How Big Data Could Challenge Planning Processes Across the Supply Chain – From S&OP Through Detailed Product Planning

Tonya Boone, Ram Ganeshan, and Nada Sanders

Preview: The authors discuss the potential of big data to improve forecasting through better understanding of consumer behavior, upgraded demand-forecasting models and more efficient supply-chain execution. However, they also document the major challenges posed for sales and operations planning and how these can be anticipated and met head on.

Key Points

1. **Copious data is available to firms both internally (clickstreams, transactions in real-time, etc.) and externally (e.g., social media).**
2. **This torrent of data can be used to understand customer behavior, manage risk, and plan supply chain activities better.**
3. **However, integration of this fine-grained data into traditional S &OP activities is fraught with challenges that require careful planning.**

INTRODUCTION

The term "big data" has dominated the popular as well as the academic press in recent years. One definition is offered by James Manyika and colleagues (2017) of McKinsey & Company: "Big data refers to datasets whose size is so large that the quantity can no longer fit into the memory that computers use for processing."

Here, we define "big data" in its most generic form: data sets that are large ("volume"); that are collected in near real-time (high "velocity"); that are present in myriad forms ("variety"); and in which there are various levels of trust ("veracity"). Especially as it relates to forecasting, big data brings with it the potential to better understand customer behavior, to more accurately predict demand, and to make supply chain execution (typically set in motion by product forecasts) more efficient.

The three of us are currently editors of a special issue of the *International Journal of Forecasting* that explores the impact of big data on forecasting. What we present here are our own ideas on the impact big data is having on forecasting and the implications and challenges it poses for sales and operations planning (S&OP). The interested reader is referred to Boone and colleagues. (2018) for a more detailed elaboration of these ideas.

BIG DATA SOURCES AND THE POTENTIAL THEY BRING

Firms are making significant investments in big data storage and applications. The number of RFID tags sold globally is projected to rise from 12 million in 2011 to 209 billion in 2021 (Manyika and colleagues, 2017). Internet of Things (IoT) investment in production is expected to double from $35 billion to $71 billion by 2020 (AT Kearney, (2017)).

Although its forecasting applications are fragmented and often idiosyncratic, there are common themes on its potential impact: big data is helping to better understand customer behavior, to improve the quality of forecasts, and to simplify supply chain coordination.

True, there is a healthy skepticism in the forecasting community over the potentially disruptive impact of big data on forecasting, as revealed in this journal's special feature section "Is Big Data 'Big Hype' for Supply Chain Forecasting in the Spring 2015 issue What we offer here is our view on the potential benefits and the challenges in reaping them.

Insight into Customer Behavior

Copious data become available as the customer moves through every stage of the decision journey. Customer clickstreams, social media interactions, and Google searchers inform us on how customers discover and evaluate products. In-store technologies such as beacons, tags, virtual rails, and engagement kiosks track customer buying behavior. Modern transactional systems not only track sales in real time but are connected to inventory and customer databases across multiple channels. Finally, technologies are enabling the rise of "omni-channel" experiences, which enable consumers to move seamlessly between physical and online stores as they evaluate, purchase, return, or seek help with products and services.

Such fine-grained data can enable *real-time personalization*. Based on customer attributes— which could include demographics, location, and browsing history—a company can personalize products, offers, and prices and “push” these to the customer (Ganeshan, 2014). At a more aggregate level, *customer micro-segmentation* can offer small groups of customers tailored products and incentives. The retailer Neiman Marcus, for example, uses behavioral segmentation matched with a multi-tier membership rewards program to identify top-spending customers. It then tailors purchase incentives for them, often resulting in significantly higher margin purchases.

