



ISSN: 0976-3031

Available Online at <http://www.recentscientific.com>

CODEN: IJRSFP (USA)

*International Journal of Recent Scientific Research*  
Vol. 8, Issue, 5, pp. 17027-17036, May, 2017

**International Journal of  
Recent Scientific  
Research**

DOI: 10.24327/IJRSR

## Research Article

### OPTIMIZING STATISTICAL FORECASTING MODELS IN REAL TIME WITH BAYESIAN COMPONENT USING 'BIG DATA' IN MARKETING

Antonio Boada<sup>1</sup>, \*Ivelis Montilla<sup>2</sup> and Francisco Jaramillo<sup>3</sup>

<sup>1</sup>CEIPA College Institution (CEIPA), 55450, Medellin, Colombia

<sup>2</sup>Simón Bolívar University (USB – SL), 1160, Camuri – Vargas State, Venezuela

<sup>3</sup>CEIPA College Institution (CEIPA), 55450, Medellin, Colombia

DOI: <http://dx.doi.org/10.24327/ijrsr.2017.0805.0268>

#### ARTICLE INFO

##### Article History:

Received 18<sup>th</sup> February, 2017

Received in revised form 10<sup>th</sup>

March, 2017

Accepted 06<sup>th</sup> April, 2017

Published online 28<sup>th</sup> May, 2017

##### Key Words:

Big Data, Marketing Data Base,  
Adjustment of Bayesian Models,  
Forecast Updating In Real Time

#### ABSTRACT

Companies possess a large amount of digital information that needs to be registered and available for enquiry on real time. That is why, it is essential the creation of a solid structure in a Big Data System that enables a company to generate a reliable data matrix of relational variables. In this article, it is intended to emphasize the relevance of the generation of such a solid structure in order to deal with historic register (hard data that derive from billing and logistics areas) and subjective information (marketing strategies) to allow the recording and enquiry of data of a company in real time. This type of system will facilitate the creation of set of schemes to store and represent non-redundant information to identify recurrent patterns for strengthening processes of estimation and simulation of product demand through statistical analysis of qualitative and quantitative variables. The Dynamic Bayesian Model techniques, which update through Dynamic tables, are suitable tools to achieve such objectives. They use Bayesian adjustment of arithmetic mean with exponential smoothing method based on product demand to generate indicators of either encouraging or inhibitory effects which function as "input" for any Multivariate Statistical Forecast Models.

Copyright © Fargab Ahmed *et al*, 2017, this is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

Throughout the history of administration, different techniques, either qualitative or quantitative, have emerged to allow the development of efficient planning processes. Since they facilitate the forecast of planning activities for near, medium or long term future, one of the most frequent techniques used for planning operations are the sales forecast.

Over the years, it has become necessary to use structural levels of Big Data to carry out adequate management activities and, to some extent, prevent labour disputes in certain work settings. For example in some work places such as: hospitals (where the entry and exit of visitors and patients are registered), police stations (where events are recorded according to exact time and date of occurrence), huge warehouses (where transactions are made through invoices that identify purchase date and, in some cases, name of the agent), among others.

The massive amount of data that is generated in those places lead to the creation of data warehouses or storages which grow in such a dizzying way that, in some cases, even the implementation of Structure query language (SQL) is

inadequate to obtain efficient results.

As far as business field is concerned, it has been observed that companies gather a huge volume of information during their labour work activities. Some of them are: sales, purchases, inventories and so on. However, while some of those data will remain available for retrieval and future use, some others will possibly just be accumulated taking the risk of getting lost due to the absence of an updated process or a change in management policies and structures of data handling.

The Planning System of Business Resources, best known as Enterprise Resource Systems (ERP), are systems of business information that integrate and handle various businesses associated with the production operations (and even strategic) as well as the distribution of aspects of a company in the production of goods and services. The main advantage of such "expert systems" is their capacity to unify the business data base in only one Big Data (figure 1), in which every area and department information come together in a single database, allowing the personnel to have confidence in the credibility and synergy of the information provided.

\*Corresponding author: **Ivelis Montilla**

Simón Bolívar University (USB – SL), 1160, Camuri – Vargas State, Venezuela



Figure 1 Big Data structure in ERP systems

Source: [http://www.complus.com.ve/?page\\_id=50](http://www.complus.com.ve/?page_id=50)

Generally, in the business world, companies need help from specialist consultants in the development of statistical applications to optimize the forecast of demand estimates according to the dynamic business needs. However, at the beginning of the process of this requirement, it is common to find that the initial data base is not analyzed in order to corroborate the effective registration of the information that can directly impact the demand depending on the characteristics of the business.

In many cases, such initial data base stores information that can represent knowledge of valuable help to get a better understanding of certain phenomena or guide future decision – making. On the other hand, the amount of data that it keeps can be of such size that makes the retrieval of it a complicated task. Therefore, there is a growing need of capacity for data analysis. In other words, it is crucial the creation of tools able to summarize efficiently such amount of information in order to retrieve the knowledge required to forecast and estimate possible sale scenarios, taking various statistical analysis of the information as reference.

Regardless of the type of business, the logistic procedure of production, importation, transportation method or storage system, making predictions along the time to achieve a future effective planning of resources and working scenarios is a challenging task.

