International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:9, No:6, 2015

Forecast Based on an Empirical Probability Function with an Adjusted Error Using Propagation of Error

Oscar Javier Herrera, Manuel Ángel Camacho

***Abstract*—**This paper addresses a cutting edge method of business demand forecasting, based on an empirical probability function when the historical behavior of the data is random. Additionally, it presents error determination based on the numerical method technique 'propagation of errors.' The methodology was conducted characterization and process diagnostics demand planning as part of the production management, then new ways to predict its value through techniques of probability and to calculate their mistake investigated, it was tools used numerical methods. All this based on the behavior of the data. This analysis was determined considering the specific business circumstances of a company in the sector of communications, located in the city of Bogota, Colombia. In conclusion, using this application it was possible to obtain the adequate stock of the products required by the company to provide its services, helping the company reduce its service time, increase the client satisfaction rate, reduce stock which has not been in rotation for a long time, code its inventory, and plan reorder points for the replenishment of stock.

***Keywords*—**Demand Forecasting, Empirical Distribution, Propagation of Error.

I. Introduction

IN the modern business world, where efficiency and resource efficacy are necessary, it is important to be exact when planning productive activities, and it is consequently vital to use ever more precise tools that allow one to find values as close to reality as possible, or values with minimal errors in calculation. Hence, for better planning and business management, there is an inevitable need to make forecasts in order to improve accuracy in the decision-making process [1].

Therefore, it is vital to use analytical techniques that take into account the use of systems variability. Now, bearing in mind the different market dynamics, the variable of demand is important in every productive management activity. Hence, the technique developed here considers the randomness of the data used in the demand estimation. This makes it a better tool to estimate forecast uncertainty in electric services, in comparison to the methods that use distribution of demand scenarios to forecast [2].

Additionally, the errors in the quantities derived from the forecast have been determined analytically through Taylor series expansion in terms of the central values, using propagation of error as a way to propagate errors in mathematical functions. This is the appropriate method to solve many simple problems [3] like the one presented in this paper.

Oscar Javier Herrera is with the Industrial Engineering Department, Universidad Cooperativa de Colombia, Av. Caracas No. 37-65, Bogotá, Colombia, (phone: 051-3107576027; e-mail: ojavierho@gmail.com).

Manuel Ángel Camacho is with the Industrial Engineering Department, Universidad Libre de Colombia, Calle 8a No. 5-80, Bogotá, Colombia, (phone: 051-3112392586; e-mail: manuel.camachoo@unilibrebog.edu.co).

II. Theoretical and Methodological Structure

First of all, it is useful to highlight the role of forecasting in the decision-making process at the management level in any business activity. Using forecasting helps to determine the number of items that need to be produced and managed, or the number of required units in order to competitively develop productive operations [4]. Furthermore, forecasts are more important than ever in the production systems controlled by the current market [5]. Thus, having highly accurate predictions is very important in order to obtain good productive decisions that allow for reduction in costs.

Consequently, forecasting is understood, in its simplest form, as *the "prediction of the future, looking at the regularity of any specific phenomenon"* [6]. Deepening the concept, a forecast is known as the estimated value of the future levels of behavior of a variable (it could be, generally, the level of demand or the level of sales) [7]. These measurements may have a strong impact on every functional area of an organization because they are used in many decision-making processes.

Generally speaking, it is not possible to imagine business management without forecasting. Managers need to predict variables such as future sales, price levels, required capacity, technological and legal trends, etc. Now, while most businesses do not have formal forecasting systems despite their advantages, other businesses have developed schemes to systematically approximate the values that the selected variables will take in the future [8].

