PREDICTION OF THE HOUSE PRICE INDEX OF TURKEY: A COMPARATIVE STUDY OF MULTIPLE LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YUSUF KEMAL ERDEKLİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

JUNE 2022

Approval of the thesis:

PREDICTION OF THE HOUSE PRICE INDEX OF TURKEY: A COMPARATIVE STUDY OF MULTIPLE LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS

submitted by **YUSUF KEMAL ERDEKLİ** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar

Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Erdem Canbay

Head of the Department, **Civil Engineering**

Prof. Dr. İrem Dikmen Toker

Supervisor, **Civil Engineering, METU**

Prof. Dr. M. Talat Birgönül

Co-Supervisor, **Civil Engineering, METU**

Examining Committee Members:

Prof. Dr. Rifat Sönmez

Civil Engineering, METU

Prof. Dr. İrem Dikmen Toker

Civil Engineering, METU

Prof. Dr. M. Talat Birgönül

Civil Engineering, METU

Asst. Prof. Dr. Güzide Atasoy Özcan

Civil Engineering, METU

Asst. Prof. Dr. Gözde Bilgin

Civil Engineering, Başkent University

Date: 16.06.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Yusuf Kemal Erdekli

Signature:

ABSTRACT

PREDICTION OF THE HOUSE PRICE INDEX OF TURKEY: A COMPARATIVE STUDY OF MULTIPLE LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS

Erdekli, Yusuf Kemal
Master of Science, Civil Engineering
Supervisor: Prof. Dr. İrem Dikmen Toker
Co-Supervisor: Prof. Dr. M. Talat Birgönül

June 2022, 96 pages

The housing market is accepted as the triggering force for the Turkish economy. Consequently, the fluctuation in house prices is an important point of interest for both policymakers in the government and decision-makers in the construction industry. In order to be able to monitor the movements in house prices effectively, it is important to understand the relationships between the macroeconomic parameters and house price movements. The main objective of this thesis is to predict the house price index (HPI) of Turkey using macroeconomic parameters by utilizing multiple linear regression (MLR) and artificial neural network (ANN) models. Among the 25 macroeconomic parameters commonly used in previous studies reported in the literature, 9 parameters are selected as the independent variables while predicting the house price index (HPI), which is chosen as the dependent variable, and the monthly time series data of these parameters are used for the time period between January 2010 and December 2019 to train the models. Subsequently, data from January 2020 to December 2021 is used to validate the models. The relationship between these macroeconomic parameters and house price movements is examined in detail, and consumer price index (CPI), unemployment rate (UR), and brent oil price (BOP) are found as the most significant parameters in determining the house price movements in Turkey. The developed MLR and ANN models provided a high level of accuracy in learning, generalizing, and converging the time series data and produced reliable prediction values. Model validation results revealed that the ANN model developed in this thesis has a higher predictive performance than the developed MLR model and also the ANN models developed in previous studies.

Keywords: House Prices, Turkey, Macroeconomic Parameters, Multiple Linear Regression (MLR), Artificial Neural Networks (ANN)

ÖZ

TÜRKİYE KONUT FİYAT ENDEKSİNİN TAHMİNİ: ÇOKLU LİNEER REGRESYON VE YAPAY SİNİR AĞI MODELLERİNİN KARŞILAŞTIRILMASI

Erdekli, Yusuf Kemal
Yüksek Lisans, İnşaat Mühendisliği
Tez Yöneticisi: Prof. Dr. İrem Dikmen Toker
Ortak Tez Yöneticisi: Prof. Dr. M. Talat Birgönül

Haziran 2022, 96 sayfa

Konut piyasası, Türkiye ekonomisi için tetikleyici güç olarak kabul edilmektedir. Buna bağlı olarak, konut fiyatlarındaki dalgalanma hem politikacılar hem de inşaat sektöründeki karar vericiler için önemli bir ilgi noktasıdır. Konut fiyatlarındaki değişimleri etkin bir şekilde izleyebilmek için makroekonomik parametreler ile konut fiyat hareketleri arasındaki ilişkileri anlamak önemlidir. Bu tezin temel amacı, makroekonomik parametreler kullanarak geliştirilen çoklu doğrusal regresyon (ÇDR) ve yapay sinir ağı (YSA) modelleri ile Türkiye'nin konut fiyat endeksini (KFE) tahmin etmektir. Literatürde daha önce yapılan çalışmalarda yaygın olarak kullanılan 25 makroekonomik parametre arasından 9 parametre bağımsız değişken olarak seçilmiş ve bağımlı değişken olarak seçilen konut fiyat endeksi (KFE) tahmin edilmeye çalışılmıştır. Bu çalışmada, modellerin eğitilmesi için seçilen parametrelerin Ocak 2010 ile Aralık 2019 aralığındaki aylık zaman serisi verileri kullanılmıştır. Ardından, modelleri doğrulamak için Ocak 2020'den Aralık 2021'e kadar olan veriler kullanılmıştır. Seçilen makroekonomik parametreler ile konut fiyat hareketleri arasındaki ilişki detaylı olarak incelenmiş ve tüketici fiyat endeksi

(TÜFE), işsizlik oranı (İO) ve Brent petrol fiyatı (BPF), Türkiye'deki konut fiyat hareketlerini belirlemede en önemli parametreler olarak bulunmuştur. Geliştirilen ÇDR ve YSA modelleri, zaman serisi verilerinin öğrenilmesinde, genelleştirilmesinde ve birleştirilmesinde yüksek düzeyde doğruluk sağlamış ve güvenilir tahmin değerleri üretmiştir. Model doğrulama sonuçları, bu tezde geliştirilen YSA modelinin, geliştirilen ÇDR modeline ve ayrıca önceki çalışmalarda geliştirilen YSA modellerine göre daha yüksek bir tahmin performansına sahip olduğunu ortaya koymuştur.

Anahtar Kelimeler: Konut Fiyatları, Türkiye, Makroekonomik Parametreler, Çoklu Doğrulsal Regresyon (ÇDR), Yapay Sinir Ağları (YSA)

Dedicated to my beloved family…

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my supervisor, Prof. Dr. İrem Dikmen Toker, and co-supervisor, Prof. Dr. M. Talat Birgönül, for the guidance, advice, criticism, encouragement, and motivation they provided throughout my thesis study. It was impossible for me to complete this thesis without their patience and support.

I would also like to thank my colleague Mr. Murat Altun for the valuable effort and time he spent for suggestions, comments, and technical assistance on my thesis.

I would like to express my special thanks to my father, Yaşar Mehmet Erdekli, and my mother, Hatice Erdekli, for bringing me where I am today and to my sister Elif Erdekli who always supported and motivated me during my thesis study.

Lastly, I wish to express my special gratitude to my deceased grandmother Fatma Erdekli for the provision of good times throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

ANN Artificial Neural Networks

ARDL Autoregressive Distributed Lag

ARMA Autoregressive Moving Average

ARIMA Autoregressive Integrated Moving Average

ART Adaptive Resonance Theory

BCI Building Cost Index

BMA Bayesian Model Averaging

BOP Brent Oil Price

BPF Bayesian Probability Framework

BPN Backpropagation Neural Network

BRSA Banking Regulation and Supervision Agency

CBOE Chicago Board Options Exchange's

CBRT Central Bank of the Republic of Turkey

CCI Consumer Confidence Index

CLI Cost of Living Index

CMPI Construction Materials Price Index

CPI Consumer Price Index

CTI Construction Turover Index

DMA Dynamic Model Averaging

DMS   Dynamic Model Selection

DSMI   Domestic Stock Market Index

DPPI   Domestic Producer Price Index

DXY   U.S. Dollar Index

EEMD   Ensemble Empirical Mode Decomposition

EER   Effective Exchange Rate

EIR   Effective Interest Rate

ER   Employment Rate

ES   Exponential Smoothing

ExSS   Explained Sum of Squares

EW   Equal-Weighted Averaging

GARCH   Generalized AutoRegressive Conditional Heteroskedasticity

GDP   Gross Domestic Product

GLM   Generalized Linear Models

GPI   Gold Price Index

GVA   Gross Value Added

HPI   House Price Index

ICC   Istanbul Chamber of Commerce

IPI   Industrial Production Index

IR   Inflation Rate

ITMA   Information-Theoretic Model Averaging

LF   Labour Force

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MCS Model Confidence Set

ML Machine Learning

MIR Mortgage Interest Rate

MLR Multiple Linear Regression

MS Money Supply

MSE Mean Squared Error

NARDL Nonlinear Autoregressive Distributed Lag

NHP Number of Housing Permits

NPE Number of Paid Employees

NPPI Non-Domestic Producer Price Index

OLS Ordinary Least Squares

POP Population

RBF Radial Basis Function Neural Network

ReLU Rectified Linear Units

RMSE Root Mean Squared Error

PPPI Residential Property Price Index

RSS Residual Sum of Squares

RW Random Walk

SPX S&P 500 Index

SVR Support Vector Regression

TURKSTAT Turkish Statistical Institute

TRY Turkish Lira

TSS Total Sum of Squares

UR Unemployment Rate

USD U.S. Dollars

VAR Vector Autoregression

VEC Vector Error Correction

VHL Volume of Housing Loans

VIX Volatility Index

WOA Whale Optimization Algorithm

CHAPTER 1

INTRODUCTION

Housing can be considered as one of the fundamental necessities of a human being, and it has a vital role in the determination of life quality. It can be defined as a product manufactured for human consumption that satisfies the needs for shelter, safety, comfort, socializing, and aesthetics (McDonald & McMillen, 2010). Housing also serves as an important and traditional investment tool for households. Although housing is a relatively illiquid investment option, it has a large portion in most individuals' investment portfolios, mainly in developed countries. Consequently, investments in housing have a significant impact on both individuals' consumption and saving habits.

Housing has essential features that make it different from the other markets of goods and services. These features include durability, spatial fixity, and heterogeneity. Durability implies that housing services have very long service lives. The demand for the new housing services determines the supply rate of new housing construction and affects the selling prices of newly constructed housing units. Spatial fixity implies that the location of a dwelling unit is fixed, and it cannot be changed by the user. Furthermore, the differences in the design, age, size, landscaping, location, and tax expenditure of the dwellings units exchanged in the housing market imply their heterogeneity (Smith, Rosen, & Fallis, 1988).

The housing market is not just a single market but also a collection of submarkets, according to its composite nature. These submarkets differentiate in housing type, quality, lifetime, location, financing, etc. In some markets, housing services are exchanged as consumption goods. In other markets, the housing stock is exchanged as an asset. On a local basis, the distinctions in the housing market include ownership and rental housing, the existence of single-family and multiple-family dwellings, second-hand housing sales, newly constructed housing sales, etc.

Housing prices are relatively higher than most other types of assets. Therefore, housing sales are considered as an important economic activity. Due to the high volume of economic activity it generates, the housing market is one of the key concerns for many parties, including the construction industry, real estate industry, banking sector, transportation sector, insurance sector, governmental institutions, etc.

In both developed and developing countries, the housing market plays a vital role in the national economy. Purchasing a housing unit is the largest single customer transaction that the vast majority of US citizens make. The Europeans and the Turkish citizens act similarly. Also, it constitutes a significant portion of wealth for most of these householders. Because the housing industry is one of the economic drivers in both developed and developing nations, occurrences in the housing market may have a ripple effect across the economy. The most famous example to explain this situation is the 2008 global financial crisis which arose from the unexpected growth of the subprime mortgage loans in the housing market. The collapse in the financial markets started in the United States with the bankruptcy of Lehman Brothers Inc., which was one of the greatest investment banks in the US, and spread all over the world just like an avalanche effect (Helleiner, 2011).

The housing market is vital for the emerging economies in which income and wealth levels show rapid increases. Turkey is an exciting example of this category. The volume of housing and real estate economy is constantly growing in Turkey, and the housing market is accepted as a triggering force for the Turkish economy. Throughout its history, the country has been plagued by a number of economic instabilities, both implicit and apparent. In this regard, the housing market's actual price appreciations have established a widespread perception among industry insiders and the general public of a high pace of upward price movements, as well as possible risk accumulations (Coskun et al., 2020).

The House Price Index (HPI), published by the Central Bank of the Republic of Turkey (CBRT), monitors the monthly price movements in the Turkish housing market. The index measures the quality-adjusted price changes of dwellings. In Figure 1.1. and Figure 1.2., the monthly movements in the house prices of Turkey and the three major cities (İstanbul, Ankara, and İzmir) for the period of January 2015 to March 2022 are monitored, respectively.

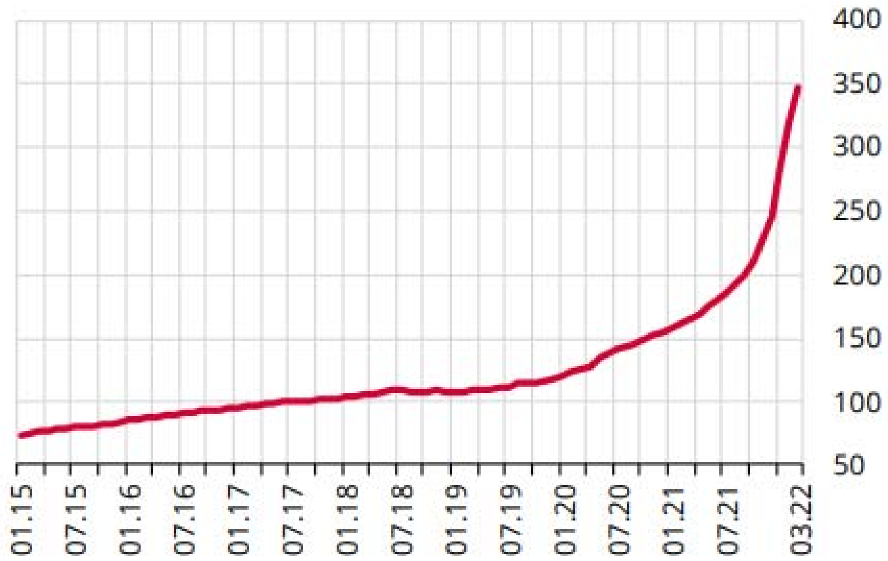

Figure 1.1. House Price Index (HPI) of Turkey (2017=100)
(Central Bank of the Republic of Turkey, 2022)

In March 2022, the HPI of Turkey increased monthly by 9.3%, recording an annual increase of 110% in nominal terms. The monthly growth of the HPI by cities was recorded as 9.8%, 9.6%, and 10.6% for İstanbul, Ankara, and İzmir, respectively, and the recorded annual increases were 122%, 111.7%, and 105.9 % in nominal terms.