### Importance of Big data

The term Big Data normally refers to the collection of a set of data which is as large and complex that is difficult to process by well-known traditional data base applications. Hence, the main purpose of the use of this technology is to find hidden knowledge from a big volume of data to turn it into useful information that leads to improve the business processes. Big data is also related to what is known as data mining which is a field of the Computer Science that attempts to discover patterns in big volumes of data. Data mining, as well as Big Data, uses

the methods of the Artificial Intelligence (AI) and the Statistics to analyze the patterns of the database that it works with (Moreno (2014).

The well-structured organized information is synonym of power today. For this reason, the Big Data system is an upward trend that defines a new competitive strategy based on a unified architecture that provides liquidity of comprehensive database, maximum scalability and security and privacy from the inside out of the company. A Big data framework is also innovation that derives from the administration of large volume of data through various procedures, namely the integration, analysis and applications that facilitate the search of statistical patterns to predict and simulate future sceneries.

There are four (4) types of supply chain strategies in which, for organization purposes, the technologies of information and communication play an important role (Chase *et al*, 2006):

Efficient supply chains: they are supply chains that employ strategies for the creation of maximum cost efficiency. In order to pursue such efficiency levels, it is necessary to detect and eliminate non-value added activities.

Protection against risks supply chains: They are supply chains (more than one supply chain) that employ strategies for sharing or keeping pooling resources so the risks of interruption of supplies are shared among each other.

Responsive supply chains: they are supply chains that employ strategies to respond with flexibility to the various customers changing needs.

Highly responsive supply chains: they are supply chains that not only respond with flexibility to meet the consumers' needs, but also provide protection against risk of shortages or supply failure, sharing among each other products in stock as well as production resources.

It is important to say that regardless the type of the supply chain system a company adopts, it is essential that it keeps a correct record and analysis of the information related to its functioning and operation. In this way, it will be possible to identify those patterns that go unnoticed to the human eye, but can easily be processed and shown by a machine.

According to Chen (2015) the essential aspect of Big Data use is to be able to turn data into knowledge and intelligence. The evidence has shown that the ideas derived from huge amount of data have the potential to transform corporate strategies and business models to improve, in this way, the commercialization. Byrnes (2012) also points out that what really matters is to identify patterns and correlations and to extract meaningful information that leads to make better decision-making.

Likewise, Moreno (2014) says that today data are no longer used to know what happened in the past; instead they are used to find out what is happening in real time. This helps to figure out what is the most useful direction in marketing to take, improving the services, reducing costs, and saving time. The possibilities are endless. Though, the question is how can the data be processed and turned into useful output that helps to make decisions in business.

Companies and corporations, sometimes, consider ERP as the all-in-one solution for managerial issues. They rely on a computer application that deals with the managing of the business processes of a company in an integrated form, based on a comprehensive database (Business Data Base). However, it is important to stress out that even though such expert systems facilitate and represent the unification of registers and data at only one level of Big Data, through a variety of reports and modules of enquiry per department, (figure 2), it is generally not enough to potentiate the prediction of the demand in real time.



Figure 2 Business objectives through a single database

Source: Forrester Research Inc.

The installation of an ERP application is, in this sense, insufficient to optimize the strategic levels of logistic planning of the future demand, what generates dissatisfaction due to the impossibility to meet the targets that have been established for future improvement of the organization.

The Big Data system deals with common issues related to collection, storage, research, sharing, analysis and visualization of data of interest for the company departments. This system is strengthened by the exponential growth, the availability and the use of information, both structured and non-structured, that robust the documentation and registration of creative strategies by the sales and marketing departments.

Big Data offers a very precise perspective in real time about the behavior of the consumer. It predicts the behavior of a future customer based on the historical data of clients that show a similar profile, it helps to hold him as long as possible.

In this sense, in practice, companies and corporations, both national and international, trust “expert consultants” of ERP systems in order to determine each variable and the characteristics of the business that have to be recorded in those systems. Though, it is the employees of the organization the ones who possess a detailed knowledge of the business.

Attaining an optimal structure from the start is crucial to register and store operating and strategic information to obtain a congruent and solid database to substantiate the predictions of the demand in time.

### Forecasting Demand of Products in Real Time

For every company of massive consumption products, the demand forecast and the simulation of future scenarios are

crucial elements for the planning and control of their resources, so that there is an efficient use of the production system, inventory and delivery of products on time (Chase, 2006). The database structure can contain information in real time of the point of sale, inventory, products in transit, market statistics, demographic information of branches and/or clients and sellers, finance, product return request and even vendors performance. This type of global structures and database, updated in real time, can help to identify the historic patterns of the operations performance (Big Data) in different periods of time. The identification of historic patterns of the operations performance permits to generate subsequent forecasting process and simulate the future scenarios to meet the corporation needs.

The historical analysis of the demand behavior is done in order to study the evolution of the past to forecast the future behavior. To this end, a reasonable margin of security is used, even checking every single department of the company which could provide useful information (quantitative and qualitative also) for making reliable forecast.

According to Boada (2011), some of the most important activities of an organization in which predictions play an important role are:

1. Estimation of the sales operation department.
2. Estimation of the demand of finished product.
3. Planning of marketing activities.
4. Programming and assignment of production task.
5. Programming of availability of financial resources.

The demand unpredictability and the supply of products constitute an ideal reference framework to understand the strategy implemented by the supply chain. The innovative products of unpredictable demand and the expanding process of supply process of fabrication or importation of underlying technology (in constant growth and change) face a great challenge these days (Chase, 2006).

