

Least Squares Method and Empirical Modeling: A Case Study in a Mexican Manufacturing Firm

RESEARCH-ARTICLE

Raúl Hernández-Molinar*, Roberto Sarmiento-Rebeles and César F. Méndez-Barrios

[Show details](#)

Abstract

Empirical modeling (EM) has been a useful approach for the analysis of different problems across a number of areas/fields of knowledge. As is known, this type of modeling is particularly helpful when parametric models due to a number of reasons cannot be constructed. Based on different methodologies and approaches (e.g., Least Squares Method, LSM), EM allows the analyst to obtain an initial understanding of the relationships that exists among the different variables that belong to a particular system or a process.

In some cases, the results from empirical models can be used to make decisions about those variables, with the intent of resolving a given problem. The investigation describes the application of EM to the estimation of shipping costs in a Mexican manufacturing firm. The results show that overall, transportation costs using an empirical model tend to be lower than costs calculated by a previous model. This demonstrates the practical and potential utility that results based on EM can have in a real-life setting.

Keywords: empirical modeling, exploratory data analysis, least squares, linearization, transportation logistics

1. Introduction

It is well known that researchers can use empirical modeling (EM) to have a better understanding of a particular problem. This type of modeling can be improved by the expert input of analysts. When investigating a particular system or process, it is always preferable to perform both exploratory/initial and confirmatory analyses of the available data and information. Nevertheless, in some cases, it is not possible to do the latter. This means that oftentimes, professionals in positions of authority have to make decisions about important variables and problems based solely on the results from initial/exploratory models.

This chapter describes the application of EM to investigate the variables associated with shipping costs in a Mexican manufacturing firm. The objective was to obtain a model that would offer a better idea of the variables and dynamics that determine those costs. To this end, the Mexican company formed a research team tasked with a complete and detailed analysis of the problem.

Using a Least Squares Method (LSM) approach, the team proposed a new model capable of estimating transportation costs of containers shipped in vessels from Europe to a port in Mexico. Using the proposed model, the firm's management was able to make comparisons between the actual costs incurred based on a previous model (formulated by the provider of the shipping service) and the estimated costs based with the new model.

The results show that in general, cost estimates from the new model tend to be lower than those of the previous model. These results allowed the Mexican firm to start new negotiations about their shipping costs with the provider of the transportation service

2. Empirical modeling: an overview

The main objective of this section consists in reviewing the concept called EM and some other concepts employed when an investigator begins the exploration of the information. Another important objective is to suggest the use of a linear model as an important resource to clarify and propose a fitted empirical model based on the observation of the data when a special transformation process of the variables is realized.

In reference [1] comments that empirical models are guided exclusively by data. Analysts attempt to find a model that reflects trends in data to make predictions instead of explaining behavior. In particular [1] underlines the potential utility of statistical approaches/tools (e.g., regression analysis) when doing EM. As is known an empirical model can aid researchers in acquiring an initial idea of the relationship between two or more variables that are representative of a particular system or process. In spite of its inherent limitations, the results obtained using empirical models can sometimes help researchers when decisions need to be made with respect to the variables that intervene in the system/process under study.

Empirical knowledge can be understood as those instances when new information/knowledge is acquired by practical/experiential means. While this type of knowledge is undoubtedly valid and useful, it should be noted that in some cases, the conjectures/conclusions we make about observed data and results are based on the analyst's own experience and interpretation. This means that sometimes, impartiality and scientific rigor in the analysis of data and results might be difficult to achieve. Consequently, inconsistencies between the real-life problem and the model proposed by the analyst can be found. It is important to consider, as reference [2] suggests, that when modeling is applied to any logistics system, flexibility must be considered.

This being said, influential thinkers and intellectuals have vigorously debated the topic of whether full certainty can be achieved with respect to the validity and representativeness of a model. For example, reference [3] argues that empirical knowledge plays an integral role in the development of so-called "scientific knowledge." This is because scientists have the opportunity to explore and confirm particular ideas/conjectures on the basis of their own empirical findings.

Under a scientific and formal context, Exploratory Data Analysis (EDA) based on empirical information requires probability and statistical concepts. However, reference [4] mentions that there exists a moment where exploratory and confirmatory data analysis must be distinguished between confirmatory nonparametric statistical data analysis, or modeling, and confirmatory parametric statistical data analysis.