Pricing decisions can now be made in near real time using a variety of new data sources – competitor pricing that can be “scrapped” from the Web or personalized pricing for customers based on their shopping profiles. Uber, for example, uses “surge” pricing, based on customer-demand characteristics. Kroger, a large US grocer is experimenting with electronic shelf edges that can personalize prices for individual customers.
https://www.wsj.com/articles/at-kroger-technology-is-changing-the-grocery-store-shopping-experience-1487646362

Improving the Quality of Demand Forecasts

Modern POS systems now can provide real-time data on transactions across selling channels. These detailed transactional datasets are helping firms estimate *actual* customer demand more accurately. It is common practice for forecasters to use sales, orders, or shipment data as the historical time series to predict future demand; however, in face of stock-outs, promotions, and supply chain disruptions, these do not give an accurate picture of the underlying customer demand (Gilliland, 2010)

For example, when demand exceeds inventory, resulting in a stock out, it is often difficult to estimate the real demand from sales data, a scenario commonly referred to as *censored demand*. Planners often manually adjust sales upwards to estimate the true demand. Now, with the ability to track granular sales and traffic data, the actual demand can be estimated from the censored sales data. The planner can use the timing of the stock-out and store traffic until the next replenishment to get better estimates of demand.

Second, these new data streams are enabling *multi-product* forecasts. Firms plan for both the assortment of products to carry and the appropriate inventory to hold for each product. These decisions are often done independently. Marketing considerations drive product assortment and, subsequently planners forecast aggregate demands. With disaggregate data, it becomes possible to forecast demands by SKU as a function of the assortment. This is because we’d have an understanding of the demand for each product in the assortment and the probability that customers would substitute between them. It is therefore possible to generate a forecast that is optimized for a portfolio of SKUs rather than a product family.

Third, these datasets provide new inputs for forecasting models, such as Google searchers and social media data, potentially reducing forecast errors. Schmidt and Vosen (2013) show how Google trends can improve aggregate consumer purchase indices. For individual product SKUs, Boone and colleagues (2015) in this journal used a case study to show how “in- sample” errors for two SKUs are reduced for a food retailer; and later (Boone and colleagues, 2018) extended the study to show “out-of-sample” error improvements for multiple SKUs over many categories.

Increased Supply Chain Coordination

On the customer-side, connected devices are now enabling firms (although the implementation is still at a very nascent stage) to evaluate customer usage patterns thus improving visibility into demand. For example, the Amazon “dash buttons” can now be placed next to the product (for example, a button to rebuy Tide detergent can be placed on the washing machine) and actuate replenishment when needed (see also Peter Catt’s (2017) commentary who gives examples of these so-called “Internet of Things”).

On the supply-side, big data has further enabled the connected supply chain. With initiatives such as Collaborative Planning, Forecasting, and Replenishment (CPFR) and Efficient Customer Response (ECR), supply chain partners can share large amounts of data and make coordinated real time decisions on product forecasts and associated replenishment. While such initiatives have focused on product and item-level forecasts for the last few years, new efforts are underway to understand the buying behavior, both to “shape” and more accurately forecast underlying customer demand.

One area that has had a special significance for supply chain management is the *forecasting of disruptions*. As a case in point, forecasting of high impact, low probability events – termed “black swans” – has significantly improved. Historically our focus was on building resiliency into the supply chain so that we could be ready to adapt to rare but inevitable events. However, big data analytics is slowly changing this notion -- the number of events that we used to consider unpredictable and purely random is getting smaller.

An excellent example of improved prediction is in weather forecasting. New radar technologies, improvements to satellite technology, as well as computer models that run on more powerful supercomputers allow forecasters to better “see” extreme weather. The key in the (big) data gathered by these radar technologies and satellites, which can then be processed within minutes creating warnings of tornadoes, hurricanes and other extreme weather events.

Big data tools bring the potential to better manage risk by by reacting faster to real time alerts on routine disruptions such as traffic or disease outbreaks. Traffic can be rerouted, for example, based on congestion to keep deliveries on time. Knowledge of outbreaks of flu can be used to determine which areas may need more supplies of medications.

THE CHALLENGE FOR AGGREGATE AND DETAILED PLANNING

The traditional demand planning process begins with historical sales data and identifies potential explanatory variables. Coupled with judgment of demand planners, forecasts are produced typically by product family over a rolling horizon.