In this respect, Paredes-Moreno (2015) highlights the importance of the implementation of Big Data in company’s areas where the highest value for the business can be provided. Some sectors consider that this process should be initiated doing an analytical study of the clients which would permit to provide a better service by understanding their real needs and being able to anticipate their future wishes and behavior.

In this way, companies will be able to optimize the distribution networks and the logistics and production areas through the use of powerful capabilities of data processing and analysis. Besides, they will be able to improve the accuracy of the forecast demand, to discover new demand patterns and to develop new services by sharing data with partners of the whole supply chain (World economic Forum, 2015).

Moreno (2014), points out that, with little investment, small companies could increase the value of their data getting the most out of them. He explains that a project cannot longer be considered unattainable because of a tight budget as it can be carried out making an investment that increases little by little. According to Gardner estimations (in TICbeat, 2013), the client relation management, the development of new products, the fraud detection, or the prediction of the consumers behavior allows the company to obtain 20% finance results over its competitors.

Finally, it is important to mention that the supply chain process begins with an optimal system of prediction that, it is essentially, based on the registration and database system of the corporation. This implies a set of functions or activities that are carried out within the organization through a joint staff effort to accomplish the company's objectives (Evans and Lindsay, 2008).

### **Relational Data Base**

Owning a unique data base with reliable, coherent and solid information fosters the study and measurement of the factors that impact and affect a close trend, rather than its extrapolation, as the main purpose of a company prediction process.

The real time update relational model has been established as the main tool for processing applications of data in companies (Silberschatz, 1998). This is a relational data base that shows the strong correspondence between tables and mathematics relations of the set theory. Through this model, the information obtained from multiple departments of a company can be linked in a single maxi-file which will facilitate the search of behavioral patterns and simulation by the real time update of this Big Data structure.

Belicove (2013) adds that a process like the one described above permits to know the direction that marketing should take, improve the quality of the services paid, and reduce costs and safe time. However, the important point is how to figure out the best ways to link and organize this information into packets in order to obtain meaningful and updated knowledge.

There are many examples that could explain the way some companies have used Big Data system to predict sales in real time. Yogesh (2014) cites one of them and shows how, through the use of Twitter information, Big Data techniques and lineal regression, predictions about classifications and sales of music product are made. Likewise, PRNewswire (2017) demonstrates that the automobile sector has also used Big data to make sales forecast until 2026.

In the field of simulations and demand prediction along the time, the extrapolation of a trend is a technique widely used in statistics. However, this technique is inadequate to handle predictions in a medium or long term. That is why it is very important that the relational factors that impact the demand of products and also affect the trends over time are well known by every company department where the corporation estimates are dealt with. Being aware of this would permit the implementation of future planning strategies that would impact the optimization of the logistics actions of production, importation, storage and distribution.

The use of a strong and reliable data base that shows complete information (either qualitative or quantitative) of a product, in real time and without the risk of data duplication, is essential to achieve successful statistical estimates of their effects on the product demand. To do this, the identification of behavioral patterns is used to create statistical modeling and tools for the simulation of future scenarios.

According to the data extracted from [http://www.marketsandmarkets.com/Market-Reports/big-data-market-](http://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html?gclid=COD8rKihvdICFdhZhgod_JgGDg)

1068.html?gclid=COD8rKihvdICFdhZhgod\_JgGDg it is expected that the Big data market goes from USD 28.65 billion in 2016 to USD 66.79 billion in 2021 at a compound annual growth rate (CAGR) of 18.45%.

### **How is the Historical Data Base of your Business?**

This is a fundamental question for every corporation department that deals with estimates and trends, since it provides the most important contribution at the moment of analyzing and substantiating the future behavior of the demand along the time.

To achieve an optimum performance of the historical information, there are a few important questions that need to be answered: 1) is your database robust enough to forecast sales? 2) Do you know the full potential of your database? 3) Do you really know the different historical factors that can affect the sale trends? 4) Do you have a full control of the factors and the planned information that affects future sales?

In order to initiate a forecast process, it is necessary to make some decisions that depend, at least to some extent, on the value of some variables of an event that will take place in the future.

The whole forecasting process of a company can be segmented in nine (9) steps. They begin and end with the communication, the cooperation, the collation and the effective data keeping. Such communication and cooperation are essential if the forecast has a desired positive effect on the decisions made (Wilson et al, 2007).

### **The steps of the forecast process in a company are**

**Specification of objectives:** It is very important to set the objectives that would be linked to the relevant decisions for forecasting. As far as business is concerned, the forecasts represent the beginning of the supply chain process.

**To determine what to forecast:** It is not enough to say that a sale forecast is required. Do you need to forecast the profits from the sale or to forecast the sale per unit of a product? Do you need an annual, semi-annual, quarterly or weekly forecast? In this respect, the Big database use is crucial because with information in detail (daily or weekly demand) it is more likely to obtain more general results (half yearly, yearly, etc).

**Identification of time dimensions:** Two types of time dimensions should be considered. In the first place, it is necessary to establish the length of the forecast horizon (usually the length of time is shorter for the inventory control and longer for finance and sales). In the second place, it is necessary to establish the urgency of the forecast making.

**Data consideration:** The data required to prepare a forecast scenario can originate from within or out of the system. Regarding the internal data, it is considered that they are, usually, easily to access and, therefore, malleable and analyzable. In this sense, an efficient Big data procedure would ensure a successful forecast process. It is recommended that it is the most detailed and disaggregated as possible.

