamental democracy (read: randomness) makes up a large portion of the lure of capitalism.
[111; 694; 5552]

3. News is, by definition, new information. News consists of something of added value that is not known a priori. But should news have a different intrinsic value in absolute terms, or should they rather be considered in relative terms? If news pieces would be considered to be of equal value, only their rate would matter: the more news pieces there exist, the more the company value might change. This would require us to specify only one stochastic process (statistical distribution) for arrival times. In reality, the value of information varies, which is challenging for automated traders. There are numerous reasons why decoding information is hard. We tackle the most important next.
[111; 687; 5496]

4. John M. Keynes's description of stock markets being analogous to a beauty contest stresses the value of market expectations: "we devote our intelligences to anticipating what average opinion expects the average opinion to be." If we would know what information the other referees (traders) possess, and how the news should be interpreted, or more precisely, what is the absolute value of news to a company whose market capitalization is known, then the relative price change is forecastable. Thus, should we know what is discounted in the price, then indeed "all animals are equal." It is not the effect of news that is the key unknown; the key unknown is market expectation.
[111; 675; 5400]

5. Consider an earnings announcement of a large-cap company. Since most sophisticated investors are well aware of the market expectation of that news before its release, the price reaction is highly dependent on it satisfying the market expectation rather than the news being positive or negative. Similarly, hard-core Nassim Taleb aficionados may also recall his character Fat Tony who gets rich (and fat) by betting against the consensus about the start of Iraq's war: oil prices were expected to head north immediately as the United States would head to Kuwait. The war started as expected, but oil prices headed south; a scenario that is not entirely irrelevant in the present turbulent market.
[111; 695; 5560]

6. There is yet another reason why I call news animals, and it is also related to Orwell: news affects the human mind. Animal spirits, a term typically attributed to Keynes, derives from the Latin term "spiritus animales." If properly captured, these "spirits of minds" may be used in building more intelligent trading strategies. One way to calculate sentiment indices is to use the flow of company news. A simple sentiment index is derivable using what is called news polarity: the difference of positive and negative news standardized properly. But only sentiment indices with a proper news weighting scheme account for the Orwellian fact that "Some animals are more equal than others."
[111; 686; 5488]

7. Automated news reading, content interpretation, and reaction is a deceivingly simple concept. The risks are rising on many fronts. First of all, as news pieces become widely automatically released, and their handling faster and more accurate, many trading strategies lose ground much faster than they did in the past. One could of course enhance their performance by making more frequent updates on technology, but it would just lead to an arms race. A more sustainable and cheaper way is to separate the trading logic from the common one-shot strategies by, for example, making a series of wiser decisions after the news hit the wire. But this may not be enough either.
[111; 670; 5360]

8. If one cannot react faster than the others to even simply interpretable structured news, the trading strategy is probably doomed or unappealing due to its risk profile. Assume that a trading strategy reacts correctly to news, but with a significant time delay. This strategy may still get into a favorable position, but it requires the price to trend longer in the wanted direction. It consequently loses money in higher transaction costs and excessive inventory risk. The position may still often turn in-the-money, but the faster guys show a healthier risk-return profile than the slower guy. Trading is largely about game theoretic decisions: how to best account for the actions of others.
[111; 692; 5536]

9. One may legitimately argue – citing the Keynes' beauty contest idea – that even with less speed the trading strategy may do well if it carefully considers the competitors' actions. Statistically, we may predict how prices react after the news arrival, but irrespectively of successful forecasting, the strategy is not as efficient as the one with the speed advantage and lower risk. Without enough speed, the most attractive source of profits sit in the time before the news. This may motivate accurate techniques for company evaluation, but it may also create a motive for Hack(ing) Data, which are Big Data that are "trolled," tampered with, or hacked deliberately to affect valuations.
[111; 688; 5504]

