AN INTEGRATED FUZZY APPROACH FOR SALES FORECASTING

(SATIŞ TAHMİNİ İÇİN ENTEGRE BULANIK BİR YAKLAŞIM)

by

Berkay AYDIN, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Supervisor: Assist. Prof. Dr. Zeynep Şener

Jun 2019

This is to certify that the thesis entitled

AN INTEGRATED FUZZY APPROACH FOR SALES FORECASTING

prepared by **Berkay AYDIN** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering** at the **Galatasaray University** is approved by the

Examining Committee:

Assist. Prof. Dr. Zeynep Şener (Supervisor)
Department of Industrial Engineering
Galatasaray University ----------

Assist. Prof. Dr. Mehtap Dursun
Department of Industrial Engineering
Galatasaray University ----------

Assist. Prof. Dr. Adnan Çorum
Department of Industrial Engineering
Bahcesehir University ----------

Date: ----------

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Assist. Prof. Dr. Zeynep Şener for her knowledge, suggestions and assistance at every stage of my study.

Jun 2019
Berkay AYDIN

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

For companies, it is of great importance to use their own resources in a way that causes minimum cost. In particular, the variability in demand in manufacturing companies makes this process more difficult. A good demand forecast companies save many costs such as overstock, overtime, inability to meet demand. In making these estimations, both a trend created from historical data and expert opinions are used. However, there is a need to develop a method to determine which criteria to consider and how to use it in demand estimation. In this study demand estimation made for a painting, will be sold in Turkey, As a result of market researches for the type of paint to be sold, a product has been determined in the concept of healthy paint. It is important to use the resources positively to determine how much new products will emerge in the market due to both raw material differences from existing products and additional marketing needs.

The relationship based on a cause and effect relation between a dependant quantitative variable and one/or more independant quantitative variable is called regression, the mathematical expression that shows this relationship is called regression model. Fuzzy regression is one of the methods that can be used in solving the problems by providing facilities to consider both quantitative and qualitative variables in decision making process. Fuzzy regression is an alternative method which is used in some situations that are not suggested for classical regression implementations such as; all/or some of the data being fuzzy due to the ambiguous system structure or the system structure's being unable to determine definite relations between variables.

The attempt to associate and manage a multidimensional issue of paint sales only with price or other paint sales can result in misleading results. Fuzzy regression model is one of the models that can be used in the sale of paint where multiple quantitative and qualitative variables are the factors. The most important reason for choosing fuzzy regression model instead of classical

regression model is insufficient data for new product. In cases where there is insufficient market data, fuzzy regression gives better results than classical regression. In this application, the fuzzy regression model was examined and used for demand estimation. However, using the fuzzy regression model alone is not sufficient for accurate estimation.

The most important factor to determine the fuzzy regression model in the most efficient way is the variables. It is critical to clarify which variables should be selected and which variables should not be included in the model because the degree of dependence of the criteria with each other is an important factor in the decision-making process. For this reason, the criteria which are considered to be important in paint and paint sales were determined by experts. Two-stage screening was performed to determine which of these criteria would be used in the fuzzy regression model.

The first of this elimination was made by eliminating the hygiene factors determined by both the consumer and the producer according to the market condition. Hygiene factors are considered to be indispensable in the market regardless of the product or service sold. For example, when buying a phone, it is not necessary to look at the feature of speech or messaging. The characteristics that the consumer considers when making a decision are non-hygiene factors such as cameras, processors. When viewed in the same way in the paint, it is a must-have for breathing, wiping, odorless, high coverage, and having TSE certificate are the hygiene factors. These criteria were removed from the evaluation and other factors that make a difference were taken into consideration. The factors that are considered to affect the sale in practice have been price, availability, healthy and environmentally friendly, application service, brand recognition, color alternative, payment collection process, communication variety and economic situation after the removal of hygiene factors.

All the criteria has not been used to determine the solution of fuzzy regression model and to facilitate the process. In order to get the best results, the most important criteria should be determined and included in the application. Therefore, the criteria to be used in fuzzy regression model were determined by using The Decision Making Trial and Evaluation Laboratory method. With this method, the factors affecting the paint sales and the interaction between the

factors were determined and the factors to be taken as the basis for fuzzy regression were determined.

The DEMATEL method is a method used to determine the affected and influencing factors in a complex structure. Basically, this method aims to draw meaningful results by visualizing complex cause and effect relationships. However, it is difficult to determine the degree of interaction between factors in these relationships. This is due to the fact that it is quite difficult to quantify the interaction between factors. Significant relationships between these factors need to be established by experts to identify and influence the affected factors. After these relations have been formed, it is necessary to make paired comparisons between the criteria. However, it is difficult to determine the extent to which a factor influences another factor when making this comparison. A integer scale has been proposed to overcome this challenge. According to this scale, the effect of one factor on another factor is expressed as linguistic variable and is expressed in five linguistic terms which are defined as too many, too, normal and very little. With DEMATEL method, the level of influence of a criterion on others and the level of relationship with others are determined. With this method, the relationship between each criterion and other criteria is shown and the criteria that are more related to other criteria can be determined.

Fuzzy regression model was applied by determining the criteria to be used in demand estimation by DEMATEL method after the criteria which are hygiene factors were separated in practice. Thus, it can be estimated how much new paint sales will be based on the status or the value of the criteria. The results will enable the company to manage a number of important processes more efficiently when evaluated by the manufacturer. It enables it to manage the factors such as raw material inventory level, long-term production capacity plan, work schedule to be performed on the supply chain side, and to manage the marketing expenses, campaigns and important features that should be told on the marketing on the sales and marketing side. All this information directs the company both to promote the product and to maximize customer satisfaction.

ÖZET

Şirketler için kendi kaynaklarını en az maliyete neden olacak biçimde kullanmak büyük önem taşımaktadır. Özellikle üretim yapan şirketlerde taleplerdeki değişkenlik bu süreci zorlaştırmaktadır. İyi bir talep tahmini şirketleri fazla stok, fazla mesai, yoka düşme gibi birçok maliyetten kurtarmaktadır. Bu tahminleri yaparken de hem geçmiş verilerden oluşturulan bir eğilimden hem de uzman görüşlerinden faydalanılmaktadır. Ancak hangi kritelerin dikkate alınacağı ve nasıl talep tahmininde kullanılacağı ile ilgili bir yöntem geliştirme gereği bulunmaktadır. Bu çalışmada Türkiye'de satışa sunulacak bir boya için talep tahmini yapılmaktadır. Satışa sunulacak boya türü için pazar araştırmaları sonucu dikkate alınmış ve sağlıklı boya konseptinde bir ürün belirlenmiştir. Çıkacak yeni ürünün hem mevcut ürünlerden hammadde farklılıkları hem de ek pazarlama ihtiyaçları nedeniyle pazarda ne kadar yer edineceğini belirlemek kaynakları olumlu kullanmak için önemlidir.

Bir bağımlı nicel değişken ile bir / veya daha fazla bağımsız nicel değişken arasındaki sebep-sonuç ilişkisine dayanan ilişkiye regresyon, bu ilişkiyi gösteren matematiksel ifadeye regresyon modeli denir. Bulanık regresyon, karar verme sürecinde hem nicel hem de nitel değişkenleri göz önünde bulundurmak için olanaklar sağlayarak sorunların çözümünde kullanılabilecek yöntemlerden biridir. Bulanık regresyon, klasik regresyon uygulamalarının önerilmediği bazı durumlarda kullanılan alternatif bir yöntemdir; Belirsiz sistem yapısından veya sistem yapısının değişkenler arasındaki kesin ilişkileri belirleyememesinden dolayı verilerin tümü ya da bir kısmı bulanık ise.

Çok boyutlu bir konu olan boya satışını sadece fiyat ile ilişkilendirmek veya diğer satış yapılan boyalarla ele alıp yönetmeye çalışmak ise yanıltıcı sonuçlara neden olabilir. Birden çok hem nicel hem de nitel değişkenin etken olduğu boya satışında kullanılabilecek modellerden en uygun olanlarından biri bulanık regresyon modelidir. Klasik regresyon modeli yerine bulanık regresyon modelinin seçilmesinin en önemli nedeni yeni ürüne ait yetersiz veriye sahip

olunmasıdır. Piyasaya dair yetersiz verinin olduğu durumlarda bulanık regresyon klasik regresyona göre daha iyi sonuçlar vermektedir. Bu uygulamada da bulanık regresyon modeli incelenmiş ve talep tahmini için kullanılmıştır. Ancak bulanık regresyon modelini tek başına kullanmak da doğru tahmin için yeterli değildir.

Bulanık regresyon modelini en verimli biçimde uygulamak için belirlenmesi gereken en önemli etmen değişkenlerdir. Hangi değişkenlerin seçilmesi, hangi değişkenlerin modele katılmaması gerektiğini netleştirmek kritiktir. Çünkü değişkenlerin birbirleri ile olan bağımlılık dereceleri karar verme aşamasında önemli bir etkendir. Bu yüzden öncelikli olarak uzmanlar tarafından boyada ve boya satışında önemli olduğu kabul edilen kıstaslar belirlenmiştir. Bu kıstaslardan hangilerinin bulanık regresyon modelinde kullanılacağını belirlemek için iki aşamalı eleme gerçekleştirilmiştir.

Bu elemenin ilki piyasa koşuluna göre hem tüketici hem de üretici tarafından belirlenen hijyen faktörlerinin elenmesi ile yapılmıştır. Hijyen faktörleri satılan ürün veya hizmet ne olursa olsun piyasada olmazsa olmaz kabul edilen faktörlerdir. Mesela günümüzde bir telefon satın alınırken konuşma özelliğinin veya mesajlaşma özelliğinin olmasına bakılmaz. Tüketicinin karar verirken değerlendirdiği özellikler kamerası, işlemcisi gibi hijyen faktörü olmayan ölçütlerdir. Boyada da aynı şekilde bakıldığında piyasada olmaz ise olmaz kabul edilen nefes alması, silinebilmesi, kokusuz olması, örtücülüğünün yüksek olması, TSE belgesi olması gibi özellikler hijyen faktörleridir. Bu ölçütler değerlendirmeden çıkartılarak farklılık yaratan diğer etmenler dikkate alınmıştır. Uygulamada satışı etkileyecek olarak değerlendirilen etmenler hijyen faktörleri çıkartıldıktan sonra fiyat, bulunabilirlik, sağlıklı ve çevre dostu, uygulama servisi, marka bilinirliği, renk alternatifi, ödeme süreci, iletişim çeşitliliği ve ekonomik durum olmuştur.

Bulanık regresyon modelinin çözüm vermesini sağlamak ve işlemleri kolaylaştırmak için belirlenen ölçütlerin tamamı kullanılmamıştır. En iyi sonucu alabilmek için geri kalan ölçütler arasından da en önemlilerinin belirlenmesi ve uygulamaya dahil edilmesi gerekmektedir. Bu yüzden bulanık regresyon modelinde kullanılacak ölçütler DEMATEL (The Decision Making Trial and Evaluation Laboratory) yöntemi kullanılarak belirlenmiştir. Bu yöntem ile boya satışını etkileyen faktörler ve faktörler arasındaki etkileşim netleştirilmiş, bulanık regresyonda esas alınacak faktörler belirlenmiştir.

DEMATEL metodolojisi, oluşan durumlarda karmaşık bir halde olan etkilenen ve etkileyen ölçütleri saptamak için kullanılan bir metoddur. Asıl olarak DEMATEL karmaşık neden sonuç ilişkilerini görsel hale getirerek anlamlı sonuçlar elde etmeyi hedeflemektedir. Ancak bu ilişkilerde etmenlerin birbiriyle etkileşimin ne kadar olduğunu belirleyebilmek oldukça zordur. Zor olmasının sebebi etkenlerin birbiriyle etkileşiminin nicel olarak belirtmenin fazlasıyla zor olmasından kaynaklanmaktadır. Etkileyen ve etkilenen ölçütlerin bulunabilmesi için bu ölçütlerin birbiriyle oluşturdukları anlamlı ilişkilerin konu üzerinde tecrübeli insanlar tarafından belirlenmesi gerekir. Bu ilişkiler belirlendikten sonra ölçütlerin kendi aralarında karşıklıklı karşılaştırmaları yapılmalıdır. Fakat bu karşılaştırmalar yapılırken bir faktörün diğer bir faktörü ne derece etkilediğini belirlemek yine fazlasıyla zordur. Bu zorluğun üstesinden gelmek için tam sayı ölçeği önerilmiştir. Bu ölçeğe göre bir faktörün diğer bir faktörü etkilemesi çok fazla, fazla, normal ve çok az olmak üzere dilsel değişken olarak tanımlanmıştır. DEMATEL yöntemi ile bir etkenin diğerlerine olan etki seviyesi ve onlarla ilişki seviyesi belirlenir. Bu yöntemle her bir ölçütin diğer ölçütlerle arasındaki ilişkisi gösterilir ve diğer ölçütler ile daha çok ilişkili olan ölçütler belirlenebilir.