It is clear that when there is no information to propose a parametric model, an exploratory analysis using empirical knowledge to obtain an initial model and solution can be justified. After this step, the analyst can judge, based on his/her expertise, whether the initial model is an adequate representation of the relationships that exists among the different variables that are part of the problem under study. Consistent with this, reference [5] also says that when it is not possible

to justify the behavior of the data, an empirical model can be utilized to obtain an initial idea vis-à-vis the nature of problem of interest.

Generally speaking, EM uses nonparametric data analysis to explore trends or behaviors within the available data. It is assumed that models based on well-defined parameters and distribution functions cannot be formulated due to incomplete data/information. This type of modeling also assumes that variables belong to sample spaces where uncertainty is present.

EM can be used to represent real-life problems that require nonanalytical methods. Examples of areas/fields where EM has proven useful include industry, science, technology, engineering, medicine, biology, and management. It should also be said that more powerful computers are of immense aid when researchers use EM, especially in those situations where high uncertainty exists.

Given the uncertainty and incompleteness associated with empirical models (along with the sometimes necessary expert input of the analysts in the definition of a model), it is evident that results and information derived from these models cannot be generalized. Adding to what has been discussed already, reference [6] notes that “Exploratory data analysis seemed new to most readers or auditors, but to me it was really a somewhat more organized form—with better or unfamiliar graphical devices—of what subject-matter analysts were accustomed to do”.

We now sum up some of the salient characteristics and benefits of EM: it is mainly based on observed empirical data. However, it can also include the expert judgment/opinion of analysts. The data involved in the empirical model belongs exclusively to the realm of the system or the process that is being investigated. This means that there is no input from variables, parameters, or principles that fall outside the scope of the problem under study. Empirical models are capable of generating feasible solutions that can be helpful when investigating a particular problem. This in turn can guide analysts when decisions have to be made with respect to the variables associated with the problem of interest.

In addition, two appendices are annexed to review issues about the modeling process and outline the general numerical method that uses least squares as criteria to select an empirical model.

3. Case Study: estimation of the total cost of transportation to create a future budget

3.1 BACKGROUND

This section discusses the case of a firm that has operations in Mexico (heretofore referred to as MF, “Mexican firm”). We now proceed to describe briefly the problem at hand: every month a sea shipment is dispatched from Europe to a port in Mexico by an affiliate of MF. Each shipment contains items that are needed for the daily operations of MF. Originally, the cost of each shipment was based on a model calculated by the company that provides the transportation service to MF. We will refer to this model as OM (“old model”). The shipping costs can vary according to the quantity of items that are being transported in the different containers included in the vessel. The combinations of items (and their respective quantities) that are transported in any shipment/container are determined according to MF’s forecasted needs.

As part of their cost-saving initiatives, MF decided to investigate whether their transportation costs could be reduced. In particular, they decided to come up with their own cost-projection model to compare its estimates with those provided by OM. In this way, a more realistic estimation of their shipping costs could be obtained. To accomplish their objective, they decided to utilize historical empirical data to calculate a new model (“NM”) that would provide a more accurate idea of the monthly costs associated with each shipment. Evidently, more accurate cost estimates can result in better budgeting decisions and its associated benefits.

To accomplish their objective, MF’s top management made the decision to conduct a detailed analysis of the situation. A research team tasked with proposing a model that would be an adequate representation of the problem was formed. One of the first and most important activities of the team was the conceptualization and understanding of the different variables upon which the monthly transportation budget depends. It was observed that the cost of a given sea shipment is a function of at least one hundred variables. These variables include the value of goods, number of pallets, sea freight charges, unitary cost, and volume of the shipped items, among others.

A key step in the research process was making sure that the data pertaining to the above variables was reliable and representative of the problem to be modeled. Reference [7] warn us about the relevance in the clarification between the forecast and the planning of the variables under study. For example, MF had information about a number of variables that were not relevant to the problem (e.g., information about items that were being shipped from the USA). This meant that the database had to be deputed in great detail. Once the database was deemed reliable, the research team began to analyze the potential relationships among the set of variables of interest. Evidently, the dependent variable (transportation/shipping cost and TC) in the modeling process has to be a function of a group of independent variables such as the ones described in the previous paragraph. It needs to be specified that the main unit of analysis is the container in which the different items are transported by sea. A maritime cargo shipment usually carries several containers.

