A GREY VERHULST MODEL FOR FORECASTING CONSTRUCTION COSTS

A Thesis Submitted to
The Graduate School of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Architecture

by
Uğurkan ERİNÇ

August 2020
İZMİR

ACKNOWLEDGMENTS

First of all, I would like to thank Prof. Dr. Serdar KALE, my esteemed thesis advisor who encouraged me in my thesis process and shared his vast experiences with me for his patience and all his help.

I would like to express my deepest affection to my dear family, my masters, my mother Kerime ERİNÇ, my father Kahraman ERİNÇ and of course my brother Onur ERİNÇ who always made their support felt, his wife Bengi ERİNÇ and my new nephew Ege ERİNÇ who is a new member of our family.

To my dear beloved Özün ALTINAY, who always gave me the greatest motivation; very special thanks for being with me and sharing life.

I would like to thank my dear friends who are always ready to support me, especially Pınar KESKİN and Kemal ÇETİN, to make a great contribution for my thesis.

ABSTRACT

A GREY VERHULST MODEL FOR FORECASTING CONSTRUCTION COSTS

Forecasting costs in construction projects has an important place in whole processes. It is very important to observe possible changes in the construction process, especially in large scale projects. Using a series of data in a functional way can be useful to make the changes that can be made at a later time in that field. Grey models have an important place in the time series forecasting model. In line with this case, there are a number of methods developed over time. In fact, when these models are examined, it is seen that they follow and complement each other in the developmental stages, but the levels of development differ with some important decompositions. The ability to model and forecast construction costs can result in more accurate cost forecasting and budgeting. This has been modeled using a residual Fourier model by analyzing the remains of the Grey Verhulst Model, conditional variability of construction cost prices. The results show that the developed model can forecast construction costs with less errors. It was observed that the deviation rate in the final values decreased up to 0.97%. The data used in the time series analysis were observed for a relatively long time. These data are calculated with the assumption that the system will be developed in the future and the forecasted values will be maintained in the future based on these time series data.

ÖZET

İNŞAAT MALİYETLERİNİN TAHMİNİ İÇİN BİR GRİ VERHULST MODELİ

İnşaat projelerinde maliyetlerin öngörülmesi, gerek tasarım gerekse uygulama süreçlerinde önemli bir yere sahiptir. Özellikle büyük ölçekli projelerde yapım sürecindeki olası değişimlerim gözlenmesi oldukça önemlidir. Bir veri serisini işlevsel biçimde kullanmak, o alandaki ileri zamanlarda yapabilen değişimleri tamnin edebilmek için işe yarayabilir. Gri modeller zaman serileri tahmin modelinde önemli bir yere sahiptir. Bu işlevsellik doğrultusunda zaman içinde yapılmış çalışmalarla geliştirilmiş bir takım yöntemler bulunmaktadır. Aslında bu modellere bakıldığında bunların da gelişim aşamalarında birbirini izlediği ve tamamladığı, fakat bazı önemli ayrışımlarla gelişmişlik düzeylerinin farklılaştığı görülmektedir. Bir inşaat projesi ilk maliyet tahminleriyle başlar. Her ne kadar planlama aşamasında tahminleri destekleyecek pek çok yöntem geliştirilmiş olsa da, inşaat faaliyetleri ve projeleri için sürelerin yanı sıra maliyetlerin erken tahminleri hala hataya açıktır. Yanlış tahminlere neden olan ana etkenlerden biri, zaman içinde ekonomik koşulların değişmesinden dolayı kaynak fiyatlarındaki değişimdir. Bu, kaynak fiyatlarındaki dalgalanmaları dikkate alarak inşaat maliyetleri trendini izlemenin ve tahmin etmenin önemini göstermektedir. İnşaat maliyetlerini modelleme ve tahmin etme kabiliyeti, daha doğru maliyet tahmini ve bütçelemeyle sonuçlanabilir. Çalışmada Gri Verhulst Modeli'nin kalıntıları üzerinden, yapım maliyet fiyatlarının koşullu değişkenliği bir kalıntılı Fourier modeli kullanılarak modellenmiştir. Sonuçlar, geliştirilen modelin daha düşük hatalarla yapım maliyetlerini tahmin edebileceğini göstermektedir. Sonuçlar değerlendirildiğinde, sapma payının %0.97'ye kadar indiği saptanmıştır. Bu veriler modeli geliştirmek ve öngörülen değerleri bu zaman serisi verilerine dayanan sistemin gelecekte de korunacağı varsayımıyla hesaplanır. Politik değişiklikler veya savaşlar gibi kontrol edilemeyen faktörler olması durumunda, uzun zaman boyunca zaman serisi verilerinin, bu faktörlerin zamana özgü yapısından dolayı farklı zamanlarda farklı kalıplara sahip olma olasılığı vardır.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

LIST OF ABBREVIATIONS

ANFIS : Adaptive-Network-Based Fuzzy Inference System
ANFIS : Adaptive-Network-Based Fuzzy Inference System
ANN-GA : Genetic Algorithm-Based Artificial Neural Network
CC : Co-efficient of Correlation
FMSE : Forecast Mean-Square Error
FWM : Fixed Weights Model
FT : Fourier Transform
GARCH : Generalized Auto Regressive Conditional Heteroscedasticity
GM : Grey Model
GMRAE : Geometric Mean of the Relative Absolute Error
GVM : Grey Verhulst Model
MAD : Mean Absolute Deviation
MAE : Mean Absolute Error
MAPE : Mean Absolute Percentage Error
MRSS : Mean Residual Sum of Squares
MSE : Mean Squared Error
MWE : Mean Week Error
MWM : Moving Weights Model
RMSE : Root Mean Squared Error
SARIMA : Seasonal Auto-Regressive Integrated Moving-Avarege
VAR : Vector Auto Regression
VEC : Vector Error Correction

CHAPTER 1

INTRODUCTION

1.1. Introduction

The cost of a construction project is an important parameter that must be analyzed by all parties, including the owner, contractor and subcontractor. An increase in construction costs during the project may have a negative impact on the overall project, such as delay or termination of the project, high project costs, and quantitative and qualitative deterioration of bid competition. Forecasting the trend in construction costs is crucial in forecasting the cost of construction projects, as well as in budget planning and assessing risks associated with coordination and costs. However, over the last few decades, many construction companies and project owners have suffered major construction losses, because final construction costs are much higher than originally anticipated at the planning stage, although they have made significant efforts to accurately forecast construction costs.

This problem exists to a large extent because it is not easy to accurately forecast construction costs due to the smallness and uncertainty such as transport and construction material costs that directly affect the construction market and socioeconomic factors such as macroeconomic environments. In addition, there are great difficulties in forecasting actual construction costs for the future in the planning process for several months or several years, because the time interval between the project planning process and the actual construction time is quite large.

Many studies have suggested forecasting methods that are based largely on causality analysis or time series analysis. In recent years, in most of the studies, methods based on time series analysis have been preferred, because independent variables should be selected precisely and accurately forecasted for causality analysis.

The data used in the time series analysis were observed for a relatively long time. These data are calculated with the assumption that the system will be developed in the future and the forecasted values will be maintained in the future based on these time series data. In the case of uncontrollable factors such as policy changes or battles, it is likely that over time, time series data will have different patterns at different times due to the time-specific nature of these factors. (Moon and Shin, 2017)

1.2. Objectives of the Study

The main objectives of the study are follows:

- To investigate forecasting models, to define how they are used, to determine which data sets are used,
- To determine the location and application of Grey Models in forecasting models,
- Evaluating the output values of forecasting models and testing their usability in construction management.

1.3. Scope and Structure of the Study

The scope of thesis presented herein, like any other research studies is defined by the above-mentioned objectives of the research. This thesis is a research for interpreting the time series models for the forecasting of construction costs and interpretation of test results. The characteristics of the models were examined and in accordance with the studies, it was determined which models gave more realistic results. This provides an opportunity for the models to be used in construction management.

This study, consists of 6 chapters aimed at defining, measuring and analyzing forecasting models and forecasting process. In the first part, (1) the reasons that require this research, (2) the aims of the study and (3) the scope of the study are prepared. In the second chapter, two main headings forecasting models determined as a result of literature review are explained. These main headings; "causal models" and "time series models". The time series models that the research focuses on are explained. In addition, a literature review is provided to gain insights into some other models. However, a forecasting method based on grey models was determined as the main objective. In addition, methods that evaluate the forecasted data are presented in this section. In the third chapter, as a result of the literature review, the use of time series forecasting models in the construction sector and the process of the models used were observed. It has been discussed how these can be improved, which methods can be used and how they will be more useful in the construction industry. The fourth chapter is the center of the researches. With the help of sampling, data preparation and statistical analysis methods, the results of the models discussed are presented and evaluated. This section is the core of the study in terms of the application of all the grey models in question such as **GM (1,1)**, **Grey Verhulst Model** and **Fourier Transformation**. In the fifth chapter, research findings on causal and time series forecasting models are interpreted and their reflection on the results are discussed. In the sixth and last chapter, the main conclusions of the study are summarized and recommendations for the role of the models in the construction sector are given.

CHAPTER 2

FORECASTING STRATEGIES AND EVALUATION METHODS

In this section, the methods used in the forecasting requirements, the forecasting method trends in many sectors, the introduction of these methods (qualifications, types, etc.), application stages and evaluation methods are mentioned.

2.1. Literature Review

Especially for large-scale projects, the cost factor has been a major concern because of its significant cost implications and long-term changes. These temporary factors, such as resource prices, can lead to under or underestimating the total project cost, as resource prices vary depending on changes in demand, market conditions and macroeconomic conditions. For example, structural steel prices in Korea have tripled from 2001 to 2008, and according to recent global economic reports, additional material costs caused by inflation can often reduce the profitability of general contractors for large-scale projects over the years. Since material costs constitute a significant portion of the total project cost, it is important to accurately estimate raw material prices in many sectors to accurately estimate the total cost (Hwang et al., 2012).

To address these issues, many researchers have attempted to accurately estimate cost changes in projects, focusing on total costs or project cost index, using forecasting techniques such as time series analysis, causal-based models such as artificial neural networks or other possible methods. Of these techniques, in particular, time series analysis is widely used because associated observations represent the time dependent lag relationship between both single and interrelated multiple series.

According to Hwang et al.; however, there are two main problems with the use of this method to forecast material cost increases. First, a large number of different materials are needed for many projects. Since the prices of various materials increase or decrease at different rates, the total material cost increase must be the sum of the price increase of each material. Second, the complex and iterative procedures of most forecasting methods require significant time and effort to identify appropriate models and to forecast price increases for each material required.

The accuracy, usage and shape of the projections that occur are determined by time interval and data availability (O'Connell, 1987). This is often difficult to do in practice and requires a fairly subjective forecast of future market conditions and inflation. The benefits of more objective methods and quantitative cost estimation models in this direction have been described for some time. Various cost models with varying complexity have been developed by the researchers. Univariate time series modeling also attracted positive attention. Time series models have been applied to estimate the behavior of project costs and sales prices (Thomas Ng. et al., 2004).

Quantitative methods are proposed for estimating cost indexes of projects. These methods can be examined in two main categories as (1) Causal and (2) Statistical methods. Statistical methods studied and analyzed use timeline analysis and curve fitting techniques to estimate the cost index of a large number of materials (Shahandashti and Ashuri, 2013).

Ashuri and Lu compared various univariate time series models to estimate various projects costs. They concluded that a seasonal autoregressive integrated moving average model and the Holt-Winters exponential correction process are the most accurate univariate time series approaches for in-sample and out-of-sample estimation. However, statistical methods are not long-term explanatory and are only efficient for short-term forecasts.

According to Hwang et al., many studies focused on the rapidly changing material market conditions and tried to address cost changing factors to make cost planning more feasible. The main issues here are to determine affecting factors and to estimate project costs accurately and simply.

The following studies are also noteworthy developments in Hwang el al. (2010). Ranasinghe (1996) presented a simple model that takes into account inflation effects and changing project costs. Akpan and Igwe (2001) developed an appropriate model for the assessment of cost overruns due to price increases, political factors and project execution factors. Trost and Oberlender (2003) presented a mathematical model to assess the accuracy of early forecastings using factor analysis and multivariate regression analysis. Sönmez (2008) developed an integrated approach to conceptual cost estimation which includes the advantages of parametric and probabilistic estimation techniques such as regression models. Finally, Shane et al. (2009) classified individual cost increase factors to evaluate the total project cost in the future.

These time series models provide systematic and time-related approaches to forecasting trends. That is, it is possible to make useful projections based on historical patterns.

Although these models are useful for addressing cost raising factors and foreseeing early in the design phase, there are some limitations to reflecting different time delays between time-varying variables and impact factors. Because most of the time-related data is dependent or actually has an autocorrelation (Lu and AbouRizk, 2009), one way to overcome these limitations is to apply time-related approaches to forecast trends in material prices.

In these methods, time series approaches have been applied to cost estimation in various projects based on time trends, past values and other differentiation factors.

2.2. Causal Forecasting Models

Causal models are mathematical models that make sense of causal relationships within some existing systems or populations. Based on statistical data, they derive inferences that can be made through causal relationships. It allows us to comment on causality epistemology and the relationship between causality and probability and makes it easier for us to analyze. In addition, focusing on some philosophical issues, applications such as logic of counterfactuals, theory of decision, and analysis of actual causation have been made (Hitchcock, 2020).

Causal modeling is an interdisciplinary field of research mainly based on the work of the American biologist and statistician Sewall Wright (1921), during the period of further use of statistical techniques developed in the 1920s. It has been developed with significant contributions from biology, medicine, computer science, econometrics, philosophy, statistics and other disciplines. Given the importance of causality in many scientific fields, there is growing interest in the use of mathematical causal models. In the 2000s, two major works; Spirtes, Glymour, and Scheines (2000) and Pearl (2009), have a decisive role in increasing its influence.

Causal models try to make estimations about the behavior of a system. In particular, a causal model examines the true value or probability of counterfactual claims about the system; estimates the effects of changes; and the probabilistic dependence or independence of the variables included in the model. With causal models, it may be easier to make the following interpretation: if we observed the results of probable correlations or experimental changes between variables, we can determine which causal models are consistent with these observations. Accordingly, the outputs of the model will guide what can be done “in principle”. For example, when excellent information about probability distribution on variables in the system is obtained, we will be able to consider how much we can deduce the correct causal structure of a system. In addition, application of causal models to the logic of counterfactuals, causality analysis and decision theory can be discussed (Hitchcock, 2020).

Accordingly, when the models developed / developing causal based are examined, the concept of "Machine Learning" will be encountered. Realizing the concept of "Machine Learning" involves working on some training data, creating a model that is trained and can process additional data to make estimations later. Various types of models used for "Machine Learning" systems have been researched (Alpaydın, 2020).

"Machine Learning" (ML) is the study of computer algorithms developed automatically through experience. It can be seen as a subset of artificial intelligence. ML algorithms create a mathematical model based on sample data known as "training data" to make estimations or decisions without being explicitly programmed. While ML algorithms are used in a wide variety of applications such as email filtering and computer vision, it is not difficult or possible to develop traditional algorithms to perform the necessary tasks, while providing many estimations (Koza et al., 1996).

ML is closely related to calculation statistics that focus on predicting using computers. Mathematical optimization study provides methods, theory and application areas to machine learning. Data mining is a relevant field of study that focuses on exploratory data analysis through unsupervised learning. Machine learning is also referred to as "predictive analytics" with the methods it applies throughout business problems (Friedman, 1998).

2.2.1. Artificial Neural Network Model

Artificial neural networks (ANNs), or connectivity systems, often called neural networks (NNs), are computational systems that vaguely sample biological neural networks that make up the animal brains (Chen et al., 2019).

The data structures and functionality of neural networks are designed to simulate relational memory. Neural networks learn by processing samples that contain a known "input" and "result" and create probability-weighted relationships stored between the two in the network's data structure. (The "input" here is more accurately called the input set because it usually contains more than one argument rather than a single value.) Therefore, the "learning" of the neural network from a particular example is the difference in the situation. Clean the sample before and after processing. After a sufficient number of samples are given, they will be able to estimate the results from the inputs using the associations created from the net sample set. If the neural network is provided with a feedback loop about the correctness of its estimations, it continues to improve their relationship, which leads to an increased level of accuracy. In short, there is a direct relationship between the number and variety of samples processed by a neural network and the accuracy of their estimates. Since neural networks do not discriminate in the form of associations, they can create unexpected associations and reveal previously unknown relationships and dependencies.

The ANN approach focused on solving problems based on the human brain. But as studies progressed, attention shifted towards performing certain tasks and led to deviations from biology. ANNs have started to be used in different jobs such as computer vision, speech recognition, machine translation, social network filtering, game board and video games, medical diagnostics, and even traditionally painting, by moving to areas where people need it. (Gatys et al.,2015).

According to Cannady (1998), an artificial neural network consists of a series of processing elements which are very interconnected and convert a set of inputs into desired outputs. The result of the conversion is determined by the properties of the elements and the weights associated with the connections between them. By changing the connections between the nodes, the network can adapt to the desired outputs. Unlike expert systems that can provide a definite answer to the user, if the reviewed features exactly match those in the code base, neural networks are trained to recognize which performs an analysis of information and forecasts that the data matches the features. While the probability of a match determined by a neural network may be 100%, the accuracy of its decisions is based on the experience of the system in analyzing examples of the specified problem.

The neural network initially gains experience by training the system to accurately identify preselected examples of the problem. The response of the neural network is reviewed and the configuration of the system is improved until the training data analysis of the neural network reaches a satisfactory level. In addition to the initial training period, the neural network also gains experience over time in analyzing the problem-related data (Cannady, 1998).