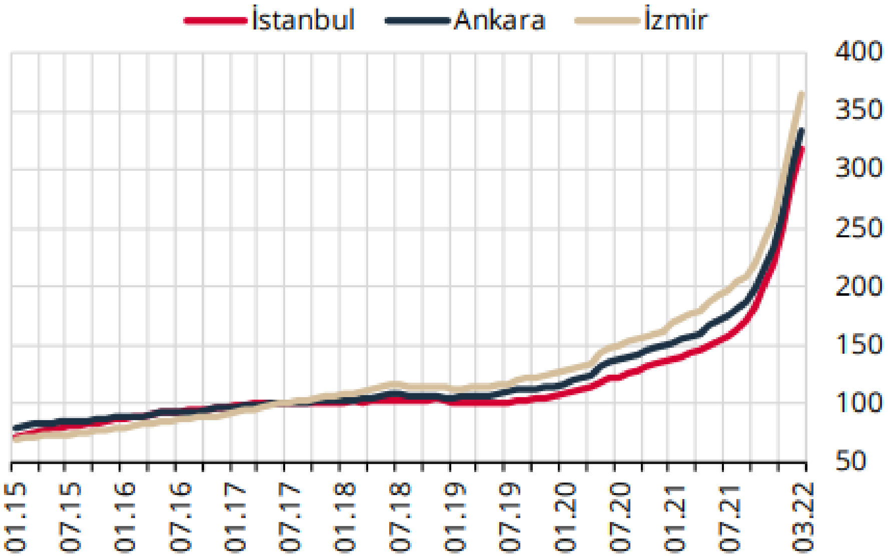

Figure 1.2. House Price Index (HPI) of Three Major Cities of Turkey (2017=100) (Central Bank of the Republic of Turkey, 2022)

Many scholars have been drawn to the housing market's price movements to explore the factors influencing housing prices in Turkey and throughout the world. Various researches on the drivers of house prices have discovered that macroeconomic and demographic variables impact housing prices. However, the same variables may not give the same results in different countries since the countries' economic, demographic, and social structures are rarely similar (Meidani, Zabihi, & Ashena, 2011).

1.1 Research Objectives and Design

This thesis focuses on predicting house prices from the movements of macroeconomic variables. The primary aims of this thesis are; to predict the house price index (HPI) of Turkey using the selected macroeconomic variables by employing multiple linear regression (MLR) and artificial neural network (ANN) models and compare the predictive performance of these two models. Due to the MLR model's ease of implementation, it is chosen to be employed in this study. The

ANN model, on the other hand, is chosen since it produces forecasts with a high level of accuracy utilizing time-series data. In Figure 1.3., the research steps of the thesis are demonstrated as a flowchart.

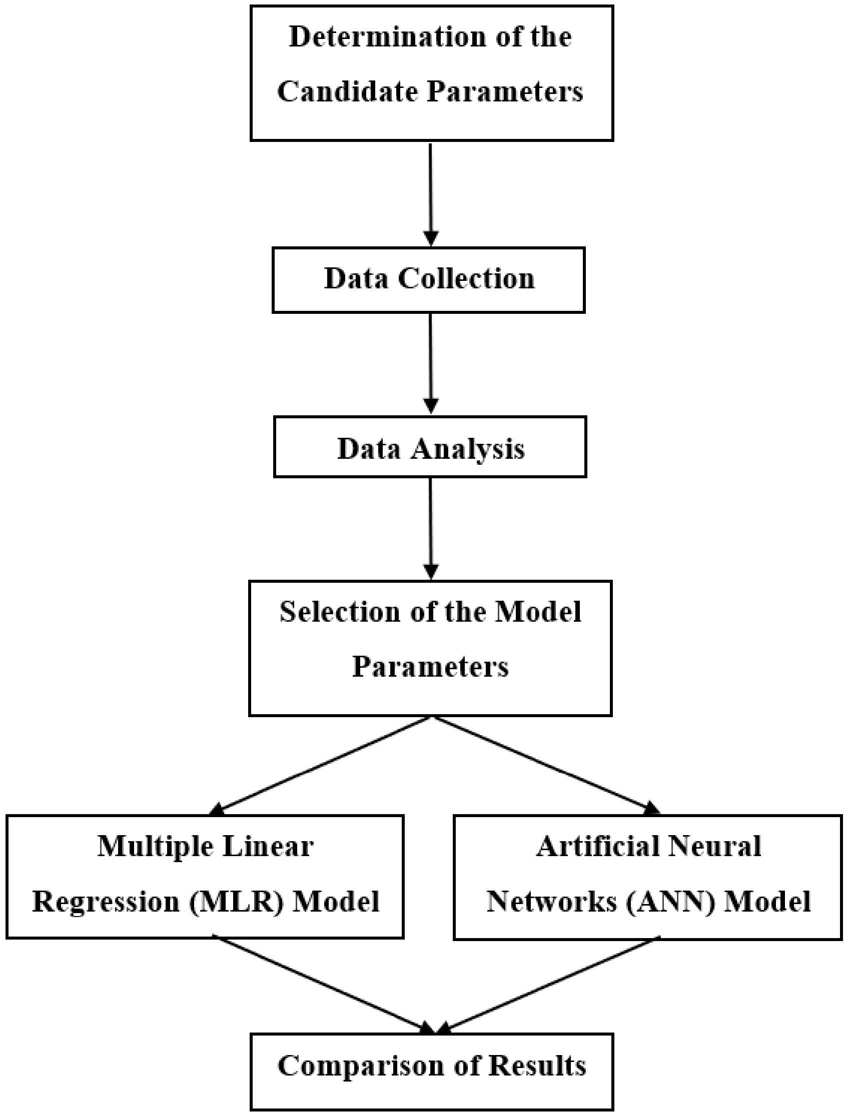

Figure 1.3. Research Methodology

1.2 The Organization of the Thesis

In Chapter 2, a literature review of studies on the macroeconomic determinants of housing prices in Turkey and throughout the world is provided. These reference studies are used to determine the candidate parameters for house price index forecasting models. Chapter 3 focuses on data collection, data analysis, and selection of the model parameters that will be employed in HPI prediction models. In Chapters 4 and 5, the multiple linear regression model and the artificial neural networks model utilized to predict the HPI are presented in detail. The findings are discussed in Chapter 6, which includes a comparison of the two models' prediction abilities. Finally, the conclusions are presented in Chapter 7.

CHAPTER 2

LITERATURE REVIEW

The determinants of house prices have been investigated in a variety of academic studies, mainly in the last few decades. These studies usually focused on the effect of macroeconomic parameters on house price movements. The parameters that have a major impact on house price movements investigated in these studies include the exchange rate, housing loan interest rate, employment rate, oil prices, consumer price index, inflation, construction material prices, gold prices, stock market index, etc. In this chapter, the empirical studies in the context of determining factors of house prices conducted in Turkey and all over the world are reviewed.

2.1 Empirical Studies that Explore and Forecast House Prices

Previous studies are listed chronologically as below:

In their early study, Case and Shiller (1990) predicted the house price movements and excess returns of housing on line with dept over the following year using a method called time-series-cross-section regression. Their time series data include a set of forecasting variables from four different cities in the United States, namely Atlanta, Chicago, Dallas, and San Francisco. The results of their study indicated that there exist positive relations between the house price movements and the parameters, including the cost of construction, the national income per capita, and the working-age population.

Case and Mayer (1996) analyzed the house price appreciation in the Boston area of the United States from 1982 to 1994 using the regression method. The evidence found in the study suggests that the changes in employment rate, population, closeness to the city center, and the number of educational facilities are related to the cross-sectional movements in the house prices.

Stevenson (2000) examined the effect of inflation on the movements of residential property prices of regional markets in the United Kingdom over a time period of 30 years. The conventional ordinary least squares models are used in this study together with causality and cointegration test. The results of the OLS tests did not provide significant evidence showing a stable and consistent relationship, and large fluctuations are observed in the results of studied regions. The results of the Engle-Granger cointegration test supported the assumption that there exists a high correlation between the housing price movements and changes in the rate of inflation. Also, the results of the causation test indicated that the increases in house prices caused the rise of the inflation rate in the United Kingdom.

By collecting the data from 130 different regions in the United States, Jud and Winkler (2002) investigated the determinants of real housing price appreciation by utilizing an MSA fixed-effects model. The findings of their study show that there is a strong correlation between the real housing price movements and changes in population, national income per capita, real effective interest rates, and construction costs. Their study also concluded that the movements in the stock market induce a strong real-time and lagged impact on house prices. They also concluded that the magnitudes of the fixed effects in certain regions have a positive correlation with restrictive policies of growth management and restrictions on the availability of lands. House price movement rates are also found to be different in various regional areas as a result of the existence of the fixed effects specific to location that represent the availability of residuals in housing price movements related to location characteristics.

In their study, McGibany and Farrokh (2004) analyzed the effect of changes in the mortgage interest rates on housing price movements in the long and short term in the United States using the Johansen and Juselius maximum likelihood framework and vector error correction (VEC) model. In line with their expectations, the results of the multivariable specification of housing demand revealed that there exists a long-term relationship between the changes in the mortgage interest rates on housing price movements, and this is a relatively inelastic response. On the other hand, the results of the Granger causality test revealed no short-term relationship existing between the house price movements and changes in mortgage interest rates.

In her master's thesis study conducted at Massachusetts Institute of Technlogy, Padilla (2005) aimed to answer to which extent the oil prices and other economic parameters can be used in the prediction of the changes in housing prices in the Calgary region of Canada and to determine what the lag time is between them. Using regression analysis, she found an excellent correlation and suggested that the oil prices, exchange rate, interest rate, and employment levels can determine up to 98% of the changes in house prices and rents. Oil prices, representing the economic viability of the city, affect the real estate industry with a lag of 7 quarters of approximately two years, while interest rates representing the financial well-being of the city affect house prices and rents with a lag of 2.5 years. The foreign exchange rate to the dollar, representing the relative global prosperity, affects real estate prices in one year.

Abelson et al. (2005) aimed to explain real house price changes in Australia in the time period between 1970 and 2003. To estimate the housing price movements in the long term, they developed a long-term equilibrium model and estimated the significant macroeconomic parameters affecting the movements of house prices. They also developed a short-term asymmetric error correction model that estimates the changes in house prices in the short term. The results of their study revealed that the consumer price index and the real disposable income could determine the real house price movements significantly and with a positive correlation in the long term. They also concluded that mortgage rates, stock market prices, unemployment rate, and housing stocks could determine the real house prices to a considerable extent but with a negative correlation in the long term.

In the study of Rapach and Strauss (2007), real housing price movements in the eight different regions of the United States (namely Indiana, Mississippi, Arkansas, Illinois, Missouri, Kentucky, and Tennessee) are investigated. They first evaluated the forecasting ability of various candidate predictors of the real housing price movements by applying the autoregressive distributed lag (ARDL) model. Particular variables, including the ratio of housing prices to income, the rate of unemployment, and the rate of inflation, are found to be useful for forecasting the real housing price movements in the eight different regions. The authors also applied different methods for combining the forecasts of individual ARDL models. They concluded that the combination of forecasts of different models was quite successful in producing accurate forecasts of real housing price movements in the eight different regions of the United States.

In their succeding study, Rapach and Strauss (2009) aimed to investigate the forecasts of the real housing price movements at the state level in the United States for the time period of 1995–2006. They evaluated forecasts taken from an autoregressive benchmark model and also the results of the models based on several national, regional, and host of state economic parameters. Their findings highlighted considerable differences in the real house price movements forecasting models over the United States, particularly in the coastal and interior states. More importantly, their findings showed that the autoregressive models, especially the models that use the data of various economic parameters, usually give relatively accurate forecasts of housing price movements for the assorted number of interior states during the selected study period.

Mikhed and Zemčík (2009) studied whether basic parameters such as housing rent, stock market prices, personal income, population, cost of buildings, and mortgage interest rate might determine the current high and subsequently swiftly declining U.S. housing prices. They started with aggregate data and ran standard unit root and cointegration tests. In order to avoid cross-sectional dependency, panel data stationarity tests were employed in the study. They concluded that panel data unit root tests offer more power than univariate testing, according to their findings.

In their global research, (Beltratti & Morana, 2010) used the factor vector autoregressive (VAR) model to examine the connections between general macroeconomic conditions and the housing market for the G-7 region. One of the primary results of the research is that the US is a substantial source of global changes, not just for real activity, nominal variables, and stock prices, but also for real housing prices. It has been demonstrated that fluctuations in G-7 housing prices are significantly predicted by global supply-side shocks in particular. Real housing prices and macroeconomic developments are shown to be correlated, although the link is bidirectional, with investment responding more strongly to housing price shocks than consumption and output.

Adams & Füss (2010) investigated the long-term influence and short-term trends of macroeconomic factors on global housing prices in another international research. They used regression and panel cointegration analysis on data from 15 nations over a 30-year period since suitable housing market data is often unavailable and of low frequency. The empirical results show that a 1 percent increase in economic activity causes a long-term rise in home prices of 0.6 percent, whereas average long-term impacts of building costs and long-term interest rates are 0.6 percent and 0.3 percent, respectively.