The research team examined a number of different types of models (e.g., linear, quadratic, and exponential) that could best fit the relationship between TC and its determinants [8]. After different tests and analyses, it was found that a linear model represented this relationship best. In particular, a linear model using the LSM was proposed. As is known, this method offers a best-fit model that minimizes the sum of the squares differences (errors) that exist between the real observations and the ideal results proposed by the model. The well-known general model is defined as follows:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (1)$$

Options

With respect to the defining function for this problem, the research team made the decision that the final set of independent variables should be the result of all those items that appeared at least once in the historical records. In other words, if an item was recorded as being shipped and received at least once, the research team decided to include it in the general model for TC. The proposed Least Squares Model has TC as the dependent variable that is a function of potentially more than one hundred independent variables.

As will be made clear later, the quantity of independent variables to include in the model to calculate TC for a given shipment and containers will depend on previous records of shipped items. Put differently, records could suggest that TC be defined by, for example, 80 items in one month, while 70 items could be used to estimate TC in the next month.

3.2. A COMPARISON BETWEEN OM AND NM ESTIMATES OF SHIPPING COSTS

We now proceed to exemplify the differences between the estimated costs using the model originally proposed by the transportation company (OM) and the model resulted from the analysis by MF's research team (NM). The results in [Table 1](#) are based on data provided by them. More specifically, the costs under the OM column reflect historical records (i.e., they are costs pertaining to completed shipments). The calculations in the NM column reflect the estimated costs had this model been used for a particular completed shipment

3.2.1. USING MLS METHOD FOR ESTIMATING THE TOTAL COST BASED ON SHIPPING PART COSTS

It is clear that linearization process is useful when several first order variables are participating in a model reference [1]. In the present study, at least hundred variables can be interacting to define the total cost of the shipment transportation. In this case, several variables (more than hundred) were considered to estimate the cost per shipment, for instance, value of goods, number of pallets, sea freight charges, volume, and unitary cost. After a serious selection process based on historical information and the expertise of the personal, a matrix considering shipment identifier and the cost of each of the parts is created. Using the historical information, a vector with the β_i coefficients is estimated using the LMS, and these are used to estimate the cost assigned to each shipment.

The LSM determines the best fit that minimizes the sum of squares magnitudes between the observed responses and those that are predicted by the model. A detailed explanation related to the method can be reviewed in references [9–12].

We know it is possible to predict the Y values by using the estimated model parameter values. We also know that the values can be generated from the following model

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad \hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (2)$$

[Options](#)

The sum of the squares deviations generated from the observed values of Y and corresponding values predicted using the regression model estimated.

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}))^2 \sum_{i=1}^n Y_i - \hat{Y}_i^2 = \sum_{i=1}^n Y_i - \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}^2 \quad (3)$$

[Options](#)

We need to recall that least squares solution consists in finding the values of estimators

$$\beta_0 \beta_1 \beta_2 \beta_k \quad (4)$$

[Options](#)

which are called least squares estimators. The minimum sum of squares is called the residual sum of squares, the sum of squares of the error, and the sum of squares due to regression. Based on the estimated values, the estimated budget is defined for each shipment.

3.3. CONSTRUCTING THE ESTIMATED BUDGET USING AN EMPIRICAL MODEL

Based on the linear model generated, an empirical model to forecast a budget considering the total cost on the budget is proposed. The coefficients estimates for determining the shipment cost per part in the corresponding container are generated using the Least Squares estimation method. [Table 1](#) shows an example for the estimation on 11 containers.

[Table 2](#) shows the estimated values generated with the MLS method for each shipment freight. It is evident that the cost associated to the land freight is constant. The estimated cost values were determined using a multiple linear model, which consider several factors were chosen by the experienced personal in the company. The empirical model that suggests the budget for the future is showed in [Figure 3](#).

Shipment freight	Container ID	OM estimates in USD	NM estimates in USD (USD)	Net difference (OM-NM)
1	1179464	2267.59	3442.27	-1174.68
2	7237802	8016.16	6661.91	1354.25
3	3311245	1871.40	1895.46	-24.06
4	9727730	7788.40	5996.43	1791.98
5	3544695	2849.20	1009.20	1839.99
6	359446	5001.89	1949.77	3052.12
7	7499748	2346.92	4122.16	-1775.25
8	1218072	5272.18	2451.45	2820.73
9	4958920	5582.10	3972.21	1609.90
10	8005021	2113.78	2570.21	-456.43
11	5503140	5578.27	4699.86	878.41
	MEAN	4310.96	3407.11	903.86
	TOTAL	48687.90	38770.94	9916.96

TABLE 1.