The most common way of combining judgment with sales data is through “judgmental adjustments” or “overrides” of statistical forecasts. For short life cycle products like apparel for example, which are plagued with long lead times sometimes of several months, planners commonly use economic optimization algorithms. These algorithms first estimate demand and error distributions from past sales and use these to commit to the season's need. The forecasts are tracked and compared with actual sales, providing a feedback loop for future forecasts.

Once generated, the forecasts are translated into demand requirements and purchase orders. Many firms that do not use detailed master schedules “fence” or lock the most immediate forecast so that its supply chain requirements such material planning, inventory allocation, and purchases, can be set.

For both S&OP and these related planning processes, big data provides some unique challenges. These concern integration of vast volumes of data; security and privacy; and unintended bias which we address next.

How to Integrate Big Data

The amount of data generated can be enormous; in some estimates the daily production is 2.5 exabytes of data equivalent to 2.5 million petabytes and 2.5 billion Gigabytes. About 90% of these data are unstructured. Social media produces 500 million Tweets a day (http://www.internetlivestats.com/twitter-statistics/#trend) and 3.5 billion Google searches (http://www.internetlivestats.com/google-search-statistics/) everyday.

Supply chains are also inundated: Amazon sells 600 items every second! Walmart collects more than 2.5 petabytes of data every hour from one million customer transactions. The question for planners is how much and which data to include in the planning process. Often these large datasets tend to be "sparse" and "transient." As more datasets are included, the complexity of data management and system support also increases.

Second, planners need to address the seeming contradiction between the increased personalization afforded by big data and the aggregate nature of the S&OP process. Traditionally, the aggregate demand plan is disaggregated into SKU variants. But big data supports a more granular approach, one that starts with detailed disaggregate planning. Big data is enabling, for example, the construction of the Ideal Customer Profile (ICP – *who* is our best customer), provides insights into how to lower the Customer Acquisition Cost (CAC), and increase the Customer Lifetime Value (CLTV). There is significant debate if the move from product (top-down) to customer (bottom-up) forecasting is even possible (see Snapp, 2017). Even if it can be done this way, however, the process can be complex and time consuming. Since both the volume and velocity of data is high, planners also need to address *how often* these plans need updating.

Third, big datasets are typically used to detect patterns and associations ("Data Mining" or "Pattern Recognition") that have predictive value. It is common to use machine learning techniques to project trends, to detect fraud (deviation from the trend), to learn association rules (recommendation engines, market basket analysis, etc.), and to segment customers.

In our experience, planners are unfamiliar with these methods. Such methods have typically been used for short-term decision making. As a case in point, when Hurricane Frances was due to hit the Florida coast in 2004, Walmart learned that customers prefer to stock up on Strawberry Pop Tarts (sales rates were 7 times the normal) and beer prior a hurricane. The stores in the area were stocked with these items, which then sold out. Such data-driven insights can translate to better customer service and higher profits.

Additionally, S&OP processes typically are not flexible enough to allow such insights to be incorporated into routine system forecasts.

Finally, the role of judgment needs to be revisited. Routine adjustments in forecast for events unplanned and unexpected (like the demand censor example alluded to earlier) are now being quantified by big data techniques. On one hand, this brings a more data-driven approach to adjustments; on the other, one has to "trust" machine-learning algorithms to make those judgments.

Security and Privacy

With the explosion of connected devices, the amount of personal data collected is substantial and some of these data are collected *without* a purpose. The Identity theft and Resource center (ITRC, 2018) reports that, in just the first three months of 2018, there have been 273 breaches with over 5 million customer records exposed.

There are other well-known data breaches – the 2017 Equifax breach exposing 140 million customers' social security numbers and more recently we saw the use of personal information from Facebook by Cambridge Analytica to help the Trump campaign. Firms that use disaggregate customer data need to be more careful about security and privacy, less they violate local and federal laws.