Uygulamada hijyen faktörü olan ölçütler ayrıştırıldıktan sonra DEMATEL yöntemi ile talep tahmininde kullanılacak ölçütlerin belirlenmesiyle bulanık regresyon modeli uygulanmıştır. Böylece belirlenen ölçütlerin durumuna veya alacağı değere göre yeni boya satışının ne kadar olacağı tahmin edilebilmektedir. Çıkan sonuçlar üretici tarafından değerlendirildiğinde bir şirket için önemli birçok süreci daha verimli yönetmesine olanak sağlar. Tedarik zinciri tarafında hammadde stok seviyesini, uzun dönemli üretim kapasite planını, yapılacak mesai planını gibi etkenleri yönetmesine olanak sağlarken, pazarlama ve satış tarafında ürüne yapılacak pazarlama harcamalarını, kampanyaları, anlatılması gereken önemli özelliklerini yönetmesine olanak sağlar. Bütün bu bilgiler hem ürünün tutundurulması hem de müşteri memnuniyetini en üst seviyeye çıkarmak için şirketi yönlendirmektedir.

1. INTRODUCTION

For many years, companies have aimed to increase customer levels and ensure customer satisfaction. The reason they did this was to prevent the bad consequences of a wrong estimate because customer demands were difficult to predict. With this approach, stock levels are evaluated independently of profit and loss and appear to be a necessity. If this requirement could not be fulfilled, customer unhappiness could lead to worse results than the cost burden of the inventory level.

As companies find it difficult to find opportunities to grow further, they are trying to increase their sales through faster access to the market, personalization and innovation. Nowadays, supply chain management has an important place to become a successful company. Despite the slowdown in stock turnover rates, increasing customer satisfaction leaves companies in a difficult position. However, the main reasons for these problems are a low performance supply chain resulting from low forecast accuracy, inefficient planning processes and slow production capabilities to respond to changing market demand. Due to these inefficiencies in the supply chain, companies' growth slows down, customer discontent increases, inventory levels increase, sales losses are experienced and old products that cannot be sold remain in the market and prevent new sales. Resources are spent in this way rather than capital marketing efficiency by using better marketing activities for product development.

A company can reduce the negative impacts of risks by specifying demand or sales expectations for products and services in the future. Demand estimation is a systematic process that includes an estimate of the demand for an organization's products and services under a range of uncontrollable and competitive forces in the future.

In the words of Evan J. Douglas (1987), “Demand estimation (forecasting) may be defined as a process of finding values for demand in future time periods.” (Douglas, 1987)

According to Cundiff and Still (1971), “Demand forecasting is an estimate of sales during a specified future period based on proposed marketing plan and a set of particular uncontrollable and competitive forces.” (Cundiff and Still, 1971)

Demand forecasts enable a company to make several business decisions such as planning purchasing raw materials, the production and the logistics process, managing funds and deciding on product price. An organization can estimate its demand by making its own estimations, called forecasting, or by taking the help of expert consultants or market research agencies.

The main challenge in predicting demand is to choose an effective technique. There is no specific way for organizations to anticipate future risks and uncertainties. Generally, there are two approaches to demand estimation.

The first approach involves estimating demand by collecting information from consumers or purchasing surveys from experts. On the other hand, the second method is to estimate the demand using historical data using statistical techniques.

In this study, a sales forecast is made for a new paint product that will be released to the market. DEMATEL method and fuzzy regression model are used in order to estimate sales. Since there is not enough market information and data for the new product, fuzzy regression model is used instead of classical regression. DEMATEL method has been integrated in order to determine the important criteria for more accurate results of the study.

The study is organized as follows. Section I aims to review the literature on forecasting method. The review is based on a search in the academic literature from 2010 to 2018. The reviewed papers are classified into two groups namely “linear time series models” and “fuzzy time series models”. In section II, taxonomy of the forecasting methods is presented by classifying the published works into two categories as "linear time series models" and "fuzzy time series models". Section III presents fuzzy regression analysis. Future works are delineated in the last section. In section IV, proposed approach and DEMATEL methodology are presented. In section V, the applicability of the proposed approach is illustrated by a case problem for forecasting the sales of a new paint product.

2. FORECASTING, FORECASTING METHODS AND REGRESSION

2.1. Forecasting

Forecasting basicly is the process of making predictions for the future based on past and present data and analyzing the most common trends. An ordinary example may be an estimate of some variables of interest in a specific upcoming date. The prediction is similar compared to the forecasting, but is used more generally. Forecasting and estimation may refer to formal statistical methods using time series, longitudinal or cross-sectional data, or alternatively to less formal trial methods. Usage may vary between application areas.

Risk and uncertainty are central to the prediction; To specify the degree of uncertainty related to estimates is generally accepted as good practice. In any case, the data must be up-to-date to ensure that the estimate is as free of error as possible. In some situations, the data used to estimate the variable of interest is estimated by itself.

2.2. Forecasting Methods

There are many forecasting techniques. They can be classified in various ways. Estimation techniques can be divided into two groups as qualitative and quantitative techniques. In quantitative techniques and qualitative techniques, the starting points are the same, that is, observation values for the event in question. Observation values from past and present observation values are estimated according to certain rules. Qualitative estimation techniques are based on historical knowledge, but are largely based on the estimator's knowledge, experience and personal opinions. In qualitative techniques, the knowledge and experience and the experience of the predictor play a significant role. Thus, in a matter of experienced, skilled person the estimate may give better results. Qualitative techniques can be said is subjective. Because different individuals may have different ratings for the same observation value. On the other hand, the estimation error cannot be measured in qualitative estimation techniques. This

does not mean that qualitative estimation techniques are completely unnecessary. As stated above, the estimator's knowledge on the subject, the more the experience, the more accurate the results in his predictions. Sometimes qualitative techniques can give better results than quantitative techniques. Moreover, it can be stated that qualitative techniques are easy to implement, and do not require much effort and time.

Although quantitative estimation techniques can also be classified in various respects, it is possible to divide them into two groups, mainly as the techniques based on the cause-effect relationships and on time series analysis techniques. Quantitative techniques based on cause and effect relationship are regression and econometric models.

2.2.1. Estimation techniques based on cause-effect relationship

In the regression technique, the help of the previous observation values determines a causal relationship between a quantitative dependent variable with a cause-effect relationship and one or more quantitative independent variables. Then, for the future values of the independent variables, the value of the dependent variable will be estimated. On the other hand, the regression technique also aims to determine the causal relationship between dependent independent variables.

Econometric models are the equation system consisting of two or more (simple or multiple) regression equations showing cause and effect relationship. Therefore, econometric models have more than one dependent variable. In econometric models, the relationships between all dependent and independent variables are examined simultaneously. Thus, the relationships between dependent independent variables are evaluated in a more realistic way. As a result, dependent variable values can be calculated with the help of independent variable values in an equation and other dependent variable values can be estimated by other equations containing this variable.

As mentioned above, estimation techniques based on cause-effect relationship are mostly used to estimate dependent variable values from independent variable values. Regression and econometric models are very limited to make future estimates. Because, in order to estimate the future values of the dependent variables with the help of regression and econometric models,

the future values of the independent variables should be determined or estimated. This is only possible with the analysis of time series of the independent variables.

2.2.2. Estimation techniques based on time series analysis

All quantitative estimation techniques, except for the regression and economic model, are based on time series of past observation values of the variable. Regression and economic models aim to estimate dependent quantitative variables from independent quantitative variables and to estimate future period values of variables with the help of historical observation values.

It can be said that all estimation techniques are based on extrapolation. The route of estimation techniques is the same. Firstly, the time series of past observation values of the variable of interest is analyzed and the main trend and characteristics of the series are determined. Then the model to reflect this trend is selected and the parameters of the model from the current time series are approximately. Then, assuming that the variable will have the same tendency in the future, the approximate values are obtained for future periods with the help of the determined model.

Time series based estimation techniques are Trend Analysis, Division Separation Technique, Correction Techniques and Box-Jenkins technique.

2.3. Regression

The relationship between a dependent quantitative variable with a cause-effect relationship and one or more independent quantitative variables is called a regression. Mathematical expression showing the relationship form is called the regression function or regression model.

Regression models can be classified in various aspects. It is divided into two as Simple Regression and Multiple Regression. The model that contains an argument is called Simple Regression, the regression with multiple arguments is called multiple regression.

On the other hand, the regression is divided into two as linear and nonlinear. Linearity and non-linearity are same both simple and multiple regression. Therefore, regression models;

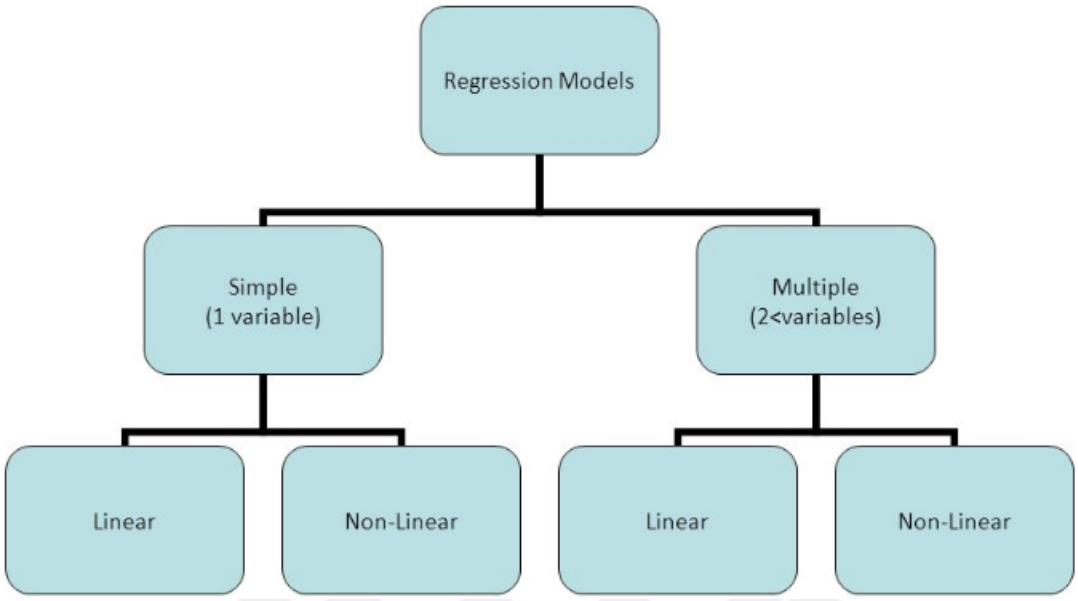


Figure 2.1. Types of regression model

It is classified in the form. Linear (simple and multiple) regression has a single, nonlinear (simple or multiple) regression has many varieties.

The Least Squares method is used to determine simple or multiple regression models. The difference between the observation values and the estimated values is called an error. Least Squares Method aims to minimize the sum of Error Frames. The error between the observation values and the estimated values is inevitable. The causes of errors are mainly;

- Measuring Errors
- The effects of qualitative variables not included in the model.

3. FUZZY SET THEORY AND FUZZY REGRESSION

3.1. Fuzzy Set Theory

Every person is confronted with situations that cannot be known in their daily lives, sometimes even as if they were presumed to be definite, but ultimately encounter situations that do not supply certainty.but ultimately unsuitable. It is only possible to make these predictions in a systematic way in advance and to make numerical predictions only after a number of assumptions and assumptions. Various studies have been carried out in engineering research and modeling so far to give certainty to acceptance and concepts with this assumption. However, as we move from large scales to small scales, we can see that the events examined go from certainty to uncertainty.