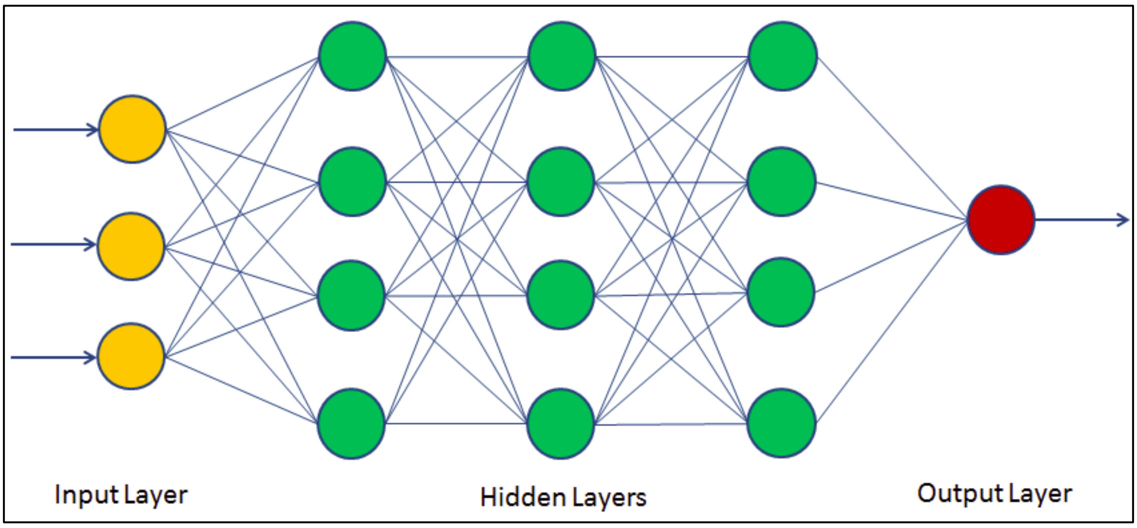


Figure 2.1. Artificial Neural Network Schema.
(Source: Navlani, 2019)

2.2.2. Genetic Algorithm (GA) Model

In the computer science workspace, genetic algorithm (GA) is a metasismatic method inspired by the natural selection process of the larger evolutionary algorithm class (EA). Genetic algorithms are widely used to produce high-quality and stable solutions to optimization problems based on biologically inspired operators such as mutation, transition, and selection. John Holland developed genetic algorithms based on Darwin's concept of evolution in 1960, and GA studies further developed by his student David E. Goldberg (Mitchell et al., 1996).

In the genetic algorithm model, an attempt is made to evolve candidate solutions (called individuals, creatures or phenotypes) towards better solutions. Candidate solutions have a number of properties (chromosomes or genotypes) that can be modified and mutated. Generally, although the solutions are represented in binary (as strings 0 and 1), it appears that other encodings can also be made (Whitley, 1994).

Evolution usually starts from a population of randomly coexisting individuals and continues with an iterative process in which each population is called a generation. In every generation, the suitability of each individual in the population is evaluated; suitability is usually the value of the objective function in the optimization problem being solved. More suitable individuals are selected stochastically from the existing population, and each individual's genome is modified (reassembled and possibly randomly mutated) to create a new generation. Next generation candidate solutions are then used in the next iteration of the algorithm. In general, the algorithm ends when a maximum number of generations are produced or a satisfactory level of suitability for the population is reached (Fernandes et al., 2020).

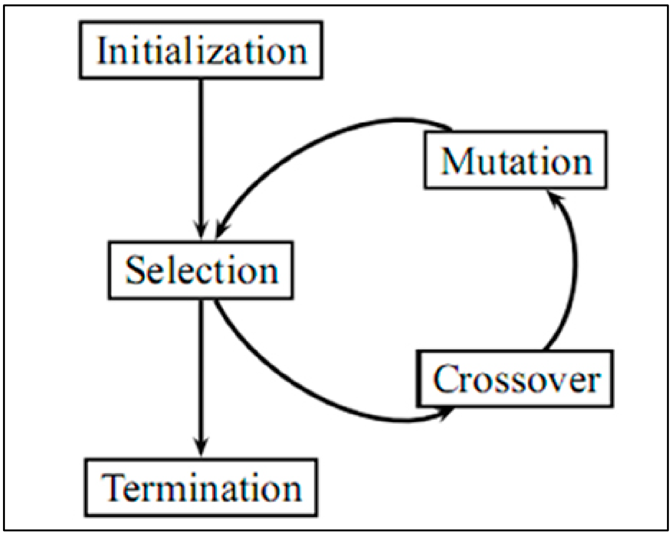


Figure 2.2. Genetic Algorithm Schema.
(Source: Epstein, 2016)

2.2.3. Decision Tree Learning Model

Decision Tree Learning is one of the predictive modeling approaches used in statistics, data mining and machine learning. Uses a decision tree (as an estimate model) to go from conclusions about an item (represented in branches) to the target value (represented in leaves) of the item. Tree models in which the target variable can take a separate set of values are called classification trees; in these tree structures, the leaves represent class labels and the branches represent combinations of features that lead to these class labels. Decision trees (typically real numbers) where the target variable can take continuous values are called regression trees. Decision trees are among the most popular machine learning algorithms given their intelligibility and simplicity (Piryonesi and El-Diraby, 2019).

In decision analysis, a decision tree can be used to visually and clearly represent decisions and decision making. In data mining, a decision tree defines data (but the resulting classification tree may be an input for decision making). This page is about decision trees in data mining (Wu et al, 2007).

Decision Trees have uses and tree-based models are known to perform well in the ML algorithms group. Although the tree's split decision on each node is optimized for the data set that it is appropriate for and rarely generalizes to other data, a more robust model can be created by combining a large number of trees created in different ways and their resulting predictions.

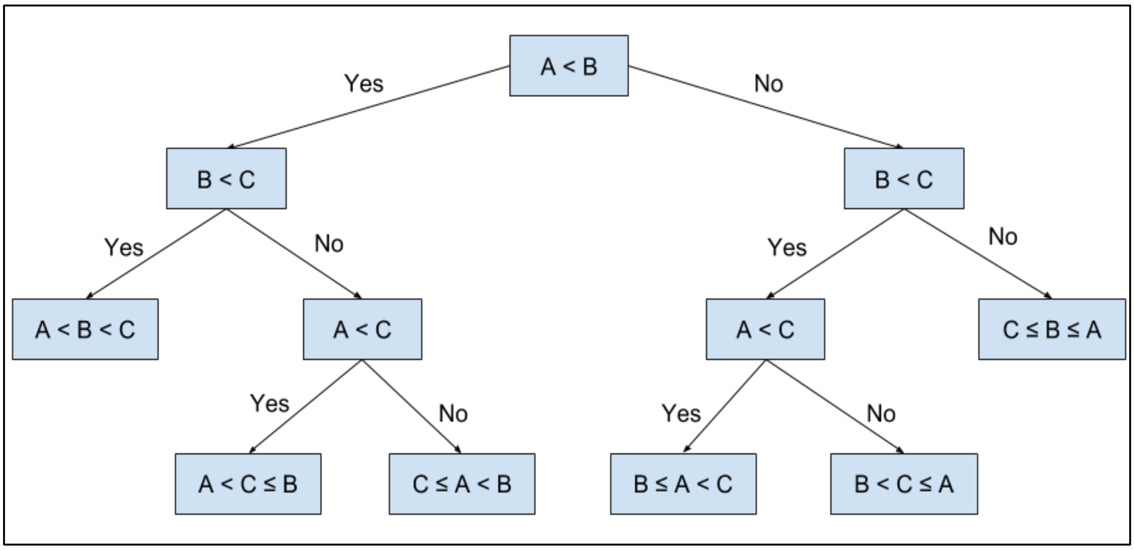


Figure 2.3. Decision Tree Schema.
(Source: Niculaescu, 2018)

2.2.4. Support Vector Machines (SVMs) Model

Support-vector machines (SVMs, as well as support-vector networks) are supervised learning models with related learning algorithms that analyze data used in machine learning for classification and regression analysis. Support Vector Machine

(SVM) algorithm, one of the popular machine learning tools, provides solutions to both classification and regression problems. Given a set of training examples, each marked as belonging to one or the other, an SVM training algorithm creates a model that assigns new samples to one category or another, making it an unlikely binary linear classifier. In the SVM model, samples are represented as points in space, so that samples of separate categories are divided into as wide a space as possible. New samples are then mapped to the same area and are estimated to belong to a category based on the space they fall into (Cortes and Vapnik, 1995).

In addition to performing linear classification, SVMs can perform a nonlinear classification using what is called the "Kernel Machine" by indirectly mapping its inputs to high-dimensional property fields (Ben-Hur et al., 2001).

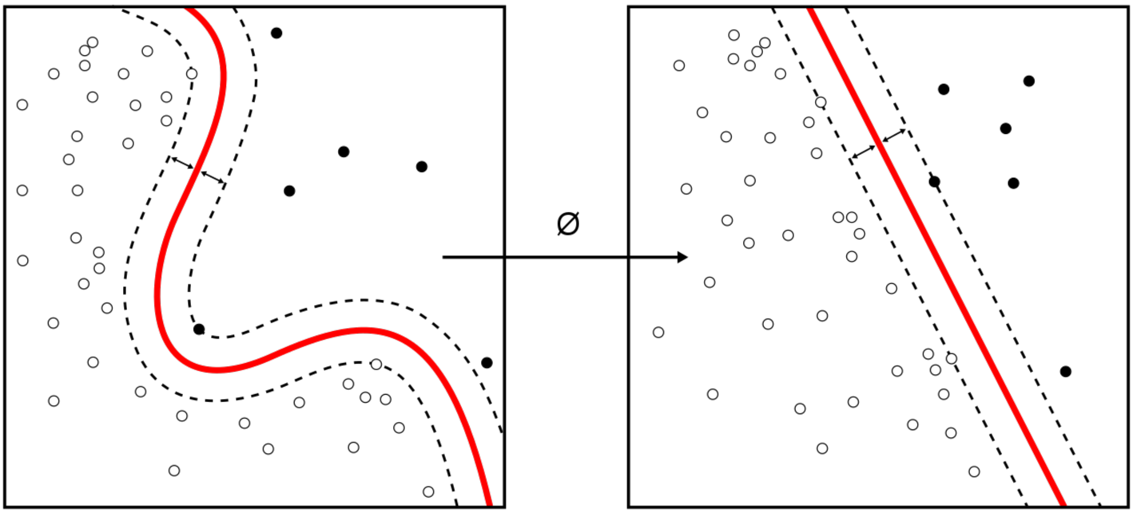


Figure 2.4. Kernel Machine Schema.
(Source: Alisneaky, 2011)

2.3. Time Series Forecasting Models

The time series is a series of data points indexed (or listed or graphed) in the order of time. More commonly, time series are a sequence of successive and preferably taken at equally spaced points. In other words, they are arrays of discrete time data. Examples of time series are the intensity of earthquake waves, the number of COVID-19 patients and the daily closing values of the Exchanges.

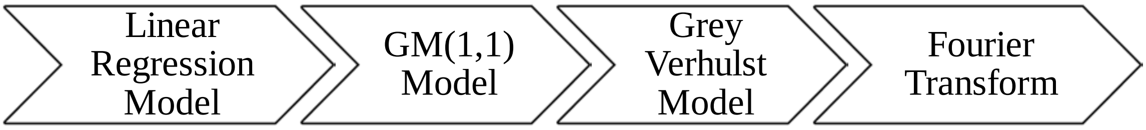
Time series are often expressed in line graphs. The time series is widely used in architecture, project management, statistics, pattern recognition, communication engineering, econometrics, finance, weather forecasting, earthquake forecasting, control engineering, astronomy, and any applied science and engineering that largely includes temporal measurements.

Time series analysis includes consistent statistics and methods for inference by processing time series data to extract hidden features of the data. Time series estimates are the use of a suitable model to forecast future values based on current known values. For example; Regression analysis is often used to test the theories that the current values of one or more independent time series affect the current value of another time series, and this type of time series analysis is not called "time series analysis". Because time series analysis focuses on comparing the values of a single time series or different dependent time series at different points in time.

Time series data has a natural temporal order. This distinguishes time series analysis from cross-sectional studies where the natural order of observations is lacking. Time series analysis is also different from those in which observations are typically evaluated by characteristics. A stochastic model for a time series will often reflect the fact that observations that are close together over time will be more closely related to separate observations. In addition, time series models often use the natural unidirectional time order, so values for a given period are expressed as derived from past values rather than future values (Lin et al., 2003).

Time series analysis can be applied to real-value, continuous data, discrete numerical data or discrete symbolic data. In order to apply, Smoothness and Exponentiality tests of the data should be performed and values that are suitable for certain intervals should be taken (Özdemir and Özdagoglu, 2017).

Many models can be used for Time Series forecastings. Among these methods; the study will focus on Linear Regression Model, Grey Model (1,1), Grey Verhulst Model and Fourier Transformation. These models will be presented and will be implemented and evaluated in the next sections. At this point, grey models will also be explained and implemented. This is the main focus of the study.



2.3.1. Linear Regression Model

Linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The state of an explanatory variable is called simple linear regression. For more than one explanatory variable, the process can be called multiple linear regression. This term differs from multivariate linear regression in which multiple associated dependent variables are forecasted rather than a single scalar variable (Freedman, 2009).

According to Codur et al., normal distribution plays an important role in linear and nonlinear regression models. In both linear and nonlinear models, the y response variable is assumed to have a normal distribution for inference. In some cases, this assumption is unrealistic. An example is when a response variable is a discrete variable, such as a number, ie the number of defects, the number of people suffering from a particular disease, or the number of occurrences of natural events involving earthquakes and hurricanes.

Generalized Linear Models allow fitting of regression models if the response is from the exponential family. In generalized linear models, the link function takes advantage of the natural distribution of the response. In particular, the wrong selection of the link function will affect the natural distribution, thus adversely affecting the results. (Codur et al., 2013)

Linear Regression models how mean expected value of a continuous response variable depends on a set of explanatory variables, where index *i* stands for each data point:

$$Y_i = \beta_0 + \beta x_i + \epsilon_i Y_i = \beta_0 + \beta x_i + \epsilon_i$$

or

$$\epsilon\ (Y_i) = \beta_0 + \beta x_i\ \epsilon(\mathrm{Y}_i) = \beta_0 + \beta x_i$$

- *Random component: Y* is a response variable and has a normal distribution, and generally we assume errors, $e_i \sim N(0, \sigma^2)$.
- *Systematic component: X* is the explanatory variable (can be continuous or discrete) and is linear in the parameters $\beta_0 + \beta x_i$. Notice that with a multiple linear regression where we have more than one explanatory variable, e.g., $(X_1, X_2, \ldots X_k)$, we would have a linear combination of these *Xs* in terms of regression parameters *β's*, but the explanatory variables themselves could be transformed, e.g., X^2, or *log(X)*.
- *Link function: Identity Link*, $\eta = g(\epsilon(Y_i)) = \epsilon(Y_i)$ identity because we are modeling the mean directly; this is the simplest link function (Jammalamadaka, 2012).

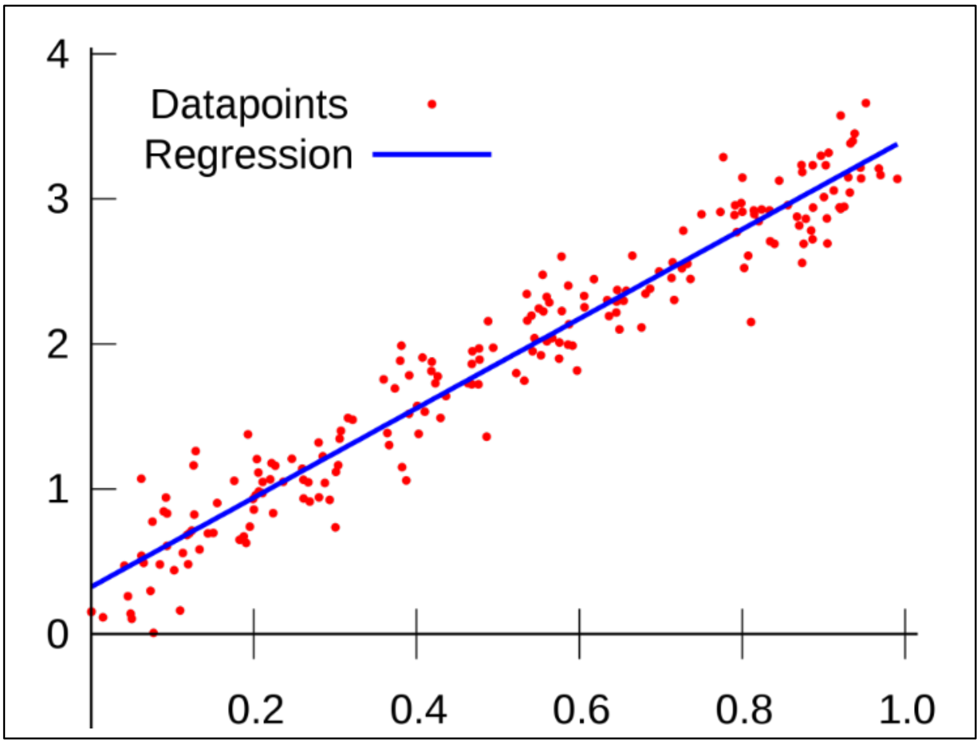


Figure 2.5. Linear Regression Model Graph.
(Source: Tran, 2018)

2.3.2. Grey System Based Forecasting

Grey system theory was founded in 1982 by Julong Deng, a new methodology that focuses on the study of problems involving small samples and weak information. It deals with ambiguous systems that contain partially known information by producing, digging and extracting useful information from the present. Thus, the operational behavior of systems and the laws of evolution can be accurately defined and monitored effectively. Uncertain systems with small sampling and weak information are widely available worldwide. This fact determines the broad applicability of grey system theory (Sifeng et al., 2011).