Hidden Biases

Another significant issue concerns hidden biases in the collection and analysis of data.

Much of the data is collected automatically through sensors, connected devices, and social media channels. Mark Graham and colleagues studied Tweets on "flood" or "flooding" to see if they could predict the impact of Hurricane Sandy. See the graphic and their research summary in: https://www.theguardian.com/news/datablog/2012/oct/31/twitter-sandy-flooding). It turns out that the vast number of Tweets came from Manhattan, giving at least the mistaken impression that it may have been the site of the most damage. It was actually New Jersey.

Another well-known example is that Google trends were overestimating the incidences of flu. It turns out people searched for terms related to the flu *after* they saw the news on TV. The Center for Disease Control, however, estimates flu incidences via field surveys.

The lesson here is that while a torrent of data may be available, they may not always represent the signal one is trying to measure. Planners using big data sources must therefore carefully differentiate the “signal” from the “noise.”

Machine learning algorithms may unintentionally target or omit certain segments of the population. In her best-selling book *Methods of Math Destruction* Cathy O’Neill (2016) describes several cases of unintentional bias in education (on how teachers are evaluated), financial services (on how certain minorities were denied services), retail (where only certain demographic were targeted).

Bias can impact the most vulnerable segments of the population. Efforts are underway to measure the bias or fairness of machine-based algorithms by constructing a “fraud score” or a “bias score.” While much of the work currently resides in research journals, we foresee a wider use as more planners begin to use these large datasets. John Podesta, President Obama’s Chief of Staff perhaps sums it up best: “The lesson here is that we need to pay careful attention to what unexpected outcomes the use of big data might lead to, and how to remedy any unintended discrimination or inequality that may result.”

CONCLUSIONS

Our intention in this paper is to point to the potential of big data to better understand the customer, improve forecast accuracy, and better execute supply chain transactions.

However, big data brings big challenges:

- The scale of the data is large and there seems to be no consensus on what and how best to integrate into the S&OP and other planning processes.

- Second, the methods and techniques needed with these big datasets are relatively new and planners will need to become familiar with them.

- Third, since the new datasets are helping address some of the uncertainties in the planning process, it is unclear if and how planners should adjust forecasts based on judgment.

- Finally, these big data sets and associated techniques are open to unintentional biases that the planners should guard against .

These are exciting times for forecasters – the new large datasets have the potential to transform and improve the S&OP process and other related detailed forecasting and planning activities. We are optimistic that as planners come to understand the insights they provide, the challenges posed by big data can be eventually overcome.

References

Boone, T., Ganeshan, R., Jain, A., and Sanders, N. Forecasting in the supply chain: consumer analytics in the era of big data. Working paper. Invited paper for the Special Issue on Forecasting and Big Data *in International Journal of Forecasting*.

Boone, T., Ganeshan, R., and Hicks, RL. (2015). Incorporating Google Trends Data Into Sales Forecasting, *Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, issue 38, pages 9-14, Summer.

Boone,T., Ganeshan, Hicks, RL, and Sanders, N. (2017) Can Google Trends Improve your Sales Forecast?, Forthcoming, *Production and Operations Management journal*.

Catt, P. (2017) Commentary: Big Data and the Internet of Things *Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, issue 45, pages 27-28, Spring.

Ganeshan, R. 2014. Clickstream Analysis for Forecasting Online Behavior, *Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, issue 33, pages 15-19, Spring.

Gilliland, M. (2010). Defining "Demand" for Demand Forecasting, *Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, issue 18, pages 4-8, Summer.

ITRC (2017). See: https://www.idtheftcenter.org/images/breach/2018/DataBreachReport_2018.pdf

Manyika J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., Dewhurst, M. (2017). *Harnessing automation for a future that works*. McKinsey.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threathens Democracy,* Crown Press, NY, NY.

Snapp, S. (2017). Is Big Data the Silver Bullet for Supply-Chain Forecasting?,"*Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, issue 45, pages 10-14, Spring.