The real world is complex. This complexity generally arises from uncertainty, lack of definite thought, and lack of judgment. In many social, economic and technical issues, uncertainties are always present due to the fact that human thought is completely immature. Computers developed by human beings cannot process such information and digital information is required for their work. Since the comprehension of a real event is not entirely possible with the inadequacy of human knowledge, human beings comment on such events in the mind and mind. Incomplete sources of information, such as complexity and uncertainty that arise in various forms, are generally referred to as fuzzy sources. It has been stated by Zadeh (1962) that the closer to the real world problems the solution will be more blurred (Zadeh, 1962). Because all of the very large sources of information cannot comprehend and interact with people at the same time and interactively. It should be emphasized that the information sources, in addition to the basic and definite information, also contain verbal information. Since a person can think verbally and can convey what he or she knows verbally to others, these statements cannot be expected to be definite.

The more we learn about a system, the better we can understand it and the degree of complexity about it decreases, but it does not disappear completely. The complexity of the systems under investigation will be more effective if there is no more or less data available. It is possible to make meaningful and useful inferences by using fuzzy logic rules from fuzzy input and output information.

3.1.1. The concept of fuzzy logic and its applications

L. Soldier Zadeh (1962) first mentioned the term "fuzzy cluster" in 1962, in his work "From circuit theory to system theory" (Zadeh, 1962). These thoughts, which can be called mathematics of uncertainty, have matured in the second study (Zadeh, 1965). In the next decade (1965-1975), 620 studies have been achieved, which could create a comprehensive bibliography from two studies (Gaines and Kohout, 1977). The number of studies increased from year to year in 1979 to 1400 (Kendall and Yager, 1979). In the following years, publications had reached an exponential rate. Those interested in fuzzy set theory have evolved from small expert groups to international communities. The field of application has increased so that it is no longer possible to monitor these publications in various disciplines.

This over-interest in fuzzy set theory and mathematics is due to the change in the problem-solving approach of the majority of theoretical and practical scientists and technocrats. The rigid system approach, which has proven its validity and strength in the development of human systems, is not so effective in model development for the natural system or semi-natural semi-human system. The "soft" system approach has become necessary for the expansion of science and technology into ecological, social and economic processes. Fuzzy set theory and its applications in expert systems are the basis of the soft system approach (Gaines, 1987).

Zadeh (1975), in his presentation for Kaufman's book, argued very ambitiously that fuzzy set theory would be the basis for the design of artificial intelligence systems in psychology, sociology, social sciences, philosophy, economics linguistics, operations research, management science and other fields (Kaufman, 1975). Indeed, studies to date confirm Zadeh's predictions. Fuzzy set theory has been applied in the following sciences and fields:

• Artificial intelligence
• Expert systems
• Control theory
• Quality control,
• Multipurpose decision making
• Product planning, selection
• Optimum system planning
• Transport, transportation
• Net work
• Games theory
• Environmental management
• Banking finance
• Agriculture

At the beginning, however, many scientists have doubts about this new theory (Arbib, 1977).

3.1.2. Uncertainties and imprecision

For logic, system, cluster, etc., fuzzy is an expression of uncertainty. In the past, probability theory was used to process uncertainties and to reach meaningful conclusions. This probability theory in mathematics and engineering is used with statistical methods in cases of uncertainty. Therefore, the concept that all uncertainty has a random character has become widespread.

The most important feature of randomness is that the chance event plays a role in the emergence of the results and that the necessary predictions and predictions cannot be done with precise accuracy. However, all known uncertainties are not random. It is easily understood that most of the uncertainties encountered in everyday life are not random. In case of non-random characters, methodologies that require numerical uncertainties such as probability and statistics cannot be used in verbal uncertainties.

The first of the two cases in which fuzzy logic is most valid is to include the opinions and value judgments of the people in the event that the incident is complex and if sufficient information is not available. The second is the situations that need human judgment, understanding and

decision. It should not be deduced that any problem encountered can be solved, even if it is complicated by fuzzy logic. However, it can be concluded that at least the idea of human thought can be better understood at least because of some verbal inferences about the examined case (Şen, 2001).

3.1.3. Methods of uncertainty

The logic structure, which contains ambiguity and is called fuzzy today, has become used in different technological devices. It was possible to include verbal ambiguities in the language of daily speech in modeling or calculation. Today's technology, washing-dishwasher, vacuum cleaner, car, elevator, artificial intelligence, modeling, such as many breakthroughs in fuzzy logic methods can be achieved in recent days.

After these explanations, it is evident that some techniques, methods, algorithms and approaches for the objective investigation of uncertainty are found on all sides and in everything, and some scientific methods with rules that can capture and quantify the uncertainty outside the conventional mathematical (differential equation, derivative). There are various uncertainty methods such as probability, statistics, stochastic, fractal, chaotic and quantum which may be applicable to these subjects. However, the oldest of these and what is necessary to understand others is "probability". These methods help in the investigation of the event rather than being normative, simplistic and dull concepts with more dynamic, uncertain and efficient approaches and opinions (Şen, 2001).

3.1.4. Fuzzy sets and membership degrees

Among the various uncertainties mentioned above, it is understood that the verbal ones always contain approximation and turbidity in terms of their inexcusable information content. It was stated by Lütfü Askerzade in 1965 that the most appropriate management science principle would be by giving different membership degrees to the cluster members in such situations of uncertainty. According to Aristotle logic, people are either tall in length or not.
However, according to the Zadeh (1975) approach, tallness has different degrees. If one of the tallest is taken as real tall, the ones that are slightly shorter than it are not tall. Even if the lengths

above and below the underlying length are not so strong, although the degree of being tall is slightly less, it still enters the tall set. In this way, we can say that the people in the whole set of people in the world have one degree of membership in length (Şen, 2001).

According to Aristotle logic and so far in the conventional classical cluster concept, if a member of a cluster belonging to the event if they belong to membership degrees 1, is assumed to be equal to 0 if they do not belong. No degree of membership can be considered between the two. It is possible to speak of different membership degrees ranging from 0 to 1 in the concept of fuzzy sets. Thus, we can already say that the membership degrees of the elements in fuzzy sets are continuous between 0 and 1. In fact, Zadeh (1972) developed fuzzy set theory, which has a broad application in clusters theory and is compatible with natural life, suggesting that the membership levels of cluster elements can vary between 0 and 1. The concept of fuzzy clusters, which has such a simple foundation, has had a great impact on technology and scientific studies especially after 1980 (Chang et al. , 2001).

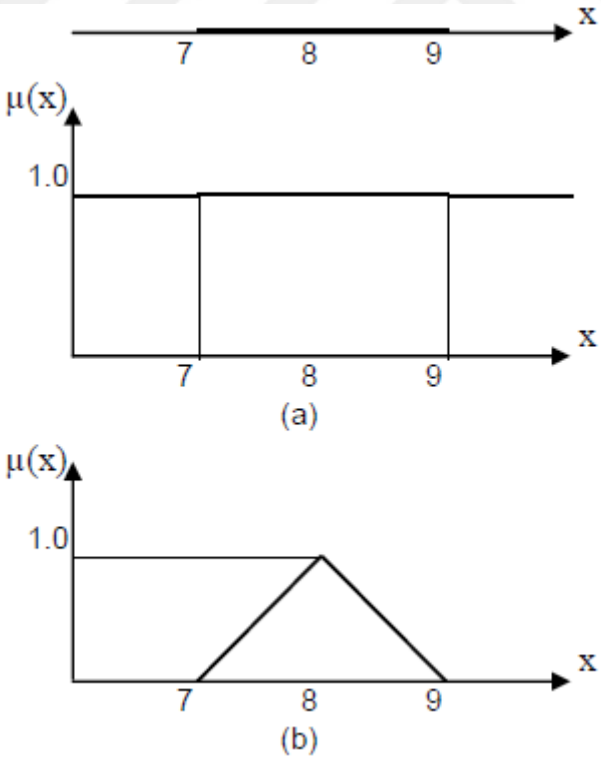


Figure 3.1. Membership degree functions (a) classical set, (b) fuzzy set (Chang et al. , 2001)

Randomness can be defined as the numerical measure of the occurrence of that event.

3.1.5. Fuzzy system

The meaning of the fuzzy is the complete and incomplete information that a researcher has if he / she is not fully known by him / her. Thus, the researcher cannot directly use the classical analytical methods and the principles of dynamic and conservation (such as conservation of energy, matter, momentum), since there is uncertainty or ambiguity in the data and information. The event and mechanism examined by the researcher can be modeled only by taking into consideration the verbal and very vague information related to them, which are only definite and present as their complement instead of equations which are accepted and assumptions in their conclusions. A fuzzy data and knowledge is involved in the investigation of incidents with the help of fuzzy principles, but the functioning of fuzzy methods is completely evident. There are generally two reasons why researchers use fuzzy systems. These can be expressed as follows:

1) Due to the fact that real world events are very complex, it is not possible to precisely control these events with definite equations. As a natural consequence of this, the investigator always prefers to resort to approximate but solvable methods, if not absolute. Already, as Einstein says, if the actual events can be expressed with mathematical equations, or if the results of the equations can be precise, or the mathematical equations can accurately describe the truth, then it is not possible to talk about real world events. Therefore, in all studies, the solutions are to some extent approximate. Otherwise, a large number of nonlinear equations need to be solved in the same time, which is known to lead to chaotic solutions which are not obvious according to current knowledge (Lorenz, 1963).

2) All theories and equations in engineering express the real world in a way. Although many real systems are nonlinear, every effort is made to accept linearity in their classical methods. For example, it was found in Hooke's law that the deformation of the material under stress is linear in the strength calculations. However, the material cannot always be expected to behave in this way, and it is therefore likely that there will be some deviations. As a natural consequence of this, by taking into account the size of the safety factor in strength sizing, the uncertainties that may be taken into consideration are taken into consideration again. The use of the safety coefficient is, in some way, the

uncertainty being introduced into the solution through the back door in a rigid manner. However, uncertainty principles are required for sizing without the need for a size such as a safety factor in the behavior of the actual material.

Today, information and the verbal data it brings are given importance. The reason for this is that people can negotiate with verbal data, not numerically, as a device. Verbal data increases day by day. It is necessary to consider oral human data in a system and in engineering systems with the numerical information given by the devices. The real issue of fuzzy systems is how to think about how to go to analysis if such information is found. A good engineering theory is expected to capture some of the important features of the phenomenon under investigation and to model it in a way that is somewhat modeling and to be controlled by solutions that are not complex in mathematics.In fact, fuzzy methods of modeling a system with approximation and very easy resolution. All numerical and verbal information obtained in engineering approaches should be able to reach meaningful solutions in the control of the examined event by participating in the solution algorithm. In this respect, the fuzzy set, logic and system principles serve to solve the problems by processing the verbal information given by the experts. However, only numerical values are used in theoretical mathematics and differential calculations. Fuzzy systems are needed to make the calculations that can be detected by computers or algorithms by digitizing the verbal information presented by people (Şen, 2001).

3.1.6. Advantages and disadvantages of fuzzy logic

Many of the problems we encounter in everyday life are not well defined, fuzzy and complex, with features close to human thinking. This is a mathematical problem with many features of this type of fuzzy control in accordance with human thinking style has a distinct advantage. It provides simple solutions to the control of indeterminate, time-varying, complex, well-defined systems. If the system is a system that can be defined by a simple mathematical model which does not have the features we just mentioned, then a traditional system approach will be sufficient. However, as the degree of complexity of the system increases, the traditional control rules, and thus the application of traditional logic, become more difficult and fuzzy logic control emerges as a good alternative. Because applying a traditional logic to a complex system is both

very difficult and costly. On the other hand, fuzzy logic control can analyze the system better than traditional logic and it is also economical.
Fuzzy control often results in a smaller software faster, because fuzzy logic can be subjected to a preprocessing and a wide range of values cannot be reduced to a small number of membership functions.

Another advantage provided by the fuzzy logic control is that it allows the user to benefit from the user's experiences directly. This is a good example of the transmission which can be changed electronically continuously. As is known, conventional automatic gear changes occur automatically when the engine reaches certain speeds. On the other hand, in a manual gear car, the driver has more freedom and the gear changes in certain situations depending on the road, load and driving style.

In the applications of fuzzy logic and fuzzy control, examples are given to traditional controllers, and fuzzy-based controllers are more successful than conventional controllers.

Another good example of fuzzy control is the fuzzy controlled air conditioner produced by Mitsubishi. This device senses the ambient conditions and sets the best operating condition and increases cooling if someone enters the room. Compared to the conventional controller, the heating and cooling times were significantly reduced, with a 20% reduction in the power used.

There are various criticisms against fuzzy logic controllers. One of these criticisms is that fuzzy logic controllers need more information about the process and need more sensors and are therefore both expensive and less reliable. This is not always true. For example, the air conditioner produced by Mitsubishi uses less sensors than the conventional controller. Similarly, the National Panasonic company released a blur-controlled washing machine with only two sensors.