According to Kayacan et al. (2010), In system theory, a system can be identified by a color representing a clear amount of information about that system. For example, if mathematical equations that describe internal properties or dynamics are not fully known, the system may be called a black box. On the other hand, if the definition of the system is fully known, it can be called a white system.

Similarly, a system having both known and unknown information is defined as a grey system. In real life every system can be considered a grey system because there is always some uncertainty. Because of the noise (and limitations of our cognitive abilities) both from inside and outside the system we are concerned about, the information we have about this system is always ambiguous and limited in scope (Lin and Liu, 2004).

There are many situations where there is a lack of information, either incomplete or inadequate. Even a simple motor control system always includes some grey features due to the system's time-varying parameters and measurement difficulties. Similarly, it is difficult to accurately estimate a region's electricity consumption due to various social and economic factors. These factors are often random and make it difficult to obtain an accurate model (Kayacan et al., 2010).

Grey models estimate future values of a time series based on only the most recent data, depending on the window size of the estimator. It is assumed that all data values to be used in grey models are positive and the sampling frequency of the time series is constant. In the simplest terms, the grey models to be formulated below can be seen as curve fitting approaches.

The main task of the grey system theory is to enforce the realistic management laws of the system using the available data. This process is known as the generation of grey sequence (Liu and Lin, 1998).

Although the existing data of the system, which is generally white numbers, is very complex or chaotic, it is always claimed to contain certain laws. If the randomness of the data obtained from a grey system is somehow softened, it is easier to obtain any special features of that system.

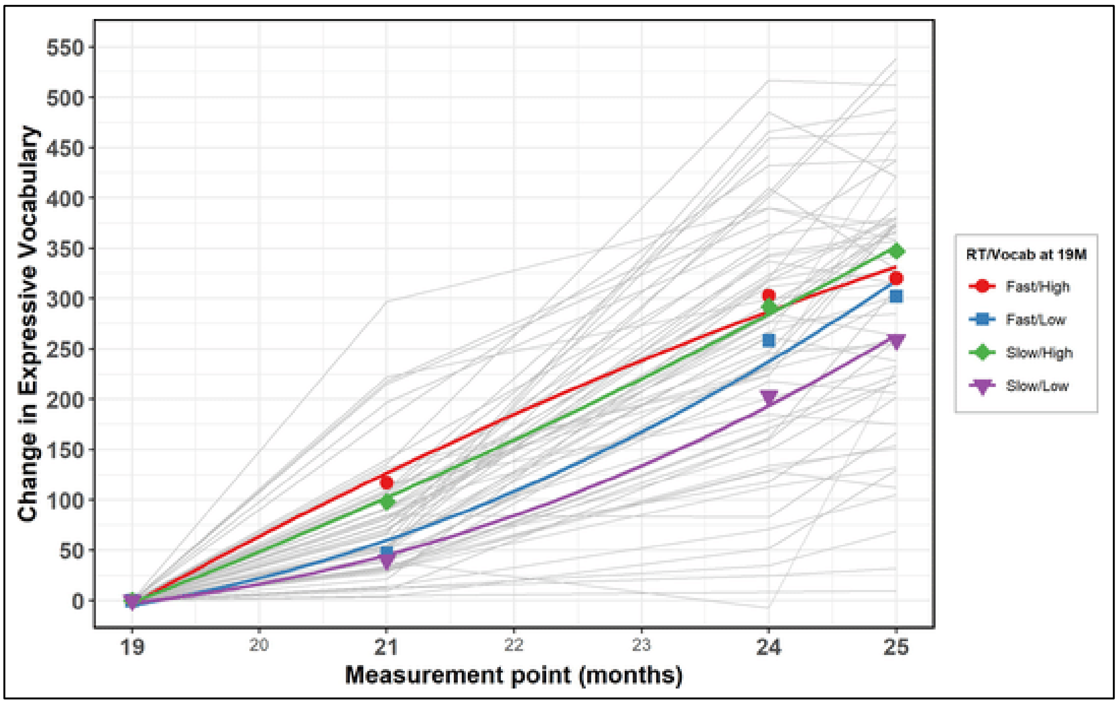


Figure 2.6. Grey System Forecasting Model Sample.
(Source: Peter et al., 2019)

2.3.3. GM(1,1) Model

GM (1,1) type grey model is the most widely used in the literature and is referred to as "Grey Model First Rank One Variable". This model is a time series forecasting model. The differential equations of GM (1,1) model have coefficients that vary with time. In other words, the model is refreshed when new data is available in the forecasting model.

GM (1,1) model can only be used in positive data series (Deng, 1989). In this paper, since all primitive data points are positive, grey models can be used to forecast future values of primitive data points.

In order to smooth the randomness, the primitive data obtained from the system consisting of GM (1,1) is subjected to an operator called Accumulation Generation Operator (AGO) (Deng, 1989). The differential equation (ie GM (1,1)) is solved to obtain the forecasted value in the n-forward step of the system. Finally, using the forecasted value, the Reverse Accumulative Manufacturing Manufacturer Operator (IAGO) is applied to find the forecasted values of the original data.

In accordance with the definitions the establishment of a difference differential equation model, GM (l, l) forecasting model.

The application of the GM(1,1) model can be shown as:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2),\ldots, x^{(0)}(n)) \tag{1}$$

Once application of Accumulation Generation Operator:

$$x^{(1)}(k) = \sum_{n=1}^{k} x^{(0)}(n) \tag{2}$$

$$x^{(1)} = x^{(1)}(1), x^{(1)}(2),\ldots, x^{(1)}(n))$$

$$= (x^{(0)}(1), x^{(1)}(1), x^{(1)}(2),\ldots, x^{(1)}(n - 1) + x^{(0)}(n)) \tag{3}$$

The albino equation $x^{(1)}$ is set up as follows:

$$\frac{dx(1)}{dt} + ax^{(1)} = b \tag{4}$$

That formula belongs to GM (1,1) model:

$$X^{(1)(}k + 1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \tag{5}$$

$$[a,b]^{T} = (B^{T}B)^{-1}B^{T}Y \tag{6}$$

Where

$$Y = [x^{(0)}(2), x^{(0)}(3),\ldots x^{(0)}(n)]^{\mathrm{T}} \tag{7}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ . & . \\ . & . \\ . & . \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{8}$$

According to Equation (5), the solution of $x^{(0)}(t)$ at time k:

$$X^{(0)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak}(1 - e^{a}) \tag{9}$$

In Equation (1), $x^{(0)}$ series consisting of the original data is shown.

In Equation (2), $x^{(1)}$ series is obtained by collecting the data by applying Accumulation Generation Operator (AGO) process.

In Equation (3), the series formed by the operation in Equation (2) is shown.

In Equation (4), the formation equation of $x^{(1)}$ series is shown.

In Equation (5), the GM (1,1) model's equation is formed.

In Equation (6), the formula for the coefficients a and b is shown.

In Equation (7) and (8), the definitions of Y and B matrices are shown.

In Equation (9), the solution of $x^{(0)}$ value forecasted at time k is shown.

After this application process, depending on the time, the variables in the series reappear as forecasted values and form a basis for future forecasts.

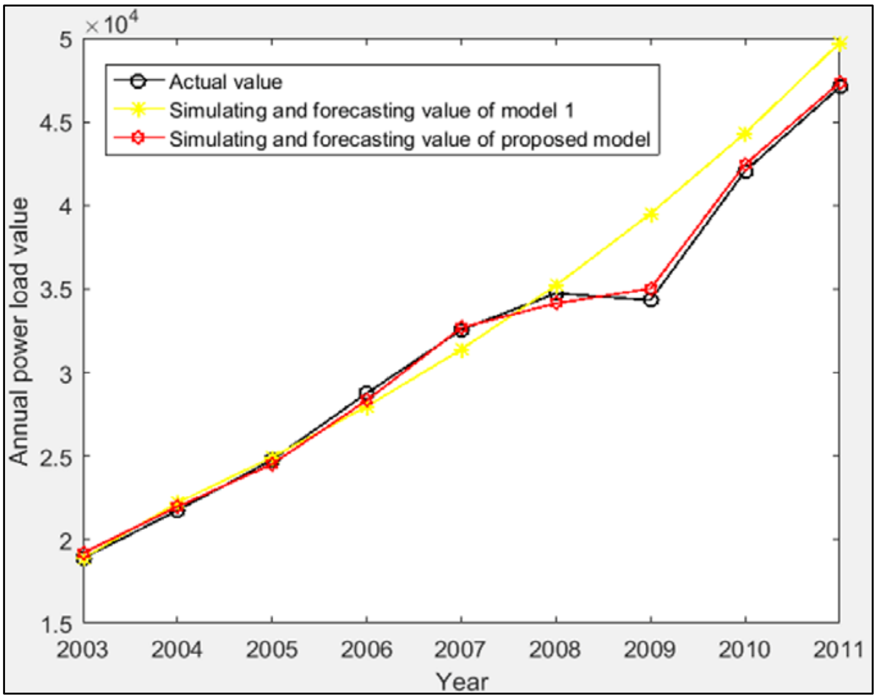


Figure 2.7. GM (1,1) Forecasting Model Sample.
(Source: Li and Wang, 2018)

2.3.4. Grey Verhulst Model

The Verhulst model was first introduced in 1837 by a German biologist, Pierre Franois Verhulst. The main purpose of the Velhulst model is to limit the entire system to a real system and is effective in defining some incremental operations such as the S curve with saturation zone (Kayacan et al.,2010).

Verhulst proposed the Verhulst model when studying the law of biological reproduction.The main idea of Verhulst model is that the number of organism grows exponentially, but the growth rate of individual organism gradually slows down due to the environmental constraints, and it finally stabilizes at a fixed value; it is mainly used to describe the S-shaped process with saturation state (Fu et al., 2019).

The Grey Verhulst model can be described as follows (Wen and Huang, 2004):

$$\frac{dx(1)}{dx} + ax^{(1)} = b(x^{(1)})^2 \tag{10}$$

Grey difference equation of Equation (10) is,

$$x^{(0)}(\mathrm{k}) + az^{(1)}(\mathrm{k}) = \mathrm{b}(z^{(1)}(\mathrm{k}))^2 \tag{11}$$

$$x^{(0)}(\mathrm{k}) = -az^{(1)}(\mathrm{k}) + \mathrm{b}(z^{(1)}(\mathrm{k}))^2 \tag{12}$$

As applied for the GM(1,1) model,

$$[\mathrm{a,b}]^T = (B^T B)^{-1} B^T Y \tag{13}$$

where

$$Y = [x^{(0)}(2), x^{(0)}(3), \ldots x^{(0)}(\mathrm{n})]^{\mathrm{T}} \tag{14}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & ((z^{(1)}(2))^2 \\ -z^{(1)}(3) & ((z^{(1)}(3))^2 \\ . & . \\ . & . \\ . & . \\ -z^{(1)}(n) & ((z^{(1)}(n))^2 \end{bmatrix} \tag{15}$$

The solution of $x^{(1)}(t)$ at time k:

$$x^{(1)}(k+1) = \frac{ax^{(0)}(1)}{bx^{(0)}(1) + \left(a - bx^{(0)}(1)\right)e^{ak}} \tag{16}$$

Applying the IAGO, the solution of $x^{(0)}(t)$ at time k:

$$x^{(0)}(k) = \frac{ax^{(0)}(1)(a - bx^{(0)}(1))}{bx^{(0)}(1) + \left(a - bx^{(0)}(1)\right)e^{a(k-1)}} * \frac{(1 - e^{a})\, e^{a(k-2)}}{bx^{(0)}(1) + \left(a - bx^{(0)}(1)\right)e^{a(k-2)}} \tag{17}$$

As can be seen, in Equation (17), if a < 0, then

$$\lim_{k\to\infty} x^{(1)}(k+1) \to \frac{a}{b} \tag{18}$$

It means that the GVM Equation is $\frac{a}{b}$ which limits the forecasting value. It is also the saturation point of forecasted $x^{(0)}(k)$ (Kayacan et al., 2010).

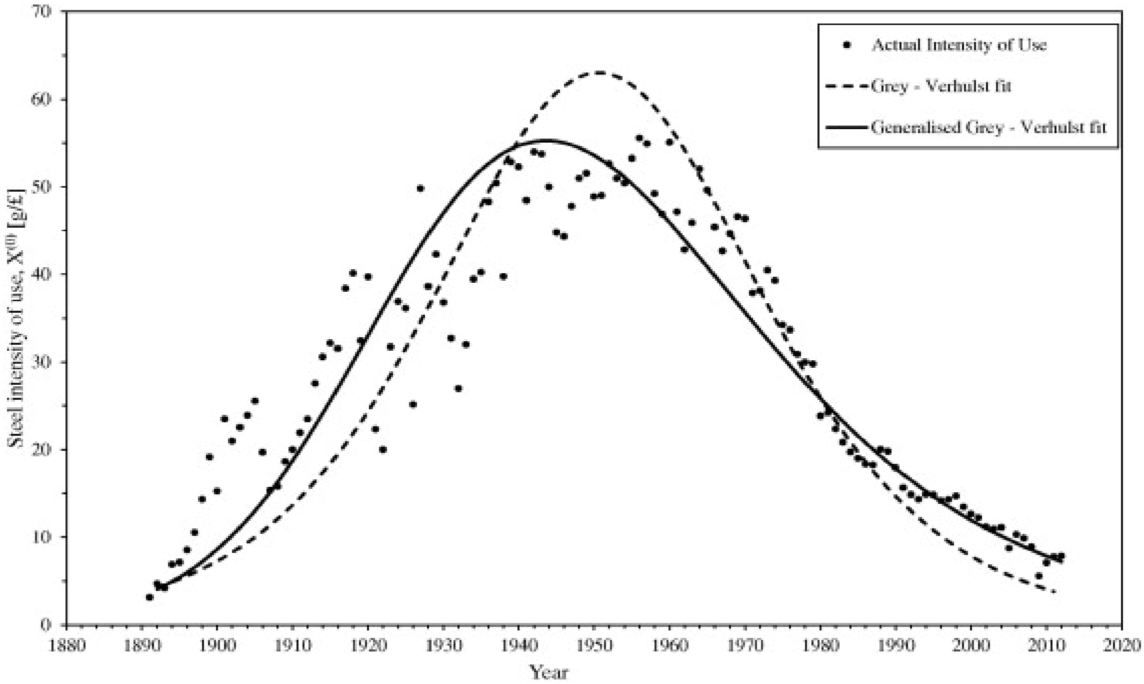


Figure 2.8. GVM Forecasting Model Sample.
(Source: Evans, 2014)

When k is sufficiently large, $x^{(1)}(k + 1)$ and $x^{(1)}(k)$ will be calculated at fairly close values. Due to this state of the GVM, it is often used to identify and forecast processes with the saturation region.

2.3.5. Fourier Residual Modification Model

In order to improve the modeling accuracy of grey models, several remedies have been discussed in the literature (Tan & Chang, 1996; Tan & Lu, 1996; Guo, Song, & Ye, 2005). In this study, fourier series have been used to modify the grey models.

Using the residual error, the modification of GM (1,1) and Grey Verhulst Models with Fourier Series can be defined as:

$$x^{(0)}(k + 1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak}\,(1 - e^{a}) \tag{19}$$

then, the error sequence of $X^{(0)}$ can be defined as:

$$\epsilon^{(0)} = (\epsilon^{(0)}(2), \epsilon^{(0)}(3), \ldots, \epsilon^{(0)}(n)) \tag{20}$$

where,

$$\epsilon^{(0)}(\mathrm{k}) = x^{(0)}(k) - x_p^{(0)}(k), \quad k = 2, 3, \ldots, n \tag{21}$$

The error residuals in Equation (21) can be expressed in Fourier series as follows:

$$\epsilon^{(0)}(k) \cong \frac{1}{2}\,a_0 + \sum_{i=0}^{z}[a_i \cos(\frac{2\pi i}{T}k) + b_i \sin(\frac{2\pi i}{T}k)] \tag{22}$$

$$T = n - 1 \;\text{ and }\; z = (\frac{n-1}{2}) - 1 \tag{23}$$

It is obvious that T will be an integer number and z will be selected as an integer number (Guo et al., 2005).

Equation (22) can be rewritten as follows:

$$\epsilon^{(0)} \cong PC \tag{24}$$

P and C matrixes can be defined as follows:

$$P = \begin{bmatrix} 1/2 & \cos\left(2\frac{2\pi}{T}\right) & \sin(2\frac{2\pi}{T}) & \cos\left(2\frac{2\pi 2}{T}\right) & \sin\left(2\frac{2\pi 2}{T}\right) & \cdots & \cos\left(2\frac{2\pi z}{T}\right) & \sin(2\frac{2\pi z}{T}) \\ 1/2 & \cos\left(3\frac{2\pi}{T}\right) & \sin(3\frac{2\pi}{T}) & \cos\left(3\frac{2\pi 2}{T}\right) & \sin\left(3\frac{2\pi 2}{T}\right) & \cdots & \cos\left(3\frac{2\pi z}{T}\right) & \sin(3\frac{2\pi z}{T}) \\ & \cdots & & & \cdots & & & \cdots \\ 1/2 & \cos\left(n\frac{2\pi}{T}\right) & \sin(n\frac{2\pi}{T}) & \cos\left(n\frac{2\pi 2}{T}\right) & \sin\left(n\frac{2\pi 2}{T}\right) & \cdots & \cos\left(n\frac{2\pi z}{T}\right) & \sin(n\frac{2\pi z}{T}) \end{bmatrix} \tag{25}$$

$$C = [a_0, a_1, b_1, a_2, b_2, \ldots a_2, b_2]^T \tag{26}$$

The least squares method can be used to solve the equation and the C matrix can be calculated:

$$C \cong (P^T P)^{-1} P^T \epsilon^{(0)} \tag{27}$$

The $x^{(0)}$ values corrected by the Fourier series should be calculated as follows:

$$x_{pf}^{(0)}(k) = x_p^{(0)}(k) - \epsilon_p^{(0)}(k), \ \ k = 2, 3, \ldots n + 1 \tag{28}$$

In mathematics, a Fourier Transform (FT) is a mathematical transform that decomposes a function (often a function of time, or a signal) into its constituent frequencies, such as the expression of a musical chord in terms of the volumes and frequencies of its constituent notes. The term Fourier Transform refers to both the frequency domain representation and the mathematical operation that associates the frequency domain representation to a function of time.