In a similar international study, Agnello and Schuknecht (2011) examined the characteristics and causes of housing price booms and busts for a group of eighteen industrialized nations from 1980 to 2007. From a historical standpoint, they discovered that current housing booms are among the longest in the last four decades. Domestic credit and interest rates, according to estimates from a Multinomial Probit model, have a considerable impact on the likelihood of booms and busts. They also discovered that financial market deregulation has greatly amplified the influence of the domestic financial sector on the incidence of booms.

Radzi et al. (2012) used an artificial neural network (ANN) model to predict residential property price values in Malaysia in their research. The individual variables were the rate of unemployment, size of the population, mortgage rate, and income level, whereas the dependent variable was the Housing Price Index (HPI). The mean absolute percentage error (MAPE) for model validation was 8%, indicating that ANN has a high degree of accuracy in its capacity to learn, generalize, and integrate time series data effectively and create accurate forecasting information.

To facilitate decision-making, Lai et al. (2013) provided a risk-based, cost-benefit analytical approach. Their method combines life cycle costing, engineering reliability tests, and forecasting using autoregression. Using historical trends with gross disposable income per capita, and the consumer price index, three series of home prices are predicted to emulate changing situations in the Melbourne, Australia, housing market. From the standpoint of an investor, it was discovered that if an asset is kept for more than ten years, there is a less than 1% chance of investment loss. Furthermore, if an asset is held for five years, a market collapse increases the likelihood of loss to 39%. In all simulated property market scenarios, increasing cost fluctuations from about 5% to 30% has a negative impact on the chance of loss. The Hasofer-Lind approach was found to be a better alternative to the more computationally demanding Monte Carlo simulation in terms of efficiency.

Hanewald and Sherris (2013) developed and examined residential property price models for banking and insurance applications, including pricing, risk management, and portfolio management. Their investigation concentrated on postcode-level housing prices in Sydney, Australia, between 1979 and 2011. Results of single-factor panel data models indicate that the market-wide housing price index explains approximately 42% and 44% of the longitudinal and cross-sectional variation in postcode-level house price increase rates. In the pricing and risk management of products exposed to house price risk, the significance of macro-financial factors, as well as geographic and sociodemographic postcode characteristics, has been established.

Ciarlone (2015) examined the features of housing price trends for a sample of 16 emerging economies from Asia and Central and Eastern Europe from 1995 to 2011 in his worldwide study. He derived results on the existence and nature of house price outperformance using regression analysis using the conventional OLS approach. According to the findings of his research, housing markets in 16 emerging nations have rarely shown major indicators of overvaluation. On the other hand, the boom period, which was marked by low global interest rates, high levels of global liquidity, and unusually low levels of risk aversion, was only partially justified by improved fundamentals and may have been fueled by extremely optimistic expectations of the part of economic agents.

A housing price prediction model was created by Park and Bae (2015), utilizing machine learning algorithms as a research method. To improve the accuracy of the property price forecast, they examined the housing data for 5359 townhouses in Virginia, USA. Then built a house price prediction model using C4.5, RIPPER, Nave Bayesian, and AdaBoost machine learning algorithms, and they compared the classification accuracy of the results. They recommended a more accurate housing price prediction algorithm to enable real estate agents or home sellers to make better decisions based on estimated property prices. Their research showed that when it comes to predicting property prices, the RIPPER algorithm consistently outperforms the other models in terms of accuracy.

Nyakabawo et al. (2015) examined the causal relationships between the real home price index and real GDP per capita in the United States using the bootstrap Granger (temporal) non-causality test and a fixed-size rolling-window estimation approach. The real house price index and real GDP per capita have a unidirectional relationship, according to the findings of the full-sample bootstrap non-Granger causality test. According to the study, full-sample causality tests are insufficient; hence a time-varying (bootstrap) rolling-window approach must be used to examine the causal relationship between these two variables. The results indicated that real GDP per capita has significant positive feedback on real home prices, even if real home prices typically follow real GDP per capita (during expansions and recessions).

In their paper, Plakandaras et al. (2015) proposed a unique hybrid forecasting method that combines Ensemble Empirical Mode Decomposition (EEMD) from signal processing with Support Vector Regression (SVR) with a machine learning background. They compared the suggested model with a Random Walk (RW), a Bayesian Autoregressive, and a Bayesian Vector Autoregressive model for the US housing price index. They asserted that this cutting-edge method might be used as a governmental early warning system for anticipating unanticipated drops in home prices.

Using the autoregressive model, Dougherty and Order (2016) explored the connection between inflation, housing costs, and the consumer price index (CPI) in the United States. They concentrated on laying a theoretical framework for accurately estimating the cost of housing before presenting some estimates of how these measurements might have influenced inflation in the 1970s, as assessed by the Consumer Price Index (CPI). They indicated that the optimal measure of property ownership cost differs considerably from the CPI. According to their calculations, between 15 to 25% of the price increase since 1968, as indicated by the CPI, might be fictitious.

Li and Chu (2017) explored the effects of macroeconomic variables on real estate price variation and developed a price volatility prediction model in their research. They used backpropagation neural network (BPN) and radial basis function neural network (RBF) models to create a nonlinear model to forecast real estate price variations in Taipei, Taiwan, using leading and synchronous economic parameters. They discovered that the RBF neural network outperforms the BPF neural network in predicting the Cathay house price index fluctuation. In contrast, when it comes to Sinyi home price index fluctuation, the BPF neural network outperforms the RBF neural network. However, the distinction is not clearly apparent.

Housing price growth rates in 30 major Chinese cities were forecasted by Wei and Cao (2017) using the dynamic model averaging (DMA) method. The DMA model's performance is compared to that of other common forecasting models in this study using recursive and rolling forecasting modes. The quantitative forecasting ability of various models is also assessed using a model confidence set (MCS) test. Empirical investigations reveal that DMA beats competing models like Bayesian model averaging (BMA), information-theoretic model averaging (ITMA), and equal-weighted averaging in both recursive and rolling forecasting modes (EW).

Aderibigbe and Chi (2018) used a predictive time series model to forecast the Florida House Price Index for the following year in their research. They took into account a variety of parameters such as sales volume, level of employment, interest rate, GDP, rate of inflation, hurricane hits, and so on. The model's forecast findings were encouraging. They were able to show that the Florida House Price Index would continue to rise but will decrease between 2023 and 2025, with a collapse in 2032.

Zheng and Tian (2018) employed Dynamic Model Averaging and Dynamic Model Selection to forecast the rate of housing prices in different levels of Chinese cities, which allows both parameters and the forecasting model to quick adjustment. The consumer price index, gross domestic product, disposable income per capita, rate of unemployment, mortgage rate, and exchange rate were all utilized as predictors. The model forecasting result demonstrated that the DMA and DMS models produce much higher predicting accuracy.

Lin and Tsai (2019) investigated the connection between the Shanghai house price index and Shanghai stock market indices, distinguishing between the A-share market, which is dominated by Chinese investors, and the B-share market, which is dominated by international investors. VAR and GARCH algorithms were used to examine monthly data from April 2003 to June 2018. They discovered that a positive housing market shock has a negative effect on the A-share market, in contrast to a positive credit-price effect between stocks and house prices established for industrialized nations, using VAR estimates. They revealed that the negative credit-price impact is determined to apply exclusively to negative home price shocks using GARCH estimates, implying that declining property prices allow more cash to be transferred to the stock market. While causation in the A-share market is one-way from home values to share prices, causality in the B-share and housing markets is two-way.

Yeap and Lean (2020) investigated the nonlinear interaction between housing supply and house price in Malaysia between 2002 and 2016. Despite the fact that housing supply has been thought to be positively and linearly connected to home price in theory, they found that the number of new houses built in Malaysia has decreased despite rising house prices. As a result, they hypothesized that housing availability and pricing are not linearly connected. They also discovered that the marginal impacts of housing prices are positive and negative, respectively, and statistically significant at the minimum and highest levels.

Zhao et al. (2020) developed a transfer function approach to increase the building cost index forecasting accuracy. In New Zealand, the suggested method is used to calculate the building cost index. The cost estimates provided by transfer function models and autoregressive integrated moving average (ARIMA) models were compared in this research to show the efficacy of the suggested strategy. The findings showed that by taking into account the time-lag causation between building costs and New Zealand housing prices, the suggested transfer function approach might obtain better results than ARIMA models. They claimed that the proposed technique might be utilized by industry experts as a realistic tool for predicting project costs and assisting them in better capturing the underlying linkages between cost and housing prices.

In order to provide reliable out-of-sample predictions for Australian log real housing prices and growth rates, Milunovich (2020) looked at the capabilities of 47 different algorithms. The algorithms were described in single and multi-equation frameworks and included classical time series models, machine learning (ML) techniques, and deep learning neural networks. Although he claimed that the length of the forecast horizon and the dependent variable (log price or growth rate) had an impact on prediction accuracy rankings, certain generalizations may be made. He found several methods that outperform the random walk with drift benchmark for forecasts one and two quarters in advance. The outcomes also showed that six of the top eight forecasts were produced using the same method, a linear support vector regressor (SVR). Simple mean forecast combinations are used to get the other two top-ranked forecasts. Forecasts provided by deep learning nets rank highly over medium and long forecast horizons, whereas linear autoregressive moving average and vector autoregression models produce reliable one-quarter-ahead predictions.

Wang et al. (2021) proposed a whale algorithm optimized support vector regression model based on the bagging ensemble learning method to address the shortcomings of a single machine learning model in house prices index prediction, such as low model prediction accuracy and insufficient generalization ability. The whale optimization process is applied to identify the best penalty factor and kernel function parameters in the SVR model, and the WOA-SVR model is subsequently created. A bagging integration approach is employed to further integrate and refine the WOA-SVR model in order to increase model generalization capabilities. The tests were done to anticipate the housing price indices of four areas, Beijing, Shanghai, Tianjin, and Chongqing, respectively, and the findings revealed that the suggested integrated model's prediction accuracy is superior in all situations to the comparison model.

As a summary, the above-mentioned studies primarily aimed at examining economic parameters that influence house prices. Researchers examined the relationship between house price movements and economic parameters in different regions and countries. They used unidentical time periods and various prediction methods to investigate these relations. The methods employed in these studies include a variety of time series forecasting methods such as regression analysis, OLS estimation, VAR model, VEC model, ARDL model, SVR model, ARIMA model, DMA model, ANN model, etc. The economic parameters commonly employed in these studies are housing/building permits, the volume of housing loans, mortgage interest rate, consumer confidence index, money supply, consumer price index, inflation rate, interest rate, stock market index, population, labor force, employment rate, unemployment rate, exchange rate, oil price, S&P500 index, gross domestic product, economic growth, income per capita, construction/building cost index, housing starts and housing sales. In this thesis, these studies will be referred to while selecting the model parameters and prediction method.

2.2 Previous Studies Regarding the Turkish Housing Market

The previous research conducted to forecast the house price index of Turkey using macroeconomic parameters or to investigate the relationship between house price fluctuations in Turkey and the movements of macroeconomic parameters are summarised as follows:

Sarı et al. (2007) investigated the relationship between housing starts and macroeconomic variables in Turkey from 1961 to 2000. This study examines the relationships between housing market activity and prices, interest rates, output, money stock, and employment utilizing the generalized variance decomposition technique, unit root test, and cointegration test. In contrast to prior findings for industrialized nations, their research showed that the monetary aggregate had a far larger and more significant impact on housing investment than employment. Interest rate, production, and price shocks all have considerable influence on changes in the Turkish housing market, according to their research.

In the study of Badurlar Öner (2008**),** dynamic effects of macroeconomic variables (i.e., gross domestic product (GDP), money supply, short-run interest rates, and exchange rates) on the house prices in Turkey are analyzed for the period 2000-2006 using multiple regression analysis. The Johansen cointegration test was used to estimate the long-run connection between house prices and macroeconomic variables. The findings of the cointegration study revealed that housing prices and macroeconomic variables had a long-term relationship. The short-run dynamic relationship between housing prices and macroeconomic variables is investigated using the Vector Error Correction (VEC) model. The findings of the VEC Granger Causality/Block Exogeneity Wald Test revealed that housing prices, interest rates, and currency rates are bi-directionally causative. Furthermore, it has been discovered that there is a one-directional correlation between GDP and money supply and house prices.

Hepşen and Baş (2009) put an attempt to investigate the dynamic causal relationships between housing market activity (private-use construction permits) and six determinants in the Turkish housing market under the new mortgage system. These determinants included the consumer price index, monetary aggregate, interest rate, industrial production index, real estate investment trusts' indices, and the volume of mortgage loans. Granger causality tests, impulse response functions, and variance decomposition models were used during the years 2002 to 2007. The industrial output index was found through causality testing to be a Granger cause of construction permits without feedback. On the other hand, it was demonstrated that changes in the volume of mortgage loans and the interest rate had feedback effects.

Gök and Keçeli (2015) used OLS estimation and a stepwise regression model to investigate the drivers of housing prices in Turkey at both the national and regional levels. National/regional gross value added (GVA), national/regional net population change, national/regional first and second-hand house sales, national/regional inflation rate, national/regional GVA per capita, net domestic migration to/from regions, and regional bank deposit/national bank deposit ratio explain annual house price changes for both national and regional levels between 2014 and 2015. The influencing elements of housing prices are investigated using ordinary least squares and stepwise regression models. The most important determinants of housing prices, according to regression analyses, are GVA per capita and net domestic migration.