Another criticism is that fuzzy logic controllers can provide high performance compared to traditional controllers by means of non-linear controller.This may be true, but it is likely that the non-linear controller will not occur with a 4-bit microprocessor.

The above described benefits have put fuzzy control into a suitable candidate for industrial applications. However, there are some difficulties encountered in practice. These can be listed as follows;

1- The rules used in fuzzy control are highly dependent on experience.

2- There is no specific method in the selection of membership functions. The most appropriate function is found by trial. This may take quite a long time.

3- A stability analysis of the audit system cannot be carried out and it is not predictable how the system will respond. The only thing to do is simulation work.

We come across many areas with data that is either clear or blurry. Fuzzy numbers or data definition are used for data that is not defined or classified. If the numerical value taken by the variable leads to a subjective definition, this data is defined as fuzzy data (Zimmerman, 1976).

3.2. Fuzzy Regression

Classical Regression is a method of estimating dependent quantitative variable values using independent quantitative variables. Classical regression is an important application tool such as making estimations based on quantitative data of existing and / or past periods, making contributions to decision support systems. Despite some important assumptions in classical regression, flexible thinking structure in real life is not included in the method (Tanaka et al., 1982).

Classical Regression is used as a statistical tool to model the relationship between variables based on available data and to define the current situation or future, and defines the relationship between the variables very sharply. Scientists try to precisely capture the effects and the amount of independent variables in the non-blur regression they use to predict the future. However, the exact mathematical models that try to reflect many ambiguous events in nature do not always work. The classical regression gives accurate results only when it is able to obtain certain data in nature. In other words, the regression which is based on the classical logic itself, leads to wrong decisions in fuzzy systems with features close to human thinking style (Tanaka et al., 1982).

Fuzzy regression was developed in response to the disadvantages of the classical regression described above. In fuzzy regression, the relationship between the dependent variable and the independent variables is not as definite as in the classical regression. For this reason, fuzzy techniques reflect the effects of independent variables in a more realistic way in uncertain cases (Tanaka et al., 1982).

Fuzzy logic and fuzzy regression enter into the work where uncertainty exists in nature and in our daily life, where mathematical methods based on traditional logic are insufficient, it increases the reliability of the system as well as making a significant decrease in costs, as well as making decisions consistent with nature (Tanaka et al., 1982).

The studies based on Tanaka, Hayashi and Watada's (1982) flexible thinking base gave Fuzzy dimension to Regression Analysis. In this way, in situations where uncertainty exists in nature and in our daily life and the classical method is insufficient, it increases the reliability of established systems and helps us to make appropriate decisions for natural thinking (Tanaka et al., 1982).

The differences between the values calculated with the classical regression and the observed values are considered to be measurement errors. In the fuzzy regression model, it is assumed that these differences are based on unstable system parameters. The proposed model is to establish a correct relationship between fuzzy coefficients and system coefficients directly entering the model (Tanaka et al., 1982).

Fuzzy regression has been proposed with fuzzy data. In contrast to the classical regression based on probability theory, fuzzy regression is based on posibility theory and fuzzy set theory. The errors between the regression model and the observed data are often assumed to be a random variable observation error with a normal distribution, with a mean "0" and constant variance, in the classical regression analysis. But in the fuzzy regression analysis, it is assumed that the same type of error is caused by the turbidity in the model structure.

Tanaka et al. (1982) in the study of fuzzy dependent variable and discrete independent variables are formulated in mathematical programming problems. The aim is to minimize the total distribution of fuzzy regression coefficients to find the value that provides the membership

value that predicts the fuzzy equivalent of the regression model. Fuzzy linear regression has been developed to model the linear relationship between variables in the case of turbidity.

According to Yang & Liu (2003), Fuzzy regression analysis is also a probability regression analysis. The deviation between the observed and predicted variables is assumed to be due to random errors, in classical regression analysis. However, these deviations are sometimes caused by unidentified system structure or fuzzy observations. The uncertainty of this type of regression model makes the model not random but rather fuzzy (Yang & Liu, 2003).

3.2.1. Literature review

Fuzzy logic is a widely used system for identifying and controlling systems and models. The first information on fuzzy principles was raised by the Azerbaijani-based Lutfu Askerzade (Zadeh, 1965). Fuzzy logic, which has become increasingly important since then, can be defined as a definite mathematical order established to explain uncertainties and to work with uncertainties. Since Zadeh's introduction of the concept of fuzzy sets, the applications that take fuzzy information into account in the regression model have been successfully applied in many disciplines. Studies on fuzzy regression can be summarized as follows.

Tanaka et al. (1982) proposed the first study of linear regression analysis with a fuzzy model. It is assumed that input and output variables are not blurred, but the system information is blurred and the objective function is based on the minimization of the spread of the estimation value of the dependent variable. The analysis is analyzed using linear programming technique. (Tanaka et al., 1982).

In fuzzy linear regression, Moskowitz and Kim (1993) determined the relationship between fuzzy parameters propagation, forms of membership functions, and H value (Moskowitz and Kim, 1993).

Chang and Lee (1996) proposed the generalized fuzzy weighted least squares method, which weighted with membership degrees and was based on interaction with the decision maker for the case of outliers (Chang and Lee, 1996).

Uemura (1996) used various fuzzy regression methods to evaluate physical tests in schools. Based on 2 approaches. First, classical regression analysis was used. In the second case, fuzzy regression was used when the data were blurred and a study was performed on the total trend. (Uemura, 1996).

Yang and Ko (1997) conducted a study on 2-step iterative algorithm of fuzzy least squares analysis for simple regression. First of all, the fuzzy classification method which gives the class membership of the observations is chosen, then these values of the memberships are used as weights. Fuzzy least squares weighted in fuzzy regression analysis are considered to be an optimization problem (Yang and Ko, 1997).

Chang (1997) used fuzzy regression for seasonal time series data (Chang, 1997).

Boreux, Pesti, Duckstein and Nicolas Paleoclimatic (1997) have made the age model prediction with a fuzzy regression model. Fuzzy regression model was used to determine the age and depth of sediments. Seaweed or regional plants are sensitive to seasonal changes. The age and depth of deposits are dependent variables and the relationship between them is usually fuzzy. A prediction model was constructed using a data set with a fuzzy linear regression model (Boreux et al., 1997).

Wang and Tsaur (2000) proposed the fuzzy least squares method developed by Tanaka to solve the non-fuzzy independent variable and fuzzy dependent variable problem (Wang & Tsaur, 2000).

Ishibuchi and Manabu (2001) mentioned some of the limitations of fuzzy regression method and made the extension of the symmetric triangular fuzzy number type coefficients to asymmetric triangular and trapezoid number (Ishibuchi and Manabu, 2001).

Chang and Bilal (2001) described the differences between fuzzy regression and classical regression.

In this study, a comprehensive literature review is done and three approaches of fuzzy regression are summarized. The first approach is based on the minimization of the blurring with

the most appropriate criterion. The second approach uses the least squares of errors as a suitable criterion and summarizes the two methods in the article. The third approach is defined as intermittent regression analysis. Numerical examples and graphical representations were used to evaluate the difference between each fuzzy regression method and the classical regression method. In the article, the main differences between the randomness type of uncertainty in classical regression model data and the turbidity of the uncertainty in traditional fuzzy regression model data were evaluated comparatively (Chang and Ayyub, 2001).

Lee and Chen (2001) presented a generalized fuzzy linear regression model and proposed a nonlinear programming model to determine fuzzy parameters (Lee & Chen, 2001).

Chang (2001) explained the main differences between fuzzy regression and statistical regression in Ayyubb studies. To fuse turbidity and randomness, hybrid fuzzy least squares regression analysis was developed and details of the method were explained. (Chang and Ayyub, 2001).

Tsaur et al. (2002) proposed a method in which fuzzy regression method was applied in the industrial area where seasonal change and annual change are important. While the mean estimation error was found to be 2,91% and 4,29% in the two time series analyzes, the mean estimation error in the fuzzy regression analysis was 1.85% (Tsaur et al., 2002).

Wu and Tseng (2002) have done a study on fuzzy examples with fuzzy parameter estimation using a fuzzy regression model. Unlike classical analysis estimates, it is related to the fuzzy measurement of the observed values. The results proved that the newly developed approach was more effective and more realistic than classical regression analysis (Wu and Tseng, 2002).

Tseng and Tzeng (2002) suggest fuzzy seasonal ARIMA (SARIMA) method combining advantageous aspects of fuzzy regression and seasonal time series ARIMA series (Tseng and Tzeng, 2002).

Tran and Duckstein (2002) present a multi-purpose fuzzy regression model that combines the central tendency and likelihood features of classical and fuzzy regression models (Tran and Duckstein 2002).

Wu and Tseng (2002) presented a fuzzy parameter estimation fuzzy regression model with the least squares approach (Wu and Tseng, 2002).

Yang and Lin (2002) proposed two predictive methods for the fuzzy input and fuzzy output variables under the least fuzzy squares approximation. Heterogeneous data control and the use of a luminescence analysis to determine outliers (Yang and Lin, 2002).

Hong, Hwang and Ahn (2004) made an estimation of the fuzzy linear regression model using the learning algorithm (Honh et al., 2004).

Lee and Chou (2004) conducted a study called Fuzzy Estimation based on the Fuzzy time series (Lee & Chou, 2004).

Hojati, Bector, Smimou (2005) have proposed a new method calculated under fuzzy thinking in two cases where only independent variables are fuzzy, both dependent and independent variables are fuzzy (Hojati, 2005).

Fuzzy regression analysis approach was developed at fuzzy intervals to create the Kwong and Bai (2005) miroscopic wrapping process. 2 fuzzy regression models were associated with 3 process parameters and 2 quality characteristics, respectively. Then, a fuzzy multi-purpose optimization problem was created. Fuzzy linear regression models are for optimization models. Optimal process parameters were reached by solving the model (Kwong and Bai, 2005).

Aydın, Karaköse and Akın (2005) proposed a predictive algorithm using fuzzy logic-based time series data mining (Aydın et al., 2005).

Nasrabadi, Nasrabadi, Nasrabay (2005) conducted a study called fuzzy linear regression analysis with multi-purpose programming approach. Fuzzy regression is criticized because it is sensitive to outliers, that all data do not affect predicted parameters and that the distribution of predicted values is even more dispersed than data in the model. A multipurpose fuzzy linear regression model was developed to eliminate these deficiencies (Nasrabadi et al., 2005).

Chen, Jia and Lee (2006) conducted a study on the modeling of heat comfort in fuzzy rooms with fuzzy regression analysis. Heat comfort is subjective and expressed in linguistic terms. Experiments were conducted to obtain the required data. The results were analyzed and individual sentiment played an important role in the model (Chen et al., 2006).

Öğüt (2006), a model for determining the number of car owners in Turkey conducted a study related to the development. Multivariable fuzzy regression model was used because of having economic, social and demographic reasons. The results showed that when the multivariate approach was added to posibilistic regression, the model yielded results not only as discrete outputs but also as output range (Öğüt, 2006).

In the Stahl study (2006), a linear fuzzy regression model was established by using fuzzy arguments and fuzzy parameters. The least squares method was used to create fuzzy parameters. The least squares method has shown to be a powerful predictor (Stahl, 2006).

In DiUrso and Santoro (2006) studies, they developed a new approach by improving the classical regression in real situations where observations are not homogeneous. In this study, clustered linear regression analysis is discussed in the context of turbidity. Fuzzy clustered linear regression analysis (FCWLR) model with symmetric fuzzy output and discrete input variables have been proposed to create (DiUrso and Santoro, 2006).

A new approach has been developed for the linear programming method of Tanaka in order to address the problems of outliers by Hung and Yang (2006). This approach has the capacity to calculate the value change behavior in the objective function when observations are destroyed (Hung and Yang, 2006).

Kandari, Soliman and Alammari (2006) conducted fuzzy voltage measurements based on fuzzy linear estimation method. In the developed algorithm, digitized voltage signal samples were used in which power quality standards were met. Voltage signals are modeled using fuzzy linear parameter estimation problems. The parameters are defined by the triangular membership function (Kandari et al., 2006).