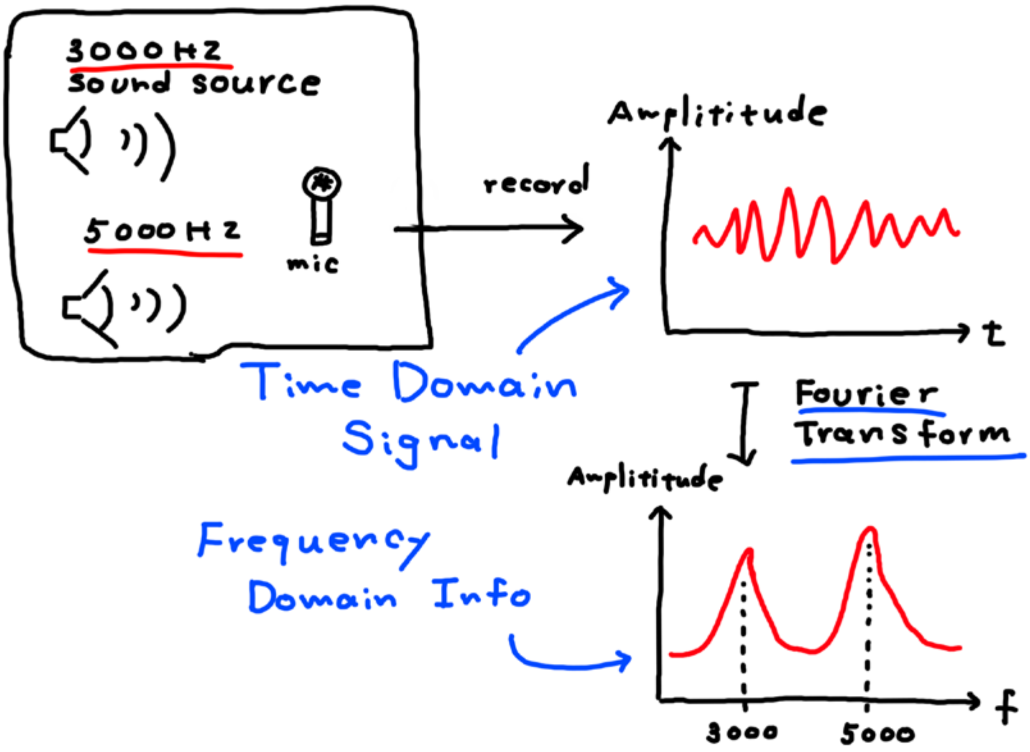


Figure 2.9. A Simple Fourier Transform Expression Chart

2.4. Evaluation Methods

All investigations and findings can be seen as fruitful and consistent, but there should always be room for doubt in science. Accordingly, three different evaluation methods, which are encountered and discussed in the literature reviews, are examined and described in the following section. These include **Mean Square Error, Root Mean Square Error and Mean Average Percentage Error**. The performance criteria of the results during the application can be evaluated with these evaluation methods.

Mean Square Error (MSE) is probably the most commonly used error metric. It penalizes larger errors because squaring larger numbers has a greater impact than squaring smaller numbers. The MSE is the sum of the squared errors divided by the number of observations.

Its formulation is as follows:

$$\text{MSE} = \frac{\sum_{t=1}^{n}(At-Ft)^2}{n}$$

Root Mean Square Error (RMSE) is the square root of **MSE**. It is effective in determining more precise errors by calculating with the square root method and comparing the performance criteria of the calculations.

Its formulation is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(At-Ft)^2}{n}}$$

Mean Average Percentage Error (MAPE) is the avarege of absolute errors divided by actual observation values. Absolute errors are formed by summing up and dividing by the number of observations, so it can be called the absolute average of the errors resulting from the calculation.

Its formulation is as follows:

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|\frac{At-Ft}{At}\right|}{n}$$

The MAPE calculation, which is one of the most widely used methods, will be evaluated in the advanced stages and the performance levels can be compared between the applied models. In this way, it will be possible to have an idea about the forecastings obtained with the values used and to make long-term and consistent estimates.

MAPE is also highly preferred in terms of evaluation accuracy, as it is a method that evaluates the forecasting made for values that change over a certain period of time, based on the difference between modeled and original data and the number of data targeted.

CHAPTER 3

TIME SERIES FORECASTING IN CONSTRUCTION

In the third chapter, the forecast models described in previous sections will be linked to the construction industry. Studies on this subject in the construction sector will be examined and presented, and the infrastructure will be established for forecastings in the next section. In line with this next chapter in the overall construction cost index in Turkey benefiting from calculations on these models will be made and examined for consistency with the current value.

Accurate cost estimation is important for managing construction projects. A construction project is generally considered successful when delivered within its budget and on time. Under these circumstances, accurate time and cost estimation has long been considered a critical function of project management in the industry.

The planning and control process also explains the importance of accurate efficiency forecasts: first, forecasts for the successful execution of construction activities inevitably require cost estimates, and second, the integrated management of construction projects must include the cost factor.

While many good forecasting models, such as regression, econometrics, and time series models, are well developed in theory and applied in practice, they all require the input of sufficient and appropriate data to generate accurate forecasts. If there is insufficient data or if the data is sufficient but does not follow certain distribution models, these models may not produce accurate forecasts.

To overcome these problems, this study aims to investigate whether the grey model can be used to estimate construction costs based on a limited amount of data, while also obtaining a degree of accuracy similar to other statistical forecasting models. Accordingly, studies in the literature on foresight in the construction industry have been researched.

3.1. Literature Review

The cost of a construction project is an indispensable parameter that must be analyzed by all stakeholders, including the owner, contractor and subcontractor. Increases in construction costs during the project may have a negative impact on the overall project, such as delay or termination of the project, high project costs and qualitative deterioration of construction, loss of confidence. Estimating the course of construction costs, estimating the cost of construction projects, as well as budget planning and coordination and assessing risks associated with costs are crucial (Moon and Shin, 2017).

Over the past few decades, many construction companies and project owners have suffered major capital losses, because total construction costs are much higher than originally planned and expected costs, although they have made significant efforts to make a positive contribution to construction costs (Touran and Lopez, 2006).

According to forecasters in construction companies, material price trends to be forecasted have the following characteristics: (1) no seasonal changes were found in most material price time series data; (2) tangible prices tend to remain constant, even if they rise once during the recession; (3) Since a large number of factors affect material prices, extensive data collection is required to make a single forecast (Williams, 1994).

According to Shane et al. (2009), this is largely a problem because it is not easy to accurately estimate construction costs due to the volatility and uncertainty in factors such as general construction material costs that directly affect the construction market. Macroeconomic conditions and socioeconomic factors are also effective in this.

In addition, there are great difficulties in forecasting the costs of construction that will continue for several months or several years in the planning process, because the time interval between the project planning process and the actual construction time is quite large.

The data used in the time series analysis were observed for a relatively long time. These data are calculated with the assumption that the system will be developed in the future and the forecasted values will be maintained in the future based on these time series data. In case of uncontrollable factors such as political changes, terrorism or wars, the time series data may have different patterns at different times due to the time-specific nature of such factors over a long period of time.

To achieve this goal, great efforts must be made to manage the construction process and cannot be done without a plan and cost control system. A control system periodically collects actual cost and scheduling data and then contradicts the schedule planned to measure work progress if it is ahead or behind the schedule and highlights potential problems (Teicholz, 1993).

Cost and schedule are two important parameters that play an important role in construction project management, and research on these parameters has been consistently proposed for the construction manager to provide appropriate methods and tools to achieve a project that will achieve its pre-construction goal. To achieve this goal, great efforts must be made to manage the construction process and cannot be done without a plan and cost control system. A control system periodically collects actual cost and scheduling data and then contradicts the schedule planned to measure work progress if it is ahead or behind the schedule and highlights potential problems (Teicholz, 1993). Cost and time are two important parameters that play an important role in construction project management, and research on these parameters has been consistently proposed for the construction manager to provide appropriate methods and tools to achieve a project that will achieve its pre-construction goal (Pewdum et al., 2009).

Many studies have focused on the rapidly changing construction material market and have attempted to address cost escalation factors to make cost planning more feasible. The main issues here are identifying escalation factors and estimating project costs accurately and simply (Hwang et al., 2012).

3.2. Used Forecasting Models in Construction Industry

There are several univariate-multivariate estimation methods that use condition-based or time series literary. Although not all of them are covered in the study, they are listed below with their descriptions:

- **Artificial Neural Network Model:**

Artificial Neural Networks (ANNs), usually simply called neural networks (NNs), are computing systems vaguely inspired by the biological neural networks that constitute animal brains.

An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron that receives a signal then processes it and can signal neurons connected to it. The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times (Chen et al., 2019).

- **Linear Regression Model:**

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are forecasted, rather than a single scalar variable (Freedman, 2009).

In linear regression, the relationships are modeled using linear forecaster functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or forecasters) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the forecasters, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis (Rencher and Christensen, 2012).

- **GM(1,1) Model:**

GM (1,1) type grey model is the most widely used in the literature and is referred to as "Grey Model First Rank One Variable". This model is a time series forecasting model. The differential equations of GM (1,1) model have coefficients that vary with time. In other words, the model is refreshed when new data is available in the forecasting model.

The GM (1,1) model can only be used in positive data series (Deng, 1989). In studies using GM (1,1), grey models can be used to forecast future values of primitive data points when all primitive data points are positive.

- **Grey Verhulst Model:**

The Verhulst model was first introduced in 1837 by a German biologist, Pierre Franois Verhulst. The main purpose of the Velhulst model is to limit the entire system to a real system and is effective in defining some incremental operations such as the S curve with saturation zone (Kayacan et al.,2010).

Verhulst proposed the Verhulst model when studying the law of biological reproduction.The main idea of Verhulst model is that the number of organism grows exponentially, but the growth rate of individual organism gradually slows down due to the environmental constraints, and it finally stabilizes at a fixed value; it is mainly used to describe the S-shaped process with saturation state (Fu et al., 2019).

- **Fourier Residual Modification Model:**

In mathematics, a Fourier Transform (FT) is a mathematical transform that decomposes a function (often a function of time, or a signal) into its constituent frequencies, such as the expression of a musical chord in terms of the volumes and frequencies of its constituent notes. The term Fourier Transform refers to both the frequency domain representation and the mathematical operation that associates the frequency domain representation to a function of time.

In order to improve the modeling accuracy of grey models, several remedies have been discussed in the literature (Tan & Chang, 1996; Tan & Lu, 1996; Guo, Song, & Ye, 2005). For example, in this study, fourier series was used to modify and improve the grey models' results used to forecast construction costs.

- **AutoRegressive Integrated Moving Average (ARIMA):**

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to forecast future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity (Box et al., 2016).

- **Vector AutoRegression (VAR):**

Vector autoregression (VAR) is a stochastic process model used to capture the linear interdependencies among multiple time series. VAR models generalize the univariate autoregressive model (AR model) by allowing for more than one evolving variable. All variables in a VAR enter the model in the same way: each variable has an equation explaining its evolution based on its own lagged values, the lagged values of the other model variables, and an error term. VAR modeling does not require as much knowledge about the forces influencing a variable as do structural models with simultaneous equations: The only prior knowledge required is a list of variables which can be hypothesized to affect each other intertemporally (Hatemi-J, 2004).

- **Vector Error Correction (VEC):**

An error correction model (ECM) belongs to a category of multiple time series models most commonly used for data where the underlying variables have a long-run stochastic trend, also known as cointegration. ECMs are a theoretically-driven approach useful for estimating both short-term and long-term effects of one time series on another. The term error-correction relates to the fact that last-period's deviation from a long-run equilibrium, the error, influences its short-run dynamics. Thus ECMs directly estimate the speed at which a dependent variable returns to equilibrium after a change in other variables (Phillips, 1985).

- **AutoRegressive Conditional Heteroskedasticity (ARCH):**

In econometrics, the autoregressive conditional heteroscedasticity (ARCH) model is a statistical model for time series data that describes the variance of the current error term or innovation as a function of the actual sizes of the previous time periods' error terms;[1] often the variance is related to the squares of the previous innovations. The ARCH model is appropriate when the error variance in a time series follows an autoregressive (AR) model; if an autoregressive moving average (ARMA) model is assumed for the error variance, the model is a generalized autoregressive conditional heteroskedasticity (GARCH) model (Engle, 1982).

- **Genetic Algorithm (GA):**

In computer science and operations research, a genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection (Mitchell, 1996). John Holland introduced genetic algorithms in 1960 based on the concept of Darwin’s theory of evolution, and his student David E. Goldberg further extended GA in 1989.

- **Fixed Weights Model (FWM):**

In statistics, a fixed effects model is a statistical model in which the model parameters are fixed or non-random quantities. This is in contrast to random effects models and mixed models in which all or some of the model parameters are random variables. In many applications including econometrics and biostatistics a fixed effects model refers to a regression model in which the group means are fixed (non-random) as opposed to a random effects model in which the group means are a random sample from a population. (Gardiner et al., 2009).

Generally, data can be grouped according to several observed factors. The group means could be modeled as fixed or random effects for each grouping. In a fixed effects model each group mean is a group-specific fixed quantity. (Ramsey and Schafer, 2002).

- **Moving Weights Model (MWM):**

A weighted average is an average that has multiplying factors to give different weights to data at different positions in the sample window. Mathematically, the weighted moving average is the convolution of the datum points with a fixed weighting function. The application is removing pixelisation(unfuzziness from a digital graphical image. In technical analysis of financial data, a weighted moving average (WMA) has the specific meaning of weights that decrease in arithmetical progression (Devcic, 2010).

Table 3.1. Classification of Forecasting Models in the Literature.

AUTHOR, DATE	CATEGORY	MODEL	COUNTRY	TIME SPAN	EVALUATION METHOD
Chau et al., 2005	Causal	Regression Analysis	China	360 Days	RMSE-CC
Chau et al., 2007	Causal	ANN-GA	China	360 Days	RMSE-CC
Thomas Ng et al., 2004	Causal-TS	Regression Analysis	UK(TPI)	8 Years	MSE-RMSE
Thomas Ng et al., 2004	Causal-TS	ARIMA	UK(TPI)	8 Years	MSE-RMSE
Hwang et al., 2012	Time Series	ARIMA	USA(CCI)	12 Months	MAPE
Hyung-Keun Park, 2004	Causal	FWM	South Korea	12 Months	MAD
Hyung-Keun Park, 2004	Causal	MWM	South Korea	12 Months	MAD
Seokyon Hwang, 2009	Causal	Regression Models	USA(CCI)	12-24 Months	MRSS
J. Scott Armstrong, 1992	-	-	USA	Annual-Quarterly	GMRAE-MAPE
Moon et al., 2017	Time Series	ARIMA	USA(CCI)	12 Months	MAE-MAPE-RMSE
Paul H. K. Ho, 2010	Time Series	GM (1,1)	China	64 Quarters	MAPE
Faghih et al., 2018	Time Series	ARIMA, VAR	USA(CCI)	36 Months	RMSE-MAE-MAPE
Shahandashti et al., 2013	Time Series	VEC	USA(CCI)	36 Months	MAPE-MSE
Ashuri et al., 2010	Time Series	SARIMA	USA(CCI)	12 Months	MAPE-MSE-MAE
Nogales et al., 2002	Causal	Dynamic Regression	Spain-USA	168 Weeks	MWE-FMSE
Sing et al., 2015	Time Series	VAR	China	14 Quarters	MAPE
Ilbeigi et al., 2016	Causal	ARCH-GARCH	USA(NAPA)	12 Months	MAPE-MSE-MAE
Rifat Sonmez, 2008	Causal	Regression Models	TR	20 Parameters	MAPE
Lowe et al., 2006	Causal	RA-ANN	UK	8 Parameters	MAPE
Sonmez et al., 2007	Time Series	Linear Regression	21 Countries	14 Parameters	MAE
Taylor et al., 2008	Time Series	SARIMA	EU(10 Countries)	24 Hours	MAE-MAPE

As a result of the literature review, as a statement to the above authors and groups to the studies; the subject is summarized by standardizing the causal or time series categories, models, countries, time spans and evaluation methods.

Chau et al. (2005-2007), causal based models were investigated and calculations were made over a 360-day period in China using Regression Analysis. RMSE and CC were used as evaluation method.

In the studies of Thomas Ng et al., (2004), causal and time series based models were investigated and calculations were made over an 8-year period in UK using Regression Analysis. MSE and RMSE were used as the evaluation method.

In the studies of Hwang et al., (2012), time series based models were investigated and calculations were made over an 12-month period in USA using ARIMA. And MAPE was used as the evaluation method.

Hyung-Keun Park (2004), causal based models were investigated and calculations were made over a 12-month period in South Korea using FWM and MWM. MAD were used as evaluation method.

In the studies of Seokyon Hwang (2009), causal based models were investigated and calculations were made over an 12 and 24-month period in USA using Regression Analysis. And MRSS was used as the evaluation method.