Dilber and Sertkaya (2016) analyzed the macroeconomic factors that can influence housing prices in Turkey in their study. The factors that have impacted the housing prices index in Turkey have been investigated using the vector autoregressive model (VAR) and variance decompositions approach, utilizing quarterly data from 2008 to 2014. The findings revealed a bidirectional association between the home price index and the currency, as well as a unidirectional relationship between interest rates and inflation.

Kolcu and Yamak (2018) investigated the short and long-run effects of income and interest rates for housing loans on house prices in Turkey. The data used in this study was monthly and covered the period from 2010 to 2017. The long-run relations between the variables were investigated in this study utilizing Pesaran and Shin's Autoregressive Distributed Lag (ARDL) bounds testing technique (1999). The presence of a long-run relationship between house prices, income, and interest rate for housing loans was discovered through ARDL bounds testing. In the long run, income has a positive influence on housing prices, according to the findings. In the long run, interest rates on housing loans have little impact on housing prices. In the near run, however, the interest rate on housing loans has a negative effect.

Afşar (2018) investigated the determinants that affect housing price variation in Turkey. This research uses Autoregressive Distributed Lag Model (ARDL) Bound Testing of Paseran on a data set spanning the years 2010 to 2017. This study's findings revealed that rising GDP and mortgage interest rates are inversely connected with increasing house prices. In contrast, there is a positive link between housing credit volume and house prices. As a result, the results indicate that GDP, mortgage interest rates, and changes in the housing industry's loan volume have a considerable impact on home prices.

Bayır (2019) explored the link between monetary policies and house prices in Turkey's economy in his research. As a method of empirical analysis, structural VAR analysis is favored. Federal funding rate, industrial production index, building permits, real house prices, and overnight interest rate were the variables employed in the study. The study spans the years 2011 to 2017, and data is collected on a monthly basis. The findings of the study revealed that monetary policy had little impact on housing prices. During the same time period, house prices only strongly reacted to the lagging values of house prices.

Bayır et al. (2019) conducted an empirical study to explore the connection between macroeconomic factors in the Turkish economy and housing prices. They used the autoregressive distributed lag (ARDL) model to evaluate the period from 2011 to

2018. House prices were the dependent variable in their analysis, whereas economic growth, rate of inflation, money supply, and exchange rate were the independent variables. This study's findings demonstrated that money supply and exchange rate are positively connected with housing prices in long-term economic growth; however, the inflation rate is adversely associated.

Using panel data analysis, Sağlam and Abidinoğlu (2020) evaluated the short-run and long-run dynamic relationship between hedonic house prices and consumer prices in Turkey based on 26 locations from 2010 to 2018. They indicate that the hedonic house price index and the consumer price index have a co-integration association, based on the empirical results. Furthermore, consumer prices have a short- and long-term impact on hedonic housing prices. A ten percent increase in the consumer price index raises the hedonic house price index by around 8.5 percent in the long term. The error-correction term validates the variables' long-run equilibrium relationship.

Canbay and Mercan (2020) examined the connections between housing prices, growth, and macroeconomic indicators in Turkey. They tested the process of housing pricing channel for Turkey in the 2010-2019 timeframe using the Vector Error Correction Model (VAR/VEC). The findings of Granger causality suggested that there exist causation linkages between interest rates and credit volume in the short and long term, as well as between credit volume and housing prices and the consumer price index. In the short and long run, there is also a causal link between growth and house prices. In the near run, however, a causal link between interest rates and growth has been shown. While a positive shock in interest rates and growth reduces housing prices, a positive shock in loan volume increases housing prices, according to the study's impulse-response analysis. House price shocks, on the other hand, boost growth and loan volume while lowering interest rates and the consumer price index.

In order to estimate exogenous credit supply, Tunç (2020) examined the impact of exogenous credit supply shocks on property prices using Turkey as an emerging economy. Using data from 26 locations over the years 2010 to 2017, he discovered that an exogenous rise in housing credits and consumer credits used had a huge and significant influence on property prices. He found that an exogenous rise in housing credits, as well as consumer credits used, has a sizable and substantial impact on pricing using a simple least-squares estimate. He asserts that the effects of quality-adjusted hedonic property values are equal.

Varlık (2020) used the NARDL (Nonlinear Autoregressive Distributed Lag) model to investigate the asymmetric impacts of economic growth on the Turkish housing market. The dependent variable in the model is the house price index, which measures changes in house prices, whereas the independent variable is the industrial production index. The conclusions of the estimation indicated that as the industrial production index increased, so did housing prices and that as the industrial production index decreased, so did housing prices. According to this, it was discovered that, in accordance with economic theory, home prices climbed during periods of economic growth and reduced during periods of economic growth. The effect of increased industrial production on house prices is found to be higher than the effect of decreased industrial production on house prices, and this effect lasts approximately seven months.

In their study, Yılmaz and Kestel (2020) employed non-parametric and various time series methods to find appropriate fits to forecast Turkey's house price index (HPI). They included macroeconomic factors relating to housing markets, such as gold prices, interest rates, and currencies, in their modeling. They created two Generalized Linear Models (GLM) and a Vector Auto-Regressive (VAR) model using the explanatory variables. After that, they built two univariate time series models. Seasonality is inherited by the HPI series. An Autoregressive Moving Average (ARMA) model, a seasonal Autoregressive Integrated Moving Average (ARIMA), and an exponential smoothing (ES) model were also used. Based on the explanatory power measure R2 values and out-of-sample error measures MSE, RMSE, and MAE, their investigation found that forecasts of Turkey's housing market index using both the seasonal ARIMA and Holt Winter models were more accurate than classical time series models.

Çetin (2021) examined the variables that influence housing prices in Turkey, as well as the long-term and causal relationships between them. The house price index, which represents property prices, was employed as the dependent variable in the analysis. Independent variables included the weighted average interest rate for housing loans, housing loans in the banking sector, CPI-based real effective exchange rate, industrial production index, construction materials wholesale price index, consumer price index, and rent index. The data is for the years 2012 through 2020. According to the results of the Autoregressive Distributed Lag Bound Test (ARDL) co-integration test, the consumer price index and industrial production index have a negative impact on housing prices, while the weighted average interest rate for housing loans and the wholesale price index for construction materials have a positive impact. The housing price index, construction materials wholesale price index, consumer price index, and industrial production index were all evaluated with the Granger causality test. There was one-way causation from the wholesale price index of building materials and the consumer price index to the house price index. The Granger causation between the industrial production index and the housing price index is proven via a two-way Granger causality test.

Asymmetric pricing and the impact of coronavirus (Covid-19) pandemic shocks on Turkey’s and Kazakhstan's house price indices (HPI) were studied by Aliefendiolu et al. (2022). A nonlinear autoregressive distributed lag model was used for the empirical investigation. The research's conclusions showed that the Covid-19 pandemic had an unbalanced long- and short-term impact on Turkey's HPI but that Kazakhstan's long-term reaction to the shock of the Covid-19 pandemic was symmetrical, having a similar long-term positive effect on both HPI markets.

Hacıevliyagil et al. (2022) studied the dynamics of the housing market in Turkey's economy and the influence of housing prices. The dynamic model averaging (DMA) approach was used to estimate monthly house price growth in their study. Google internet searches are included in the study due to the growing usage of information technology. In the period 2010-2019, twelve independent variables were employed, with the Residential Property Price Index as the dependent variable. Bond yields, mortgage levels, foreign direct investments, unemployment rate, industrial production, exchange rates, and the Google Trends index were all predictors of the Residential Property Price Index, according to the findings of the study.

The above-mentioned empirical studies constitute the literature on macroeconomic determinants of the house price index in Turkey. The main objective of these studies was to investigate the house price fluctuation in Turkey and detect the parameters that have an effect on this. For this purpose, researchers examined the impact of economic parameters on house price movement using various methodologies in different time periods. The empirical methods utilized in these studies include simple and multiple regression analysis, variance decomposition, VEC model, Granger causality test, ARDL model, NARDL model, VAR model, ARIMA model, ARMA model, DMA model, and panel data analysis. The economic parameters commonly employed in these studies are housing permits, the volume of housing loans, mortgage interest rate, consumer confidence index, money supply, consumer price index, inflation rate, interest rate, stock market index, population, employment rate, unemployment rate, exchange rate, gold price, gross domestic product, gross value added, economic growth, income per capita, construction cost index, and housing sales.

Findings from literature constitute the basis of the model, as will be discussed in Chapter 3.

CHAPTER 3

DATA COLLECTION AND ANALYSIS

3.1 Candidate Parameters for the Prediction Models

Following a thorough review of the literature, candidate parameters for predicting the house price index were identified. Table 3.1 summarizes the candidate parameters and the reference studies in which they were employed in the prediction of the house price index.

Table 3.1. Candidate Parameters for House Price Index Prediction Models

Parameters	Reference Studies	
Housing Permits/ Building Permits	▪ Case and Mayer, 1996 ▪ Rapach and Strauss, 2007 ▪ Rapach and Strauss, 2009	▪ Ciarlone, 2015 ▪ Li and Chu, 2017 ▪ Bayır, 2019
Volume of Housing Loans	▪ Padilla, 2005 ▪ Hepşen and Kalfa, 2009 ▪ Agnello and Schuknecht, 2011 ▪ Ciarlone, 2015	▪ Li & Chu, 2017 ▪ Afşar, 2018 ▪ Canbay and Mercan, 2020 ▪ Tunç, 2020
Effective Mortgage Interest Rate	▪ Case and Shiller, 1990 ▪ Jud and Winkler, 2002 ▪ McGibany and Nourzad, 2004 ▪ Abelson et al., 2005 ▪ Rapach and Strauss, 2007 ▪ Mikhed and Zemčík, 2009 ▪ Rapach and Strauss, 2009 ▪ Radzi et al., 2012 ▪ Hanewald and Sherris, 2013 ▪ Lai et al., 2013	▪ Ciarlone, 2015 ▪ Park and Bae, 2015 ▪ Li and Chu, 2017 ▪ Wei and Cao, 2017 ▪ Aderibigbe and Chi, 2018 ▪ Zheng and Tian, 2018 ▪ Kestel and Yilmaz, 2020 ▪ Milunovich, 2020 ▪ Hacıevliyagil et al., 2022 ▪ Özgüler et al., 2022
Construction Material Cost Index	▪ Yeap and Lean, 2020	
Industrial production index	▪ Varlık, 2020	▪ Çetin, 2021
Cost of Living Index	▪ Kestel and Yilmaz, 2020	
Consumer Confidence (Economic) Index	▪ Rapach and Strauss, 2007 ▪ Rapach and Strauss, 2009	▪ Kestel and Yilmaz, 2020
Producer Price Index	▪ Stevenson, 2000	▪ Zhao et al., 2020
Money Supply / Money Stock	▪ Sarı et al. 2007 ▪ Badurlar Öner, 2008	▪ Bayır et al., 2019

Table 3.1. (Cont'd)

Consumer Price Index	▪ Case and Shiller, 1990 ▪ Jud and Winkler, 2002 ▪ McGibany and Nourzad, 2004 ▪ Abelson et al., 2005 ▪ Mikhed and Zemčík, 2009 ▪ Beltratti and Morana, 2010 ▪ Adams and Füss, 2010 ▪ Lai et al., 2013 ▪ Dougherty and Order, 2016	▪ Wei and Cao, 2017 ▪ Li and Chu, 2017 ▪ Zheng and Tian, 2018 ▪ Milunovich, 2020 ▪ Sağlam and Abdioğlu, 2020 ▪ Çetin, 2021 ▪ Wang et al., 2021 ▪ Aliefendioğlu et al., 2022 ▪ Hacıevliyagil et al., 2022
Inflation Rate	▪ Stevenson, 2000 ▪ Abelson et al., 2005 ▪ Rapach and Strauss, 2007 ▪ Mikhed and Zemčík, 2009 ▪ Rapach and Strauss, 2009 ▪ Beltratti and Morana, 2010 ▪ Hanewald and Sherris, 2013	▪ Gök and Keçeli, 2015 ▪ Plakandaras et al., 2015 ▪ Dilber and Sertkaya, 2016 ▪ Dougherty and Order, 2016 ▪ Bayır et al., 2019 ▪ Kestel and Yilmaz, 2020
Effective Interest rate	▪ Case and Shiller, 1990 ▪ Jud and Winkler, 2002 ▪ McGibany and Nourzad, 2004 ▪ Padilla, 2005 ▪ Sarı et al. 2007 ▪ Badurlar Öner, 2008 ▪ Hepşen and Kalfa, 2009 ▪ Mikhed and Zemčík, 2009 ▪ Beltratti and Morana, 2010 ▪ Adams and Füss, 2010 ▪ Agnello and Schuknecht, 2011 ▪ Radzi et al., 2012 ▪ Lai et al., 2013	▪ Plakandaras et al., 2015 ▪ Dilber and Sertkaya, 2016 ▪ Li and Chu, 2017 ▪ Afşar, 2018 ▪ Bayır, 2019 ▪ Kolcu and Yamak, 2018 ▪ Zheng and Tian, 2018 ▪ Canbay and Mercan, 2020 ▪ Yeap and Lean, 2020 ▪ Çetin, 2021 ▪ Wang et al., 2021 ▪ Hacıevliyagil et al., 2022
Domestic Stock Market Index/ Equity Price Index	▪ Abelson et al., 2005 ▪ Jud and Winkler, 2002 ▪ Mikhed and Zemčík, 2009 ▪ Rapach and Strauss, 2009 ▪ Beltratti and Morana, 2010 ▪ Agnello and Schuknecht, 2011 ▪ Hanewald and Sherris, 2013	▪ Ciarlone, 2015 ▪ Plakandaras et al., 2015 ▪ Lin and Tsai, 2019 ▪ Milunovich, 2020 ▪ Hacıevliyagil et al., 2022 ▪ Özgüler et al., 2022
The Population	▪ Case and Shiller, 1990 ▪ Case and Mayer, 1996 ▪ Jud and Winkler, 2002 ▪ Rapach and Strauss, 2007 ▪ Mikhed and Zemčík, 2009 ▪ Rapach and Strauss, 2009	▪ Adams and Füss, 2010 ▪ Agnello and Schuknecht, 2011 ▪ Radzi et al., 2012 ▪ Gök and Keçeli, 2015 ▪ Plakandaras et al., 2015 ▪ Wang et al., 2021
Working Age Population/ Labor Force	▪ Case and Shiller, 1990 ▪ Rapach and Strauss, 2007 ▪ Rapach and Strauss, 2009	▪ Agnello and Schuknecht, 2011 ▪ Aderibigbe and Chi, 2018
Employment Rate	▪ Case and Shiller, 1990 ▪ Case and Mayer, 1996 ▪ Padilla, 2005 ▪ Rapach and Strauss, 2007 ▪ Sarı et al., 2007	▪ Rapach and Strauss, 2009 ▪ Adams and Füss, 2010 ▪ Radzi et al., 2012 ▪ Aderibigbe and Chi, 2018