10. Straightforward news reaction is unlikely to lead to profits, generally, because news does not necessarily lead to expected behavior: thus the quote "Some animals are more equal than others." Should the wins and losses be of similar magnitude, then the bad losses would be averaged out in the long-run. Unfortunately, the bad losses are worse than the wins if no risk precautions are made. In reality, text-book cases are surprisingly rare: price reaction to news has often started before the news arrives, and it may not create any significant ex post reaction. The explanations include: the news may have already been released elsewhere, more privately, or its release may be controlled.
[111; 689; 5512]

11. Large market capitalization stocks are the early natural candidates for the archetypal news-based trading strategies. It is still far easier to grasp what information is reflected in the prices of large-caps. Their much higher news count and liquidity also allow to take better advantage of any profit opportunities. The fast trader is more likely to trade large-caps while the slower trader seeks to take advantage of any longer persisting price dislocations usually more prevalent in small-caps. From the angle of the slow (or dishonest) guy, trading should be more profitable and less risky in small-caps, but their smaller liquidity limits the upside to take more serious advantage of any information edge.
[111; 710; 5680]

12. Slower but smarter news-based trading strategies should, by the above made simple arguments, be concentrated mainly on mid- or small-cap stocks. To identify the required discrepancy between the true and market price, one could use a few standard financial metrics, their advanced forms, or a set of their combinations. In any case, the best chance for the slower trader should be found in more precise valuation principles and forecasts of available liquidity. Even for high-frequency traders – among whom there exist slower traders as well – the smarts of algorithms will play an increasingly crucial role as the competition intensifies, and this game is not only played against the other traders.
[111; 698; 5584]

13. Obviously, as the smarts of automated trading algorithms keep rising, the better positioned or faster traders can take advantage of the slow trader in several ways. When news arrive, volatility of the relevant asset often increases. At the same time, its order-book typically shifts to an active state that allows good possibilities for the right traders with the right setup; it increases their motivation to either take advantage of the rapid information flow – or outrightly to control it. The game is then moved away from trading desks towards data centers where servers are placed. In capitalism, we should expect democratic randomness and high market efficiency, but they are far from guaranteed.
[111; 702; 5616]

14. Relevant to us, information (in news) can only vary between two extremes: a preset structure with entropy 0 (communism) and a random structure with entropy 1 (capitalism). Information theory – founded by Claude Shannon in the same dark World War II aftermath as Orwell wrote "Animal Farm" – declares that a message with a high entropy value takes a longer time to send due to its more complex nature; which basically just means more randomness, or freer capitalism in our terminology. In WWII, news complexity was also found relevant as by then Alan Turing worked to break the German cipher of the Enigma machine just an hour away from London in Bletchley Park.
[111; 661; 5288]

15. News complexity warrants some explanation here. Structured news typically comes with a title and a body of a certain length. The context must be known a priori or deduced. We emphasize these two points in the structure of this article, which has 23 paragraphs with 111 words (prime numbers) in each of them. Similarly, the cipher machines of WWII had several wheels each with numbered or alphabetical positions. To understand messages perfectly, one had to understand their context well. Machines could, in theory, both read and write more efficiently than humans do, but machines still suffer from many unsolved linguistic issues, and most of them have to do with context dependence.
[111; 684; 5472]

16. Turing's Enigma problems are not much unlike the problems in decoding news – or Morse code. If we limit the size of the news packets, which is practically necessary due to the inefficient rules of Morse, we must then know quite exactly what “SOS” means. But this is simple context dependence. Those who have seen the sci(fi) movie Interstellar, may recall a harder context dependence problem when pilot Cooper tries to send scientific news on wormholes to his daughter back on Earth using only the inefficient Morse code, some help from gravity, and his age-old wristwatch. Whether the future trading algorithms can interpret news as smartly as Cooper's daughter will have significance.
[111; 686; 5488]

17. To crack the code of financial news, we would need the same news received several times; we would need depth, in the jargon of cryptography. In WWII, a major breakthrough in decoding the German Lorenz cipher – with a total of 16*10^18 initial wheel settings, more precisely 12 wheels each with a prime number of settable lugs around the circumference – was only allowed by a small human mistake: a message of 4000 characters was sent twice with the same exact key and slightly modified content. William Tutte, Thomas Flowers, and the Colossus computer at Bletchley Park cracked it. In financial markets, though, context does not stay constant. News itself changes the key.
[111; 672; 5376]