J. Scott Armstrong (1992), compared the performance criteria with GMRAE and MAPE by evaluating the data calculated in the USA.

In the studies of Moon et al. (2017), time series based models were investigated and calculations were made over an 12-month period in USA using ARIMA. MAE, MAPE and RMSE were used as the evaluation method.

Paul H. K. Ho (2010), time series based models were investigated and calculations were made over a 64-quarter period in China using GM (1,1). MAPE was used as evaluation method.

In the studies of Faghih et al. (2018), time series based models were investigated and calculations were made over an 36-month period in USA using ARIMA and VAR. MAE, MAPE and RMSE were used as the evaluation method.

Shahandashti et al. (2013), time series based models were investigated and calculations were made over an 36-month period in USA using VEC. MAPE and MSE were used as the evaluation method.

In the studies of Ashuri et al. (2010), time series based models were investigated and calculations were made over an 12-month period in USA using SARIMA. MAPE, MSE and RMSE were used as the evaluation method.

Nogales et al. (2002), causal based models were investigated and calculations were made over an 168-week period in Spain and USA using Dynamic Regression. MWE and FMSE were used as the evaluation method.

In the studies of Sing et al. (2015), time series based models were investigated and calculations were made over an 14-quarter period in China using VAR. MAPE was used as the evaluation method.

Ilbeigi et al. (2016), causal based models were investigated and calculations were made over an 12-month period in USA using ARCH and GARCH. MAE, MAPE and MSE were used as the evaluation method.

In the studies of Sonmez (2008), causal based models were investigated and calculations were made over an 20-parameter period in Turkey using Regression Models. MAPE was used as the evaluation method.

Sonmez et al. (2007), time series based models were investigated and calculations were made over an 14-parameter period in 21 Countries using Linear Regression. MAE was used as the evaluation method.

In the studies of Lowe et al. (2006), causal based models were investigated and calculations were made over an 8-parameter period in UK using Regression Models. MAPE was used as the evaluation method.

Taylor et al. (2008), time series based models were investigated and calculations were made over an 24-hour period in 10 EU Countries using SARIMA. MAE and MAPE were used as the evaluation method.

CHAPTER 4

RESEARCH METHOD

4.1. Use of Time Series Forecasting Models in the Models

In line with various types forecating strategies and evaluation methods discussed in the previous section, on the basis of Turkey; **Linear Regression Model**, **GM (1,1)**, **Grey Verhulst Model** and finally **Fourier Transform with residual values** will be applied. The application steps of these models will be presented at this chapter, presentation and evaluation of the resulting values will also be discussed in the next chapter. In the light of this method, it is aimed to test the efficiency of the models used.

4.2. Time Series for Forecasting of Construction Costs Figures

As noted previously in earlier chapters, statistical data is indisputably important for the forecasting ability. In this study, all statistical data in the country and play an important role for Turkey, which is responsible for creating, TurkStat datas will be used.

The source of this data is provided from, which is Turkey's institutions engaged in statistical data and planning, Turkish Statistical Institute (TurkStat, TÜİK in Turkish). A range of these data will be taken in the near term and firstly will be subjected to Quasi-Smoothness and Quasi-Exponentiality tests to determine their suitability for the models, and then these data will be processed through the models.

As seen above, the data in the table out from Tüik, to apply the model as a data set with the nearest maturity and importance, exchange of the construction cost index in Turkey in 2019 is based. In this direction, modeling will be done on the first 9 months of data and the consistency of forecasting the next months or a longer period will be tested.

Table 4.1. Construction Cost Index and Rate of Change, 2015-2020

İnşaat maliyet endeksi ve değişim oranı, 2015-2020
Construction cost index and rate of change, 2015-2020
[2015=100]

Yıl Year	**Ocak** January	**Şubat** February	**Mart** March	**Nisan** April	**Mayıs** May	**Haziran** June	**Temmuz** July	**Ağustos** August	**Eylül** September	**Ekim** October	**Kasım** November	**Aralık** December
						Endeks - Index						
2015	97,13	97,65	98,27	99,14	100,17	100,03	100,83	101,35	102,25	101,56	101,01	100,60
2016	108,19	108,12	109,39	110,26	112,11	111,60	111,52	111,90	112,19	113,06	115,83	118,90
2017	124,69	125,09	126,43	126,84	127,26	127,06	128,16	129,51	130,94	132,76	136,09	138,14
2018	144,92	146,60	149,08	152,10	156,58	160,17	162,78	172,71	182,87	182,57	176,85	173,57
2019	184,83	186,51	189,25	192,27	195,51	193,97	192,76	191,35	190,23	190,36	190,32	192,25
						Bir önceki aya göre değişim (%) - Monthly rate of change (%)						
2015	-	0,54	0,63	0,89	1,04	-0,14	0,80	0,52	0,89	-0,67	-0,54	-0,41
2016	7,54	-0,06	1,17	0,80	1,68	-0,45	-0,07	0,34	0,26	0,78	2,45	2,65
2017	4,87	0,32	1,07	0,32	0,33	-0,16	0,87	1,05	1,10	1,39	2,51	1,51
2018	4,91	1,16	1,69	2,03	2,95	2,29	1,63	6,10	5,88	-0,16	-3,13	-1,85
2019	6,49	0,91	1,47	1,60	1,69	-0,79	-0,62	-0,73	-0,59	0,07	-0,02	1,01
						Bir önceki yılın aynı ayına göre değişim (%) - Annual rate of change (%)						
2015	-	-	-	-	-	-	-	-	-	-	-	-
2016	11,39	10,72	11,32	11,22	11,92	11,57	10,60	10,41	9,72	11,32	14,67	18,19
2017	15,25	15,70	15,58	15,04	13,51	13,85	14,92	15,74	16,71	17,42	17,49	16,18
2018	16,22	17,20	17,92	19,91	23,04	26,06	27,01	33,36	39,66	37,52	29,95	25,65
2019	27,54	27,22	26,95	26,41	24,86	21,10	18,42	10,79	4,02	4,27	7,62	10,76
						On iki aylık ortalamalara göre değişim oranı (%) - Rate of change in twelve months moving averages (%)						
2015	-	-	-	-	-	-	-	-	-	-	-	-
2016	-	-	-	-	-	-	-	-	-	-	-	11,92
2017	12,26	12,69	13,06	13,38	13,51	13,69	14,04	14,47	15,04	15,54	15,77	15,63
2018	15,71	15,85	16,06	16,48	17,30	18,34	19,37	20,87	22,83	24,52	25,54	26,26
2019	27,17	27,96	28,65	29,12	29,18	28,66	27,82	25,77	22,61	19,78	17,95	16,77

TÜİK, İnşaat Maliyet Endeksi, Ocak 2020
TurkStat, Construction Cost Index, January 2020

4.3. Applications of Time Series Forecasting

According to the information in the Table 4.1, shows Construction Cost Index values for the values given above for the period between 2015-2020 year period in Turkey. Among these, the models will be applied by taking the Construction Cost Index January-December values for 2019. First, data over the first nine months (January-September) will be modeled and forecasts for the next three months (October-December).

At this point, the Quasi-Smoothness and Quasi-Exponentiality states of the referenced values will be tested and it will be seen whether they are within the standard limits. For this, the following methods will be used respectively.

The time series used is as follows:

$$x^{(0)} = (184.83, 186.51, 189.25, 192.27, 195.51, 193.97, 192.76, 191.35, 190.23) \quad (1)$$

Generate $x^{(1)}$ by AGO,

$$x^{(1)}(k) = \sum_{n=1}^{k} x^{(0)}(n)$$

$$x^{(1)} = (184.83, 371.34, 560.59, 752.86, 948.37, 1142.34, 1335.10, 1526.45, 1716.68) \quad (2)$$

Check quasi-smoothness on $x^{(0)}$

$$\rho(k) = \frac{x^{(0)}(k)}{x^{(1)}(k-1)}, \qquad k = 3, 4, \ldots, n$$

When $k > 3$ and $0 < \rho(k) < 0.5$,

$$\rho(3) = \frac{x^{(0)}(3)}{x^{(1)}(2)} = \frac{189.25}{371.34} \cong 0.5096 \quad (3)$$

$$\rho(4) = \frac{x^{(0)}(4)}{x^{(1)}(3)} = \frac{192.27}{560.59} \cong 0.3430 \quad (4)$$

$$\rho(5) = \frac{x^{(0)}(5)}{x^{(1)}(4)} = \frac{195.51}{752.86} \cong 0.2597 \quad (5)$$

$$\rho(6) = \frac{x^{(0)}(6)}{x^{(1)}(5)} = \frac{193.97}{948.37} \cong 0.2045 \quad (6)$$

$$\rho(7) = \frac{x^{(0)}(7)}{x^{(1)}(6)} = \frac{192.76}{1141.34} \cong 0.1687 \quad (7)$$

$$\rho(8) = \frac{x^{(0)}(8)}{x^{(1)}(7)} = \frac{191.35}{1335.10} \cong 0.1433 \quad (8)$$

$$\rho(9) = \frac{x^{(0)}(9)}{x^{(1)}(8)} = \frac{190.23}{1526.45} \cong 0.1246 \quad (9)$$

The quasi-smoothness values must be $0 < \rho(k) < 0.5$,

Looking at Equation (3) to (9), it is seen that the ρ values are in the range. Thus, the quasi-smoothness test was provided.

Quasi-exponentiality must also be tested, check quasi-exponentiality on x(1)

When $k > 3$, $1 < \sigma(k) < 1.5$,

$$\sigma(3) = \frac{x^{(1)}(3)}{x^{(1)}(2)} = \frac{560.59}{371.34} \cong 1.5096 \quad (10)$$

$$\sigma(4) = \frac{x^{(1)}(4)}{x^{(1)}(3)} = \frac{752.86}{560.59} \cong 1.3430 \quad (11)$$

$$\sigma(5) = \frac{x^{(1)}(5)}{x^{(1)}(4)} = \frac{948.37}{752.86} \cong 1.2597 \quad (12)$$

$$\sigma(6) = \frac{x^{(1)}(6)}{x^{(1)}(5)} = \frac{1142.34}{948.37} \cong 1.2045 \quad (13)$$

$$\sigma(7) = \frac{x^{(1)}(7)}{x^{(1)}(6)} = \frac{1335.10}{1142.34} \cong 1.1687 \quad (14)$$

$$\sigma(8) = \frac{x^{(1)}(8)}{x^{(1)}(7)} = \frac{1526.45}{1335.10} \cong 1.1433 \quad (15)$$

$$\sigma(9) = \frac{x^{(1)}(9)}{x^{(1)}(8)} = \frac{1716,68}{1526.45} \cong 1.1246 \quad (16)$$

The quasi-exponentiality values must be $1 < \sigma(k) < 1.5$,

Looking at Equation (10) to (16), it is seen that the σ values are in the range. Thus, the quasi-exponentiality test was provided.

The quasi-exponentiality and quasi-smoothness of $x^{(0)}$ is satisfied. Therefore, $x^{(0)}$ values can be applied on models.

Table 4.2. Quasi-smoothness and quasi-exponentiality process of the data set.

Months		**y**	**x^1**	**z^1**	**Q. Smoothness**		**Q. Exponentiality**		*√checked*
January	1	**184,83**	184,83						
February	2	**186,51**	371,34	185,67					
March	3	**189,25**	560,59	280,30	*ρ(3) =*	0,5096	*σ(3) =*	1,5096	
April	4	**192,27**	752,86	376,43	*ρ(4) =*	0,3430	*σ(4) =*	1,3430	
May	5	**195,51**	948,37	474,19	*ρ(5) =*	0,2597	*σ(5) =*	1,2597	
June	6	**193,97**	1142,34	571,17	*ρ(6) =*	0,2045	*σ(6) =*	1,2045	
July	7	**192,76**	1335,10	667,55	*ρ(7) =*	0,1687	*σ(7) =*	1,1687	
August	8	**191,35**	1526,45	763,23	*ρ(8) =*	0,1433	*σ(8) =*	1,1433	
September	9	**190,23**	1716,68	858,34	*ρ(9) =*	0,1246	*σ(9) =*	1,1246	
October	10	190,36	1907,04	953,52	*ρ(10) =*	0,1109	*σ(10) =*	1,1109	
November	11	190,32	2097,36	1048,68	*ρ(11) =*	0,0998	*σ(11) =*	1,0998	
December	12	192,25	2289,61	1144,81	*ρ(12) =*	0,0917	*σ(12) =*	1,0917	
2019					$0 \leq \rho \leq 0,5$		$1 \leq \sigma \leq 1,5$		
2015=100					***in the range***		***in the range***		

4.3.1. Linear Regression Model Application

Construction cost data for January-September 2019 are determined as fixed data. Linear Regression Model will be applied by calculating the received fixed data via Excel.

The first nine months will be used as reference data for the model, and the remaining three months will be envisaged. In this way, the reliability of the model will be tested by comparing the values we have with the forecasted values.

Linear Regression Model was applied and formulated via Excel over the first 9 months data set. In this direction, calculations were made depending on the time variable.

Table 4.3. Application Steps of the Used Linear Regression Model.

Months	**x^0**	**y**		**Formula**	**Forecast**
January	1	**184,83**		y = 0,7473x + 187,01	187,76
February	2	**186,51**	1,68		188,50
March	3	**189,25**	2,74		189,25
April	4	**192,27**	3,02		190,00
May	5	**195,51**	3,24		190,75
June	6	**193,97**	-1,54		191,49
July	7	**192,76**	-1,21		192,24
August	8	**191,35**	-1,41		192,99
September	9	**190,23**	-1,12		193,74
October	10	190,36	0,13		**194,48**
November	11	190,32	-0,04		**195,23**
December	12	192,25	1,93		**195,98**
2019					
2015=100					

The necessary formula for forecasting the next months with the index of the first nine months based on the values based on the months has been reached and forecasts for the next months have been made.

Table 4.4. Results and Evaluation Steps of the Used Linear Regression Model.

Dev.1	**Dev.2**	**Abs.Dev.Range**	**Mape1**	**Abs.Dev.Range**	**Mape2**	**Mse**	**Rmse**
2,93		0,0158	**1,17**		**1,43**	**9,73**	**3,12**
1,99		0,0107					
0,00		0,0000					
-2,27		0,0118					
-4,76		0,0244					
-2,48		0,0128					
-0,52		0,0027					
1,64		0,0086					
3,51		0,0184					
	4,12			**0,0217**			
	4,91			**0,0258**			
	3,73			**0,0194**			

Mape values were calculated by evaluating the forecastings reached for the first nine months and the output for the next three months. Mape1 represents the model's performance criterion in the first nine months, while Mape2 is the model's performance criterion for the results of the last three months out of sample. Mse and Rmse performance criterion were also calculated for the Linear Regression Model.

4.3.2. GM (1,1) Model Application

Construction cost data for January-September 2019 are determined as fixed data. GM(1,1) Model will be applied by calculating the received fixed data via Excel.

The first nine months will be used as reference data for the model, and the remaining three months will be envisaged. In this way, the reliability of the model will be tested by comparing the values we have with the forecasted values.

GM (1,1) Model was applied and formulated via Excel over the first nine months data set. In this direction, calculations were made depending on the time variable.

Table 4.5. Application Steps of the Used GM (1,1) Model (Stage 1).

Months		**y**		**x^1**	**z^1**	**[B]**	
January	1	**184,83**		184,83			
February	2	**186,51**	1,68	371,34	185,67	-185,67	1
March	3	**189,25**	2,74	560,59	280,30	-280,30	1
April	4	**192,27**	3,02	752,86	376,43	-376,43	1
May	5	**195,51**	3,24	948,37	474,19	-474,19	1
June	6	**193,97**	-1,54	1142,34	571,17	-571,17	1
July	7	**192,76**	-1,21	1335,10	667,55	-667,55	1
August	8	**191,35**	-1,41	1526,45	763,23	-763,23	1
September	9	**190,23**	-1,12	1716,68	858,34	-858,34	1
October	10	190,36	0,13	1907,04	953,52		
November	11	190,32	-0,04	2097,36	1048,68		
December	12	192,25	1,93	2289,61	1144,81		
2019							
2015=100							

The $x^{(1)}$ series was created over the $x^{(0)}$ values and then the $z^{(1)}$ series was created. Then, the B matrix was created and the process process of the GM (1,1) model was applied.

Table 4.6. Application Steps of the Used GM (1,1) Model (Stage 2).

$[B]^T$							
-185,67	-280,30	-376,43	-474,19	-571,17	-667,55	-763,23	-858,34
1	1	1	1	1	1	1	1
$[B^TB]^{-1}$	0,00	0,00					
	0,00	0,82					
B^Ty	#######		**$(B^TB)^{-1}\ B^Ty =$**		**$(a,b)^T$**	-0,0045	
	1531,85					189,1249	
$[B^TB]^{-1}\ B^Ty$		-0,0045					
		189,12					

a and b coefficients were obtained by forming $[B^TB]^T$ and B^Ty over the B matrix and its transpose.

Table 4.7. Results and Evaluation Steps of the Used GM (1,1) Model.

x^(1)(k)	**x^(0)(k)**	**Dev.1**	**Dev.2**	**Abs.Dev.Range**	**Mape1**	**Abs.Dev.Range**	**Mape2**	**Mse**	**Rmse**
184,83	184,83				**1,24**		**1,88**	**20,63**	**4,54**
375,22	190,39	3,88		0,0208					
566,47	191,25	2,00		0,0106					
758,58	192,11	-0,16		0,0008					
951,57	192,98	-2,53		0,0129					
1145,42	193,86	-0,11		0,0006					
1340,16	194,73	1,97		0,0102					
1535,77	195,61	4,26		0,0223					
1732,27	196,50	6,27		0,0330					
1929,66	**197,39**		**7,03**			**0,0369**			
2127,94	**198,28**		**7,96**			**0,0418**			
2327,12	**199,18**		**6,93**			**0,0360**			

By applying the GM (1,1) equation on the constant values of the first referenced period, forecasted $x^{(0)}$ values were reached.