Table 3.1. (Cont'd)

Unemployment Rate	▪ Case and Mayer, 1996 ▪ Abelson et al., 2005 ▪ Rapach and Strauss, 2007 ▪ Rapach and Strauss, 2009 ▪ Hanewald and Sherris, 2013 ▪ Ciarlone, 2015 ▪ Plakandaras et al., 2015	▪ Wei and Cao, 2017 ▪ Aderibigbe and Chi, 2018 ▪ Zheng and Tian, 2018 ▪ Kestel and Yilmaz, 2020 ▪ Milunovich, 2020 ▪ Hacıevliyagil et al., 2022 ▪ Özgüler et al., 2022
Real Effective Exchange Rate of the Local Currency vs. US Dollars	▪ Abelson et al., 2005 ▪ Padilla, 2005 ▪ Badurlar Öner, 2008 ▪ Beltratti and Morana, 2010 ▪ Agnello and Schuknecht, 2011 ▪ Hanewald and Sherris, 2013 ▪ Dilber and Sertkaya, 2016	▪ Zheng and Tian, 2018 ▪ Bayır et al., 2019 ▪ Kestel and Yılmaz, 2020 ▪ Milunovich, 2020 ▪ Wang et al., 2021 ▪ Hacıevliyagil et al., 2022 ▪ Özgüler et al., 2022
Real Oil Price	▪ Padilla, 2005 ▪ Beltratti and Morana, 2010	▪ Plakandaras et al., 2015
Gold Price	▪ Kestel & Yilmaz, 2020	▪ Özgüler et al., 2022
S&P 500 Stock Market Index	▪ Jud and Winkler, 2002 ▪ Rapach and Strauss, 2007	▪ Mikhed and Zemčík, 2009 ▪ Rapach and Strauss, 2009
Gross Domestic Product / Economic Growth	▪ Case and Shiller, 1990 ▪ Jud and Winkler, 2002 ▪ Rapach and Strauss, 2007 ▪ Badurlar Öner, 2008 ▪ Mikhed and Zemčík, 2009 ▪ Rapach and Strauss, 2009 ▪ Beltratti and Morana, 2010 ▪ Adams and Füss, 2010 ▪ Agnello and Schuknecht, 2011 ▪ Gupta and Miller, 2012 ▪ Hanewald and Sherris, 2013 ▪ Lai et al., 2013 ▪ Ciarlone, 2015	▪ Nyakabawo et al., 2015 ▪ Plakandaras et al., 2015 ▪ Li and Chu, 2017 ▪ Wei & Cao, 2017 ▪ Aderibigbe and Chi, 2018 ▪ Zheng and Tian, 2018 ▪ Bayır et al., 2019 ▪ Canbay and Mercan, 2020 ▪ Milunovich, 2020 ▪ Wang et al., 2021 ▪ Aliefendioğlu et al., 2022 ▪ Özgüler et al., 2022
Gross Value Added	▪ Gök and Keçeli, 2015	
Income Per Capita	▪ Case and Shiller, 1990 ▪ Jud and Winkler, 2002 ▪ Abelson et al., 2005 ▪ Rapach and Strauss, 2007 ▪ Mikhed and Zemčík, 2009	▪ Radzi et al., 2012 ▪ Lai et al., 2013 ▪ Afşar, 2018 ▪ Kolcu and Yamak, 2018
The Real Construction Cost/ Construction Cost Index	▪ Case and Shiller, 1990 ▪ Jud and Winkler, 2002 ▪ Mikhed and Zemčík, 2009 ▪ Adams and Füss, 2010 ▪ Ciarlone, 2015	▪ Plakandaras et al., 2015 ▪ Yeap and Lean, 2020 ▪ Zhao et al., 2020 ▪ Çetin, 2021
Building Cost Index	▪ Mikhed and Zemčík, 2009	▪ Zhao et al., 2020
Housing Starts	▪ Case and Shiller, 1990) ▪ Rapach and Strauss, 2007 ▪ Rapach and Strauss, 2009	▪ Wei and Cao, 2017 ▪ Yeap and Lean, 2020
Housing Sales	▪ McGibany and Nourzad, 2004 ▪ Rapach and Strauss, 2007	▪ Rapach and Strauss, 2009 ▪ Gök and Keçeli, 2015

3.2 Data Collection

In the previous part, the determined candidate parameters that can be utilized in house price index prediction models were listed. It is clear that the parameters to be included in the model depend on data availability. These parameters are searched in the databases of relevant institutions in Turkey that publish data periodically. The time-series data of the candidate parameters used in this study are imported from the databases of the Central Bank of the Republic of Turkey (CBRT), Banking Regulation and Supervision Agency (BRSA), and Turkish Statistical Institute (TURKSTAT). The database of Investing.com, which is a reliable real-time financial data platform, is also used for obtaining the time series data of some variables. The collected time series data is monthly data that covers a time period of twelve years, from January 2010 to December 2021. All data series are complete. In other words, no gaps or missing values are included in the dataset. A total of 144 observation values exist in each data set, and the values are seasonally adjusted. The candidate parameters Gross Domestic Product (GDP), Gross Value Added (GVA), Construction Cost Index, Building Cost Index (BCI), Housing Starts, and Housing Sales are not included in this study since they are not available in the local databases of Turkey as a complete monthly time series data for the selected time period. Different from the reference studies, the monthly time series data of four additional parameters, Construction Turnover Index (CTI), Number of Paid Employees (NPE), S&P 500 Volatility Index (VIX), and US Dollar Index (DXY), are decided to be included in this study.

3.2.1 House Price Index (HPI)

The House Price Index (HPI) statistics monitor the price movements in the housing market. The Residential Property Price Index (RPPI), published by the Central Bank of the Republic of Turkey (CBRT), covers the indicators constructed for monitoring the price movements in Turkey. It measures the quality-adjusted price changes of dwellings in Turkey. The value of this index in the base year (currently 2017) is arbitrarily set at 100. The monthly time series data of this index is exported from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period and used as the House Price Index (HPI) in this study.

3.2.2 Number of Housing Permits (NHP)

Housing permits refer to the legal approvals given by local jurisdictions before the construction of a new housing unit. The number of Housing Permits (NHP) parameter included in this study is published by the Turkish Statistical Institute (TURKSTAT), under construction statistics monitors the monthly number of dwelling units in Turkey according to construction permits. The monthly time series data of this parameter is exported from the Data Portal of the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.3 Volume of Housing Loans (VHL)

The Volume of Housing Loans (VHL) parameter included in this study indicates the cumulative volume of housing (mortgage) loans of consumers in Turkey in terms of million Turkish Liras. The monthly time series data of this parameter is obtained from the Monthly Banking Sector Data published by the Banking Regulation and Supervision Agency (BRSA) for the reference time period.

3.2.4 Mortgage Interest Rate (MIR)

The Mortgage Interest Rate (MIR) parameter used in this study monitors the weighted average of the effective yearly interest rate of bank loans for housing in Turkey over the Turkish Lira currency. The monthly time series data of this parameter is exported from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.5 Construction Materials Price Index (CMPI)

The Construction Materials Price Index (CMPI) used in this study is the wholesale prices index of construction materials published by the Istanbul Chamber of Commerce (ICC). It monitors the construction material prices in Istanbul, Turkey. The value of this index in the base year (1968) is arbitrarily set at 100. The monthly time series data of this index is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.6 Construction Turnover Index (CTI)

The Construction Turover Index (CTI) used in this study monitors the total monetary volume of works (turnover) in the construction of buildings in Turkey. The value of this index in the base year (currently 2015) is arbitrarily set at 100. The monthly time series data of this index is exported from the Data Portal of the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.7 Number of Paid Employees (NPE)

The Number of Paid Employees (NPE) parameter used in this study indicates the total number of paid employees working in the construction of buildings in Turkey. The monthly time series data of this parameter is taken from the paid employee statistics published by the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.8 Industrial Production Index (IPI)

The Turkish Statistical Institute (TURKSTAT) publishes the Industrial Production Index (IPI), a financial indicator that analyzes the real production output of the manufacturing, mining, and utility sectors. The majority of the variance in national production over the course of the business cycle is explained by this index, together with other industrial indices. The value of this index in the base year (currently 2015) is arbitrarily set at 100. The monthly time series data of this index is exported from the Data Portal of the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.9 Cost of Living Index (CLI)

The Cost of Living Index (CLI) used in this study is published by the Istanbul Chamber of Commerce (ICC). It monitors the cost of living for wage earners in Istanbul, Turkey. The value of this index in the base year (1995) is arbitrarily set at 100. The monthly time series data of this index is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.10 Consumer Confidence Index (CCI)

The Consumer Confidence Index (CCI) that is used in this study is a financial indicator that forecasts how households will spend and save in the future based on responses to questions about their expected financial situation, attitudes toward the state of the economy as a whole, unemployment, and savings ability. The monthly time series data of this parameter is taken from the consumer tendency statistics published by the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.11 Domestic Producer Price Index (DPPI)

The Domestic Producer Price Index (DPPI) utilized in this study tracks the average changes in prices domestic producers in Turkey receive for their products. The value in the base year (2003) is arbitrarily set at 100. The monthly time series data of this index is exported from the Data Portal of the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.12 Non-Domestic Producer Price Index (NPPI)

The Non-Domestic Producer Price Index (NPPI) utilized in this study tracks the average changes in prices non-domestic producers receive for their products. The value in the base year (2010) is arbitrarily set at 100. The monthly time series data of this index is exported from the Data Portal of the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.13 Consumer Price Index (CPI)

The Consumer Price Index (CPI) utilized in this study represents the cost of a market basket of consumer goods and services weighted on average and purchased by Turkish households. The measured CPI fluctuates to reflect changes in prices over time. The monthly percent change of this index is accepted as the inflation rate in Turkey. The value in the base year (2003) is arbitrarily set at 100. The monthly time series data of this index is exported from the Data Portal of the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.14 Effective Interest Rate (EIR)

The Effective Interest Rate (EIR) is the annual rate of interest that an investor can earn in a year. The effective interest rate used in this study is the maximum of the effective yearly interest rate applied by the state banks of Turkey for Turkish Lira deposits for up to three months. The monthly time series data of this parameter is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.15 Money Supply (MS)

The entire amount of money owned by the public at a specific time is referred to as the money supply (also known as the money stock) in macroeconomics. The Money Supply (MS) parameter used in this study indicates the M2 level monthly monetary aggregates and counterpart items in Turkey in billion Turkish Liras. The monthly time series data of this parameter is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.16 Population (POP)

Population typically refers to the number of people in a single area. The Population (POP) parameter used in this study monitors the population of Turkey at 15 years of age and over in thousand people. The monthly time series data of this parameter is taken from the seasonally adjusted main labor force statistics published by the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.17 Labor Force (LF)

The term "labor force" or "workforce" refers to the group of people who are either employed or unemployed. The labor force of a country includes both the employed and the unemployed human beings that are capable of working. The Labour Force (LF) parameter used in this study monitors the 15 years of age and over labor force of Turkey in thousand people. The monthly time series data of this parameter is taken from the seasonally adjusted main labor force statistics published by the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.18 Employment Rate (ER)

The employment rate is a ratio used in statistics to measure the percentage of the working-age population of a nation that is employed (data are frequently provided for ages 15 to 64). The Employment Rate (ER) parameter used in this study monitors the employment rate of Turkey's working-age population (15 years of age and over labor force) in percentage. The monthly time series data of this parameter is taken from the seasonally adjusted main labor force statistics published by the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.19 Unemployment Rate (UR)

The unemployment rate is a statistical ratio that tracks the percentage of a nation's population that is neither self-employed nor in paid labor during the reference period but is still currently available for employment. The Unemployment Rate (UR) parameter used in this study monitors the unemployment rate of Turkey's working-age population (15 years age and over labor force) in percentage. The monthly time series data of this parameter is taken from the seasonally adjusted main labor force statistics published by the Turkish Statistical Institute (TURKSTAT) for the reference time period.