**Statistical Model selection:** once it is got an effective and efficient database, in terms of query and real time update, that is to say, a data base of accurate consistency, a company can

start to assess and select the most suitable statistical model for the business and the forecast distribution. Sometimes, some companies prefer to perform adjustments and build an efficient Big data and implement it along the time (not taking into account the past) and when they obtain enough amount of records (usually over a calendar year) they start to seek for an appropriate statistical model.

**Model evaluation:** Once the methods have been selected, it is necessary to do an evaluation of its operation along a certain period of time.

**Forecast preparation:** In order to make a final forecast, it is recommended to perform more than one type of prediction. For example, it would be possible to implement a statistical model and an expert judgement in agreement with the projected trends.

**Forecast presentation:** For an optimal use of forecast reports, it is necessary that they are presented clearly, providing extra information about the way the results were obtained so they seem trustworthy. In this regard, Wilson *et al* (2007) says that the forecaster should be able to communicate his findings using a language that is easily understood by the managers in charge. The objective is that, in consensus meetings, the managers reach agreements to make the company work in correspondence with the given forecast. These agreements are usually done in Sales and Operation Plan meetings (S&OP).

**Monitoring the results:** The deviations between the forecasts and the real events should be discussed openly, objectively and positively, gaining learning and evolving from the gotten experience.

In order to make an optimal Big data use, it is important to consider the variables that directly or indirectly impact the behavior and fluctuation of the demand. It is also important to optimize the forecast performance of the demand along the time

The database that is used to predict the demand is, in some cases, different from billing or shipping data (even though it is mainly fed from those data). Since, it has to contemplate the "optimum service level" scenario, it has to register or estimate what is being called the Unmet Demand or Sold out. Apart from that, when making predictions, it is necessary to determine the information that comes from previous planning and their results in terms of demand, as well as the variables that, in fact, affect the demand of products along the time (quantitative and/or qualitative variables, operational and/or strategic variables).

A database can be defined as a unified set of information that comes out of an IT project and that will be shared by the different organization users. As far as the forecast perspective is concerned, a database should contain basic information which must be included in a finalized product (planning and original strategies) and that should be registered along the time maintaining the periodicity for every single item.

According to Boada (2013), it is also important that there is synergy among some areas of the company such as marketing, billing, logistic and shipping to agree criteria in specific aspects. This implies:

- The sold and/or shipping of units.

- The estimated units.
- The offer price and the real price.
- The sales code and/or the manufacturing code.
- The business unit to which a certain product belongs.
- Unmet demand or Sold out products.

Moreover, the qualitative information, which in some occasions is not effectively registered in the computer systems, is of great importance to understand, analyze and, even, quantify the historical impacts in future actions in every period along the time:

- Package-like grouping (Simultaneous sale of related products)
- Categories according to product use.
- Classification according to zone of use or application.
- Type of product according to physical characteristics.
- Product geometry
- Item tone
- Presentation strategy and/or Publicity
- Advocacy strategy and/or special offers.
- Number of pieces that compound the product.
- Sales method.
- Use of contingency and/or additional products to boost sale
- Product weight
- Product dimension
- Units sold per seller tile
- Size and/or any other variable of classification
- Binary variables (presence or absence of a particular phenomenon. For example, the presence or absence of publicity of a product on Television)

Occasionally, it is difficult to register qualitative information because it is done through an initial parametrization of possible fields that facilitate the generation of a list of options to select from (no open fields). These registers are useful to understand, analyse and, even, quantify the historical impacts in future actions by collecting and assessing similar scenarios statistically and valuing their evolution along the time.

These qualitative records could sometimes suffer different changes depending on the strategies used by the marketing and sales departments to estimate and record future scenarios. This practice shows the importance of keeping a register updated in real time so the demand planner gets to know and value the fluctuations of the inherent strategies of the product.

The collaboration of the strategic departments of marketing and sales is extremely important to keep the future Big data enriched with updated and planned information. It is considered that the greater the lack of information in the database, the more difficult it will be to apply efficient statistical techniques to optimize the demand prediction along the time. Knowing and owning a large quantity of strategic information is crucial. It contributes with the improvement of the predictive power of a company in order to reduce undesired effects of estimation (low sales) which would generate excess of unwanted inventory, or low estimation of sales (over-sold) that would generate sold out products, affecting the company's service and the final supply chain consumer (Boada y Mayorca, 2012).

Making forecast at the right time is, then, vital and essential to generate hedging strategies that are continuously assessed during the performance of a company.

Forecasting future scenarios based on historical data is like driving a vehicle through an unknown road in the darkness, but only taking as guidance the experience of traveling a road before. This thought of comparison entails to answer some questions that demand planners constantly make the whole company: is the traveled road dark or clear enough? If the traveled road is clear enough, do you know how to study and analyze it? These questions are made to minimize surprises and undesirable scenarios (or scenarios not previously expected).

Once the demand plan department establishes possible future demand scenarios, corporate meetings of Sales and Operation Plan (S&OP) are held to set commitments to execute the plan and establish hedging strategies in predicted scenarios, until an unknown date (since future scenarios are being dealt with). Thus, the registration of marketing information in real time is crucial to get estimations of good quality in correspondance with the strategies planned.

Because it is a strategic department, planning changes and adjustments are normally performed by the marketing department as part its daily business activity. The registration of information of the products and the merchandising strategies are decisive to attain subsequent estimations through statistical forecasting models that go from classical statistical models such as Regression to Bayesian models.