18. The problem of context is related to a statistical fact: the more observations of good quality we possess, the better prepared we are to make sound inference. After necessary filtering operations for innocently erroneous observations in Bug Data – a subset of Big Data that may or may not intersect with Hack Data – news and company categorizing, we are typically left with only a handful of good observations compared to the original larger news dataset. Non-parametric estimation and machine learning techniques are typically quite data intensive, though: they tend to require much data to fit a model properly. Their learning phase may also require long time periods. So the key is changed.
[111; 693; 5544]

19. Stationary time periods with the same key are rare. The true underlying data generating process (DGP) is more likely non-stationary, which causes problems from a standard statistical point of view as the structure of the DGP is inherently more time-varying and complex with high entropy. This may be partly solved by switching regimes, local stationarity, and alike, but once we jack up flexibility we face the curse of dimensionality: relative lack of data points. While high-frequency traders may have thousands of price observations in a day for an active stock, we may only have a few news. As markets go through time periods of changing contexts, estimation is in fact difficult.
[111; 684; 5472]

20. Social media firms, such as Twitter, have made many fresh observations available, alleviating the problem of data scarcity. But with a tradeoff: unfiltered social media data are noisy. Probably in more than 99 percent of the cases the social news effect is not as expected. Thus, the potentially wrong trading decisions must be identified immediately; social Big Data need to be quickly filtered or aggregated on-line otherwise they fall in the category Bug Data. But as social media gets more widely and smartly applied, and news written by machines (robo-journalism), data may become better organized and hold part of the answer, or at least act as a valuable extra information source.
[111; 687; 5496]

21. News is going to be generated ever more frequently, for more companies, and data reliability will be the core question. Big Data methods are going to be applied all around, but they only make sense when they are not fed false data. As financial markets ideally approach extreme efficiency in terms of speed and information content, the problem we have emphasized – the "true" value of a company – is closer to being solved. Information sources that did not exist in the past are efficiently made available to serve markets. People are willingly or unknowingly part of an aggregation machine. Its downside: individuals in a capitalistic society become influenced by Big Data Brother.
[111; 682; 5456]

22. Big Data Brother can affect automated news-based trading in myriad ways. Clearly, news may be released efficiently by robo-journalists to serve only the special interests of certain groups, whether illegally or legally, or they may be tampered with in the news preprocessing stage. Social media data can also be easily used to analyze individuals or groups of people and news can be used to affect the state of mind of public consciousness like some terrorist groups are already doing. Trading based on Hack Data is led by animals of a different sort than the genuine, and market efficiency is compromised with a society-entropy nowhere near its maximum of the range [0,1].
[111; 673; 5384]

23. Yet, if one does not consider the unsettling deeper questions of truly (falsely) efficient markets, democracy, and entropy, basic trading guidelines will stay intact irrespectively of who is in control, unless the society collapses way back to communism where present-day trading is not meaningful. In an old influential book on military strategies written around in the six century before the birth of Christianity, Sun Tzu coins the phrase "If you know your enemies and yourself, you will not be imperiled in a hundred battles." Similarly, for the future trading strategies it will remain essential to know your opponents well: "All [minds] are equal, but some [minds] are more equal than others." □
[111; 699; 5592]

BIG DATA
BUG DATA
HACK DATA

[bio.txt]

Tommi A. Vuorenmaa, PhD, is the Founder and President of *Triangle Intelligence*, a company advancing start-up projects related to automated high-frequency trading (HFT) and its research. Tommi is a frequent conference speaker, author of HFT related articles, and a book "*Lit and Dark Liquidity with Lost Time Data: Interlinked Trading Venues around the Global Financial Crisis*."

E-mail: tommiavuorenmaa@triangleintelligence.com
Company: http://triangleintelligence.com
Main company project: http://hft.exchange
Personal: http://tommiavuorenmaa.net
Blog: http://versustakes.co