Mape values were calculated by evaluating the forecastings reached for the first nine months and the output for the next three months. Mape1 represents the model's performance criterion in the first nine months, while Mape2 is the model's performance criterion for the results of the last three months out of sample. Mse and Rmse performance criterion were also calculated for the GM (1,1) Model.

4.3.3. Grey Verhulst Model Application

Construction cost data for January-September 2019 are determined as fixed data. Grey Verhulst Model will be applied by calculating the received fixed data via Excel.

The first nine months will be used as reference data for the model, and the remaining three months will be envisaged. In this way, the reliability of the model will be tested by comparing the values we have with the forecasted values.

Grey Verhulst Model was applied and formulated via Excel over the first nine months data set. In this direction, calculations were made depending on the time variable.

Table 4.8. Application Steps of the Used GVM (Stage 1).

Months		**y**		**x^1**	**z^1**	**[B]**	
January	1	**184,83**		184,83			
February	2	**186,51**	1,68	371,34	185,67	-185,67	34473,35
March	3	**189,25**	2,74	560,59	187,88	-187,88	35298,89
April	4	**192,27**	3,02	752,86	190,76	-190,76	36389,38
May	5	**195,51**	3,24	948,37	193,89	-193,89	37593,33
June	6	**193,97**	-1,54	1142,34	194,74	-194,74	37923,67
July	7	**192,76**	-1,21	1335,10	193,37	-193,37	37390,02
August	8	**191,35**	-1,41	1526,45	192,06	-192,06	36885,12
September	9	**190,23**	-1,12	1716,68	190,79	-190,79	36400,82
October	10	190,36	0,13	1907,04	190,30		
November	11	190,32	-0,04	2097,36	190,34		
December	12	192,25	1,93	2289,61	191,29		
2019							
2015=100							

The $x^{(1)}$ series was created over the $x^{(0)}$ values and then the $z^{(1)}$ series was created. Then, the B matrix was created and the process process of the Grey Verhulst Model was applied.

Table 4.9. Application Steps of the Used GVM (Stage 2).

$[B]^T$							
-185,67	-187,88	-190,76	-193,89	-194,74	-193,37	-192,06	-190,79
34473,35	35298,89	36389,38	37593,33	37923,67	37390,02	36885,12	36400,82
$[B^TB]^{-1}$	0,015205	7,949E-05					
	7,95E-05	4,157E-07					
B^Ty	-1012,66			$(a,b)^T$	-0,30		
	189911,2				0,00		
$[B^TB]^{-1} B^Ty$		-0,300865					
		-0,0015552					

The order and content of the Grey Verhulst Model enables the forecasted values to be reached with the a and b coefficients reached through the B matrices as described in the previous chapter.

Table 4.10. Results and Evaluation Steps of the Used GVM.

$\hat{x}^{(0)}(k)$	Dev.1	Dev.2	Abs.Dev.Range	Mape1	Abs.Dev.Range	Mape2	Mse	Rmse
184,83				**0,97**		**0,99**	**5,16**	**2,27**
187,00	-0,49		0,0026					
188,63	0,62		0,0033					
189,86	2,41		0,0125					
190,79	4,72		0,0242					
191,47	2,50		0,0129					
191,98	0,78		0,0040					
192,37	-1,02		0,0053					
192,65	-2,42		0,0127					
192,86		**-2,50**			**0,0131**			
193,01		**-2,69**			**0,0142**			
193,13		**-0,88**			**0,0046**			

By applying the Grey Verhulst Model equation on the constant values of the first referenced period, $x^{(0)}$ forecasted values were reached.

Mape values were calculated by evaluating the forecastings reached for the first nine months and the output for the next three months. Mape1 represents the model's performance criterion in the first nine months, while Mape2 is the model's performance criterion for the results of the last three months out of sample. Mse and Rmse performance criterion were also calculated for the Grey Verhulst Model.

4.3.4. Fourier Residual Modification Model Application

In the previous title, the foracasting that emerged from the Grey Verhulst Model application was compared with the original values and the deviations were calculated.

In this application, how to reach a more ideal forecasting by using these deviations. The first nine months will be used as reference data for the model, and the remaining three months will be envisaged. In this way, the reliability of the model will be tested by comparing the values we have with the forecasted values.

Table 4.11. Application Steps of the Fourier Transform (Stage 1).

Months		**y**	**xˆ(0)(k)**	**Dev.1**	***P***	(8*7)					
January	1	184,83	**184,83**								
February	2	186,51	**187,00**	-0,49	0,5	0,00	1,00	-1,00	0,00	0,00	-1,00
March	3	189,25	**188,63**	0,62	0,5	-0,71	0,71	0,00	-1,00	0,71	0,71
April	4	192,27	**189,86**	2,41	0,5	-1,00	0,00	1,00	0,00	-1,00	0,00
May	5	195,51	**190,79**	4,72	0,5	-0,71	-0,71	0,00	1,00	0,71	-0,71
June	6	193,97	**191,47**	2,50	0,5	0,00	-1,00	-1,00	0,00	0,00	1,00
July	7	192,76	**191,98**	0,78	0,5	0,71	-0,71	0,00	-1,00	-0,71	-0,71
August	8	191,35	**192,37**	-1,02	0,5	1,00	0,00	1,00	0,00	1,00	0,00
September	9	190,23	**192,65**	-2,42	0,5	0,71	0,71	0,00	1,00	-0,71	0,71
October	10	190,36	192,86								
November	11	190,32	193,01								
December	12	192,25	193,13								
2019											
2015=100											

The table above shows the stage of forming the P matrix with the Fourier formulation.

Table 4.12. Application Steps of the Fourier Transform (Stage 2).

P ^{T}	(7*8)						
0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
0,00	-0,71	-1,00	-0,71	0,00	0,71	1,00	0,71
1,00	0,71	0,00	-0,71	-1,00	-0,71	0,00	0,71
-1,00	0,00	1,00	0,00	-1,00	0,00	1,00	0,00
0,00	-1,00	0,00	1,00	0,00	-1,00	0,00	1,00
0,00	0,71	-1,00	0,71	0,00	-0,71	1,00	-0,71
-1,00	0,71	0,00	-0,71	1,00	-0,71	0,00	0,71

Table 4.13. Application Steps of the Fourier Transform (Stage 3).

$(P^TP)^{-1} P^T$							
0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25
0,00	-0,18	-0,25	-0,18	0,00	0,18	0,25	0,18
0,25	0,18	0,00	-0,18	-0,25	-0,18	0,00	0,18
-0,25	0,00	0,25	0,00	-0,25	0,00	0,25	0,00
0,00	-0,25	0,00	0,25	0,00	-0,25	0,00	0,25
0,00	0,18	-0,25	0,18	0,00	-0,18	0,25	-0,18
-0,25	0,18	0,00	-0,18	0,25	-0,18	0,00	0,18

By applying the Fourier equation to P matrices, the *a* series is calculated and combined with the Grey Verhulst Model outputs, resulting in more realistic results.

Table 4.14. Application Steps of the Fourier Transform (Stage 4).

C	***Xpf***	**y**	**xˆ(0)(k)**	**Dev.1**	**Dev.2**
		184,83	**184,83**		
1,7748	**185,22**	186,51	**187,00**	-0,49	1,29
-2,0899	**190,72**	189,25	**188,63**	0,62	-1,47
-2,0370	**191,90**	192,27	**189,86**	2,41	0,37
-0,1550	**190,94**	195,51	**190,79**	4,72	4,57
0,2288	**191,24**	193,97	**191,47**	2,50	2,73
0,3793	**191,61**	192,76	**191,98**	0,78	1,15
-0,5448	**192,91**	191,35	**192,37**	-1,02	-1,56
		190,23	**192,65**	-2,42	

Table 4.15. Results and Evaluation Steps of the Fourier Transform.

Abs.Dev.	**Mape1**	**Abs.Dev.**	**Mape2**	**Mse**	**Rmse**
	0,97		**0,97**	**4,51**	**2,12**
0,0026		0,0069			
0,0033		0,0078			
0,0125		0,0019			
0,0242		0,0234			
0,0129		0,0141			
0,0040		0,0060			
0,0053		0,0082			
0,0127					

CHAPTER 5

RESEARCH FINDINGS AND DISCUSSIONS

5.1. Causal Based Forecasting Models

Causal models are seen as mathematical models that make sense of causal relationships within some existing systems or populations. Based on statistical data, they derive inferences that can be made through causal relationships. It allows comment on the epistemology of causality and the relationship between causality and probability and facilitates its analysis.

Causal models try to make predictions about the behavior of a system. In particular, a causal model examines the true value or probability of counterfactual claims about the system; predicts the effects of changes; and the probabilistic dependence or independence of the variables included in the model.

With causal models, it may be easier to interpret the following: By observing the results of possible correlations or experimental changes between variables, it can be determined which causal patterns are consistent with these observations. Accordingly, the outputs of the model will guide what can be done "in principle".

"Machine Learning" (ML) is the study of computer algorithms that are automatically developed through experience. It can be seen as a subset of artificial intelligence. Machine learning algorithms create a mathematical model based on sample data known as "training data" to make estimations or decisions without being explicitly programmed. While machine learning algorithms are used in a wide variety of applications such as email filtering and computer vision, it seems difficult or possible to develop traditional algorithms to perform the required tasks while providing many estimations.

Connection systems, called artificial neural networks (ANNs) or often called neural networks (NNs), can be viewed as computational systems that ambiguously sample the biological neural networks that make up animal brains. The data structures and functionality of neural networks are designed to simulate relational memory. Neural networks learn by processing samples that contain a known "input" and "result" and are thought to create stored probabilistic relationships between the two in the data structure of the network.

In the field of computer science, the genetic algorithm (GA) is a metasismatic method inspired by the natural selection process of the larger evolutionary class of algorithms (EA). Genetic algorithms are widely used to generate high quality and stable solutions to optimization problems based on biologically inspired operators such as mutation, transition, and selection.

Decision Tree Learning is one of the predictive modeling approaches used in statistics, data mining and machine learning. A decision tree (as a forecast model) can be used to navigate from the results about an item (represented in the branches) to the target value of the item (represented in the leaves). Tree models in which the target variable can take a separate set of values can be called classification trees; In these tree structures, leaves represent class tags and branches are expressed as combinations of traits that lead to these class tags.

Support vector machines (SVMs and support vector networks) are represented as supervised learning models with corresponding learning algorithms that analyze data used in machine learning for classification and regression analysis. The Support Vector Machine (SVM) algorithm, one of the popular machine learning tools, can be deduced to provide solutions to both classification and regression problems.

5.2. Time Series Forecasting Models

A time series is a series of data points that are indexed (or listed or graphed) in order of time. More commonly, the time series is a consecutive sequence and preferably taken from equidistant points. In other words, they can be called discrete time data strings. Examples of time series are the intensity of earthquake waves, the number of COVID-19 patients, and the daily closing values of the Stock Markets.

Time series are usually expressed in line charts. It can be easily said that time series is widely used in architecture, project management, statistics, pattern recognition, communication engineering, econometrics, finance, weather forecasting, earthquake forecasting, control engineering, astronomy, and any applied science that includes largely temporal measurements.

Time series analysis includes consistent statistics and methods for inference by processing time series data to extract hidden properties of the data. Time series estimates are the use of a suitable model to forecast future values based on current known values. For example; Regression analysis is often used to test theories that current values of one or more independent time series affect the current value of another time series, and such time series analysis is not called "time series analysis". It can be said that time series analysis focuses on comparing the values of a single time series or different dependent time series at different points in time.

Linear regression can be called a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The state of an explanatory variable can be called simple linear regression. For more than one explanatory variable, the process can be called multiple linear regression. This term is positioned differently from multivariate linear regression in which multiple associated dependent variables are forecasted instead of a single scalar variable.

In system theory, a system can be identified by a color that represents a clear amount of information about that system. For example, if the mathematical equations describing the internal properties or dynamics are not known exactly, the system may be considered a black box. On the other hand, if the description of the system is known exactly, it can be called a white system.

Similarly, a system with both known and unknown information can be defined as a grey system. In real life, any system can be considered a grey system because there is always some possibility of uncertainty. The cause of concern is the noise (and the limitations of cognitive abilities) that come from both inside and outside the system, so the knowledge we have about this system is always vague and limited in scope.

Grey models can be said to forecast future values of a time series based only on the most recent data based on the estimator's window size. It is assumed that all data values to be used in grey models are positive and the sampling frequency of time series is constant. In the simplest terms, the grey models to be formulated below can be viewed as curve fitting approaches. The main task of the grey system theory can be said to apply the system's realistic management laws using available data. This process is considered the creation of the grey system.

The GM (1,1) type grey model can be said to be one of the most widely used models in the literature and is called "Grey Pattern First Order One Variable". This model is a time series forecasting model. The differential equations of the GM (1,1) model have time-varying coefficients. In other words, the model is refreshed when new data is available in the forecast model. The GM (1,1) model can only be used in positive data series, and when all primitive data points are positive, grey models can be used to forecast future values of the primitive data points.

The main purpose of the Velhulst model can be summarized as limiting the whole system to a real system and defining some incremental operations such as the S-curve which is the saturation region. Verhulst proposed the Verhulst model while studying the law of biological reproduction, and the main idea of the Verhulst model is that the number of organisms increases exponentially, but the growth rate of the individual organism gradually slows down due to environmental constraints and eventually stabilizes at a constant value; It is mainly used to describe the saturation state and the S-shaped process. It is demonstrated through case analysis that traditional modeling simulation and forecasting accuracy are significantly improved by the Grey Verhulst model.

Several solutions have been discussed in the literature to improve the modeling accuracy of grey models. In mathematics, a Fourier Transform (FT) can be expressed as a mathematical transformation that decomposes a function (usually a function of time or a signal) into its constituent frequencies, such as the expression of a music chord in terms of the volumes and frequencies of their sounds. constituent notes. It can be said that the term Fourier Transform refers to the mathematical operation that relates both the frequency domain representation and the frequency domain representation to a time function.

Time series forecasting models have been developed with their univariate and residual values, and they have become more and more preferred in large areas due to their accuracy in forecasting.

In the study conducted on values selected as a fixed data set, it was found that the most efficient time series forecasting model was Grey Verhulst Model and the results could be further improved by Fourier Transformation.

5.3. Evaluation Results Review

The results of the values applied with Grey Verhulst Model in accordance with the applications; MSE, RMSE and MAPE, which were used in the study, were found reliable in the short term.

The performance criteria of the applied models were examined with different evaluation methods. Three different assessment methods encountered and discussed in literature reviews are the subject of the study. These include Mean Square Error, Root Mean Square Error, and Mean Mean Percent Error. Performance criteria of the results during the application were evaluated by MSE, RMSE and MAPE evaluation methods.

The MSE, RMSE and MAPE results of the forecasted values decreased gradually in LRM, GM (1,1), GVM and FT applications and decreased to 0.97% levels.

MAPE values were calculated over two separate estimates, model fitting and posterior forecasting, and presented separately as Mape1-Mape2. In this way, the performance criteria between the period in which the actual data were taken as input and the period when forward-term forecasts made were compared.

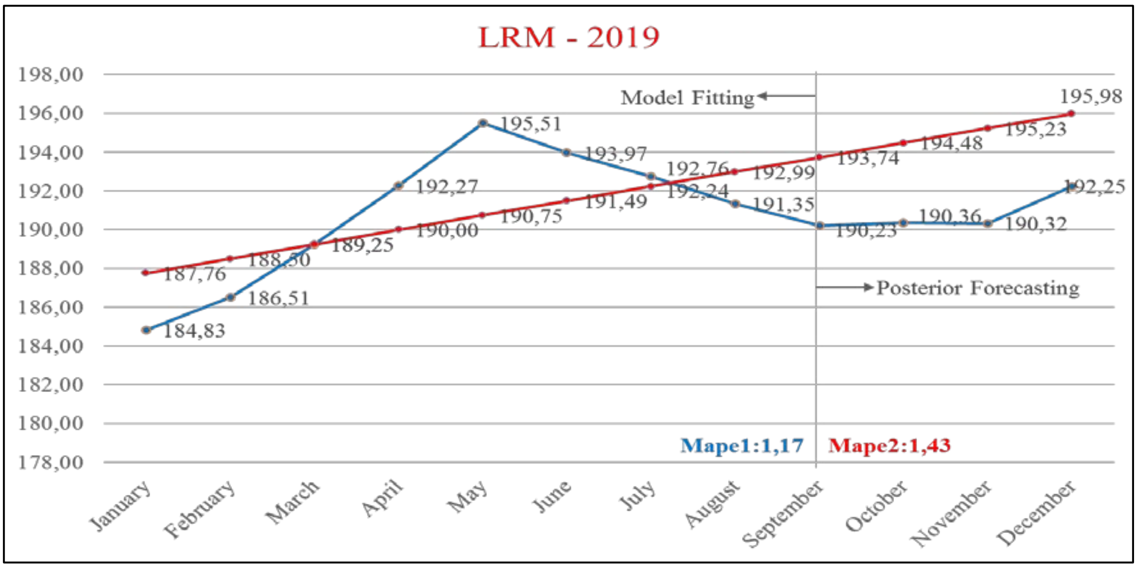


Figure 5.1. Model Fitting and Posterior Forecasting Chart of LRM (2019).

In LRM Modeling, over Predecessor and Successor Modeling results, Mape1 **1,17**; Mape2 is calculated as **1,43**.