Marketing variables and their automatic updating are essential as well. There is what is called time variables to establish the inherent strategies of every sale product, the number of distributors and sellers in the area, the advertising strategies represented by qualitative variables as well as concept promotion in each specified time, using either a direct promotion or packages of a different concept that boost the sales and keep the profit margin that the corporation required.

Dealing with sales prices characterized by historical and predictive aspects and their corresponding real time update are also important in the forecasting field. It is worth to remember that the sale price as an absolute variable represents a mixture of various indicators: length of time elapsed since the product was offered the last time, inflation historically accumulated in the country, variation of production cost, among other aspects.

In this regard, the relations of sale price in correspondance with the inflation rate are required to verify whether the price increases less than in proportion with the inflationary price. Such situation affects the performance of the demand because the demanded amount may increase without necessary presenting any offer at all. The same thing can happen during advertising campaigns as the product price may remain constant while the inflation continues its course.

In the same way, it is possible to register binary information (existence or absence) which determine variables like, for example, existence or absence of different size product, existence or absence of additional variables that boost or inhibit the demand, existence or absence of external situations that contradictory affect the demand in spite of the marketing strategies implemented with anticipation, among other new

aspects that escape from the initial structure of the Big Data for marketing.

In the forecasting field, there are different statistical models that can successfully be applied to each Big Data. This is possible depending on the evolutionary behavior of the demand in correspondance with the causal variables of marketing.

However, despite of implementing classic or modern statistical tools to perform a direct modeling of products and being able to predict units at a SKU level, it is vital and necessary that the residual results of the statistical model of prediction are continuously updated along the time, guaranteeing an automatic and continuous adjustment in real time.

#### ***Automated Prediction Tool: Use of Big Data and Bayesian Components to Establish Automatic Adjustment of Estimates per Product***

Each individual product shows the average performance of the demand along the time in correspondance with its behavior and being driven by the implementation of the promotional strategies available through the marketing techniques.

Through the original average performance of the product, it is possible to quantify the level or approximate "weight" that characterizes every item in a given time. In other words, it is possible to determine the tendency of the product behavior, whether the product has an upward trend, stable trend or downward trend, regardless of the promotional strategies established by marketing.

Statistical models such the Bayesian components can facilitate an automatic continuous assessment of the product performance along the time, provided that the registers of the promotional strategies are adequately recorded on real time through automatic Big Data systems.

Even though, statistical models typically group products with similar characteristics, the Bayesian system can be used to make individual predictions per item. This can be done because the level of sales that is figured out by the lineal dynamic Bayesian model is basically individual and independent and it is based on the performance of the average demand along the time.

Consequently, the Bayesian statistics provides an ideal theoretical framework for modeling data, as it permits to deal with issues with both axiomatic transparency and flexibility. This facilitates the development of congruent inferences.

That is why, the Bayesian components are, to some extent, mainly implemented in correspondance with the average demand of items, but excluding the promotional strategies impacts of marketing (either demand booster or inhibitory).

The Bayesian component permits to determine the way in which the average levels of demand, per individual sale product, are adjusted (tendency) in presence of the following independent sceneries that Pericchi (2002) pointed out:

- Insertion of a new market competitor (entailing a product or a company).
- Variation in the amount and length of sale campaigns.
- Extreme instable situation of a particular country.
- Upgrade level of sale derived from product redesign.

- Sale price excessive variation.
- etc.

The scenarios mentioned above can cause an alteration on the general tendency or product average, according to its demand level. Through the Dynamic Bayesian model is possible to determine, with few historical campaigns, the effect that any of those scenarios cause and which are not originally taken into consideration by classical statistical models of estimation (this has been assessed through multiple regression statistical models by Boada, 2013) and that can be used to adjust the evolution of the quality of the demand along the time (West, 1989). In the same way, the Dynamic Bayesian model takes into consideration subtle and continuous changes in the product trend that can subsequently be automatically adapted to minor modifications of the sales levels (Boada, 2000).

The lineal model with Bayesian components can be used in wide database updating itself in real time, grouping products in categories according to their similarities in terms of physical characteristics (for example: deodorants, shampoos, liquid perfumes, etc). Likewise, the assessment of the tendency could be made per individual product, being immediately updated after obtaining the real demand information, figuring out the capacity of the reaction and the adjustment of estimation better than any other estimator.

In order to predict future demand, a business Big Data system should be able to consider, on one hand, the implications of the promotional impacts which can previously be registered through the marketing area in real time. And, on the other hand, it ought to consider the tendency of the performance of the family of the product (as a group) and the tendency of the product itself (as an individual item). This can be done through a bayesian model which is automatically self-updated by the registration of the historical information. Once the evolution of the tendency of the average performance of the family of products, as a group, and the product itself, as an individual

item, is established in correspondance with a bayesian model, it will be possible to adjust the predictions according to the short term changes of such tendency, being even plausible to visualize the information graphically through an automatized application (see figure 3).

This automatic tool of product estimation for companies of direct sale style, or catalog sale, uses a major statistical model to estimate the units per catalog according to the enhancing or inhibitory impact of the causal variables, determined by the classical statistics (as, for example, the multiple regression models). On the other hand, this system uses a bayesian component of residual adjustments in order to continuously evaluate, in real time, the disparity between predicted value (Demand Prediction) and real value (Real Demand) to generate a component of Bayesian adjustment that evolves according to the performance of the demand of every product along the time. Interestingly, this component is based on the predictive residues that results from the establishment of any major predictive model that could be developed through classical statistics or neural networks, paths models, among other statistical-mathematics models.