According to [9]; "the formal forecasting systems have two types of advantages: 'classic advantages' and 'world-class advantages'. Regarding the former, forecasting systems help with the reduction of stocks and fixed capital, the reduction of undelivered purchase orders, the establishment of a realistic basis for planning, the construction of reasonable business scenarios, the possibility to analyze the joint impact of various variables, and the overall capacity to explain reality. Regarding 'world-class' practices, these are the working methods implemented by worldwide successful businesses, no matter the location of their operations. Here, forecasting is aligned with continued learning based on the business' mission and its values and aspirations, in response to efficiency standards". These estimates are not one hundred percent accurate, which causes some errors. These errors measure the accuracy of the forecasting model that has been used, comparing the predicted values with the real or observed values. If F(t) is the forecast in the period of time t, and A(t) is the real demand for the period of time t, the forecast error (or deviation) is defined as: *real demand – forecasted value* = A(t) – F(t). The most used forecasting method is *Mean Absolute Deviation* (MAD), whose value is calculated by adding the absolute value of the individual forecast errors and dividing the result by the total of data periods (n). There is also the *Mean Squared Error* (MSE), which is the second method of measuring the global forecast error, and it is defined as the average of the square of the errors (predicted values minus observed values). Finally, another method frequently used is the *Mean Absolute Percentage Error* (MAPE), which is calculated by the average of the errors and is expressed as a percentage of the real values.

Regarding the theory of numerical errors, these errors are generated by the use of approximations, in order to represent the mathematical operations and quantities [10]. Taking as a basis for this work the propagated error whose error is possible in the output (response) generated in subsequent steps because of the occurrence of a previous error [11]. This is the type of error obtained in the output (response) generated in successive stages, due to the occurrence of a previous error. This is taken as a technique to determine the way numbers can be propagated in mathematical functions, specifically in the forecasts used in productive management. This type of error allows a function $f(x)$, with an independent variable x, where x' is an approximation of x. Therefore, this analysis attempts to evaluate the effect of the discrepancy between x and x' in the value of the function, as:

$$\Delta f(x') = |f(x) - f(x')| \quad (1)$$

The difficulty to evaluate $\Delta f(x')$ lies in the fact that $f(x)$ is unknown, because x is unknown. This difficulty is overcome if x' is close to x and $f(x')$ is continuous and differentiable. If these conditions are satisfied a Taylor series is used to calculate $f(x)$ close to $f(x')$.

$$f(x) = f(x') + f'(x')(x - x') + f''(x')(x - x')^2 + \cdots \quad (2)$$

Removing the second term and those of higher order and rearranging them gives:

$$f(x) - f(x') \equiv f'(x')(x - x') \; or \; \Delta f(x') = |f'(x')|(x - x') \quad (3)$$

where $\Delta f(x') = |f(x) - f(x')|$ represents an estimation of the error of the function y $\Delta(x') = |x\text{-}x|$, represents an estimation of the x error. Equation (4) provides the ability to approximate the error in $f(x)$ giving the derivative of a function with an estimation of the error in the independent variable.

Probability Distribution

Given that probability distribution indicates all the range of values that can be represented as the result of an experiment, if performed, it can be said that the Probability Distribution describes the probability of an event taking place in the future. It is a fundamental tool for forecasting because a scenario of future events can be designed, taking into consideration the current trends of natural phenomena [12]. Thus, the distribution of empirical probability essentially depends on the probability of the occurrence of the respective random variable, which takes statistical values obtained via measurements in some type of random experiment [13]. This uses for its generation the weight of this probability through the establishment of ranges in which the respective percentage is established, taking into account the accumulated probability [14], [15].

The procedure to follow is the generation of the respective random numbers in order to later establish the random variable by determining in which probabilistic range it is located, and thus the corresponding value of the variable is assigned to it. Some properties of empirical probability [16];

Given the sample space $S = \{S_1, S_2, \ldots, S_n\}$ and the event $\{s\}$ i and the empirical probability $P(s_i)$, then: $0 \leq P(S_i) \leq 1$

The empirical probability of each result is a number between 0 and 1. $P(S_1) + P(S_2) + \ldots + P(S_n) = 1$

The empirical probabilities of all results add up to 1. *If* $E = \{e1, e2, \ldots, en\}$, *then* $P(E) = P(e1) + P(e2) + \ldots + P(en)$.

The empirical probability of an event E is the sum of the empirical probabilities of the individual results in E.