3.2.20 Effective Exchange Rate (EER)

A measure of a currency's strength in relation to another currency or a basket of currencies is the effective exchange rate. The Effective Exchange Rate (EER) parameter used in this study monitors the strength of the U.S. Dollar (USD) relative to the Turkish Lira (TRY). It considers the US dollar selling price of the Central

Bank of the Republic of Turkey (CBRT). The monthly time series data of this parameter is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.21 Domestic Stock Market Index (DSMI)

A stock index, also known as a stock market index, is an index used in finance to monitor changes in stock prices on a whole stock market or a segment thereof. This makes it easier for investors to compute the performance of the market by comparing current stock price levels with earlier ones. The Domestic Stock Market Index (DSMI) used in this study monitors the monthly closing price of the BIST 100 index, which is the domestic stock market index of Turkey. The value of this index in the base month (January 1986) is arbitrarily set at 0.01. The monthly time series data of this index is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.22 Brent Oil Price (BOP)

Brent Blend, London Brent, and Brent petroleum are the other common names for Brent Crude Oil. Due to its low density and low sulfur concentration, this grade is referred to as light and sweet. The most commonly used worldwide reference for crude oil prices is Brent. Two-thirds of the world's supply of crude oil that is traded globally are priced using it. The Brent Oil Price (BOP) parameter used in this study monitors the Europe Brent Petrol Spot Price in terms of U.S. dollars per barrel. The monthly time series data of this parameter is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.23 Gold Price (GP)

Gold has been used as money throughout history and, until very recently, was a relative basis for currency equivalents unique to economic areas or nations. In the second part of the 19th century, several European nations adopted gold standards; however, these were briefly halted during the financial crisis associated with World War I. Central banks of leading countries and the International Monetary Fund play an essential role in the gold price. The Gold Price (GP) parameter used in this study indicates the free market (London) selling price of 1 ton of gold in U.S. Dollars. It is actually the global gold exchange index (XAUUSD) that monitors the price of 1 ounce (≈ 28.35 grams) of 24 karat gold in terms of U.S. Dollars. The monthly time series data of this parameter is obtained from the Electronic Data Delivery System of the Central Bank of the Republic of Turkey (CBRT) for the reference time period.

3.2.24 S&P 500 Index (SPX)

The Standard and Poor's 500, commonly known as the S&P 500, is an index used to measure the performance of 500 significant businesses that are listed on American stock exchanges. The monthly closing values of the S&P 500 Index (SPX) are used in this study. The monthly time series data of closing values of this index is obtained from the database of Investing.com for the reference time period.

3.2.25 Volatility Index (VIX)

The Volatility Index of the Chicago Board Options Exchange (CBOE) is denoted by the ticker symbol and common name VIX. Based on options for the S&P 500 index, VIX is a successful predictor of the stock market's anticipated volatility. It is sometimes referred to as the fear index or fear gauge and is calculated and released by the CBOE in real-time. The Volatility Index (VIX) parameter used in this study monitors the monthly closing values of the S&P 500 VIX index. The monthly time series data of this index is obtained from the database of Investing.com for the reference time period.

3.2.26 U.S. Dollar Index (DXY)

The U.S. Dollar Index (DXY), commonly known as a basket of US trading partners' currencies, is a gauge of the value of the U.S. dollar in relation to a group of other currencies. The Index climbs when the worth (strength) of the U.S. dollar relative to other currencies increases. The U.S. Dollar Index (DXY) used in this study monitors the monthly closing values of the DXY. The monthly time series data of this index is obtained from the database of Investing.com for the reference time period.

Table 3.2. summarizes the candidate parameters included in the collected data for the forecasting models with their abbreviations and the data sources from which the monthly time series data of these parameters are obtained.

Table 3.2. Data Collected for House Price Index Prediction Models

	Name of the Parameter	Abbreviation	Data Source
1	House Price Index	HPI	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
2	Number of Housing Permits	NHP	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
3	The Volume of Housing Loans	VHL	Banking Regulation and Supervision Agency (BRSA) - Monthly Banking Sector Data
4	Mortgage Interest Rate	MIR	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
5	Construction Materials Price Index	CMPI	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
6	Construction Turnover Index	CTI	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
7	Number of Paid Employees	NPE	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
8	Industrial Production Index	IPI	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
9	Cost of Living Index	CLI	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
10	Consumer Confidence Index	CCI	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
11	Domestic Producer Price Index	DPPI	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
12	Non-Domestic Producer Price Index	NPPI	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
13	Consumer Price Index	CPI	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
14	Effective Interest Rate	EIR	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
15	Money Supply	MS	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
16	Population	POP	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
17	Labor Force	LF	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics

Table 3.2. (Cont'd)

18	Employment Rate	ER	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
19	Unemployment Rate	UR	Turkish Statistical Institute (TURKSTAT) - Data Portal for Statistics
20	Effective Exchange Rate	EER	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
21	Domestic Stock Market Index	DSMI	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
22	Brent Oil Price	BOP	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
23	Gold Price Index	GPI	Central Bank of the Republic of Turkey (CBRT)- Electronic Data Delivery System
24	S&P 500 Index	SPX	Investing.com - Financial Markets Real-time Data Platform
25	Volatility Index	VIX	Investing.com - Financial Markets Real-time Data Platform
26	US Dollar Index	DXY	Investing.com - Financial Markets Real-time Data Platform

3.3 Data Analysis

The House Price Index (HPI) parameter is selected as the output variable of the forecasting models employed in this thesis. A total of 25 parameters are selected as candidate input variables of the HPI forecasting models, and time-series data of these parameters are collected from the mentioned data sources. In order to evaluate the association between the 25 candidate input variables, a correlation test is employed.

3.3.1 Correlation Test

The correlation coefficient represents the covariance of the two variables divided by the multiplication of their standard deviations. It takes values in the range of -1 and 1. Having a correlation coefficient close to 0 implies a weak correlation between the two variables. Whereas having a correlation coefficient close to -1 or 1 implies a strong negative or positive correlation between the two variables (being exactly equal to 1 represents an unrealistic perfect correlation). As a measure of the linear correlation between parameters, the Pearson correlation coefficient, which was developed by Karl Pearson in the 1880s, is selected. The correlation matrix of the 25 candidate input variables, which is generated in Python, is shown in Figure 3.1.

	CMPI	CTI	VHL	NHP	NPE	MIR	CCI	CLI	CPI	DPPI	...	LF	ER	UR	EER	DSMI	BOP	GP	SPX	VIX	DXY
CMPI	1.000	0.939	0.961	-0.164	0.276	0.994	-0.816	0.997	0.996	0.974	...	0.829	0.333	0.703	0.981	0.924	-0.512	0.478	0.983	0.136	0.631
CTI	0.939	1.000	0.959	-0.063	0.468	0.943	-0.723	0.944	0.942	0.909	...	0.913	0.548	0.636	0.920	0.908	-0.518	0.306	0.960	-0.017	0.687
VHL	0.961	0.959	1.000	-0.080	0.475	0.966	-0.716	0.963	0.958	0.905	...	0.894	0.464	0.664	0.932	0.920	-0.596	0.378	0.973	0.050	0.698
NHP	-0.164	-0.063	-0.080	1.000	0.279	-0.184	0.308	-0.172	-0.181	-0.179	...	-0.047	0.161	-0.250	-0.185	-0.061	-0.018	-0.191	-0.123	-0.255	-0.028
NPE	0.276	0.468	0.475	0.279	1.000	0.271	0.014	0.273	0.254	0.190	...	0.630	0.855	-0.053	0.203	0.389	-0.376	-0.239	0.391	-0.444	0.524
MIR	0.994	0.943	0.966	-0.184	0.271	1.000	-0.818	0.996	0.995	0.957	...	0.848	0.338	0.742	0.974	0.911	-0.546	0.456	0.976	0.135	0.660
CCI	-0.816	-0.723	-0.716	0.308	0.014	-0.818	1.000	-0.827	-0.831	-0.829	...	-0.632	-0.148	-0.717	-0.846	-0.691	0.452	-0.375	-0.764	-0.218	-0.567
CLI	0.997	0.944	0.963	-0.172	0.273	0.996	-0.827	1.000	0.999	0.976	...	0.836	0.336	0.718	0.986	0.926	-0.524	0.462	0.981	0.133	0.644
CPI	0.996	0.942	0.958	-0.181	0.254	0.995	-0.831	0.999	1.000	0.979	...	0.828	0.324	0.718	0.989	0.922	-0.507	0.471	0.979	0.139	0.629
DPPI	0.974	0.909	0.905	-0.179	0.190	0.957	-0.829	0.976	0.979	1.000	...	0.746	0.262	0.635	0.987	0.918	-0.387	0.520	0.957	0.154	0.519
NPPI	0.971	0.906	0.909	-0.178	0.184	0.955	-0.827	0.974	0.977	0.997	...	0.737	0.244	0.639	0.992	0.912	-0.394	0.527	0.952	0.171	0.514
IPI	0.846	0.885	0.891	0.063	0.553	0.840	-0.608	0.848	0.841	0.808	...	0.863	0.602	0.488	0.818	0.842	-0.480	0.264	0.874	-0.085	0.648
EIR	0.994	0.942	0.966	-0.184	0.270	1.000	-0.820	0.996	0.995	0.958	...	0.848	0.337	0.744	0.974	0.911	-0.549	0.451	0.977	0.137	0.662
MS	0.988	0.921	0.945	-0.158	0.229	0.977	-0.809	0.988	0.988	0.986	...	0.766	0.254	0.667	0.988	0.926	-0.473	0.530	0.968	0.185	0.566
POP	0.947	0.953	0.979	-0.109	0.486	0.962	-0.733	0.951	0.945	0.873	...	0.944	0.536	0.700	0.906	0.883	-0.625	0.301	0.955	-0.001	0.770
LF	0.829	0.913	0.894	-0.047	0.630	0.848	-0.632	0.836	0.828	0.746	...	1.000	0.762	0.649	0.780	0.792	-0.637	0.042	0.868	-0.192	0.838
ER	0.333	0.548	0.464	0.161	0.855	0.338	-0.148	0.336	0.324	0.262	...	0.762	1.000	0.112	0.272	0.401	-0.349	-0.350	0.434	-0.518	0.613
UR	0.703	0.636	0.664	-0.250	-0.053	0.742	-0.717	0.718	0.718	0.635	...	0.649	0.112	1.000	0.699	0.554	-0.689	0.070	0.659	0.166	0.680
EER	0.981	0.920	0.932	-0.185	0.203	0.974	-0.846	0.986	0.989	0.987	...	0.780	0.272	0.699	1.000	0.904	-0.478	0.490	0.957	0.176	0.592
DSMI	0.924	0.908	0.920	-0.061	0.389	0.911	-0.691	0.926	0.922	0.918	...	0.792	0.401	0.554	0.904	1.000	-0.421	0.432	0.949	-0.006	0.530
BOP	-0.512	-0.518	-0.596	-0.018	-0.376	-0.546	0.452	-0.524	-0.507	-0.387	...	-0.637	-0.349	-0.689	-0.478	-0.421	1.000	0.176	-0.491	-0.157	-0.871
GP	0.478	0.306	0.378	-0.191	-0.239	0.456	-0.375	0.462	0.471	0.520	...	0.042	-0.350	0.070	0.490	0.432	0.176	1.000	0.402	0.404	-0.182
SPX	0.983	0.960	0.973	-0.123	0.391	0.976	-0.764	0.981	0.979	0.957	...	0.868	0.434	0.659	0.957	0.949	-0.491	0.402	1.000	-0.001	0.635
VIX	0.136	-0.017	0.050	-0.255	-0.444	0.135	-0.218	0.133	0.139	0.154	...	-0.192	-0.518	0.166	0.176	-0.006	-0.157	0.404	-0.001	1.000	-0.066
DXY	0.631	0.687	0.698	-0.028	0.524	0.660	-0.567	0.644	0.629	0.519	...	0.838	0.613	0.680	0.592	0.530	-0.871	-0.182	0.635	-0.066	1.000

Figure 3.1. Correlation Matrix of the Candidate Input Parameters

3.3.2 Selected Input Parameters

After the evaluation of correlation coefficients, it is decided to eliminate the parameters that have a correlation coefficient value greater than 0.85. After this elimination, nine parameters are left, and these parameters are selected as the input parameters for the models. The correlation matrix of the selected nine input parameters is given in Figure 3.2.

	NHP	CCI	CPI	IPI	ER	UR	BOP	GP	VIX
NHP	1.000	0.308	-0.181	0.063	0.161	-0.250	-0.018	-0.191	-0.255
CCI	0.308	1.000	-0.831	-0.608	-0.148	-0.717	0.452	-0.375	-0.218
CPI	-0.181	-0.831	1.000	0.841	0.324	0.718	-0.507	0.471	0.139
IPI	0.063	-0.608	0.841	1.000	0.602	0.488	-0.480	0.264	-0.085
ER	0.161	-0.148	0.324	0.602	1.000	0.112	-0.349	-0.350	-0.518
UR	-0.250	-0.717	0.718	0.488	0.112	1.000	-0.689	0.070	0.166
BOP	-0.018	0.452	-0.507	-0.480	-0.349	-0.689	1.000	0.176	-0.157
GP	-0.191	-0.375	0.471	0.264	-0.350	0.070	0.176	1.000	0.404
VIX	-0.255	-0.218	0.139	-0.085	-0.518	0.166	-0.157	0.404	1.000

Figure 3.2. Correlation Matrix of the Selected Input Parameters

Symbols and explanations of the selected input and output parameters of the HPI forecasting models are demonstrated in Table 3.3. Time series data of the model parameters, including 144 monthly observation values from January 2010 to December 2021, is given in Appendix A.

Table 3.3. Selected Output and Input Parameters for the Models

Symbol	**Name of the Parameter**	**Abbreviation**
Y (Output)	House Price Index	HPI
X1 (Input)	Number of Housing Permits	NHP
X2 (Input)	Consumer Confidence Index	CCI
X3 (Input)	Consumer Price Index	CPI
X4 (Input)	Industrial Production Index	IPI
X5 (Input)	Employment Rate	ER
X6 (Input)	Unemployment Rate	UR
X7 (Input)	Brent Oil Price	BOP
X8 (Input)	Gold Price	GP
X9 (Input)	Volatility Index	VIX

Descriptive statistics, including the mean value, standard deviation, minimum value, 25 percentile value, 50 percentile value, 75 percentile value, and maximum value of the selected output and input parameters, are demonstrated in Figure 3.3.