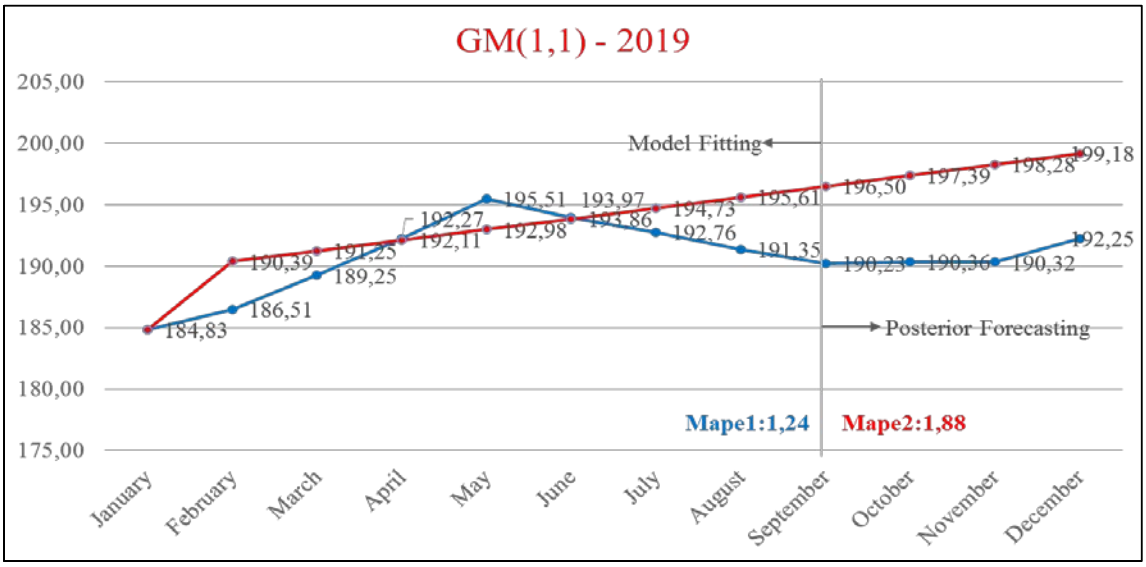


Figure 5.2. Model Fitting and Posterior Forecasting Chart of GM (1,1) (2019).

In GM (1,1) Modeling, over Predecessor and Successor Modeling results, Mape1 **1,24**; Mape2 is calculated as **1,88**.

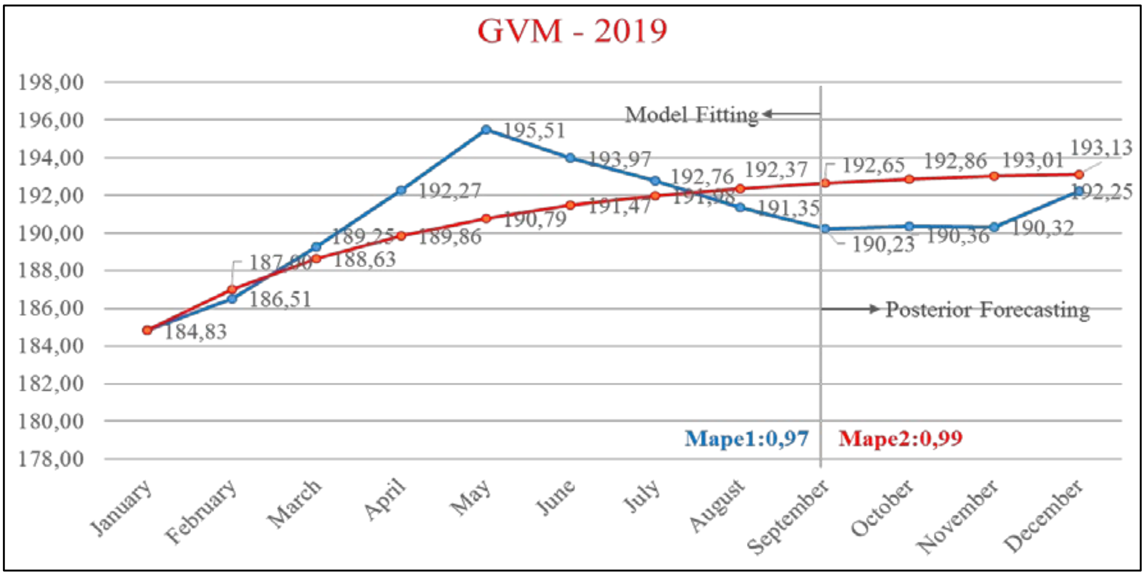


Figure 5.3. Model Fitting and Posterior Forecasting Chart of GVM (2019).

In GVM Modeling, over Predecessor and Successor Modeling results, Mape1 **0,97**; Mape2 is calculated as **0,99**.

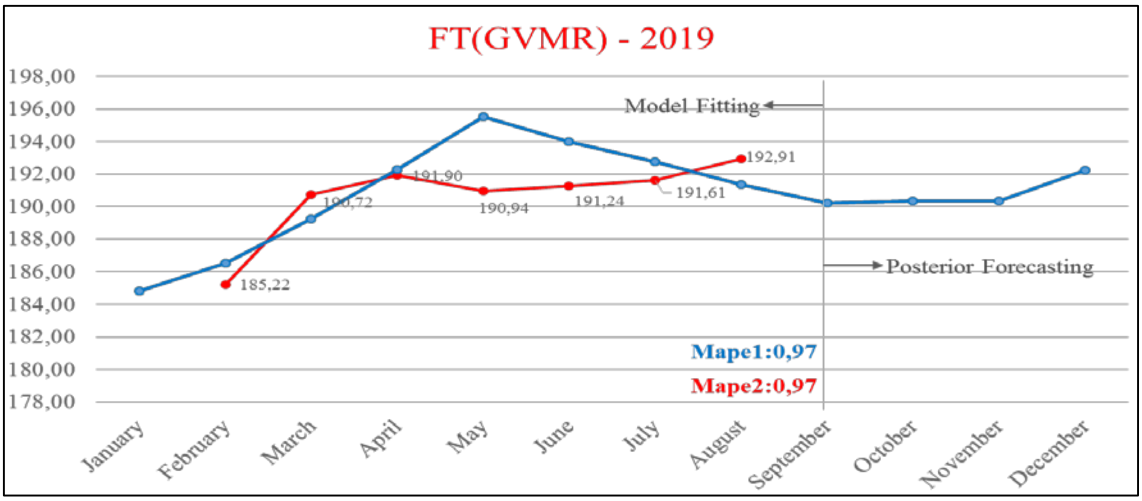


Figure 5.4. Model Fitting and Posterior Forecasting Chart of FT(GVMR) (2019).

In FT(GVMR) Modeling, over Predecessor and Successor Modeling results, Mape1 **0,97**; Mape2 is calculated as **0,97**.

5.4. Discussions

As a result of all literature reviews and researches, it has been seen that causal based forecasting models are used in many areas. Same time; It was also found that especially multivariate ones were affected by different components and could not forecast the results clearly.

On the other hand, it has been observed that time series based forecasting methods can be evaluated under a separate title as they become widespread as science develops. Generally, in forecasts made over numerical data, time series have come to the fore and they have clarified their usage area.

Causal based models were researched and informed. Then, time series based models, which are the focus of the study, were directed. In this direction, as explained in detail in previous chapters, from primitive to complex; Linear Regression Model, Grey Model (1,1), Grey Verhulst Model and Fourier Transform Models are examined and followed through Turkey Construction Cost Index data in 2019 was applied.

When the results obtained from the applied models were evaluated, the data determined as data set (depending on the structure of the time series) and the most realistic forecasting results came out of Fourier Transformation (with GVM residues).

In this direction, it will be beneficial to make cost estimation in future term or large budget projects to complete the project on time and with quality.

The data for the models are taken from TurkStat, the 2019 Construction Cost Change Index is used. Firstly, the first nine-month period of the year was modeled to build the models (Predecessor Stage), and then the cost estimates of the last three months were made with the trend (Successor Stage).

Considering the performance criteria, it can be determined that the MAPE level has decreased to 0.97% in the forecasting made over the FT (GVMR) and this margin of deviation, which is less than 1%, provides a consistent basis for short-medium term forecastings. In this way, based on the results of the models, both the performance criteria of the models and their applicability degree on this index were measured and it was seen that mostly consistent results were obtained in the processed data.

Table 5.1. Out-of-Sample Forecasting Performance Criteria of the Models (2019).

Model	**Linear**	**GM (1,1)**	**GVM**	**Fourier Residual**
MAPE1	1.17	1.24	0.97	0.97
MAPE2	**1.43**	**1.88**	**0.99**	**0.97**

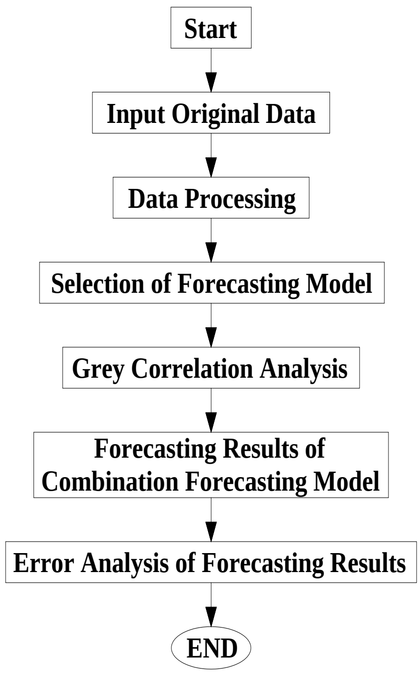


Figure 5.10. Flowchart of Grey Correlation based Forecasting Combination Analysis.

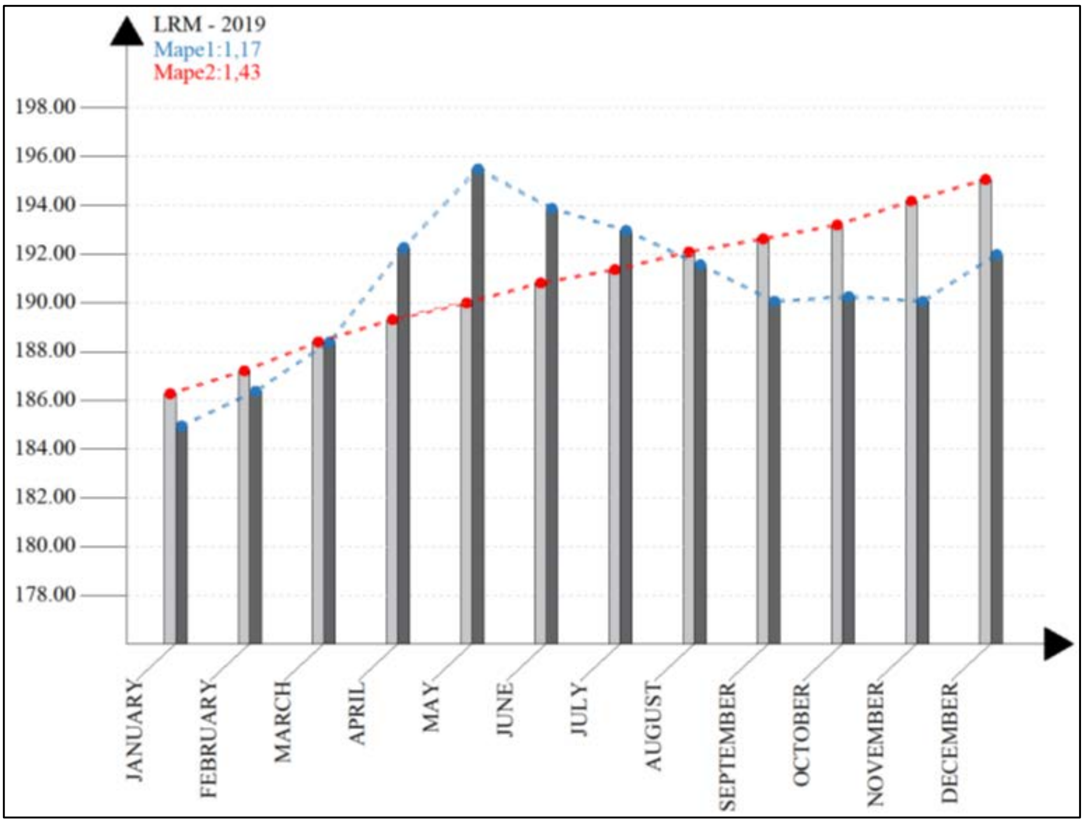


Figure 5.5. Comparative Bar Chart of LRM (2019).

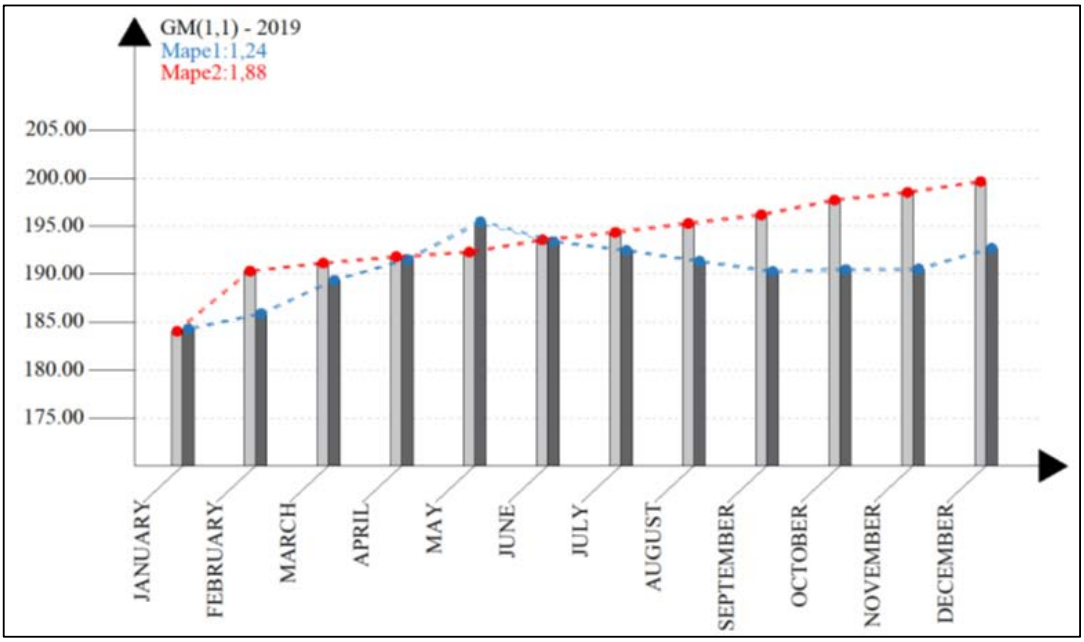


Figure 5.6. Comparative Bar Chart of GM (1,1) (2019).

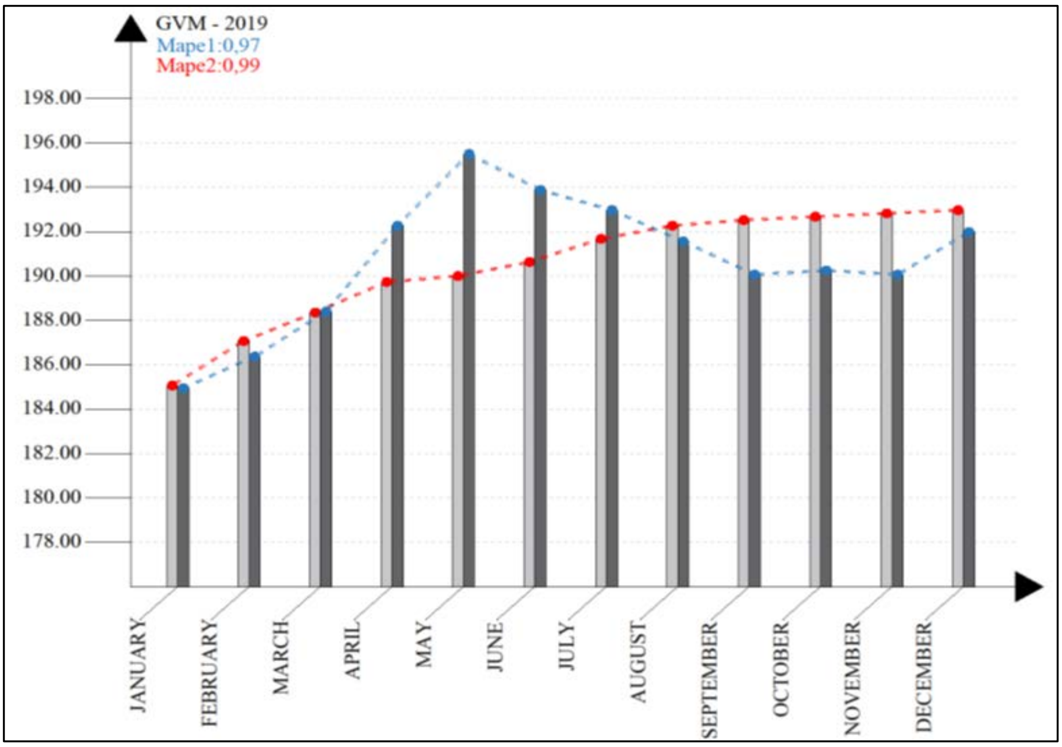


Figure 5.7. Comparative Bar Chart of GVM (2019).