A comparison between historical records of shipping costs (OM column) and estimated costs using NM for 11 containers.

[Figure 1](#) also illustrates the calculations made in [Table 1](#).

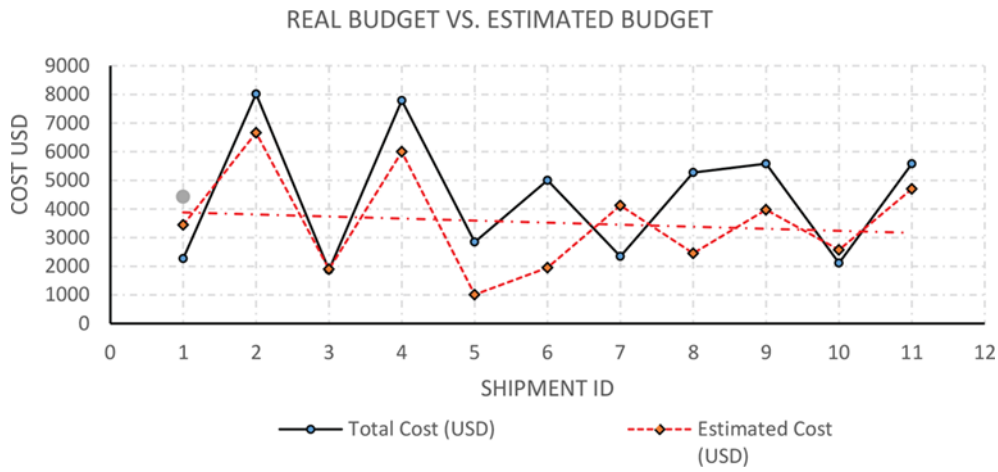


FIGURE 1.

Real and estimated budget.

From the 11 comparisons between OM and NM estimates, it can be observed that the net difference is negative in four instances. However, the cumulative net difference shows that overall, NM offers a lower estimate of the shipping costs (savings of \$9,916.96 in the total budget). This suggests that from MF’s perspective, their proposed model (NM) could be used to obtain lower estimates of their transportation costs. This overall difference is made clear once the LSM estimates of both OM and NM are calculated.

[Figure 2](#) illustrates the difference between these estimates. These two linear models have been estimated based on the OM and NM values. It is clear that NM estimates are, in general, lower than OMs. As was said before, this suggests that from MF’s perspective, the use of NM’s calculations would benefit them in the long run.

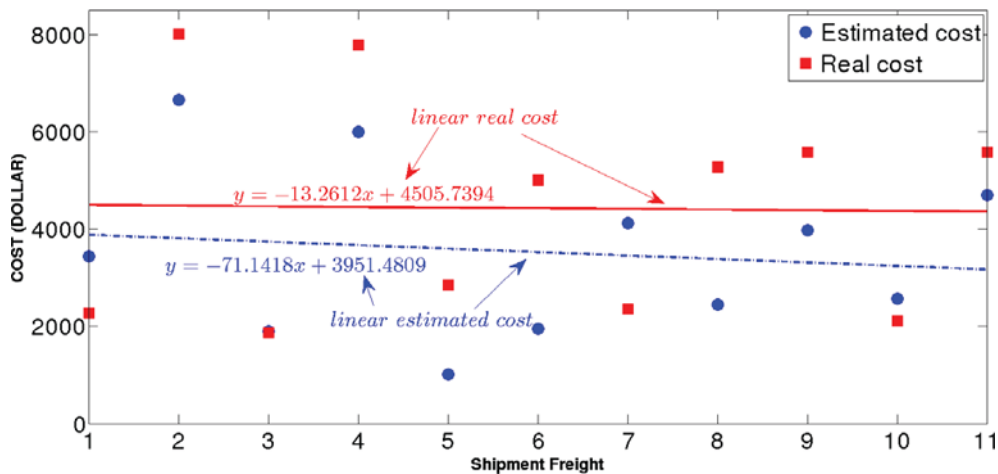


FIGURE 2.

Comparison between real and estimated budget.

In order to probe the validity of the proposed model (NM) we can observe that in most of the cases the goodness of the model is associated with well-balanced residual values above or below a reference axis. This permits to be sure that there is no overestimation or underestimation of the predicted values.

4. Conclusion and further research

This case shows that EM can help in the forecasting process. Undoubtedly, modeling is usually a very common tool given the complexity and accuracy required in transportation problems as it is mentioned in references [2,14–19]. The described case also shows that the selection of the model is very important in any planning activity.