These statistical components of adjustment derive from the bayesian statistics and provide the automatic application of a suitable theoretical framework to model data in real time. This permits to deal with update problems faster than the way classic statistics does because it handles them with an axiomatic transparency and flexibility simultaneously (Pericchi, 1990). As a consequence, this system permits to development inferences as congruent as possible, being applied on predictive residues of automatic update under historical information and component of "adjustment" in scenarios of prediction.

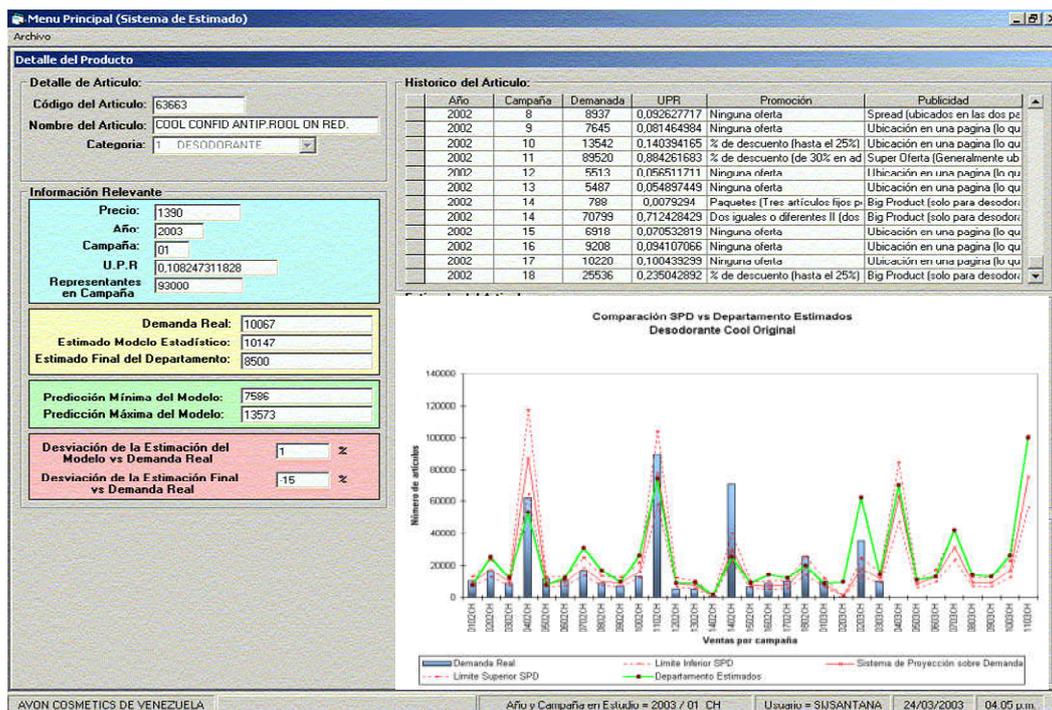


Figure 3 Demand Projection System (DTS) Estimation Software for companies of direct product sale (Boada, 2013)

These Bayesian components are implemented as a consequence of the limitations of the classical prediction models that can predict the promotional impacts on the demand of products, but cannot effectively show the tendency of the demand changes when exogenous variables appear because they are unknown or belong to a particular macroeconomic environment of a certain geographical region, as for example, the insertion or the removal of a Market competitor, modification of the fiscal policy, particular economic situation of a certain country, etc (Boada, 2000).

The Bayesian component is used as a complement of the model that is eventually calculated through a study of residual analysis that is obtained by any classical model of prediction in presence of Big data, which keep the formality of assumptions of the randomness in the residues of  $\varepsilon \sim N_n(0, \sigma^2 \cdot I_n)$ .

### Linear Dynamic Bayesian Model of Order 1

The Linear Dynamic Bayesian Model of order 1 matches the complementary stage of any statistical predictive model that is directly performed on the residues that come from a major statistical model which could be calculated by the classical statistics, neural networks or any other statistical modeling technique.

Through this Bayesian model, it is possible to achieve a predictive distribution that will evolve along the time and whose adjustments can be added to future prediction results "t+1" that have originally been provided by any predictive model (as for example the multiple regression model).

#### Theorem 1

For every length of time (years, months, days)  $t = 1, 2, 3, \dots$ ; the following equations of observation and system are considered (West 1989, Pericchi, 2002):

#### Equation of Observations

$$y_t = \mu_t + v_t; \text{ where, } v_t \sim N(0, V_t)$$

#### Equation System

$$\mu_t = \mu_{t-1} + \omega_t; \text{ where, } \omega_t \sim N(0, W_t)$$

Taking as initial information  $\Pi(\mu_0|D_0) \sim N(m_0, C_0)$ , where the assumptions of randomness and stability of residuals of the exponential smoothing model used are kept, then  $m_0 = 0$ ,  $y C_0 = 1$ .

In the former definition, the terms of the successions  $\{v_t\}$  y  $\{\omega_t\}$  are mutually independent. For the distribution  $(\mu_0|D_0)$ ,  $D_0$  is the initial information,  $m_0$  is an estimate for the level of the series and  $C_0$  is the uncertainty on the average  $m_0$ .