The calculation of the empirical probability when the results are equally probable in an experiment, the empirical probability of an event E is:

$$P(E) = \frac{Number\ of\ favorable\ results}{Total\ number\ of\ results} = \frac{n(E)}{n(S)} \quad (4)$$

("Favorable results" are the results in E)

III. Implementation Context

The business where the present work was developed is called Hombresolo S.A. and is located in Bogota, Colombia. This is a maintenance services business (electric, hydro-sanitary and civil engineering) whose main focus is 100% satisfactory customer service. This is why, through the determination of a *Forecast based on an Empirical Probability Function,* we pursued the creation of the adequate stock of products required by the company to provide its services, to help the company reduce its service time, increase the client satisfaction rate, reduce stock which has not been in rotation for a long time, code its inventory, and plan reorder points for replenishment of stock.

Regarding customer service, the company has a high demand for immediate services, which are mainly urgent or high-priority needs (which means that they must be tended to in 8 days or less). The service is coordinated by professionals, whose work is assigned according to their areas of specialty.

The practical use of establishing a forecast based on an empirical probability function, was based on the random behavior of the elements or products required for the company's services (without presenting cycles or detectable trends), which are not easily predictable, as shown in Fig. 1.

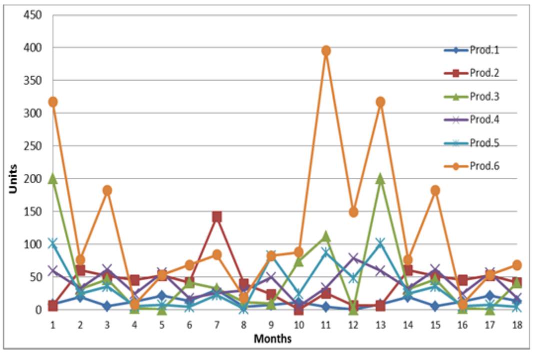

Fig. 1 Historical Data of Products 2010-2011

This behavior suggests the use of a probabilistic behavior, not defined by any known probability function. This led to the use of an empirical probability function, which was established specifically to obtain the forecast of two aspects essential to this research that respond to the working conditions of the company as seen here:

1) *Level of demand adjustment* is the value that indicates the level of increase or decrease in demand, established by the analyst, depending on the possibility of winning contracts or other factors. Therefore, a scale of five levels of change is given as: (5) maximum increase, (0) no variation, and (-5) maximum decrease.
2) *Quantity variation factor* is a value of increase or an addition of demand compared to current demand, and it depends directly on the level of adjustment previously established by the analyst.

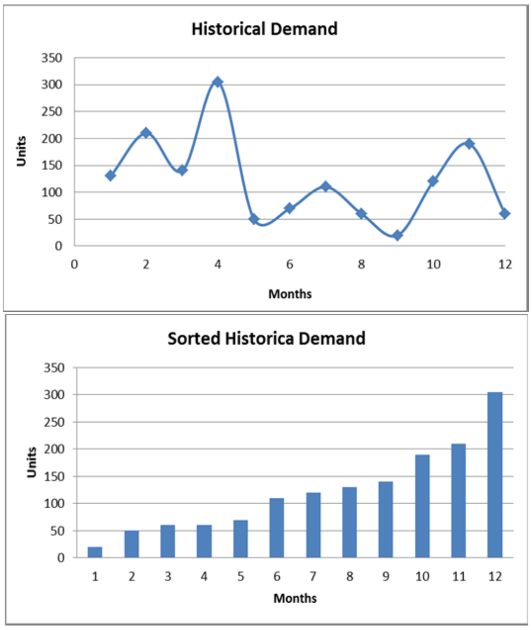

Fig. 2 Historical Demand Data for 2011 and Historical Data rearranged to suit the reference product

In order to simplify the development of this work, only one product will be assessed in the procedures. This is the most important product for the company due to its high levels of rotation, as well as its higher demand and income production. First, the historical data of the demand will be established for the year 2011. These values will be rearranged incrementally in order to be able to observe the ranges of demand in which the data shifts. This is an initial factor that will allow for the determination of the central value of the random variable that will be used in the forecasting process, which can be seen in Fig. 2. Then, considering the above, the *level of demand adjustment* was established as shown in Fig. 3.