Parameter	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
# of Observations	144	144	144	144	144	144	144	144	144	144
Mean Value	91.55	67503.81	88.47	308.81	100.86	44.66	10.76	75.68	1419.47	18.50
Standard Deviation	41.17	41477.11	6.13	117.60	20.46	1.88	1.69	26.57	235.18	6.96
Minimum Value	45.40	9622.00	68.90	174.07	56.84	40.60	8.00	14.85	1070.85	9.51
25% Value	56.78	46498.50	82.70	215.86	85.81	43.10	9.48	53.34	1239.25	13.74
50% Value	85.10	60931.00	90.25	271.91	99.88	45.20	10.50	71.55	1324.85	16.59
75% Value	109.03	79009.75	92.87	398.23	114.96	46.13	11.90	105.06	1630.36	20.56
Maximum Value	247.40	330735.00	97.97	686.95	165.56	47.70	14.10	126.59	1964.40	53.54

Figure 3.3. Descriptive Statistics of the Selected Output and Input Parameters

An extreme outlier is defined as a value in a dataset that lays outside of 3 times the interquartile range value below 25 percentile value or above 75 percentile value (Neter et al., 1993). Since major outliers are not observed in most of the nine input parameters' time series data, outlier treatment is not applied.

3.3.3 Data Split for Training and Validation of the Models

In order to evaluate the predictive performance of the models utilized, the time series dataset, which includes 144 observation values of all parameters, should be split to generate training and validation sets by using an optimal ratio for data split. In line with the recommendations of Raykar and Saha (2015), the optimal ratio for data split is determined as 5/6:1/6, which means 5/6 of the data is reserved for training and 1/6 of data for validation. The dataset is divided into six folds, including 24 observation values each. The multiple linear regression model is trained using five folds of data as the training set, and the predictive performance of the established model is tested using the remaining one fold of data as the validation set. The artificial neural networks model is developed similarly using the five folds of data as a training set, and the remaining one fold of data is used as the validation set to measure the predictive performance of the model. In line with the requirements of the artificial neural networks model, the training set is also validated using the cross-validation of 5 folds of data.

CHAPTER 4

PREDICTION OF HPI USING MULTIPLE LINEAR REGRESSION

In this chapter, the multiple linear regression model is introduced for the prediction of the house price index. Specifically, the concepts and background of the model, basic information on regression parameters, the measures of strength of the relationship between the outcome and explanatory variables, and the assumptions underlying the model are discussed.

4.1 Background of the Multiple Linear Regression (MLR) Model

Regression analysis is a statistical method that is used for investigating the relationship between an outcome variable (dependent variable, denoted by Y) and a series of explanatory variables (independent variables, denoted by X_1, X_2, … X_n where n is the number of explanatory variables). It is one of the most commonly used statistical tools since it provides simple methods to establish a functional relationship between the variables. The result of the regression analysis represents how the outcome variable moves with the change of other variables when one variable is changed while others are kept constant. The relationship among variables is described in equational form or a model connecting the dependent variable and one or multiple explanatory variables (Arkes, 2019; Chatterjee & Hadi, 2012; Chatterjee & Simonoff, 2013).

The relationship between Y and X_1, X_2, … X_n can be expressed by the regression model:

$$Y = f\,(X_1, X_2, \ldots X_n) + \varepsilon \tag{4.1}$$

where the random error term ε represents the discrepancy in the estimation, it accounts for the failure of the regression model to fit the data exactly.

4.1.1 The Simple Linear Regression Model

In linear algebra, a straight line is represented by the following equation:

$$y = a + b\,x \tag{4.2}$$

where:

- x is the horizontal-axis variable,
- y is the vertical-axis variable,
- a is the y-intercept,
- b is the slope of the straight line.

Similarly, in the Simple Linear Regression Model, there is a single X and Y variable. However, not all points lay in a straight line. The Simple Linear Regression line represents the straight line that best fits the given data points.

The Simple Linear Regression Model is represented by the following equation:

$$Y_i = \beta_0 + \beta_1 \times X_i + \varepsilon_i \tag{4.3a}$$

Note that the given equation describes n data points, and the i subscript refers to each individual point in the data sample (i = 1, 2, …, n). Therefore the given equation for the simple linear regression model actually represents n separate equation for each observation:

$$Y_1 = \beta_0 + \beta_1 \times X_1 + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 \times X_2 + \varepsilon_2$$

$$....$$

$$Y_N = \beta_0 + \beta_1 \times X_N + \varepsilon_N \tag{4.3b}$$

The components of the Equations (4.3a and 4.3b) are:

- The dependent variable (Y) is also called the response variable, regressand, outcome, or simply Y variable.
- The independent variable (X) is also called the explanatory variable, treatment variable, regressor, or simply X variable. The name independent variable is not commonly used because, in practice, the explanatory variables are rarely independent of each other.
- The coefficient of the explanatory variable (β_1), which is the slope of the regression line, represents how the dependent variable Y moves with a unit change in the independent variable X.
- The intercept term (β_0), which is also known as the constant, is the y-intercept value of the regression line. It is simply the expected value of Y when X is equal to zero.
- The error term (ε), which is also known as the residual, represents the vertical distance of an individual data point from the fitted regression line. It is simply the absolute value of the difference between the actual Y value and the predicted Y value. This term exists since the regression line cannot perfectly fit all data points with a hundred percent accuracy.

Equation 4.3a is considered the true or theoretical regression equation. However, due to the randomness involved in sampling, the true regression equation cannot be known. Without knowing the true coefficient estimates, it is not possible to find the true error term. Thus, using the data, the estimated regression equation is produced as follows:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 \times X_i + \hat{\varepsilon}_i \qquad (4.4a)$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \times X_i \qquad (4.4b)$$

where the hats over $Y, \beta_0, \beta_1, and\ \varepsilon$ show that these values are estimated or predicted values.

Note that the predicted value of Y does not include the error term (residual).

4.1.2 Ordinary Least Squares Method

The perfect linear regression model reveals the anticipated link between the predictor variables and the outcome, which was unknown before model construction. Estimating this connection or the unknown parameters (*β)* is the primary objective of the linear regression study. In order to provide an accurate estimate, this estimating procedure needs a rule or criterion that is based on reality. The most common one is the Ordinary Least Squares (OLS) method of estimation (Arkes, 2019). The "Least Squares" part refers to minimizing the sum of the square of residuals across all observations in the data frame.

While using the OLS method, the following equation is used to estimate the slope of the regression line:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{4.5}$$

To derive the coefficient of estimate for β_0, the centroid feature of OLS is used:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} \tag{4.6a}$$

In other words, the regression line under OLS goes through the point that has average X and Y values, and the equation 3.6a can be reorganized as follows:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{4.6b}$$

Therefore, the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ gives the estimated regression equation $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \times X_i$ (Eq 3.4b).

4.1.2.1 Total Variation of the Regression Model in OLS Method

In OLS, the Total Sum of Squares (TSS) is used to examine the variation in the outcome variable. TSS is defined as the sum of the squared deviations from the mean value of the outcome variable:

$$TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \quad (4.7)$$

TSS is similar to the variance of a variable, except that it is not divided by the number of total observations minus one. Then, var (Y) × (n – 1) = TSS.

The total variation (TSS) can be decomposed into two parts:

- Explained Sum of Squares (ExSS) is the total variation that is explained by the regression model.
- Residual Sum of Squares (RSS) is the total variation that remains unexplained by the model. It is the sum of squares of the residuals.

$$TSS = ExSS + RSS \quad (4.8)$$

In the OLS method, the linear regression models are established by finding the set of coefficients that maximizes the amount of total variation explained by the model (ExSS). At the same time, this minimizes the total variation that remains unexplained by the model (RSS). Consequently, “Least Squares” refers to the minimization of the sum of squares of the residuals (RSS).

4.1.3 Measuring the Quality of Fit in Regression Analysis

The square of Pearson product-management correlation coefficient (R^2) is one of the most commonly used goodness-of-fit measures in linear regression analysis. It is also known as the coefficient of determination. R-squared can be defined as the proportion of the total variation in the outcome variable Y that is expressed by the explanatory variable(s) X. R-squared (R^2) can be written as:

$$R^2 = \frac{ExSS}{TSS} = \frac{TSS-RSS}{TSS} = 1 - \frac{RSS}{TSS} \quad (4.9)$$

Having closer data points to the regression line results in a lower RSS value, and the value of ExSS gets closer to the value of TSS. And consequently, the higher the R-squared value becomes.

For the purpose of forecasting using regression models, the R-squared value is used as the indicator of the quality of fit, and obtaining a higher R-squared value is desired.

In Multiple Linear Regression models, where there exist more than one explanatory variables, the same formulas are applicable for TSS, ExSS, RSS, and R-squared.

4.1.4 Multiple Linear Regression Analysis

The simple linear regression model is established by using a single explanatory variable. Whereas, in the multiple linear regression (MLR) model, at least two explanatory variables are used. This is the primary difference between the simple and multiple models.

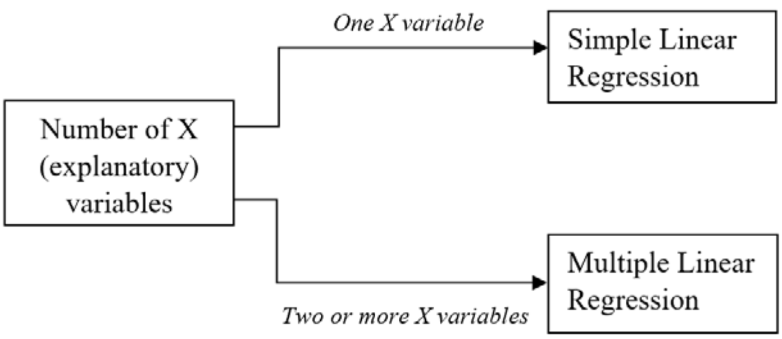

Figure 4.1. Simple and Multiple Linear Regression

The multiple linear regression model is represented by the following equation:

$$Y_i = \beta_0 + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \ldots + \beta_K \times X_{Ki} + \varepsilon_i \qquad (4.10a)$$

Where K refers to the total number of explanatory variables.

In the MLR model, one component of the equation represents a set (a vector) of several variables. For instance, a vector X_i represents the set of multiple explanatory variables. Using linear algebra, the MLR model can be represented by the following equation in matrix form:

$$\underset{n \times 1}{\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}} = \underset{n \times K}{\begin{bmatrix} 1 \; X_{11} \; X_{21} \ldots X_{K1} \\ 1 \; X_{12} \; X_{22} \ldots X_{K2} \\ \cdot \\ \cdot \\ \cdot \\ 1 \; X_{1n} \; X_{2n} \ldots X_{Kn} \end{bmatrix}} \times \underset{K \times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}} + \underset{n \times 1}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}}$$

$$Y = X \times \beta + \varepsilon \qquad (4.10b)$$

where each row represents one observation, and the matrix represents all of the n observations.

For observation 1 the equation can be written as:

$$Y_1 = (1 \times \beta_0 + \beta_1 \times X_{11} + \beta_2 \times X_{21} + \ldots + \beta_K \times X_{K1} + \varepsilon_1) \qquad (4.10c)$$

In MLR models, the explanatory variables are classified into two types:

- The key explanatory (X) variable(s) is the variable that identifies the causal effect. This is often called treatment.
- The control variables exist in the model to help identify the causal effects of the key explanatory variables.

The coefficient estimates are made using the same concept used in the simple linear regression model. These estimates are carried out by minimizing the sum of the squares of residuals (RSS) across all observations in the data frame. That is, the estimates of $\hat{\beta}_0$, $\hat{\beta}_1$, ... , $\hat{\beta}_K$ are selected to minimize $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ or $\sum_{i=1}^{n}(\hat{\varepsilon}_i)^2$.

4.1.5 Assumptions of Regression Models

A list of several assumptions of regression models referred to as "Gauss Markov assumptions" or a similar set of assumptions called "Classical Regression Model assumptions" are often used in typical textbooks of regression analysis (Arkes, 2019).

Assumption 1: The average of residuals (ε) equals zero. This is a rule of the OLS method, and the intercept is automatically arranged to ensure this.

Assumption 2: The residuals are identically and independently distributed. This means that observation has no correlation with another observation.

Assumption 3: The residuals are normally distributed. If the residuals were not normal, this would not mean that the estimates for coefficients are biased. However, the test for significance might be off, and this would be a problem.

Assumption 4: The residuals are homoskedastic. That is, the variance of the residual term (ε) is not correlated with the values of the explanatory variables.

Assumption 5: The key explanatory variable(s) are not correlated with the residuals.

4.2 Application of the MLR Model to HPI Forecasting

Using the selected nine explanatory variables, an OLS Multiple Linear Regression model is developed to predict the HPI of Turkey for the time range of January 2010 to December 2021. As the programming language, Python is used in the development process of the model. The model is developed using the training set that consists of 120 monthly observation values and validated using the training set that includes randomly split 24 monthly observation values.