Wang, Zhang and Mei (2007) have developed fuzzy nonparametric regression based on simplified local linear techniques. (Wang et al., 2007)

Ge., H., Wang, S. (2007), using non-symmetrical fuzzy triangular coefficients, conducted a study on the degree of adaptation in the fuzzy linear regression and the dependency between input noise. (Ge and Wang, 2007)

The relationship between He, Chan and Wu (2007), productivity, customer satisfaction and profitability was tried to be determined using the classical regression and new fuzzy regression approach. 22 samples were selected from Hong Kong and statistical values of three variables were given. Many models were initially estimated using the known least squares method (OLS). A new constraint was added to the existing FLR method and an "improved FLR" method was developed. The data were solved by the statistical regression model, FLR model and newly developed FLR method and the results were compared (He et al., 2007).

By Sánchez (2007), Tanaka's classic fuzzy regression method and Sherman's rights protection plan method are combined and developed to create a new protection method (Sananchez, 2007).

Huang and Tzeng (2008) have developed a two-step new fuzzy part regression analysis method to calculate product life time and annual order quantity of products. (Huang and Tzeng, 2008)

Arulchinnappan et al. (2011) demonstrated fuzzy linear regression model for estimation of reverse osmosis permeating parameters conditions. They observed that fuzzy regression model was effectively taken on non-crisp, fuzzy and crisp data. They examined the variables that contribute to the deterioration of membrane. They concluded that the fuzzy linear regression model had accuracy with efficient prediction model. (Arulchinnappan et al., 2011)

Azadeh et al. (2011) proposed fuzzy regression–analysis of variance (ANOVA) algorithm to estimated and predicted electricity consumption in uncertain environment. They compared the fuzzy regression–analysis of variance algorithm with sixteen fuzzy regression models. They observed that there was no clear cut as to which of the recent fuzzy regression model was suitable for given set of actual data with respect to electricity consumption .and also difficult to model uncertain behavior of electricity consumption with conventional time series and proper

fuzzy regression could be an ideal substitute for such cases. They suggested that the fuzzy regression–analysis of variance algorithm used to identified the optimum model with lowest error. (Azadeh et al., 2011)

Abdullah et. al. (2012) developed the fuzzy regression model used for road accident in Malaysia over the period of 1974 to 2007 used three predictors. They introduced the threshold level h=0.5, 0.9 were accounted and also model structured was the road accident uses registered vehicles, population and road length as variables based on one response variable was road accident. They showed that the variables of registered vehicles and population provide higher impact to the number of road accident. (Abdullah et. al., 2012)

Akdemir et al. (2013) studied the effect of the variation of metrological conditions on the total burned area in hectares, by used fuzzy linear regression analysis based on Tanaka's fuzzy regression model. They observed that temperature, wind speed and rainfall were significant on total burned area. (Akdemir et al., 2013)

Azadeh et al. (2013) implemented fuzzy regression model was applied to gas consumption in Iran. They compared with conventional regression methods and fuzzy regression model to the crisp data. They observed that fuzzy algorithm was better performance than conventional regression analysis. (Azadeh et al., 2013)

Muzzioli, Ruggier and Debaets (2015) investigated the linear fuzzy regression methods of Tanaka et al., Savic, Pedrycz, Ishibuchi and Nii and applied on finance problem and estimated of the smile function. They evaluated both the in-sample pricing performance of the different estimation methods and the out-of-sample forecasting performance of the moments, by using as a benchmark the standard cubic spine interpolation. They suggested that in sample, by using fuzzy regression was better estimation than the classical approach based on cubic spines. (Muzzioli et al., 2015)

Lately, Yoon and Choi (2019) examine the impact of satisfaction with family, friends, school, and government on life satisfaction, and to identify the impact of knowledge, attitude, and practice of sharing on life satisfaction. (Yoon and Choi, 2019)

3.2.2. Fuzzy linear regression model

Over the years, classical regression has been practiced in almost every field of science.
The purpose of the regression is to explain the change of a dependent Y variable in f(x) depending on the explanatory variables x in Y= f(x). The use of classical regression is based on the fact that non-observable error terms are mutually independent and have the same distribution and some definite assumptions about the given data. Consequently, classical regression can only be applied if the data are distributed by a model and the relationship between x and y is definitively defined. Fuzzy linear regression (FLR) was first developed by Tanaka et al. in 1982 and some definite assumptions of classical regression were relaxed (Tanaka et al., 1982).

In the absence of rigid assumptions of classical regression analysis, many researchers have conducted research in the field of linear fuzzy regression because of the applicability of fuzzy regression to real-life problems.

Since the development of the first fuzzy regression model, it has been criticized by Celmins, Chang et al. and Redden ve Woodall. After entering the first method, many fuzzy regression methods have been used to minimize blur. Tanaka and Ishibuchi (1992) developed quadratic membership functions to produce fuzzy coefficients (Ishibuchi and Tanaka, 1992). For fuzzy regression, Chang and Lee (1996) have developed an unrestricted model of width. Tanaka et al. (1982) developed a regression that compared the results of fuzzy regression with classical regression. In these developed methods, linear programming problem solving technique has been used to obtain minimum turbidity as a criterion of conformity (Chang and Ayyub, 2001).

Four of the fuzzy regression models developed so far are summarized below. In the next section, a numerical sample study will be made on the fuzzy regression models described and the results will be interpreted according to the decision maker.

3.2.3. Fuzzy regression model of Tanaka

It is assumed that the deviation between observed and predicted data in fuzzy regression is due to system turbidity or blur of regression coefficients. The aim of fuzzy regression is to

find a regression model suitable for all observed fuzzy data. Different fuzzy regression models are created depending on the use of appropriate criteria. Tanaka et al. (1982) formed the first fuzzy linear regression model with a fuzzy model. According to this method, regression coefficients are fuzzy numbers. Because the regression coefficients are fuzzy numbers, the predicted variable Y is the fuzzy number. Fuzzy regression analysis with single independent X variables is summarized below. A_0 fuzzy interaction coefficient and A_i fuzzy slope coefficient (Tanaka et al., 1982).

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X$$

Each fuzzy parameter $\tilde{A}_i = (c_i, s_i)$ is expressed as symmetrical triangular membership functions with central value c_i and distribution value s_i. Other membership function shapes can also be used.

According to this approach, the fuzzy coefficients $\tilde{A}_i (\ i = 0,1)$calculate the predicted fuzzy output to have a minimum fuzzy width of the $\tilde{Y}$ to achieve the aim of belief H.. The term "H" is referred to as a measure of "the degree of compatibility" measuring data between the data and the regression model (Tanaka et al., 1982).

A basic fuzzy linear regression model is assumed as follows;

$$\tilde{Y} = \tilde{A}_0 X_0 + \tilde{A}_1 X_1 + \cdots + \tilde{A}_N X_N = \tilde{A} X$$

X = $(X_0, X_1, \ldots, X_N)^T$ is the independent variable vector.
$\tilde{A} = (\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_N)^T$ membership function is the vectors of the blurred coefficient $\tilde{A}_j = (c_j, s_j)$ in the symmetrical triangular fuzzy number structure defined as follows:
c_j is the central value and s_j is the distribution value.

$$\mu_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|c_j - a_j|}{s_j}, & c_j - s_j \leq a_j \leq c_j + s_j, \forall j = 1,2, \ldots, N \\ 0 \end{cases} \quad (3.1)$$

Thus, the fuzzy regression model can be written as follows;

$\widetilde{Y_i} = (c_0, s_0) + (c_1, s_1)x_{1i} + (c_2, s_2)x_{2i} + \cdots + (c_N, s_N)x_{Ni}$

Above is the fuzzy regression analysis, the parameter distribution of the relationship between the input and output data predicts certain input and output data which are fuzzy functions.
By adding Extension groups, the membership of the fuzzy number $\tilde{Y}_i$ is calculated as follows;

$$\mu(Y_i) = \begin{cases} 1 - \frac{|Y_i - X^t c|}{s^t|X|}, & X \neq 0 \\ 1, & X = 0, Y \neq 0, \\ 0, & X = 0, Y = 0, \end{cases} \quad \forall I = 1,2,\dots,M \tag{3.2}$$

$s^t = (s_0, s_1, \dots, s_N), \qquad c = (c_0, c_1, \dots, c_N)$

Each dependent fuzzy variable value is calculated as $\tilde{Y}_i = (Y_i^L, Y_i^{h=1}, Y_i^U)$ i = 1,2,…,M. The lower limit of $\tilde{Y}_i$ fuzzy number is $Y_i^L = \sum_{j=0}^{N}(c_j - s_j)X_{ij}$: the central value of $\tilde{Y}_i$ is $Y_i^{h=1} = \sum_{j=0}^{N} c_j X_{ij}$ and the upper value of $\tilde{Y}_i$ is $Y_i^U = \sum_{j=0}^{N}(c_j + s_j)X_{ij}$.

To achieve fuzzy regression analysis to minimize blur, the objective function is adapted to minimize the overall distribution of fuzzy number $\tilde{Y}_i$;

(3.3)

$$mins^t|X| = min\sum_{j=0}^{N}\left(s_j\sum_{i=1}^{M}|X_{ij}|\right),$$

Constraints require that each Y_i observation value be bound to $\tilde{Y}_i$ with a minimum degree of H. So, $\mu Y_i \geq H \ (i = 1,2,\dots,M)$

(3.4)

$$1 - \frac{|Y_i - X^t c|}{s^t|X|} \geq H, \forall i = 1,2,\dots,M$$

Membership Value

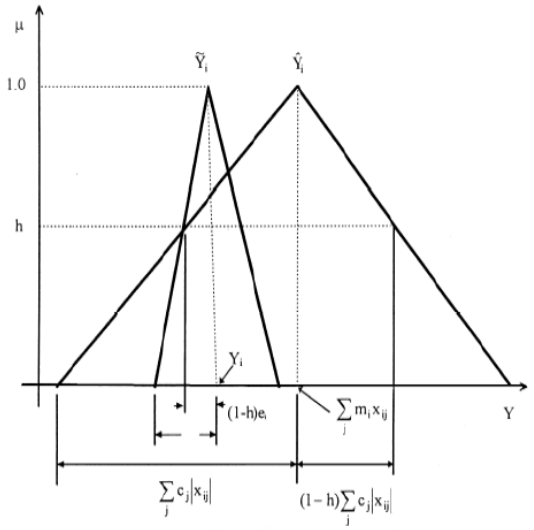


Figure 3.2. The degree of conformity of $\tilde{Y}_i$ required to obtain fuzzy data $\tilde{Y}_i$ (Chang et al., 2001)

Each observed data set, which may be fuzzy $\tilde{Y}$ and specific Y_i data, must fall to the predicted $\tilde{Y}$ value at the H level as shown in figure 3.2. Determination of H in fuzzy regression was left to the analyst. There is no specific criterion for the choice of H. The recommended H level is usually 0.5 (Uras, 1998).

To calculate the fuzzy coefficient $\tilde{A}_i = (c_i, s_i)$, the following linear fuzzy regression model developed by Tanaka et al. (1982) are formulated:

$$\text{Min } Z = \sum_{j=0}^{N}\left(s_j \sum_{i=1}^{M} \left| x_{ij} \right| \right) \tag{3.5}$$

subject to

$$\sum_{j=0}^{N} c_j x_{ij} + (1-H)\sum_{j=0}^{N} s_j \left| x_{ij} \right| \geq y_i$$

$$\sum_{j=0}^{N} c_j x_{ij} - (1-H)\sum_{j=0}^{N} s_j \left| x_{ij} \right| \leq y_i$$

$x_{i0} = 1$, $i = 1, 2, \ldots, M$,

$s_j \geq 0$, $j = 0, 1, \ldots, N$.

