



The Promise and Pitfalls of Big Data

Implementing a Big Data project might make sense for your organization. But before you start investing in the time and resources required, make sure the effort will ultimately enable you to differentiate between the beneficial “signals” and the distracting “noise.”

Dr. Lapide is a lecturer at the University of Massachusetts' Boston Campus and is an MIT Research Affiliate. He received the inaugural *Lifetime Achievement in Business Forecasting & Planning Award* from the Institute of Business Forecasting & Planning. He welcomes comments on his columns at llapide@mit.edu.

The latest hype around “Big Data” is fueled by the vast amounts of information being generated by the Internet. There appears to be boundless enthusiasm for solving the most pressing business problems by leveraging lots of data streams.

The concept of Big Data is not new. It really began with the invention of the printing press, which enabled information to be generated at an exponentially growing rate. Since the dawning of computers and later the Internet, Big Data got even bigger and grew even faster. Luckily, computing capabilities have kept pace so that the data could be more easily and accurately assembled and analyzed.

Picking the right data streams is extremely important, since not all data is information. Information supports improved decision-making, and not all data is useful for that. Generally, data becomes information when it is used within a decision-support system that has an underlying business model and principles imbedded within. However, can all the data now available really provide information significant enough to improve business operations? Or is Big Data just more data?

Implementing the new Big Data is a big deal. And supply chain managers will no doubt struggle with the question of whether to implement it within their companies. They'll need to decide whether it makes sense to expend the enormous amounts of time, money, and other resources required to begin this effort. And they need to know whether this might distract them

from pursuing other opportunities requiring significantly less effort. Even more important, they'll need to carefully assess whether implementing Big Data will increase their “signal-to-noise ratio” enough to yield benefits sufficient to cover the large investments. Signal-to-noise ratio? What's that?

The Signal and the Noise

I just finished a book written by Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*. I recommend it as a must-read for forecasters, planners, and supply chain managers. The author uses the concept of a “signal” versus “noise” from electrical engineering. Whenever an electrical signal is transmitted, spurious noise distorts it along its path to a receiver. So engineers need to focus on developing receivers that pull out the noise in order to understand the original signal. The lower the signal-to-noise ratio, the harder it is for the receiver to pull out enough noise. In the business world, the signal of interest is the “truth”; the receiver is a manager trying through analytic means to decipher what is true from among myriad confounding (noisy) data signals.

Silver discusses “the promise and pitfalls of (the fashionable term) Big Data”. The promise is whether volumes of data will “obviate the need for theory, and even the scientific method”, while the pitfall is that too much data might be distracting and provide little knowledge about the truth.

Silver provides useful insights about glean-

ing information from data. The author researched prediction across a wide swath of arenas and describes how each successfully and unsuccessfully tackles problems leveraging various types of data. He offers nuggets of advice for managers to help filter out the signals in Big Data—specifically, those that improve prediction from the noise that might confuse and not be fruitful.

Silver gives a good overview discussing, for example, the successes and failures in sports and gambling predictions, including baseball, chess, and poker. He also addresses prediction in the social sciences such as economies and political elections. Some discussion deals with areas virtually impossible to predict, such as terrorist attacks, financial market bubbles, earthquakes, and global climate change. Lastly, Silver discusses some of the successes in forecasting the spread of infectious diseases and the weather.

This book should offer some comfort to managers in that our business forecasting and planning issues are less problematic than for earthquakes, terrorism, and global climate change!

Lessons from Downstream Data

Big Data is certainly not new to supply chain management. The industry has been working on the use of downstream demand signals—a Big Data concept—since 1992 when Wal-Mart began offering POS data to suppliers via RetailLink. Suppliers have been evaluating how to get maximum value out of the voluminous amount of data that the retail giant offers, as well as the data other retailers now provide.

The industry has conducted a great many pilots using downstream data, believing there is great value to be obtained from these signals. Piloting and other efforts have largely shown that while there is value to be gained, it is not worth implementing downstream data as “big-deal” projects, which would include the development of big Demand Signal Repositories (i.e., data bases).

Supply chain managers feel that the downstream data is too detailed and cumbersome to process, especially for all products and for an entire customer base. Instead, they favor focusing on a few elements of downstream data, often from major customers and important products. For example, many feel that instead of assembling a large amount of detailed POS and inventory data, aggregated data streams will suffice. The lesson learned is that it is better to focus on a few signals, and treat the rest of the downstream data as noise.

Search for a Few Good Signals

I’ve come across a multitude of industry examples of good portending signals. I have also seen others that, while seemingly predictive, were impossible to derive real value from. A few of my favorite examples of both types are discussed below.

- While I was a graduate student, a CEO came to visit our campus and was asked how he was able to manage effectively given all the information he received. He said he first looked at one metric: interplant shipments. If these were too high, there was a mismatch between supply and demand; either some plants were producing too much or too little, or certain sales territories needed to be better aligned. Interplant shipments that were too low indicated inventory excesses or insufficient sales (other misalignments of supply and demand). In either case, he instinctively knew he had to look deeper at other metrics only when this one indicated a supply-demand imbalance.

- The early identification of “winners” and “losers” in the book, media, and software industries is critical to profitability. To do this, these businesses capture sales during the introduction of a product at a few retail outlets. For example, many track sales at some trendy stores in New York City and Los Angeles to get early readings of their winning and losing product introductions. (Of course, many now track Amazon sales as well). These predictive signals go a long way towards continuing the successful launch of a winner, and a rapid phase-out of a loser.

- During a panel I moderated on downstream demand signals, a VP from a cosmetics company lamented about the lack of good demand signals for his trendy products. He gave an example where Lady Gaga had worn a very unusual nail polish at one of her concerts. The sales of the nail polish immediately took off, the company ran out of the product, and their suppliers quickly ran out of materials to make more. The company scurried about to try to get supply, yet missed a lot of sales opportunities. These types of sales spikes are noise rather than signals because they are impossible to get demand signals from which they could be predicted. In essence, they are equivalent to the prediction of earthquakes in this VP’s business.

In summary, if you are thinking about implementing Big Data at your company, make sure to first identify a few good predictive signals before building an expensive database to house them. Looking at lots of noisy data will confuse things and decrease your signal-to-noise ratio, rather than increasing it. You’ll definitely get your signals crossed and get too little information out of the data to support improved decision-making. ☺☺