Despite some special programs that are able to generate the proposed models automatically, it has been made clear when information is not available or practically unknown, EM is an option that could help in the generation of structure, method, and formal knowledge. It is important to recall that the main objective in this approach is to find the best model that can represent the relationship between the variables under study, and EM is useful to do it.

The empirical model proposed is pioneering the decisions in the corporation, and it has been implemented with success. There is still interest in the improvement of criteria to upgrade the multiple linear models to estimate the containers' cost, but until now this proposal has given good results. Although this is a novel and simple approach, it is possible to mention that the combination of available data with the experience of personnel has been helpful for decision-makers.

The LSM is used as an algorithm to generate estimates for a new model that the MF has been considered sufficient and pertinent to produce significant savings. The case study has been helpful to propose the relevant data to study and estimate relations in assigning the shipping cost, based also on the experience and knowledge of the company experts. The method helped in the construction of one empirical model supported for a linearization process and has provoked significant changes in the planning process of each monthly budget.

The model proposed in this research has provided successful results, however, the team continues using other exploratory data techniques to improve it. It is expected that in the near future, it would be possible to release other options to propose better forecast of the shipment freight budget. Further studies can be conducted using parametric models generated with statistical tools or through a deep analysis using polynomials to suggest more effective transformations.

The model to forecast the shipment freight budget proposed in this research has provided successful results; this conducts to better profits and sustainable growth.

Furthermore, the research team continues using other exploratory data techniques to improve the model. It is expected that in the near future it would be possible to release other options to propose better forecasts. Also, further studies can be conducted using parametric models generated with statistical tools or through a deep analysis using polynomials to suggest more effective transformations.

Appendix A

A.1. THE MODELING PROCESS

A model can be conceptualized as a mathematical description which is generated using knowledge, experience, and experts opinions, but based on data that were registered previously. As references [8,13] indicate, the data help in identifying the geometric or physical tendency of a potential model and those values that correspond to characteristic values representing

relevant parameters. An appropriate model suggests adjustment, or simplicity under a practical approach, and this must be conducted based on the good quality of the used information.

In general, the modeling process requires the consideration of the following issues:

- The knowledge of the system where the proposal will be applied.
- The definition of the objectives related to the activity of the system under study
- The identification of those variables that participate into the model
- A clear definition of the measurement system to quantify the variables to be revised.
- The analysis of models, algorithms, or processes that are more appropriate to get the objectives
- To achieve a detailed process of analysis of the obtained results that support the resulting alternatives
- To construct a detailed report indicating the way that the solution must be applied.

During the analysis of modelling process, the main idea is to elaborate a predictive model that helps to propose a better solution, and consequently to suggest an improvement in the indicators of the system. In order to do this is convenient for the identification of those trends or feasible models, which can be used as a reference during the process.

Another important aspect in modeling is to guarantee that data is representative of the problem under study. This requires a deep analysis of the relationships between the variables or specific sources, and to clearly point out the obtained empirical model destination.

One of the advantages in using EM is that they can conduct the right answer most of the time and does not require very formal information. This can be useful when a solution must be implemented promptly because the empirical model will be based only in the available information.

However, there is confusion about the goodness of using theoretical models instead of empirical models. It is not possible to declare that one type of model is better than the other because it depends of the specific context they are applied. The empirical models are useful when a theoretical model is not available. It is clear that the objective is to model scenarios with the best performance in order to solve a given problem or a simulation.

It is very common to apply empirical models when certain events in nature are not characterized by theoretical models, as those related to climate, air, environmental contamination, shipping, lifetime in active products, friction mechanisms trends, and etcetera.

Sometimes the use of data is not easy or is very expensive because they require long time to be obtained or not available for special causes. When this occurs, the EM is a practical option to create scenarios to simulate the behavior of the variables of interest.

It is known that many scientific, social, or engineering observations are generated through experimentation or observing the situation under study. Records of these values are stored in a data base. The information is analyzed and reported using several types of plots of the associated points.

With the available information, the investigators can apply different methods to propose formulas (equations) to formally represent the behavior of data. In most of the cases, the adjustment process considers the possibility of determining a function, to use transformed data that must be fitted to the observed values.

This approach indicates that it is very likely to propose similar results to those that a process sample would represent. Based on this, the researcher would be able to promptly represent the variables tendency under study.