3.2.4. Fuzzy least squares regression

This section describes the development of fuzzy least squares regression analysis (FLSRA). Different perspectives were developed by Celmins, Diamond, Savic and Pedrycz and Chang and Ayyub (2001) in the fuzzy least squares regression. Celmins defined the appropriate measurement between fuzzy data and a model and used this measurement as a model compliance criteria. Diamond has developed a method of least squares. Savic and Pedrycz developed an integrated approach to FLSRA by fusing the minimum fuzzy criteria into the classic regimen. Chang and Ayyub (2001) emphasized reliability issues such as standard deviation and correlation coefficient. This section summarizes the two methods of FLSRA developed by Celmins, Savic and Pedrycz (Chang and Ayyub, 2001).
Least Squares Regression Using Minimum Fuzzy Criterion:

Savic and Pedrycz (1991) formulated the fuzzy regression method by combining the least squares principle and the minimum turbidity criterion. The method is done in 2 consecutive steps. The first step uses the classical regression to find the fuzzy center values of fuzzy regression coefficients. The second step uses the minimum turbidity criterion to find the fuzzy intervals of fuzzy regression coefficients. In the first step, a regression line is developed using the available information on the central values of fuzzy observations. Fuzzy data acts as simplified specific data and regression analysis is performed as a classical regression. The results of the first step are used as the central value of the central coefficients of fuzzy regression coefficients. In the second step, the fuzzy coefficients are calculated using the minimum turbidity criterion. The width of the fuzzy coefficients is calculated using Equation 3.6 and 3.7 (Savic and Pedrycz, 1991).

$$\min S = ns_0 + s_1 \sum_{i=1}^{n} |X_i|$$

st

$s_0 \geq 0, s_1 \geq 0$

$$\sum_{j=0}^{1} c_i X_{ij} + (1-H) \sum_{j=0}^{1} s_j |X_{ij}| \geq Y_i + (1-H) e_i \; for \; i = 1-n \qquad (3.6)$$

$$\sum_{j=0}^{1} c_i X_{ij} - (1-H)\sum_{j=0}^{1} s_j \left|X_{ij}\right| \geq Y_i - (1-H)e_i \; for \; i = 1-n \tag{3.7}$$

Range Regression :

According to this method, fuzzy data and fuzzy regression coefficients act as spacing numbers. Interval operations are added to the fuzzy regression and thus are referred to as interval regression analysis. Fuzzy regression coefficients are calculated as if all fuzzy outputs are in a fuzzy regression model. An interval regression model for specific X and specific Y is shown in Figure 3.3. The following linear programming formulation for $\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X$ is used to solve fuzzy regression coefficients $\tilde{A}_0 = (c_0, s_0)$ and $\tilde{A}_0 = (c_1, s_1)$.

$$\min ns_0 + s_1 \sum_{i=1}^{n} X_i$$

st

$$s_0 \geq 0, s_1 \geq 0 \tag{3.8}$$

$$(c_0 - s_0) + (c_1 - s_1)X_i \leq Y_{i,L} \quad i = 1-n \tag{3.9}$$

$$(c_0 + s_0) + (c_1 + s_1)X_i \geq Y_{i,L} \quad i = 1-n \tag{3.10}$$

$Y_{i,L}$ and $Y_{i,U}$ are the lower and upper limit values for each fuzzy data, respectively. The objective function in Equation 3.8 results in a minimum of total fuzzy magnitudes. Equation 3.9 and 3.10 are used to limit the observed fuzzy data to a fuzzy regression model. The above formulation is called the minimization problem by Ishubichi (Ishibuchi, 2005).

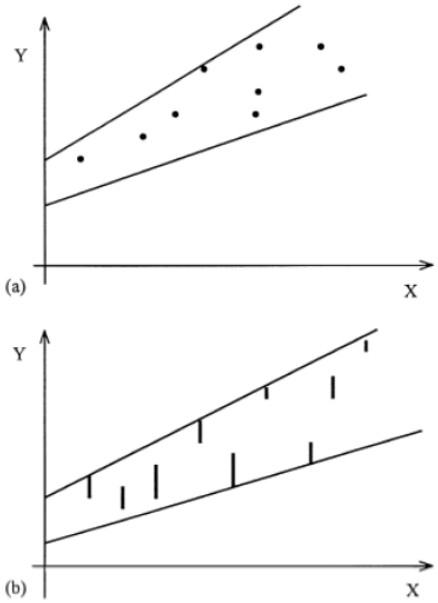


Figure 3.3. Range Regression Models (a) for specific X and specific Y values, Range Regression model (b) For specific X and fuzzy Y values

3.2.5. Classical regression and fuzzy regression analysis

Regression analysis can be defined as a statistical tool that demonstrates the relationship between independent quantitative variables and dependent quantitative variables. However, a healthy analysis requires collecting precise information from nature. However, this is often impossible. In such cases, we can compensate for the shortcomings of regression analysis using fuzzy regression.

Because classical regression analysis with fuzzy data will bring us far and distorted results. In our world, where blurring and complexity are increasing day by day, neither businesses nor other organizations can tolerate distant, meaningless and deviant consequences.

Therefore, fuzzy data and fuzzy regression analysis can be performed by using standard data.

Another advantage of fuzzy regression is that it offers thousands of alternatives to the decision maker. Thus, the analyst will be able to choose the most suitable model for the enterprise.

Fuzzy regression is based on system blur and gives different results than classical regression. The difference between fuzzy regression and classical regression results from differences in views between observed values and estimated values. Deviations in classical regression are random errors in observations. In fuzzy regression, deviations are caused by fuzzy errors caused by system turbidity. Both the fuzzy and the classic regression only take into account the total uncertainty. In fact, there are two different types of uncertainty in the randomness and turbidity regression analysis. In the classical regression, probability theory is used to model random errors, and the results are presented in classical regression equations. On the other hand, fuzzy set theory can be used to model fuzzy errors and the results are presented in fuzzy regression equations.

Classical regression is a suitable tool for certain observed data analyzing random errors between observed and estimated values. In the classical regression, unlike fuzzy errors in fuzzy data, certain data are considered to contain random errors. Fuzzy regression linguistic explanations: excellent, very good, as well as provides a way to model the observed fuzzy data. Fuzzy regression is needed if data contains turbidity. However, in fuzzy regression, if the data are forced to be approached to a discrete state, fuzzy regression results may approach classical regression results. However, such a feature has not yet been used in fuzzy regression methods. The reason for this is that fuzzy regression uses the hypothesis of system turbidity instead of the random assumption in classical regression.

The blur of the data is used as a substitute for the randomness of the data rather than by giving it a hint of the data. Based on this assumption, fuzzy regression models are the results of data blur and system turbidity. However, different compliance criteria should be used in fuzzy regression models. As a result, fuzzy regression may provide a fuzzy model, assuming it contains system blur, even if all data is specific.

If fuzzy regression is not sufficient for classical regression analysis if the data set is not provided for statistical regression analysis, if the regression model has a weak representation, if there is human judgment (input or output is fuzzy number), the error terms are better than the classical regression if it is related to system uncertainty. (Özelkan and Duckstein, 2000)

Generally, there are certain circumstances under which the application of fuzzy regression is appropriate, but statistical regression is not – for example, when the observations of the dependent variable are obtained as fuzzy numbers (Tanaka, Uejima and Asai, 1982), or some data are collected by measurements (i.e., crisp), yet other data are estimated qualitatively (i.e., fuzzy) (Johnson and Ayyub, 1993). Such cases often occur when human expert knowledge is the main source of the data or information needed for modeling. (Kim et al., 1996)

The objectives of regression analysis are to make predictions and describe/explain phenomena. The experiments showed that in terms of predictive capability, statistical regression is generally superior to fuzzy regression. However, with respect to descriptive capability, performance, and therefore choice of method, depends on various factors associated with the data set, proper specificity of the model, and the nature of the process being investigated. Interestingly and somewhat counter-intuitively, fuzzy regression should not be used when the data quality is bad (i.e., when there exist outliers or high variability in the data). (Kim et al., 1996)

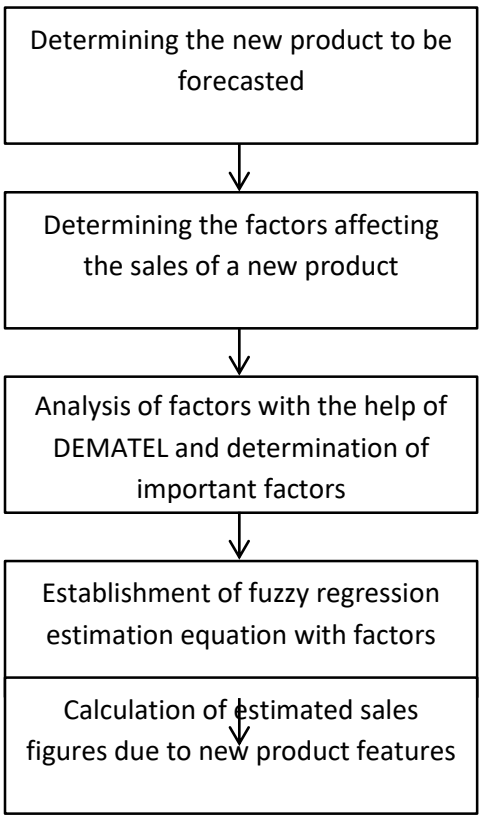
Although it is conventional and generally superior to use statistical regression analysis for predicting and describing phenomena from data, there are cases when fuzzy regression analysis is not only appropriate but indeed superior in describing a process. In general, one of the characteristics of the problems suitable for statistical regression is that they are open to experimentation. This, in essence, means that sufficient data must be available for deriving a valid statistical relation among variables. Often, in practice, this is not the case. Another important characteristic of problems amenable to statistical regression is the validity of the assumptions in order for the linear regression model to be considered in practical applications (Savic and Pedrycz, 1991). Fuzzy regression provides a viable, alternative approach for modeling situations which fail to satisfy one or both of the above characteristics. (Kim et al., 1996)

4. PROPOSED APPROACH FOR SALES FORECASTING

The forecasting of sales is considered as one of the most crucial processes in order to succeed in competitive markets. Especially, effective prediction of a new product's sales volume is critical for accurate production and marketing planning. The lack of historical data and the vagueness that exists between factors affecting the sales makes fuzzy regression suitable for sales forecasting problems concerning new products or services.

This study proposes an integrated approach based on Decision Making Trial and Evaluation Laboratory (DEMATEL) and fuzzy linear regression in order to forecast the volume of sales. DEMATEL methodology is employed to determine the variables influencing the sales performance in manufacturing companies. Then, a fuzzy linear regression model which predicts the sales using explanatory variables determined by DEMATEL method is built.

Steps to take when application:



4.1. DEMATEL Methodology

The DEMATEL methodology (Fontela and Gobus, 1976) developed by "the Science and Human Affairs Program of the Battelle Memorial Institute of Geneva" between 1972 and 1976 (Tzeng et al., 2011). Based on the graph theory, the DEMATEL method can divide multiple factors into a cause-effect group and it enables the decision maker to visualize influences between criteria with a network relationship map (Yang et al., 2011).

The method begins by generating the initial direct influence matrix *A*. The elements a_{ij} of the matrix *A* represent the direct influence of each factor *i* exerts on each factor *j*, evaluated by a decision maker. The matrix A is normalized by using (4.1) and it is named as the matrix D (Tzeng et al., 2011), (Yang et al., 2011), (Hsu et al., 2012).

$$D=d.A, \tag{4.1}$$

where

$$d= \min\left[1/\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|, 1/\max_{1\le i\le n}\sum_{i=1}^{n}|a_{ij}|\right]$$

The total relation matrix *T* is defined as $T = D(I - D)^{-1}$, where *I* is the identity matrix.

Define *r* and *c* be *n* x 1 and 1 x *n* vectors representing the sum of rows and sum of columns of the total relation matrix *T*, respectively. Suppose r_i be the sum of i^{th} row in matrix *T*, then r_i represents both direct and indirect effects given by criteria *i* to the other criterias. If c_j describes the sum of j^{th} column in matrix *T*, then c_j represents both direct and indirect effects by criteria *j* from the other criterias (Yang et al., 2011).

The degree of importance for a factor *i* is considered as equals to the sum ($r_i + c_j$) when $j= i$ (Hsu et al., 2012), (Wang and Chen, 2012). A network relationship map which explains the structural relations among factors can be obtained by setting up a threshold value which is determined by the decision makers (Yang et al., 2011).

Additionally, the difference (r_i - c_j) represents the net effect that factor *i* contributes to the system. A factor *i* is a net causer if (r_i - c_j) is positive, and when (r_i - c_j) is negative, factor *i* is a net receiver (Yang et al., 2011).

5. ILLUSTRATIVE SALES FORECASTING PROBLEM

This work aims to forecast the volume of sales using an integrated approach based on DEMATEL and fuzzy linear regression.

The degree of dependence of the criteria for the new paint is an important issue in the process of selling. This study has focused on a new paint product will occur in Turkey. The DEMATEL method was used to determine the dependent weight values of the criteria. Research results with focus groups were used for the new product type.

The types of products used are as follows:

Most healthy paint: It is a type of paint that does not have any special additive (biocide or volatile component) in its structure and which has the lowest level of damage to the environment in places where it is applied, has no negative effect on indoor air quality.

Paint applied on one layer: Classic wall paints are applied with at least 2 layers. There are also waiting times between layers. Instead, it is a type of paint that can be applied as a single layer and thus saves time and labor.