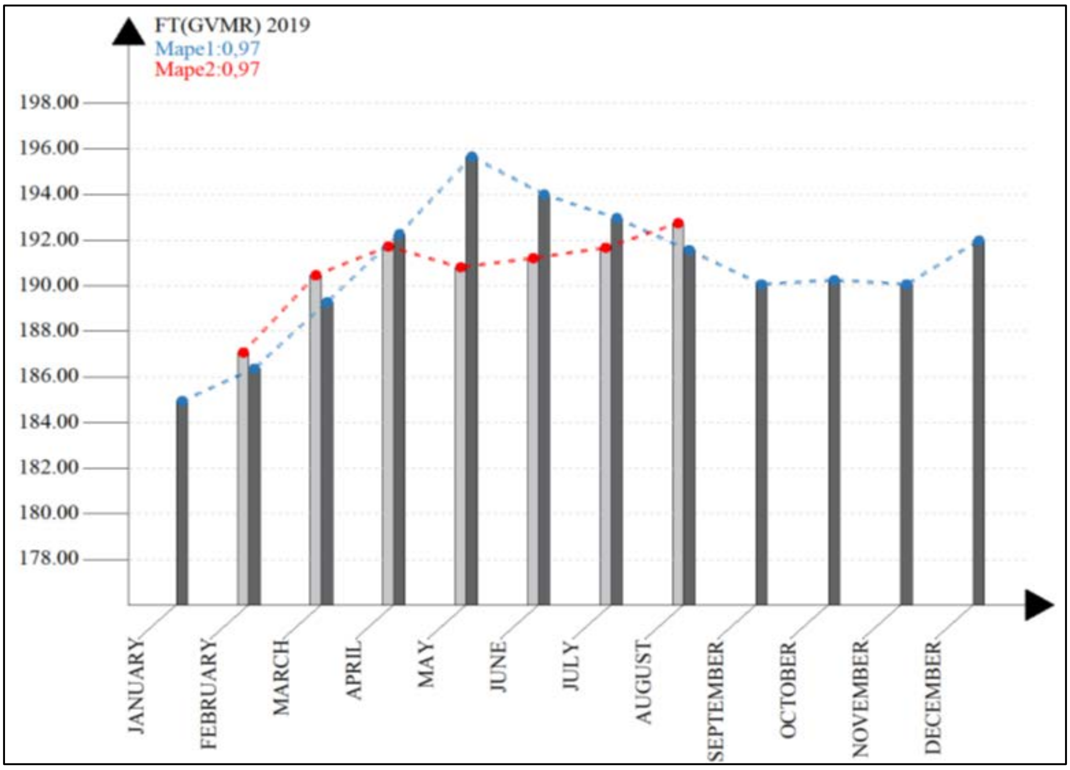


Figure 5.8. Comparative Bar Chart of FT(GVMR) (2019).

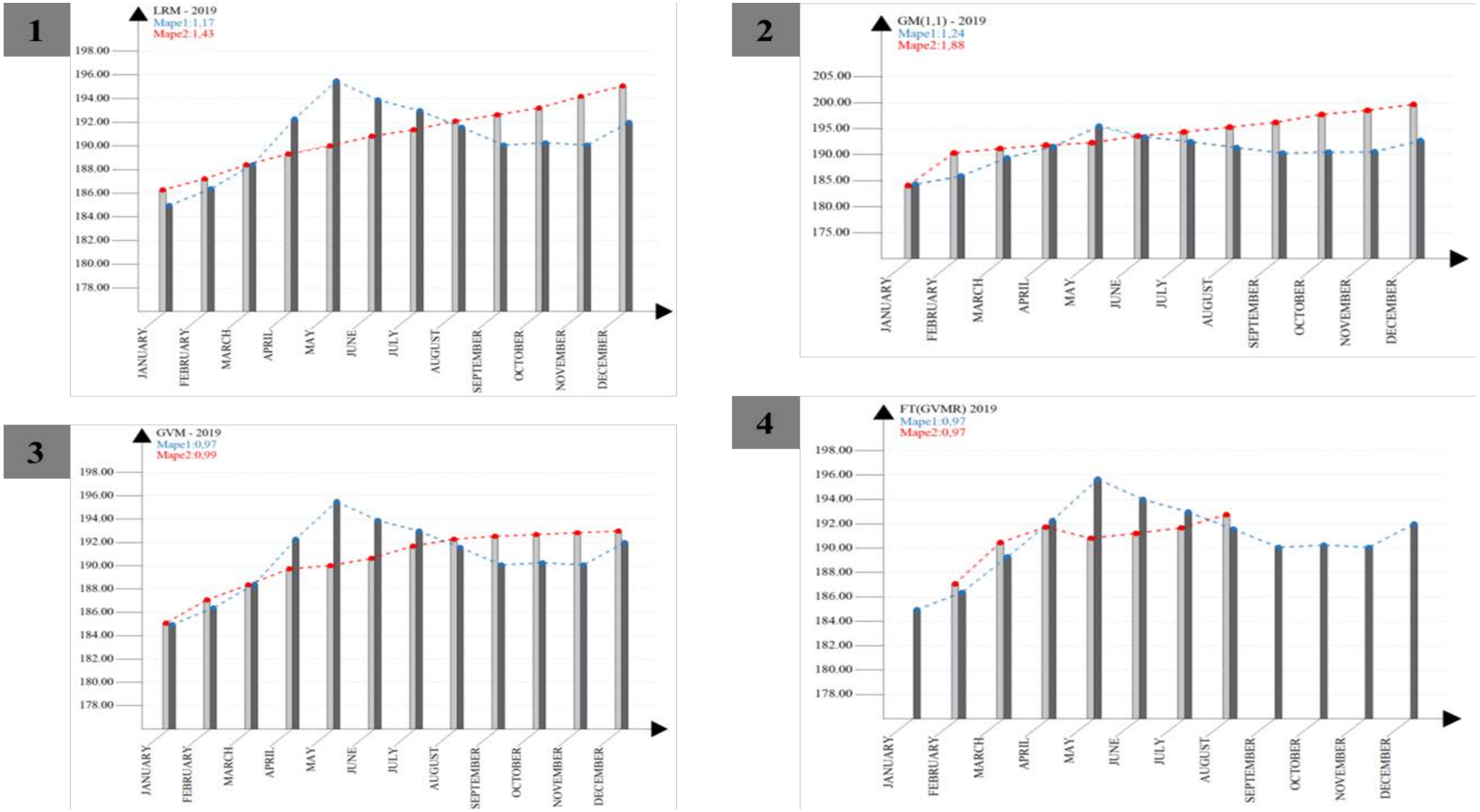


Figure 5.9. Comparative Bar Charts of All Models (2019)

CHAPTER 6

CONCLUSIONS

Forecasting costs in construction projects has an important place in both design and implementation processes. It is very important to observe possible changes in the construction process, especially in large scale projects.

Using a series of data in a functional way can be useful to make the changes that can be made at a later time in that field. Grey models have an important place in the time series forecasting model. In line with this functionality, there are a number of methods developed over time. In fact, when these models are examined, it is seen that they follow and complement each other in the developmental stages, but the levels of development differ with some important decompositions. Although many methods have been developed to support forecasts during the planning phase, early forecasts of costs as well as periods for construction activities and projects are still error-prone. One of the main factors leading to false forecasts is the change in resource prices due to changes in economic conditions over time. This shows the importance of monitoring and forecasting the construction costs trend, taking into account fluctuations in resource prices.

The ability to model and forecast construction costs can result in more accurate cost forecasting and budgeting. This has been modeled using a residual Fourier model by analyzing the remains of the Grey Verhulst Model, conditional variability of construction cost prices. The results show that the developed model can forecast construction costs with less errors. The data used in the time series analysis were observed for a relatively long time. These data are calculated with the assumption that the system will be developed in the future and the forecasted values will be maintained in the future based on these time series data. In the case of uncontrollable factors such as political changes or battles, it is likely that over a long period of time, the time series data may have different patterns at different times due to the time-specific nature of these factors.

6.1. Limitations and Future Works

- In line with research and applications, causal based forecasting models are also frequently used, but time series forecasting methods are widely preferred, especially in numerical fields such as economics and cost management. This trend can be attributed to the fact that models are applied more consistently and not affected by different components, especially depending on the single variable, ie time.
- From another point of view, time series based forecasting methods and trends formed by these series are more widely used in economy and cost management. Because of they are based on science and computation and can also be developed with each other, positive results can be obtained with these methods.
- It is seen that these models are becoming widespread and there is a need to deduce and benefit from big data. In this direction, studies can be carried out to develop more efficient results by developing models.

REFERENCES

Armstrong, Scott. (2001) "Evaluating Forecasting Methods. In Principles of Forecasting: A Handbook for Researchers and Practitioners" (Ed. J. Scott Armstrong), Kluwer.

Armstrong, Scott, and Fred Collopy, (1992) "Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons." International Journal of Forecasting 8, no. 1: 69-80.

Ashuri, Baabak, and Jian Lu. (2010) "Time Series Analysis of ENR Construction Cost Index." J. Constr. Eng. Manage., 136(11): 1227-1237.

Ashuri, Baabak, and Jian Lu. (2010) "Forecasting ENR Construction Cost Index: A Time Series Analysıs Approach." Construction Research Congress.

Chau, Kwok Wing, Kwok Wu, Yan-Song Li. (2005) "Comparison of Several Flood Forecasting Models in Yangtze River." J. Hydrol. Eng., 10(6): 485-491.

Diebold, Francis, and Lutz Kilian. (2000) "Unit-Root Tests Are Useful for Selecting Forecasting Models." Journal of Business & Economic Statistics 18, no. 3: 265-73.

Ding, Scheng, Yunkun Dang. (2014) "A new optimized grey verhulst model and its application." 2014 IEEE International Conference on Systems, Man, and Cybernetics.

Faghih, Sayed Amir Mohsen, and Hamed Kashani. (2018) "Forecasting Construction Material Prices Using Vector Error Correction Model." Journal of Construction Engineering and Management 144, no. 8.

Friedman, Jerome. (1998) "Data Mining and Statistics What's the Connection?" Stanford Linear Accelerator Center, Stanford University.

Fu Yimu, Xin Liu, Wenjing Chang, Lu Xu, Nan Wang, Shan Gao, Nan Xu, Jinghong Yang. (2019) "Application of grey Verhulst model in mid-long term load forecasting." The 31th Chinese Control and Decision Conference.

Gao, Pumei, Jun Zhan, Jiefang Liu. (2019) "Fractional Order Reverse Accumulative Non-Homogeneous Discrete Grey Forecasting Model and Its Application." The Journal of Grey System Volume 31, No.2.

Guo, Zijian, Jian Ye, and Xiangqun Song. (2005) "A Verhulst Model On Time Series Error Corrected For Port Throughput Forecasting." Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 881 - 891.

Gopnik, Alison, Henry Wellman. (2012) "Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory." Psychological Bulletin, 138(6).

Ho, Paul. (2010) "Forecasting Construction Manpower Demand by Gray Model." J. Constr. Eng. Manage., 136(12): 1299-1305.

Hwang, Seokyon. (2009) "Dynamic Regression Models for Prediction of Construction Costs." J. Constr. Eng. Manage., 135(5): 360-367.

Hwang, Seokyon, and Liang Liu. (2010) "Contemporaneous Time Series and Forecasting Methodologies for Predicting Short-Term Productivity." J. Constr. Eng. Manage., 136(9): 1047-1055.

Hwang, Seokyon. (2011) "Time Series Models for Forecasting Construction Costs Using Time Series Indexes." Journal of Construction Engineering and Management 137, no. 9: 656-62.

Hwang, Sungjoo, Moonseo Park, Hyun-Soo Lee, and Hyunsoo Kim. (2012) "Automated Time-Series Cost Forecasting System for Construction Materials." Journal of Construction Engineering and Management 138, no. 11: 1259-69.

Ilbeigi, Mohammad, Alireza Joukar, and Baabak Ashuri. (2016) "Modeling and Forecasting the Price of Asphalt Cement Using Generalized Auto Regressive Conditional Heteroscedasticity." Construction Research Congress 698-707.

Ilbeigi, Mohammad, Baabak Ashuri, and Alireza Joukar. (2017) "Time-Series Analysis for Forecasting Asphalt-Cement Price." Journal of Management in Engineering, no. 1.

Jie, Cui, Ma Hongyan, Hu Hongliang and Yang Zhengya. (2013) "On Property Of Derived Grey Verhulst Model Under Multiple Transformation Of Modelling Data." Grey Systems and Intelligent Services, IEEE International Conference.

Kayacan, Erdal, Baris Ulutas, and Okyay Kaynak. (2010) "Grey System Theory-Based Models in Time Series Prediction." Expert Systems with Applications 37, no. 2:1784-89.

Kordnoori, Shaghayegh, Hamidreza Mostafaei, and Shirin Kordnoori. (2014) "The Application of Fourier Residual Grey Verhulst and Grey Markov Model in Analyzing the Global ICT Development." Hyperion Economic Journal Year II, no. 1(2).

Li, Cheng, and Jixiang Li. (2016) "A Solution of Gm(1,1) Based on Excel." Business and Management Research 5, no. 3.

Li, Yuhong, and Yue Chen. (2009) "Grey System Theory Based Parallel Combination Forecast Method and Its Application." IEEE International Conference on Grey Systems and Intelligent Services: 408-14.

Lin, Yi, and Sifeng Liu. (2004) "A Historical Introduction to Grey Systems Theory." IEEE International Conference on Systems, Man and Cybernetics: 2403-08.

Long, Scott, and Laurie Ervin. (2000) "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." The American Statistician 54, no. 3: 217-24.

Lowe, David, Margaret Emsley, and Anthony Harding. (2006) "Predicting Construction Cost Using Multiple Regression Techniques." J. Constr. Eng. Manage. 132(7): 750-758.

Moon, Taenam, and Do Hyoung Shin. (2017) "Forecasting Construction Cost Index Using Interrupted Time-Series." KSCE Journal of Civil Engineering 22, no. 5: 1626-33.

Ng, Thomas, Sai On Cheung, Martin Skitmore, and Toby Wong. (2004) "An Integrated Regression Analysis and Time Series Model for Construction Tender Price Index Forecasting." Construction Management and Economics 22, no. 5: 483-93.

Nogales, Francisco Javier, Javier Contreras, Antonio Conejo, and Rosario Espínola. (2002) "Forecasting Next-Day Electricity Prices by Time Series Models." IEEE Transactions On Power Systems 17, no. 2: 342-48.

Park, Hyung-Keun. (2004) "Cash Flow Forecasting in Construction Project." KSCE Journal of Civil Engineering 8, no. 3: 265-271.

Parker, Philip. (1994) "Aggregate diffusion forecasting models in marketing: A critical review." International Journal of Forecasting 10: 353-380.

Pearl, Judea. (2009) "Casuality, Models Reasoning and Inference" Second Edition, New York

Petersen, Maya, and Mark J. van der Laan. (2014) "Causal Models and Learning from Data: Integrating Causal Modeling and Statistical Estimation" Epidemiology Vol. 25, no. 3 pp. 418-426

Pewdum, Wichan, Thammasak Rujirayanyong, and Vanee Sooksatra. (2009) "Forecasting Final Budget and Duration of Highway Construction Projects." Engineering, Construction and Architectural Management16, no. 6: 544-57.

Rencher, Alvin, and William F. Christensen. (2012) "Methods of Multivariate Analysis" Third Edition, Utah.

Shahandashti, Seyed Mohsen, and Baabak Ashuri. (2013) "Forecasting Engineering News-Record Construction Cost Index Using Multivariate Time Series Models." Journal of Construction Engineering and Management 139, no. 9: 1237-43.

Shahandashti, Seyed Mohsen, and Baabak Ashuri. (2016) "Highway Construction Cost Forecasting Using Vector Error Correction Models." Journal of Management in Engineering 32, no. 2

Sing, Michael, Dylan Edwards, Henry X. Liu, and Peter Love. (2015) "Forecasting Private-Sector Construction Works: Var Model Using Economic Indicators." Journal of Construction Engineering and Management 141, no. 11

Shao, Jun. (1993) "Linear Model Selection by Cross-Validation." Journal of the American Statistical Association, 88, no. 422:486-494.

Sifeng, Lui, Jeffrey Forrest, and Yang Yingjie. (2012) "A Brief Introduction to Grey Systems Theory." Grey Systems:Theory and Application 2, no. 2:89-104.

Sonmez, Rifat. (2008) "Parametric Range Estimating of Building Costs Using Regression Models and Bootstrap." Journal of Construction Engineering and Management 134, no. 12:1011-16.

Sonmez, Rifat, Arif Ergin, and M. Talat Birgonul. (2007) "Quantitative Methodology for Determination of Cost Contingency in International Projects." Journal of Management in Engineering 23, No. 1:35-39.

Taylor, John and Patrick McSharry. (2008) "Short-Term Load Forecasting Methods: An Evaluation Based on European Data." IEEE Transactions on Power Systems 22: 2213-2219.

Taylor, Richard, and Paul Bowen. (1987) "Building price-level forecasting: an examination of techniques and applications." Construction Management and Economics 5: 21-44.

Wang, Xiaochen, Peng Guo, and Xiaobin Huang. (2011) "A Review of Wind Power Forecasting Models." Energy Procedia 12: 770-78.

Wang, Zheng-xin, Yao-guo Dang, and Si-feng Liu. (2009) "Unbiased Grey Verhulst Model and Its Application." Systems Engineering - Theory & Practice 29, no. 10: 138-44.

Xiaojun, Guo, Liu Sifeng, and Fang Zhigeng. (2013) "Study On A Grey Verhulst Self-Memory Model and Application." IEEE International Conference on Grey systems and Intelligent Services.

Yongqing, Chen, and Huang Jiatao. (2009) "A Grey Verhulst Forecasting Model for the Diffusion of ISO9000 Standard in Guangxi province in China." The 1st International Conference on Information Science and Engineering, 4640-43.

Zhang, Peter, Butje Eddy Patuwo, and Michael Hu. (2001) "A simulation study of artificial neural networks for nonlinear time-series forecasting." Computers & Operations Research 28, 381-396.