A.2. DESCRIPTION OF THE MODELING PROCESS

In general, the modeling process can be described in several steps. Readers interested in this topic can also review in reference [1]:

1. **Definition of problem to be solved** Most of the times this step is not formally considered. However, it is necessary to establish clearly (by written) the main objective of the generation of a modeling process. It is common that this objective changes during the searching of the solution and one must be careful in order to avoid redundancies while modeling. Normally the definition of one or more questions should be sufficient to have a reference related with the main objective. The core idea is to answer questions that are easy to comprehend.
2. **Identification and selection of the model** The investigator must select the models based on the previous knowledge or experience. It is important to consider the feasibility and the possible adjustment to data tendency. Considering that the information collected by the investigators through experimentation and observation is called empirical data, there are some scenarios that can help to understand the selection process: for example, studies that use lifetime in bearings, clinical trials, medical information, contaminant emissions, residual waters, fertilizers, or insecticides; other examples are related to costs of transportation or air conditioner failure times in flying hours of airplanes.
3. **Definition of variables** It must be considered that a variable is the observation that can have a numerical value and that this value belongs to a variable sample space. Also they are called quantiles. It is desirable that a characteristic value can be determined based on the behavior of the studied values . The idea during the modeling process is to make assumptions about the most important variable or variables in the model. This will help to detect those variables that are not useful to represent the problem under consideration. Once the variables have been identified, it is important to use specific symbols to recognize them.
4. **Calibration Model.** All models must be calibrated using available data; for instance, data can be related to lifetime on mechanisms, humans, or products. The calibration activity requires a very careful analysis making comparisons

with the expected external responses. Also a thorough assessment process is required, given the importance of replication of the results in a systematic way.

5. **Validation Model.** Once the model has been calibrated, the model is validated to confirm that the behavior is well adjusted to data. Common statistical test can be achieved, for instance, the Kolmogorov–Smirnov test or a chi-squared test to guarantee a goodness of fit. Sometimes, the test is realized based on the experience or previous knowledge.
6. The validation is mandatory to review the tests that permit the verification of previously defined assumptions. Given the nature of the problem, there is no enough knowledge to describe the real analyzed system, accurately. It is necessary to take into account that all models are conceptualized based on a set of assumptions generated in the Step 1 or during the modeling process.
7. The adjusted models are used to compare against the corresponding phenomena, for example: using a well-validated model can lead in applying the property of unbiasedness. If there are other models, it is possible to analyze them at this moment. If in Steps 4 or 3, a model is not the appropriated; one can seek for other feasible models. There is the possibility that more than one model could be used, and there is an interest in propose them.
8. **Selection of the model** The valid models are chosen, and they are analyzed. Some criteria are generated to select the better option. It is possible to use the results of the tests or the comparison with other related models. It is important to consider that the models proposed can be used in the future.
9. **Implementation of the proposed model(s)** The analysis of results based on the model selected will help to simulate several scenarios useful to generate the final reports. A process of polishing is suggested in this step.

In case more or other data are collected or given the context has been changed, the iteration in the modelling process can be repeated. In general, the steps above mentioned can be summarized as is illustrated in [Figure A.1](#)