Non-dripping paint: In the application of classical paints, it is protected against paint drops by covering the various faces on the ground and the surrounding area. As the name suggests, the non-drip dye is a more dense consistency and a product type that can be used without dripping around during its application.

Multi purpose paint: It can be applied safely on different surfaces other than walls, metal, plastic, wood etc.

Scent paint: It is a type of paint that emits a pleasant smell for a certain period when applied to the walls in the interior.

Research has been conducted to determine which product type to use.

The content of the survey is as follows:

WHO? : The people who participated in the survey were selected from the people who had their paint done in the last year. (ABC1C2 , %50 Male - %50 Female)

HOW?: In this study, quantitative CAPI face to face interview technique was used. The questionnaire to be used in the study was used for field studies. The questionnaire interviews lasted an average of 23 minutes.

HOW MANY?: Representing a total of 821 interviews across 15 provinces of Turkey:

• Makes everything: 192

• Make a brand decision and buy it, but paint it to the painter: 316

• Deciding on the brand and leaving the purchase and painting to the painter: 110

• Leave everything to the painter: 203

The product type results were surveyed and the "Most Healthy Paint" concept in all segments was significantly more attractive than the other concepts.

		make everything themself	buy themself, paint to the painter	buy and paint to the painter	leave everything to the painter
N=821		N=192	N=316	N=110	N=203
Most healthy paint	39%	35%	31%	45%	50%
Paint applied on one layer	19%	24%	23%	13%	12%
Non-dripping paint	15%	16%	15%	15%	14%
Multi purpose paint	14%	14%	16%	11%	12%
Scent paint	13%	12%	15%	12%	12%

The ratio of those who say "1-Most Attractive" for each concept is reported.

Question: Which one of these promotional ideas did you find most attractive? Could you please sort?

Figure 5.1. Survey Results

The factors influencing the sales which are determined by market research are price, availability, healthy and environment friendly (represents the level of special additive and the negative effect on indoor air quality), painting service, reputation of the company, number of color alternatives, payment collection process, communication variety, economic risk

situation. The criteria taken into account in this forecasting problem are summarized in Table 5.1.

Table 5.1: Criteria influencing the volume of sales

	Definition
Criterion 1 (C1)	Price
Criterion 2 (C2)	Availability
Criterion 3 (C3)	Healthy and environment-friendly
Criterion 4 (C4)	Painting service
Criterion 5 (C5)	Reputation
Criterion 6 (C6)	Number of color alternatives
Criterion 7 (C7)	Payment collection process
Criterion 8 (C8)	Communication variety
Criterion 9 (C9)	Economic risk situation

Criterias Definition**:**

1. *Price:* Sales price to the consumer.
2. *Availability:* Availability of the product in the market
3. *Be Healthy*: It is a type of paint that does not have any special additive -biocide or volatile component- in its structure and which has the lowest level of damage to the environment in places where it is applied, has no negative effect on indoor air quality
4. *Giving Painting Service:* Supporting the application of the product with the application service
5. *Reputation :* Consumer's perception of brand reliability
6. *Color Alternatives:* Big scala size of color alternatives
7. *Payment Collection Process:* the method and the type of payment when purchasing the product

8. *Communication Variety:* communication diversity for new product through channels such as campaigns, dealers, sales clerks, sales person etc.
9. *Economic Risk Situation:* general economic situation of the country

A team of manufacturer experts indicates the influence which is express the influence of the i-th object on the j-th object, an 4-degree scale is used, where: 0 – no influence, 1 – low influence, 2 – medium influence, 3 – high influence, 4 – very high influence. The initial direct influence matrix is shown in Table 5.2. This matrix is normalized using (5.1) and the total relation matrix, given in Table 5.3, is calculated.

Table 5.2: Initial direct influence matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	0	0	0	2	3	0	0	2	1
C2	1	0	0	2	3	0	0	0	0
C3	4	2	0	3	4	2	0	0	0
C4	4	3	0	0	4	0	0	0	0
C5	3	2	2	3	0	4	0	0	0
C6	2	4	0	0	3	0	0	0	0
C7	1	0	0	0	0	0	0	3	3
C8	1	0	0	0	0	0	4	0	3
C9	4	2	0	1	0	0	4	4	0

$$D=d.A, \tag{5.1}$$

where

$$d= \min\left[1/\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|, 1/\max_{1\le i\le n}\sum_{i=1}^{n}|a_{ij}|\right]$$

The total relation matrix T is defined as $T = D(I - D)^{-1}$, where I is the identity matrix.

Table 5.3: Total relation matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	0,101	0,065	0,022	0,157	0,217	0,046	0,043	0,133	0,081
C2	0,129	0,058	0,022	0,156	0,220	0,046	0,005	0,016	0,010
C3	0,355	0,220	0,037	0,271	0,375	0,179	0,014	0,043	0,026
C4	0,298	0,216	0,031	0,104	0,314	0,066	0,012	0,036	0,022
C5	0,294	0,220	0,119	0,249	0,188	0,249	0,011	0,036	0,022
C6	0,180	0,251	0,024	0,084	0,244	0,051	0,007	0,022	0,013
C7	0,118	0,029	0,003	0,030	0,029	0,006	0,083	0,214	0,200
C8	0,123	0,030	0,003	0,031	0,031	0,006	0,260	0,093	0,209
C9	0,296	0,141	0,009	0,114	0,093	0,020	0,278	0,291	0,100

Table 5.4 shows the sum ($r_i + c_j$) and Table 5.5 shows the difference ($r_i - c_j$) calculated for each criterion.

Table 5.4: Sum (rows + colums)

	sum ($r_i + c_j$)
C1	2,759
C2	1,891
C3	1,790
C4	2,294
C5	3,100
C6	1,546
C7	1,426
C8	1,671
C9	2,028

Table 5.5: Difference (rows - colums)

	difference (r_i - c_j)
C1	-1,031
C2	-0,568
C3	1,248
C4	-0,095
C5	-0,324
C6	0,207
C7	-0,001
C8	-0,096
C9	0,660

According to Table 5.5, the criteria 3, 6 and 9 are net causer and the criteria 1, 2, 4, 5, 7 and 8 are net receiver.

The degree of importance for a factor *i* is considered as equals to the sum ($r_i + c_j$) when $j = i$. According to Table 5.4, the most important criterion for sales performance is the reputation. The price of the product, the painting service provided by the company and economic risk situation of the country are the other critical factors.

In order to built a fuzzy linear regression model which predicts the sales of a new paint using explanatory variables determined by DEMATEL method whose the sum ($r_i + c_j$) when $j = i$, greater than 2, namely the price (C1), the painting service (C4), the reputation (C5), and the economic risk situation (C9), this work uses hypothetical data related to 10 paints. The data set is given in Table 5.6.

Table 5.6: Data related to critical factors

	Sales (Dependent variable)	Price ($)	Painting service	Reputation	Economic Risk Situation
Paint 1	48764	37	3	7	3
Paint 2	74428	20	5	4	2
Paint 3	85097	32	4	8	1
Paint 4	29180	36	2	6	4
Paint 5	51008	45	4	5	2
Paint 6	22105	75	3	4	2
Paint 7	31680	43	3	7	3
Paint 8	28052	75	2	5	4
Paint 9	26331	56	3	6	4
Paint 10	27447	70	2	5	3

The fuzzy linear equation is expressed in such a way to relate the sales (monthly), which is selected as the dependent variable, to explanatory variables: price (x_1), painting service (x_2) evaluated using an integer scale 1 to 5, reputation (x_3) evaluated using an integer scale 1 to 10, and economic risk situation (x_4) evaluated using an integer scale 1 to 5.

$$\tilde{y}_i^* = \tilde{A}_0 + \tilde{A}_1 x_{i1} + \tilde{A}_2 x_{i2} + \tilde{A}_3 x_{i3} + \tilde{A}_4 x_{i4} \tag{5.2}$$

The H value is equals to 0.5. The center and spread values of fuzzy parameters of the resulting predictive equation obtained by using fuzzy linear regression (5.3) are shown in Table 5.7. The resulting predictive equation is as follows:

$$\text{Min } Z = \sum_{j=0}^{N}\left(s_j \sum_{i=1}^{M}\left|x_{ij}\right|\right) \tag{5.3}$$

subject to

$$\sum_{j=0}^{N} c_j x_{ij} + (1-H)\sum_{j=0}^{N} s_j \left| x_{ij} \right| \geq y_i$$

$$\sum_{j=0}^{N} c_j x_{ij} - (1-H)\sum_{j=0}^{N} s_j \left| x_{ij} \right| \leq y_i$$

$x_{i0} = 1$, $i = 1, 2, \ldots, M$,

$s_j \geq 0$, $j = 0, 1, \ldots, N$.

Table 5.7: Fuzzy parameters H=0.5

	Value
c_0	40943.420
s_0	0
c_1	-208.3473
s_1	46.0909
c_2	9291.041
s_2	0
c_3	2149.728
s_3	0
c_4	-11046.75
s_4	0

Volume of Sales= 40943.420 + (-208.3473, 46.0909) Price + 9291.041 Painting service +2149.728 Reputation -11046.75 Economic risk situation (5.4)

The *H* value is equals to 0.0. The center and spread values of fuzzy parameters of the resulting predictive equation obtained by using fuzzy linear regression (5.3) are shown in Table 5.8.

Table 5.8: Fuzzy parameters H=0.0

	Value
c_0	40943.420
s_0	0
c_1	-208.3473
s_1	23.5454
c_2	9291.041
s_2	0
c_3	2149.728
s_3	0
c_4	-11046.75
s_4	0

Volume of Sales= 40943.420 + (-208.3473, 23.5454) Price + 9291.041 Painting service (5.5)
+2149.728 Reputation -11046.75 Economic risk situation

The *H* value is equals to 0.7. The center and spread values of fuzzy parameters of the resulting predictive equation obtained by using fuzzy linear regression (5.3) are shown in Table 5.9.

Table 5.9: Fuzzy parameters H=0.7

	Value
c_0	40943.420
s_0	0
c_1	-208.3473
s_1	53.0649
c_2	9291.041
s_2	0
c_3	2149.728
s_3	0
c_4	-11046.75
s_4	0

Volume of Sales= 40943.420 + (-208.3473, 53.0649) Price + 9291.041 Painting service (5.5)
+2149.728 Reputation -11046.75 Economic risk situation

The resulting fuzzy linear regression equation can be used for sales forecasting of a new paint. As an example, consider a new paint whose characteristics are shown in Table 5.10.

Table 5.10: Data related to sales forecasting

	Price ($)	Painting service	Reputation	Economic Risk Situation
New Paint	95	4	8	5

For H=0.0,

The volume of monthly sales predicted using the fuzzy linear regression equation (5.4) is (20278.665, 2236.813). Thus, the predicted sales volume is in the interval [18041.852, 22515.478].

For H=0.5,

The volume of monthly sales predicted using the fuzzy linear regression equation (5.4) is (20278.665, 4378.636). Thus, the predicted sales volume is in the interval [15900.029, 24657.300].

For H=0.7,

The volume of monthly sales predicted using the fuzzy linear regression equation (5.4) is (20278.665, 5041.166). Thus, the predicted sales volume is in the interval [15237.499, 25319.830].

6. CONCLUSION

The forecasting of sales is a very critical process leading important results that are essential when making strategic decisions in competitive markets. Especially, effective prediction of a new product's sales volume is critical for accurate production and marketing planning.

Fuzzy regression analysis can be applied effectively to estimate relationships among variables and to make predictions when the data is incomplete and/or fuzzy. The lack of historical data and the vagueness that exists between factors affecting the sales makes fuzzy regression more suitable than other classical techniques for sales forecasting problems concerning new products.

This research proposes a combined DEMATEL and fuzzy linear regression approach for sales forecasting. It is necessary to determine the variables to implement the fuzzy regression model in the most efficient way. The degree of dependence of the criteria with each other is an important factor in the decision-making process. The factors to be evaluated, such as TSE (Turkish Standards Institute), breathable, wiping, odorless, high coverage, which are considered as hygiene factors, are not included. Because hygiene factors are considered to be indispensable in the market. The criteria that are considered to affect the sale in practice have been price, availability, healthy and environmentally friendly, application service, reputation, color alternative, payment collection process, communication variety and economic risk situation. DEMATEL (The Decision Making Trial and Evaluation Laboratory) methodology is employed to determine the variables influencing the sales performance in manufacturing companies. With this method, the factors affecting the painting sales and the interaction between the factors were determined. In order to built a fuzzy linear regression model which predicts the sales of a new paint using explanatory variables determined by DEMATEL method whose the sum ($r_i + c_j$) when $j = i$, greater than 2, namely the price (C1), the painting service (C4), the reputation (C5), and the economic risk situation (C9).