For each  $t$ , we will assume that  $D_{t-1}$  gathers the whole information of  $D_0, V_t, W_t, Y_t, Y_{t-1}$ , then, the only new information for each "t" will be  $D_t = \{Y_t, D_{t-1}\}$  (West, 1989). In this sense, since it is a system generated from the residuals, we will take  $W_t = 1$ , keeping the levels of randomness between  $\mu_t$  and  $\mu_{t-1}$ ; however, since "y<sub>t</sub>" corresponds to an equation of observation created according to the values of learning system, a variance  $V_t = 12$  will be used subjectively to generate a Bayesian filter of learning "A<sub>t</sub>" that converges at 20% of the difference between the predicted value and historic value. This value would correspond to the percentage of learning and

automatic update that will be kept along the real time to adjust the estimates due to the use and update of Big data.

#### Theorem 2 (West 1989 and Pericchi 2002)

Posteriori of  $\mu_{t-1}$ :

In this section, the Bayesian study is initiated based on the data derived from residuals of the demand of the product that it is being studied.

$$\Pi(\mu_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1})$$

Where,  $m_0 = 0$  y  $C_0 = 1$ , keeping the assumption of randomness of residuals previously argued (which is obtained through any predictive statistical model that the company can use).

Priori of  $\mu_t$ :

With this distribution, the behavior of  $\mu_t$  is simulated with the data on time "t-1"

$$\Pi(\mu_t|D_{t-1}) \sim N(m_t, R_t)$$

Where  $R_t = C_{t-1} + W_t$ . that specifically for this complementary model of residuals would be  $R_t = C_{t-1} + 1$ .

Predictive  $y_t$ :

With this distribution, it is predicted the value of the error on time "t" with updated data at "t-1".

$$\Pi(y_t|D_{t-1}) \sim N(f_t, Q_t)$$

Where  $f_t = m_{t-1}$ ; and additionally  $Q_t = R_t + V_t$ .

This is considered the most remarkable stage of the procedure because a prediction of the current error is obtained from the difference between the predicted values gotten from the predictive statistical model and the real demand of the product provided by the historic data until t-1.

Where  $f_t$  is obtained from update and historic information until time t-1.

Posteriori for  $\mu_t$  (Closing cycle)

Once the historic data is obtained until the time t, the posteriori distribution can be calculated for the errors along the time, therefore closing the cycle of the Linear Dynamic Bayesian Model.

For this section, the following terminologies are calculated:

$e_t$ : indicates the fault or difference between the original error on time t and the estimates by the previously described model

$$e_t = \text{Original}_t \text{ Error} - f_t$$

In this aspect, Original Error is defined as the difference between the predicted value calculated by any predicted model (multiple regression, exponential, series of time, neural networks, among others) and the real value of the demand obtained on real time through the use of Big data.

At: Terminology used as a percentage of filter that will indicate the portion of fault or difference that must be added as learning knowledge to the new value of mt.

$$A_t = \frac{R_t}{Q_t}$$

If it is used a constant value of  $V_t = 12$  the generated filter by  $A_t$ , will approximately converge at 20%. In this sense, the data a posteriori of  $m_t$  will take as information the previous value  $m_{t-1}$  plus  $A_t\%$  of  $e_t$ .

Then, the distribution for the error on time t, will be:

$$\Pi(\mu_t|D_t) \sim N(m_t, C_t)$$

Where:

$$m_t = m_{t-1} + A_t \cdot e_t$$

$$C_t = A_t \cdot V_t$$

In this sense, the Bayesian component could include a factor of adjustment in presence of the residuals that derived from any predictive statistical model which is established in a wide data base structure, with real time update like Big Data, that permits to generate predictions and updates of tendencies instantly to estimate future values continuously, benefiting any corporation of massive consumption.

## CONCLUSIONS

Experts' consultants or even specialized business personnel spend, in occasions, significant amount of time trying to get to know and deal with the most modern statistical techniques to process and predict data in the most efficient way. However, very often, they do not take into consideration whether the information handled in the forecast area is, in fact, the most optimal to support the different strategies of projection of the demand.

A marketing analysis allows to take strategic decisions and to predict the future based on accurate, real and timely information. In this particular case, the Big data and the data mining tools provide a significant support. They help to assess own products, inform if a particular aspect has to be redesigned, communicate about the necessity to adapt to the clients' needs in a specific, personalized and trustworthy way (loyalty) each more, contribute with the improvement of sales and perform their projection in real time.

Through the Big data use is possible to unify operative and strategic information in a single data base. It is a tool that assist the strategic management areas (Marketing and Sales Departments) and the operative management areas (Logistics and Finance Departments) of a company to achieve an effective communication in order to identify patterns of historical performance that permit to assess future scenarios of the demand behavior, establishing future hedging strategies and preparation to minimize unexpected impacts that could generate negative results of over sale or down sale of a product. The Big data changes in accordance with the type of company that requires it. It adapts to the company's nature, structure of work and the way the information is handled. Thus, there are not specific rules to manipulate the structure and volume of data to be registered. However, the demand planner should be able to identify those variables that, whether registered or not in the system, generate impacts on the demand fluctuation (enhancing or inhibiting it) in correspondance with the type of register (quantitative or qualitative).

In this sense, the use and manipulation of information is achievable in real time through an automatic development and continuous use of Big data. Statistical techniques are, then, carried out, on one hand, to predict promotional impacts (performed through classical statistics techniques). On the other hand, they are used to adjust the upward or downward tendency of every product in real time implementing an additional component such as the lineal dynamic bayesian model in an automatic application of Big data (An example of this is the Projection System of the Demand (PSD) applied and installed in Avon cosmetics company of Venezuela).