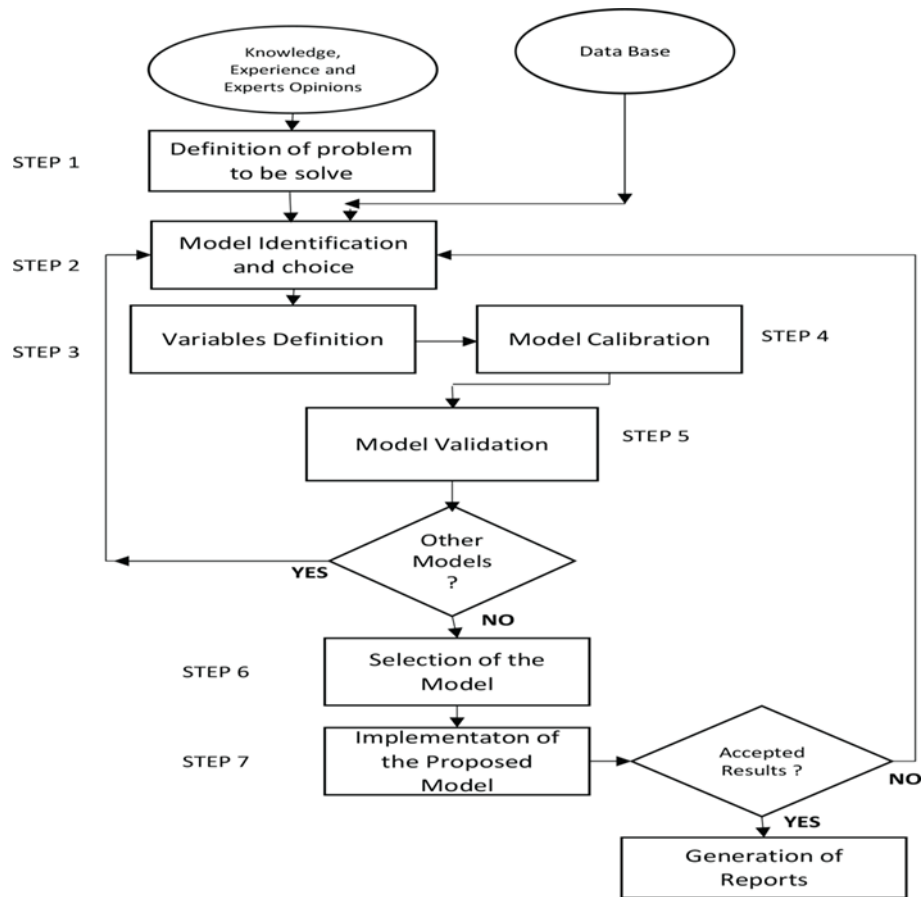


FIGURE A.1.

The Modeling Process.

Appendix B

B.1. TYPES OF EMPIRICAL MODELS

It is common to employ theoretical frameworks (based on mathematical and statistical concepts) [1,8,13] to construct models following the use of data base to define constant values. This represents characteristic values (parameters) considering a defined model.

This process sometimes is denominated the fitting model. Although the model does not adjust well to the observed data, they would be accepted; assuming the presence of some errors; and the definition is useful to explain the tendency of the studied situation. When we use this type of processes, the models are called analytical models.

In EM, it is considered that the use of data (observations) is based on sample observations or data that are coming from experiments or simple observations of the studied reality. This leads to the seeking for some trends or additional knowledge. The searching is oriented to explain the presence of certain dependent variables.

In other words, in EM, the main idea is to get rapidly the best tendency of the information and use it to find and propose a model that would be useful to make some decisions that contribute in the solution of a specific problem. **Table B.1** shows some types of typical and useful transformations to create empirical models.

Type of model	Function	Mean features
Linear	$y = ax + b$	Simply and easy to use.
Power	$y = ax^b$	The function is called “power” given an increase of x by a factor of t causing an increase of y by the power t^b of t (for $b > 0$).
Quadratic	$y = ax^2 + bx + c$	Used to adjust data when data have minimum and maximum values that can be used as limits of a range of values
Cubic	$y = ax^3 + bx^2 + cx + d$	Used when a minimum or a maximum value can be determined. The selection depends on the context we are analyzing.
Quartic	$y = ax^4 + bx^3 + cx^2 + dx + e$	Used to adjust data when data have minimum and maximum values that can be used as limits of a range of values. When we deal with polynomial models, it is important to consider the significance of the complexity and the precision of the intervals under study.
Exponential	$y = ab^x$ or $y = ae^{kx}$	Has constant percent (relative) rate of change (a constant quotient of two consecutive y values).
Logarithmic	$y = a + b \ln(x)$	If $b > 0$, then the function is increasing and concave down. If $b < 0$, then the function is decreasing and concave up.

TABLE B.1.

Types of linear transformations.

B.2. LINEARIZATION PROCESS IN EMPIRICAL MODELING

To propose the best model based on the obtained observations (data values), in EM it is very helpful to linearize a data set, transforming and adjusting a simple model (linearized) based on a transformation processes assuming the simulation of a continuous variable x . Some models can be linearized using the obtained data.

If the functions have some of these forms, the linearization process can be achieved transforming the models considering the relationship with a linear model [1,8]. Keep in mind that if $y = ax^b$, then $\ln(y) = \ln(a) + b\ln(x)$; $\ln y = \ln a + b \ln x$. ; So if y is a power function, $\ln(y)$ is a linear function of $\ln(x)$.

In modelling certain situations, there is a special interest in some aspects associated with the nature of the values that correspond to the variables under study. Sometimes, the linearization is achieved for a set of variables interacting simultaneously, using a numerical algorithm. One algorithm is called LSM, which is based on the minimization of the corresponding residuals.