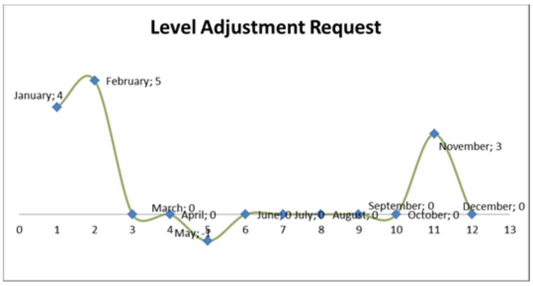

Fig. 3 Values of Level of Demand Adjustment for each forecasted month for 2012

Here, one can observe the estimation of the increase in value of the demand established by the analyst, depending on the probability of winning a contract etc.: increases of 4, 5, and 3 times in January, February and November, respectively, with 2011 as the base year for the demand. Likewise, a decrease of 1 is observed in May; while in the rest of the months, the levels of demand of 2011 remain constant. These values are established according to the probabilities of new business or contracts that the analyst has projected for a service in 2012.

The variation factor for the referenced product, projected for 2012, is 20, meaning it is established by the stock management expert and by the marketing area, and there is an increase of 20 times the demand in comparison to that of 2011.

Now, according to the ranges in which the historical demand shifts and to the frequencies of the values in each range, the procedure of an empirical distribution is established in order to find the new values of the demand forecast, as shown here:

TABLE I
EMPIRICAL DISTRIBUTION DATA MANAGEMENT IN ORDER TO FIND THE DEMAND

Range.	Variable	Freq.	Prob.	Accum.	Ranges Rx
0-50	25	2	16.7%	16.7%	0.0 - 16.7
51-100	75	3	25.0%	41.7%	16.8 - 41.7
101-150	125	4	33.3%	75.0%	41.8 - 75
151-200	175	1	8.3%	83.3%	75.1 - 83.3
201-250	225	1	8.3%	91.7%	83.4 - 91.7
251-300	275	0	0.0%	91.7%	--
301-350	325	1	8.3%	100.0%	91.8 - 100

Finally, after generating random numbers (Rx) between 0 and 1, and considering the Level of Adjustment of Demand

and the variation factor, the forecast value (Ft) is established by adding the variable obtained in the empirical distribution: the final value in the demand variation, which is obtained by multiplying the level of adjustment value by the variation factor. This is why, as a matter of example, 75 units are given for the first level of demand forecasted for January 2012, according to the use of the empirical distribution, and then 80 are added (which comes from multiplying the variation factor 20 by the level of adjustment 4). Then, 155 units are obtained as the final forecasted demand. In Table II the following data is presented:

TABLE II
DEMAND FORECAST FOR 2012

Month	Random	Ft
January	0.35534	155
February	0.54053	225
March	0.52473	125
April	0.93969	325
May	0.49260	105
June	0.50057	125
July	0.46153	125
August	0.26857	75
September	0.35663	75
October	0.57106	125
November	0.56700	185
December	0.38043	75

Fig. 4 shows the behavior of the historical data and the forecast for 2012.

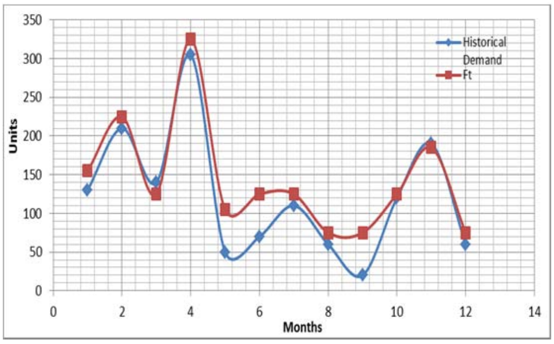

Fig. 4 Demand vs Forecast Ft

In this paper the numerical method technique *Propagation of Error* was used, in order to calculate the error of the forecast. In this method it is necessary to establish a mathematical function that, in this case, relates the historical demand data with the forecast data. Fig. 4 shows that the behavior of these two variables is interrelated through time (months), because when there is a variation in one variable (demand), the other (forecast) varies too. The smaller the difference is among these two variables, the smaller the forecast error.