Then, a fuzzy linear regression model which predicts the sales using explanatory variables determined by DEMATEL method is built. According to the results, the increase in product price and economic risk situation negatively affects the sales of new products. When we look at the paint service and reputation, they have a positive effect on the sale of paint. This predicted

fuzzy linear regression equation can be used to sales forecasting when the criteria have been evaluated.

Future research will focus on determining the conditions under which the proposed product can be reached, based on the sales volume of the same product under different conditions, to the targeted sales and profitability.

REFERENCES

Abdullah, L., Zamri, N. (2012). *Road Accident Models with Two Threshold Levels of Fuzzy Linear Regression,* Journal of Emerging Trends in Computing and Information Sciences vol.3, No.2, pp. 225-230

Akdemir, H. G., Tiraki, F. (2013). *Using fuzzy linear regression to estimate relationship between forest fires nd meteorological condition*, vol.8, No. 2, pp. 673-683

Arbib, M. A. (1977). *Book reviews in bulletion of the American mathematicial society*, 83(5): 446-451

Arulchinnappan, S., Rajendran, G. (2011). *A Study on Reverse Osmosis Permeating Treatment For Yarn Dyeing Effluent Using Fuzzy Linear Regression Model*, African Journal Of Biotechnology. Vol. 10 No.78, pp. 17969-17972

Aydın, İ., Karköse, M., Akın, E. (2005). *Bulanık mantık tabanlı zaman serisi veri madenciliği kullanan tahmin algoritması*, II. Mühendislik Bilimleri Genç Araştırmacılar Kongresi, 47-56

Azadeh, O., Seraj, M. (2011). *An Integrated Fuzzy Regression–Analysis of Variance Algorithm For Improvement Of Electricity Consumption Estimation In Uncertain Environments*, Int J AdvManuf Technol, vol. 53, pp. 645–660

Azadeh, A., Sohrabi, P., Ebrahimipour, V. (2013). *A fuzzy regression approach for improvement of consumption estimation with uncertain data*, Int. J. Industrial and systems Enginneering, Vol.13, No.1, pp. 92-109

Boreux, J. J., Pesti, G., Duckstein, L., Nicolas, J. (1997). *Age model estimation in paleoclimatic research: fuzzy regression and radiocarbon uncertainties*, Palaeo Palaeogeography, Palaeoclimatology, Palaeoecology, 128(1-4): 29-37

Chang, P. T., Lee, E. S. (1996). *A generalized fuzzy weighted least-squares regression, Fuzzy Sets and Systems*, 82(3): 289-298

Chang, P. T. (1997). *Fuzzy seasionality forecasting*, Fuzzy Sets and Systems, 90(1): 1-10

Chang, Y. H. O., Ayyub, B. M. (2001). *Fuzzy regression methods – a comparative assessment*, Fuzzy Sets and Systems, 119(2): 187-203

Chen, K., Jia, Y., Lee, S. (2006). *Fuzzy adaptive networks in thermal comfort*, Applied Mathematics Letters, 19(5): 420-426

Cundiff, E.W., Still, R.R. (1971). *Basic Marketing: Concept, Decisions and Strategy*, p. 68

Douglas, E. J. (1987). *Managerial Economics: Analysis and Strategy*

D'Urso, P., Santoro, A. (2006). *Fuzzy clusterwise linear regression analysis with symmetrical fuzzy output variable*, Computational Statistics & Data Analysis, 51(1): 287-313

Fontela, E., Gobus, A. (1976). *The DEMATEL observer*, Battelle Geneva Research Center

Gaines, B. R., Kohout, L. J. (1977). *The fuzzy decade: a bibliography of fuzzy systems and closely related topics*, Int. J. Man-Machine Studies, 9: 1-68

Ge, H., Wang, S. (2007). *Dependency between degree of fit and input noise in fuzzy linear regression using non-symmetric fuzzy triangular coefficients*, Fuzzy Sets and Systems, 158(19) : 2189-2202

He, Y. Q., Chan, L. K., Wu, M. L. (2007). *Balancing productivity and consumer satisfaction for profitability: statistical and fuzzy regression analysis*, European Journal of Operational Resarch, 176(1): 252-263

Hojati, M., Bector, C. R., Smimou, K. (2005). *A simple method for computation of fuzzy linear regression*, European Journal of Operational Research, 166(1): 172–184

Hong, D. H., Hwang C., Ahn C. (2004). *Ridge estimation for regression models with crisp inputs and gaussian fuzzy output*, Fuzzy Sets and Systems, 142 (2): 307-319

Hsu, C. H., Wang, F. K., Tzeng G. H. (2012). *The best vendor selection for conducting the recycled material based on a hybrid MCDM model combining DANP with VIKOR*, Resources, Conservation and Recycling, vol. 66, pp.95-111

Hung, W. L., Yang, M. S. (2006). *An omission approach for detecting outliers in fuzzy regression models*, Fuzzy Sets and Systems, 157(23): 3109- 3122

Huang, C., Tzeng, G. (2008). *Multiple generation product life cycle predictions using a novel two-stage fuzzy piecewise regression analysis method*, Technological Forecasting and Social Change, 75(1) : 12-31

Ishibuchi, H., Tanaka, H. (1992). *Fuzzy regression analysis using neural networks*, Fuzzy Sets and Systems, 50(3): 257-265

Ishibuchi, H., Manabu, N. (2001). *Fuzzy regression using asymmetric fuzzy coefficient and fuzzified neural Networks*, Fuzzy Sets and Systems, 119 (2): 273-290

Kandari, A. M., Soliman, S. A., Alammari, R. A. (2006). *Power quality analysis based on fuzzy estimation algorithm: Voltage flicker measurements*, Electricial Power & Energy Systems, 28(10): 723-728

Kaufmann, A. (1975). *Introduction to the theory of Fuzzy Sub-sets*, Academic Pres, 495- 496

Kendall, A., Yager, A. (1979). *Bibliographs of fuzzy sets, their applications and related topics*, Advances in Fuzzy Set Theory and Applications, 621-744

Kim, K.J., Moskowitz, H. and Koksalan, M. (1996). *Fuzzy versus statistical linear regression*, European Journal of Operational Research, 92, 417-434.

Kutay, F. (1989). *Zaman Serilerinde Tahmin Teknikleri ve Box-Jenkis Modelleri*, 1-99

Kwong, C. K., Bai, H. (2005). *Fuzzy regression approach to process modelling and optimization of epoxy dispensing*, International Journal of Production Research, 2359 – 2375

Lee, H. T., Chen, S. H. (2001) *Fuzzy regression model with fuzzy input and output data for manpower forecasting*, Fuzzy Sets and Systems, 119(2): 205-213

Lee, H. S., Chou, M. T. (2004). *Fuzzy time-series based on adaptive expectation model for TAIEX forecasting*, International Journal of Computer Mathematics, 34(2): 781- 789

Moskowitz, H., Kim, K. (1993). *On assessing the h value in fuzzy linear regression*, Fuzzy Sets and Systems 58(3): 303-327

Muzzioli, S., Ruggier, A., Debaets, B. (2015). *A Comparison of Fuzzy Regression Methods For The Estimation Of The Implied Volatility Smile Function*, Fuzzy Sets And Systems, vol.266, pp. 131–143

Nasrabadi, M. M., Nasrabadi, E., Narabady, A. R. (2005). *Fuzzy linear regression analysis: a multi-objective programming approach*, Applied Mathematics and Computation, 163(1): 245-251

Öğüt, K. S. (2006). *Modelling car ownership in turkey using fuzzy regression*, Transportation Planning and Technology, 29:33, 233-248

Özelkan, E. C., Duckstein, L. (2000). *Multi-objective fuzzy regression: a general framework*, Computers & Operations Research, 27(7-8): 635-652

S'anchez, J. A. (2007). *Claim reserving with fuzzy regression and taylor's geometric separation method*, Insurance Mathematics&Economics, 40(1): 145-163

Savic, D., Pedrycz, D. (1991). *Evaluation of fuzzy regression models*, Fuzzy Sets and Systems 39(1): 51-63

Stahl, C. (2006). *A strong consistent least-squares estimator in a linear fuzzy regression model with fuzzy parameters and fuzzy dependent variables*, Fuzzy Sets and Systems, 157(19): 2593 – 2607

Şen, Z. (2001). *Bulanık Mantık ve Modelleme İlkeleri*, Bilge Kültür Sanat Yayınevi, 9- 27,

Tanaka, H., Uejima, S., Asai, K. (1982). *Linear regression analysis with fuzzy model*, IEEE, Trans, System Management, Sybernet, 12(6): 903 -907

Tran, L., Duckstein, L. (2002). *Multiobjective fuzzy regression with central tendency and possibilistic properties*, Fuzzy Sets and Systems, 130(1): 21-31

Tsaur, R.C., Wang, H.F., Yang, J.C. (2002). *Fuzzy regression for seasonal time series analysis*, International Journal Of Information Technology&Decision Making, 165-175

Tseng, F.M., Tzeng, G.H. (2002). *A fuzzy seasonal ARIMA model for forecasting*, Fuzzy Sets and Systems, 126 (3): 367-376

Tzeng, G. H., Chen, W. H., Yu, R., Shih, M. L. (2010). *Fuzzy decision maps: a generalization of the DEMATEL methods*, Soft Computing, vol. 14, pp. 1141-1150

Uemura, Y. (1996). *Evaluation of physical tests by fuzzy regression analysis*, Cybernetics and Systems, 27:43

Wang, F. H., Tsaur, R. C. (2000). *Insight of a fuzzy regression model*, Fuzzy Set and Systems, 355-369

Wang, N., Zhang, W., Mei, C. (2007). *Fuzzy nonparametric regression based on local linear smoothing technique*, Information Sciences, 177(18):3882-3900

Wang, C.H., Chen, J. N. (2012). *Using quality function deployment for collaborative product design and optimal selection of module mix*, Computers & Industrial Engineering, vol. 63, pp. 1030-1037

Wu, B., Tseng, N.F. (2002). *A new approach to fuzzy regression models with application to business cycle analysis*, Fuzzy Sets and System, 130 (1): 33-42

Yang, M.S., Ko, C.H. (1997). *On cluster-wise fuzzy regression analysis, IEE Transaction on Systems*, Man and Cybernetics Part B: Cybernetics, 27(1):1

Yang, M.S., Lin, T.S. (2002). *Fuzzy least squares regression analysis for fuzzy inputoutput data*, Fuzzy Sets and Systems, 126 (3): 389-399

Yang, M.S., Liu, H.H. (2003). *Fuzzy least-squares algorithms for interactive fuzzy linear regression models*, Fuzzy Sets and Systems, 135(2): 305–316

Yang, J. L., Tzeng, G. H. (2011). *An integrated MCDM technique combined with DEMATEL for a novel cluster-weighted with ANP method*, Expert Systems with Applications, vol. 38, pp. 1417- 1424

Yoon, H. S., Choi, S. H. (2019). *The Impact on Life Satisfaction of Nursing Students Using the Fuzzy Regression Model*, International Journal of Fuzzy Logic and Intelligent Systems 2019; 19(2): 59-66

Zadeh, L.A. (1962). *From circuit theory to system theory proceedings of institue of radio engineering*, Academic Press, Newyork, 856-865

Zadeh, L.A. (1965). *Fuzzy sets information and control*, Academic Press, Newyork, 8: 338-353

Zimmerman, H.J. (1976). *Description and optimization of fuzzy systems*, International J. General Systems, 4

BIOGRAPHICAL SKETCH

Berkay AYDIN was born on March 12, 1991 in Istanbul and now currently living in Küçükyalı, İSTANBUL.

He took his primary education at the Göztepe Primary School. Berkay completed his high school education at Galatasaray High School. In 2014, he graduated from Galatasaray University, Department of Industrial Engineering.

Since 2015 he has been working as a strategic planning and project specialist at the leading paint company of Turkey, BETEK BOYA VE KİMYA SANAYİ A.Ş.

His paper, A Combined DEMATEL and Fuzzy Regression Approach for Sales Forecasting, has been accepted by the 2019 8th International Conference on Economics and Finance Research (ICEFR 2019).