Figure B.1 shows the form of a model using x as predictor variable and y as explained or response variable.

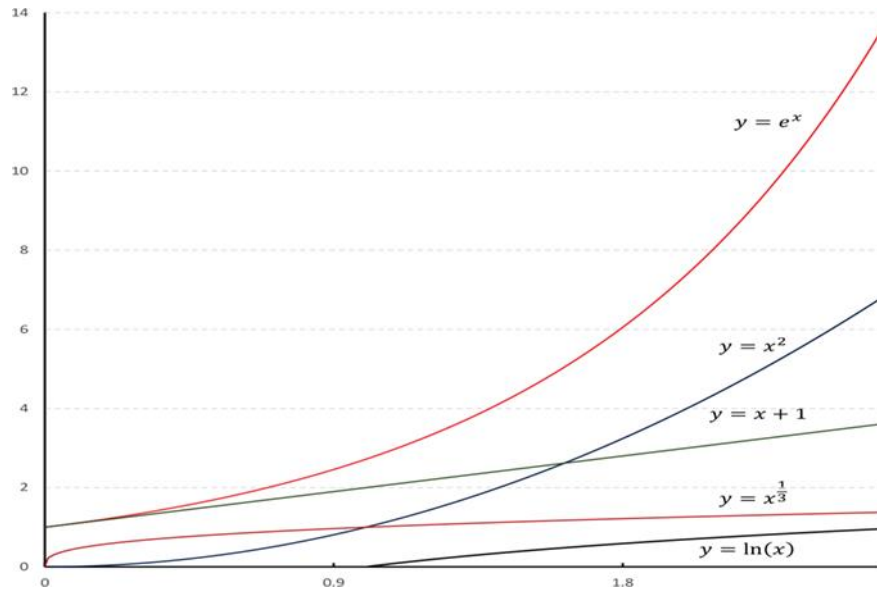


FIGURE B.1.

Several examples of functions.

B.3. PARAMETERS COMPUTATION WHEN USING LSM

In order to compute the parameters $a; b; c, \dots$, shown in **Table B.1**, the following general procedure can be adopted. Let's consider the function

$$f(x, p) = \sum_{j=0}^n a_j x^j \quad (5)$$

Options

along with the following cost function

$$J(p) = \sum_{i=0}^m (y_i - f(x_i, p))^2 \quad (6)$$

Options

Where p is the vector of parameters to be determined, i.e., $p = [a_0 \ a_1 \ \dots \ a_n]^T$. Then, to determine parameter p , the best fit to the set of data (x_i, y_i) , corresponds to the parameter p which minimizes the cost function J , and we know from calculus that such a parameter must satisfy:

$$\nabla_p J(p) = 0 \quad (7)$$

Options

Based on this:

$$\sum_{i=0}^m x_j^k y_i + \sum_{k=0}^n a_k \sum_{i=0}^m x_j^{k+1} = 0, \text{ for } j=0, 1, \dots, n \quad (8)$$

Options

Since this holds for each $j \in \{0, 1, \dots, n\}$, then, such a equation can be structured in a convenient way:

$$\begin{pmatrix}
 \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 & \sum_{i=1}^m x_i^3 & \dots & \sum_{i=1}^m x_i^n \\
 \sum_{i=1}^m x_i y_i & \sum_{i=1}^m x_i^2 y_i & \sum_{i=1}^m x_i^3 y_i & \dots & \sum_{i=1}^m x_i^n y_i \\
 \sum_{i=1}^m y_i & \sum_{i=1}^m y_i^2 & \sum_{i=1}^m y_i^3 & \dots & \sum_{i=1}^m y_i^n
 \end{pmatrix}
 \begin{pmatrix}
 a_0 \\
 a_1 \\
 a_2 \\
 \vdots \\
 a_n
 \end{pmatrix}
 =
 \begin{pmatrix}
 \sum_{i=1}^m y_i \\
 \sum_{i=1}^m x_i y_i \\
 \sum_{i=1}^m x_i^2 y_i \\
 \vdots \\
 \sum_{i=1}^m x_i^n y_i
 \end{pmatrix}
 \tag{9}$$

[Options](#)

REMARK

It is clear from the above expression that such an equation can be written as a system of linear equations $Ap=B$. Then, to solve it for a large amount of parameters, several numerical methods can be applied (e.g., Gauss–Seidel, Jacobi, between others).