The propagation of errors technique requires, as said before, that the combination of a dependent variable and an independent variable, or more, propagates the error in a mathematical function. In this specific work, the forecast used in productive management is considered. The ascending correlation value between the historical demand and its forecast was 0.9607; therefore, a linear regression analysis between these two variables was performed. The analysis considers the dependency of a variable in function on another explanatory variable, with the goal of exploring or quantifying the behavior of the first from a set of known values or fixed values of the second [13]. In Fig. 5 and based on this comparison, for a 12-month period, an equation (5) was established for the demand forecast, with a proportion of the adjustment of the dependent variable (forecast) in a linear form of 92.3% in correlation with the quantity demanded. Therefore, for January 2012, the forecast (or approximation of the real value) Ft *(x´)* is 155 units and the error $\Delta(x')$ is 17.4%. It was obtained by calculating the absolute difference, or error difference, between the total of the historical data and the total of the predicted data in terms of percentage of the total of the historical demand.

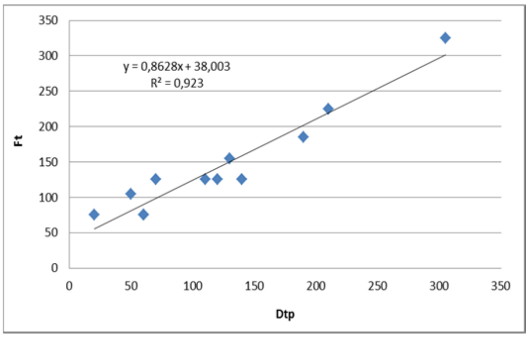

Fig. 5 Demand vs. Forecast and its regression function

Now, considering the forecast function obtained through the regression analysis and the forecast used for January 2012, with its respective variation based on the previously-obtained error, the estimation of the error of the forecast function $\Delta f(x')$ was calculated. This was done according to the theory of error propagation, as was explained in the theory section of this paper. In order to get to this point, the respective forecast function was derived, verifying that it was continuous and differentiable, and was then multiplied by the forecast error. The data and the obtained result are presented here:

- Forecast function:

$$f(x) = 0.8628x + 38.003 \quad (5)$$

- Dependent variable approximation: $x' = 155$
- Error estimation of x; $\Delta(x') = 0.174$

Calculating the variation of the real value through the derivative of the forecast function the following value is obtained:

$$\Delta f(x') = 23.3$$

Finally, the function value for the forecasted value for the month of January is obtained, along with its error calculation:

$$f(155) = 171.7 \pm 23.3$$

Likewise, the rest of the values for the other months of the year are obtained as shown in Table III:

TABLE III
FORECAST FT. AND MONTHLY ERROR

Month	Final Forecast		Error
January	171.7	±	23.3
February	232.1	±	33.8
March	145.9	±	18.8
April	318.4	±	48.8
May	128.6	±	15.8
June	145.9	±	18.8
July	145.9	±	18.8
August	102.7	±	11.3
September	102.7	±	11.3
October	145.9	±	18.8
November	197.6	±	27.8
December	102.7	±	11.3

IV. CONCLUSION

The result of this research was the establishment of a novel technique of forecasting demand based on empirical probability distribution and, most importantly, the use of the procedure of error propagation as a numerical method technique to effectively calculate error, establishing system in decision, two control elements for changing demand when it is random. On the one hand, the forecast based on empirical probability distribution was well-evaluated through the different calculations of forecast error, when there is randomness but also when one can anticipate increases or decreases in demand can be provided (in certain periods) along with their respective variation size. These were included in the model via the *adjustment level* for the former and the *variation factor* for the latter. These values are given by the owner of the process, who are usually the stock analysts and commercial agents of a business. On the other hand, it was observed that using the propagation of error technique to calculate the interval of variation of error in the forecast function obtained good results for the desired estimation of variation, since the forecasted real value for January 2012 was 155 units of the reference product, which was within the limit assigned with the propagation of error method (148.8 and 195 units). This can diminish uncertainty in future prediction in businesses and can help make better decisions in the productive area, therefore serving as a mechanism for increasing efficiency in different businesses and in the productive sectors in general. It is important to highlight that the type of demand forecast used influences this estimation of error. In this case in particular, the method of empirical distribution was used.