Considering there is a lot of qualitative information, the

following questions emerged, how can qualitative information be registered in a Data base? After answering this query, it is necessary to go on with the following one, how is it possible to identify the patterns of behavior and quantify the impact of such qualitative variables in the demand of products?

The use of predictive statistic models in business field is feasible as long as the quality of information is available in real-time. This is a mandatory requirement, mainly, in the strategic area of marketing which has to keep the planning changes up-to-date in every period of time.

Among the various advantages of Big data use, it can be highlighted the breadth of information that it can handle and its capacity to update it in real-time, facilitating the creation of congruent strategies and mathematics techniques to boost the benefits of updating estimates. Therefore, the creation of solid structures of updating such as the Bayesian components is extremely important to get the most out of a Big data system.

## References

- Belicove, Mikal E. «Cómo aprovechar el Big Data en tu empresa». En SoyEntrepreneur.com. 22 de octubre de 2013. Recuperado de <http://www.soyentrepreneur.com/25933-como-aprovechar-el-big-data-en-tu-empresa.html>
- Boada, A. (2016). Aplicación del modelo lineal dinámico bayesiano, como procedimiento de actualización automática para modelos estadísticos predictivos. Revista Dimensión Empresarial. Indexada en Scielo, Scielo Social Science, EBSCO, Econolit, Dialnet, Publindex, Latindex, Ideas, Econ Papers, RePec. No. Especial 14(2) (2016). Pp 1-27. ISSN 1692-8563 & 2322-956X.
- Boada, A (2016). "Sistema Forecast". Predicción automatizada en empresas de venta directa. Revista Opcion. Indexada en Revicyhluz, World List of Social Science Periodicals (UNESCO), Ulrich's International, Periodicals Directory, The Library of Congress USA, Rezulcyt, Latindex, CLASE, CERPE, RedALyC, Dialnet, DOAJ (Directory of Open Access Journal), e-revistas, Scielo, Scielo Social Science, Scimago Journal & Country (SCOPUS). Volumen 32, No. Especial 11 (2016). ISSN 1012-1587. Pp 121 - 142
- Boada, A & D. Vasconcelos. (2013). Modelo estadístico de regresión múltiple, columna vertebral para predecir en empresas multinacionales con estilo de venta por catálogo. Revista Lasallista. Indexada en el Publindex. Latindex. Scielo. Artículo Invitado. Vol. 10(1), pp. 112 - 127. Observaciones: ISSN 1794-4449
- Boada, A. & R. Mayorca (2011). Planificación de Demanda, en Empresas con Estilo de Venta por Catálogo. Revista Lasallista. Indexada en el Publindex. Latindex. Scielo. Artículo Invitado. Vol. 8(2), pp. 124 - 135. Observaciones: ISSN 1794-4449
- Byrne, R. F. (2012). Driving profitable growth with Big Data and better forecasts. Supply Chain Europe, 21(1), 40-41. Recuperado de <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=77943142&lang=es&site=e-host-live>
- Chase, R., Robert J., & Nicholas A. Administración de la producción y operaciones para una ventaja competitiva. Mc Graw Hill. Décima Edición. 2006.

- Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems*, 32(4), 4-39. doi:10.1080/07421222.2015.1138364. Recuperado de <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=114514526&lang=es&site=ehost-live>
- Evans, J. and William L. Administración y control de la calidad. Cengage Learning. Séptima Edición. 2008.
- Moreno, J. P. (2014). Una aproximación a Big Data. *Revista de Derecho UNED*, (14), 471-505. Recuperado de: <http://search.ebscohost.com/login.aspx?direct=true&db=fua&AN=98510180&lang=es&site=ehost-live>
- Paredes-Moreno, A. (2015). Big Data: Estado de la cuestión. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, Vol. 2, Num. 1, pp. 38-59. Recuperado de [www.ijisebc.com](http://www.ijisebc.com)
- Pericchi, L. (2002). Análisis de Decisión, Inferencia y Predicción Estadística Bayesiana. Universidad Simón Bolívar. Edición corregida.
- Silberschatz, A., Henry F., & S. Sudarshan. Fundamentos de bases de datos. Mc Graw Hill. Tercera Edición. 1998.
- TICbeat. Big data (2013): Los datos como valor de negocio. Recuperado de <http://www.ticbeat.com/bigdata/big-data-los-datos-como-valor-de-negocio/>
- West, M. & Jeff H. (1989). Bayesian Forecasting and Dynamic Models. Springer - Verlag.
- Wilson, H., Barry K., & Galt J. (2007). Business forecasting with accompanying excel-based forecast software. Mc Graw Hill. Fifth Edition.
- World Economic Forum (2015). Three ways big data can improve your supply chain. Recuperado de <https://www.weforum.org/agenda/2015/05/3-ways-big-data-can-improve-your-supply-chain/>
- Yogesh, J., Rand, W., & Monner, D. (2014). Part F Digital Marketing and Social Media: Predicting the Charts: Using Big Data from Social Media to Forecast Market Potential. AMA Summer Educators' Conference Proceedings, 25F-20-F-21. Recuperado de <http://search.ebscohost.com/login.aspx?direct=true&db=bth>

**How to cite this article:**

Fargab Ahmed et al. 2017, Optimizing Statistical Forecasting Models in real Time With Bayesian Component Using 'Big Data' In Marketing. *Int J Recent Sci Res.* 8(5), pp. 17027-17036. DOI: <http://dx.doi.org/10.24327/ijrsr.2017.0805.0268>

\*\*\*\*\*