The final result of the methodology carried out in the business was the adequate stock of the products required to provide services. This reduced the service time, increased the client satisfaction rate, reduced stock which had been out of rotation for quite some time, coded its inventory, and helped plan reorder points for replenishment of stock.

To sum up, applying this technique on a real case, where pertinent information is at hand, shows that numerical method techniques are very useful for improving levels of calculation within most management operations in real systems, such as planning, programming, and controlling productive activities.

In the case of this particular paper, improvement in the levels of trust for the forecast estimation were achieved for a group of specified products, using a traditional technique in order to define different strategies for productive operations and management operations. This gives the decision-makers greater flexibility to face the natural dynamic of the productive systems.

This paper presents great opportunities for research, for example, if the multiple regression functions are linear or not to apply the technique of error propagation, in addition to assessing its use with other forecasting techniques.

REFERENCES

[1] J. M. Fernández Güell. *El diseño de escenarios en el ámbito empresarial* (Book style). Ediciones Pirámide, España. 2004
[2] J. W. Taylor, R. Buizza. *Using weather ensemble predictions in electricity demand forecasting*, International Journal of Forecasting, Volume 19, Issue 1, 2003, pp 57-70.
[3] J. Swain, L. Taylor. *Numerical construction of likelihood distributions and the*. Nuclear Instruments and Methods in Physics Research, 1998, pp 153-158
[4] L. S. Ensenada. "*Importancia Estrategica del pronóstico de Demanda*". Mexico, BC. 2010
[5] D. Sipper, R. L. Bulfin. "*Planeación y Control de la Producción*". México D.F.: McGraw-Hill, 1998
[6] S. G. Makridakis. "*Pronósticos, Estrategia y Planificación para el siglo XXI*". Madrid (España): Ediciones Días de Santos, S.A. 1993
[7] O. J. Herrera. "*Optimización de Sistemas Logísticos*". Memorias, Diplomado Gerencia en Logística (pp. 89-99). Neiva, Huila: Unidad de desarrollo Empresarial UDE. 2011
[8] S. Makridakis, S. C. Wheelwright. "*Forecasting Methods for Management*". New York: John Wiley & Sons, fifth Edition. 1990
[9] E. Yacuzzi, G. Paggi. *Diseño e implementación de un sistema de pronóstico de ventas en Whirlpool Argentina*, Serie Documentos de Trabajo, Universidad del CEMA: Área: negocios, No. 209. 2002
[10] S. C. Chapra, P. Raymond. "*Métodos Numéricos para Ingenieros*". México: Mc Graw-Hill. 1995
[11] N. Hurtado. "*Métodos numéricos aplicados a la ingeniería*". México: Continental. 1997
[12] S. Murray, J. Schiller, R. A. Srinivasan. "*Probabilidad y Estadística*". Mc Graw-Hill. 2011
[13] D. Peña Sánchez de Rivera. *Fundamentos de Estadística* (1ª edición). Alianza Editorial. 2008, pp. 688
[14] R. Delgado de la Torre. "*Probabilidad y Estadística para Ciencias e Ingenierías*". Galicia: Delta publicaciones. 1ra Ed. 2008
[15] D. C. Montgomery, G. C. Runger. *Probabilidad y estadística aplicadas a la ingeniería*, Segunda edición. Limusa Wiley. 2002
[16] S. Waner, S. R. Costenoble, *Finite Mathematics and Applied Calculus*, February 2000. (ref. of June 7, 2012). Available on Web: http://www.zweigmedia.com/RealWorld/tutindex.html.