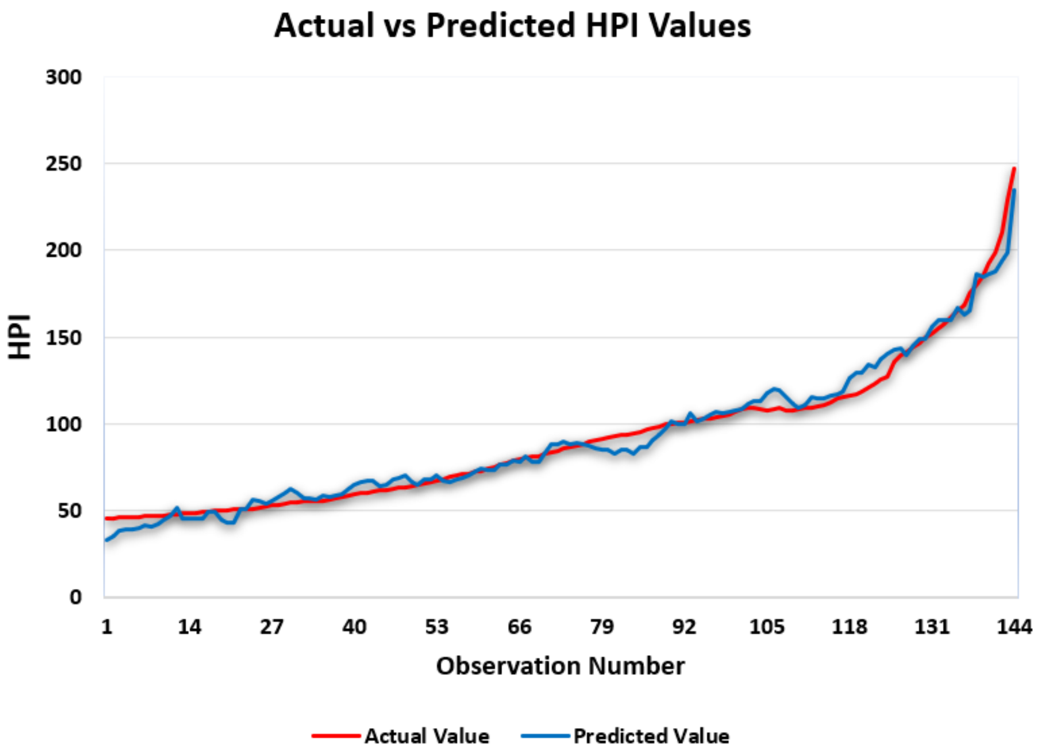

Figure 4.2. OLS Regression Model – Actual vs. Predicted HPI Values

A line chart of the actual HPI values and the predicted HPI values obtained from the developed OLS regression model is demonstrated in Figure 4.2, and the structure of the developed OLS model is illustrated in Figure 4.3.

As mentioned before, the R-squared value is used as an indicator of the quality of fit. In the developed model, an R-squared value of 0.976 is obtained, implying that the quality of fit of the developed OLS regression model is pretty good. Moreover, Prob (F-statistics) shows the probability that the null hypothesis is true. The obtained Prob (F-Statistics) value of the developed model is very close to zero. This implies that the developed OLS regression model is meaningful, and the explanatory variables used in the model have the power to predict the output variable HPI to a certain extent.

Dep. Variable:	Y	**R-squared:**	0.976
Model:	OLS	**Adj. R-squared:**	0.974
Method:	Least Squares	**F-statistic:**	502.4
Date:	Mon, 06 Jun 2022	**Prob (F-statistic):**	5.17e-85
Time:	01:02:04	**Log-Likelihood:**	-394.53
No. Observations:	120	**AIC:**	809.1
Df Residuals:	110	**BIC:**	836.9
Df Model:	9		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	92.2677	42.809	2.155	0.033	7.430	177.106
X1	2.152e-05	1.75e-05	1.232	0.221	-1.31e-05	5.61e-05
X2	0.3242	0.201	1.615	0.109	-0.074	0.722
X3	0.4215	0.021	19.820	0.000	0.379	0.464
X4	-0.0010	0.082	-0.012	0.991	-0.163	0.161
X5	-1.4087	0.709	-1.986	0.049	-2.814	-0.003
X6	-6.0497	0.870	-6.955	0.000	-7.773	-4.326
X7	-0.1683	0.040	-4.221	0.000	-0.247	-0.089
X8	-0.0108	0.006	-1.939	0.055	-0.022	0.000
X9	-0.2504	0.122	-2.058	0.042	-0.491	-0.009

Omnibus:	27.612	**Durbin-Watson:**	2.158
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	56.326
Skew:	0.944	**Prob(JB):**	5.87e-13
Kurtosis:	5.775	**Cond. No.**	5.55e+06

Figure 4.3. Structure of OLS Regression Model

P-value is an indicator of the significance level of the parameters included in the model. The smaller the P-value, the higher the significance level of a parameter. The determination level of the P-value is usually specified as 0.10. Having a P-value smaller than this level implies that the parameter is significant in predicting the output. According to the results of the developed model, the input parameters X3 (Consumer Price Index), X6 (Unemployment Rate), and X7 (Brent Oil Price) have the lowest P-values, which implies that these are the most significant parameters in the prediction of the HPI. By looking at the P-values, the following parameters are X9 (Volatility Index), X5 (Employment Rate), and X8 (Gold Price), respectively. These parameters are also found to be significant in the prediction of HPI since they have P-values smaller than 0.10.

To compare the actual and predicted values, and evaluate the predictive performance of the developed model, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) values are calculated for training and validation sets as measures of forecast accuracy.

Table 4.1. shows the calculated MAE, MAPE, and RMSE values of the developed OLS regression model for training and validation sets. For the training set, MAE, MAPE, and RMSE values are calculated as 4.86, 5.76, and 6.48, respectively. On the other hand, for the validation set, MAE, MAPE, and RMSE values are obtained as 3.10, 4.20, and 4.23, respectively.

Table 4.1. Measures of Accuracy for the OLS Regression Model

	Training	**Validation**
MAE	4.86	3.10
MAPE	5.76	4.20
RMSE	6.48	4.23

In data science, noise refers to variations in the dependent variable that cannot be explained by explanatory variables (Gupta & Gupta, 2019). Obtaining higher accuracy scores in the validation set's findings could indicate that the training set has more noise than the validation set. This is not surprising since the time-series data collection used in the model's development includes only 144 observation values.

4.3 Comparison with Previous Studies

In previous studies, Padilla (2005) predicted the house prices in Canada using a regression model. She obtained a MAPE value of 4% to evaluate the predictive performance of her model. Palakandras et al. (2015) introduced different models, including random walk and bayesian autoregressive models for house price index prediction in the US. The MAPE values obtained for evaluating the accuracy of their models were in the range of 4.73-6.59 for training and 5.35-11.93 for validation. Zhao et al. (2020) introduced autoregressive integrated moving average (ARIMA) models using three different datasets to predict the house price index of New Zealand. The MAPE values obtained for validation of these models were in the range of 2.09-5.77. When these values are compared with the OLS regression model developed in this thesis, it can be observed that the model outperforms or provides similar predictive performance values to the previous predictive models of the house price index introduced in different countries.

CHAPTER 5

PREDICTION OF HPI USING ARTIFICIAL NEURAL NETWORKS

Artificial neural network models are commonly preferred tools for forecasting with time-series data. In this chapter, the ANN model is introduced for predicting the house price index, and the fundamentals of the model are discussed.

5.1 Background of the Artificial Neural Networks (ANN) Model

An artificial neural network is an information processing system that attempts to mimic basic information processing methods used by the human brain. Since the human brain is able to perform complex tasks, artificial neural networks developed based on the human brain have also been found useful for solving complex problems (Samarasinghe, 2007).

5.1.1 Biological Neurons

The primary mission of the human brain is information processing. It works as a system that receives information from the environment by sensors, interprets them, and generates the final result. The fundamental processing elements of this complex system are called neurons.

As shown in Figure 5.1., a biological neuron consists of four basic components, namely dendrites, cell body (or soma), axon, and synapses. Dendrites are responsible for receiving signals, weighted by connection strength, from axons of the other neural cells and transmitting these signals to the cell body. These weighted signals are accumulated and further processed by the cell body. After that, the processed signals are transmitted to the other neurons via axons. The signals move across the neural cells via electrochemical reactions. The synapses have the role of releasing a chemical transmitter that enters the dendrite. This chemical transmitter increases or decreases the electrical potential of the cell body. The cell body collects the inputs, and when a threshold level is reached, it sends an electrical impulse to the axon, usually known as firing. When these impulses reach the synapses, the cycle continues. The connection strengths (input weights) are adjusted by the synapses through this process, and this is accepted as the basis for learning in the human brain.

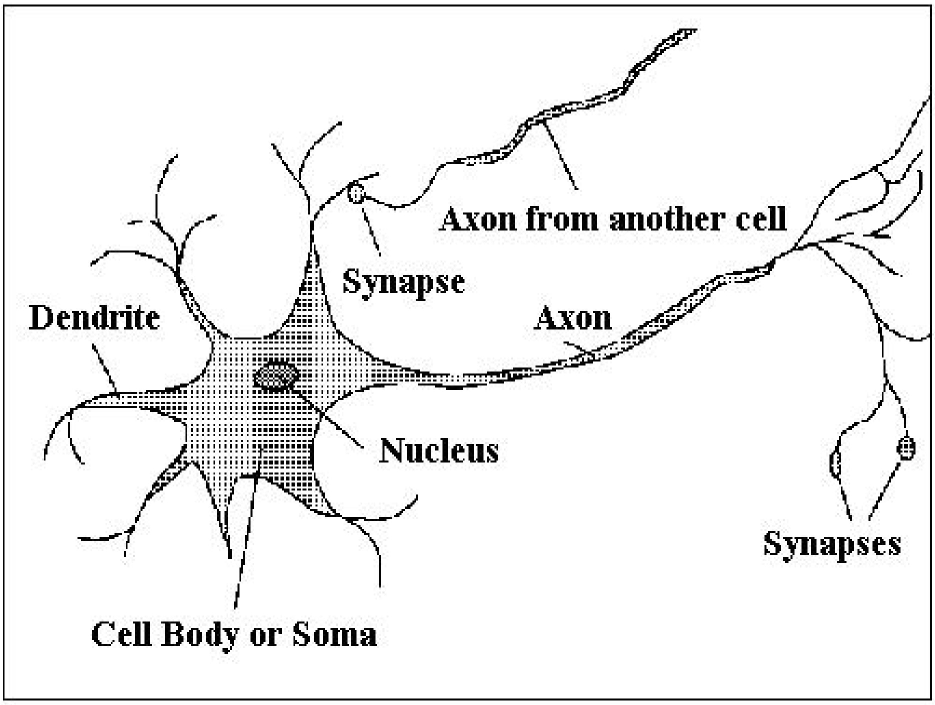

Figure 5.1. A Simple View of a Biological Neuron

5.1.2 Historical Background of ANN

In the 1940s, researchers started to carry out the initial scientific studies on Artificial Neural Networks. McCulloch and Pitts (1943) developed the first neural network model that explains the activity of a neuron in a computational or mathematical way.

In the study, the majority of their effort is concentrated on the behavior of a simple biological neuron. They made several assumptions about how the neurons work. Their model uses an activation function called the hard threshold activation function that returns one as output if the result is over the threshold value and 0 otherwise.

In 1949, Donald Hebb introduced the theory that explains the learning procedure of artificial neural networks in his book (Hebb, 1949). His premise was that the strength of the connection between two neurons increases if they work simultaneously. This theory provided a basis for artificial neural networks that have the ability to learn and adapt.

Using the facilities of IBM, Rochester et al. (1956) attempted to make computer simulations of artificial neural networks regarding Hebb's theory of learning and concept formation.

Together with several other researchers, Rosenblatt (1958) developed a large class of artificial neural networks named perceptrons. After the development of the perceptron model, the studies and developments on artificial neural networks have gained momentum.

Widrow and Hoff (1960) proposed the Adaline Learning Algorithm. In this algorithm, a single-layer neural network exists with multiple nodes where each node receives multiple inputs and generates a single output. It is based on the McCulloch and Pitts neuron and closely related to the perceptron learning rule. Adaline algorithm uses the delta rule for training to minimize the mean-squared error (MSE) between the actual and desired output. The weights and the bias (threshold) are adjustable.

The development of perceptrons was accepted as a success in the early years. However, the mathematical proof of convergence of iterative learning under suitable assumptions revealed the limitations of learning in a perceptron model (Minsky and Papert, 1969). The conclusions resulted in the disenchantment of researchers in the field, and a considerable prejudice was activated.

The 1970s and 1980s were quiet decades for scientific progress in the artificial neural networks field. Yet, some researchers did not lose interest and continued working on the topic.

Werbos (1974) developed and used the algorithm for propagating information about the errors at the output layer back to hidden layers. This algorithm is named the back-propagation learning algorithm. Although this algorithm is the most well-known and widely used in artificial neural networks today, its popularization of it took several years.

Together with David Tank, John Hopfield has developed a number of neural networks using fixed weights and adaptive activation functions. The Hopfield nets can serve as associative memory networks and can be used in the solution of constraint satisfaction problems (Hopfield, 1982, 1987; Hopfield and Tank, 1985, 1986; Tank and Hopfield, 1987).

The Boltzmann machines, developed by Ackley, Hinton, and Sejnowski (1985), are stochastic and generative neural networks capable of learning internal representations to solve complex combinatoric problems, and they are accepted as the fundamentals of the early optimization techniques used in artificial neural networks.

Carpenter and Grossberg (1988) developed the Adaptive Resonance Theory (ART). This theory describes a number of artificial neural network models that use supervised and unsupervised learning algorithms and are used in the solution of problems such as prediction and pattern recognition.

After the 1980s, the field of artificial neural networks started to gain significant attention from researchers, and a large number of studies have been carried out in the field.

5.1.3 Structure of an Artificial Neuron

Modeling of an artificial neuron is based on the information processing concept of a biological neuron. The type of connections made between the neurons (also referred to as the architecture), the process used to determine the weights of the connections (often referred to as the training or learning algorithm), and the activation function are the most important features of the artificial neural networks (Fausett and Elwasif, 1994).

Figure 5.2. demonstrates the computing process of a model neuron. The inputs (signals) are represented by x(n), and each of them is weighted by multiplying by a connection weight indicated by w(n). Each input can be considered as the output of another neuron in the network. The weighted inputs are then accumulated by the summation function Σ in the cell body and further processed by the activation function φ (or f(Σ)) to produce the output (y).

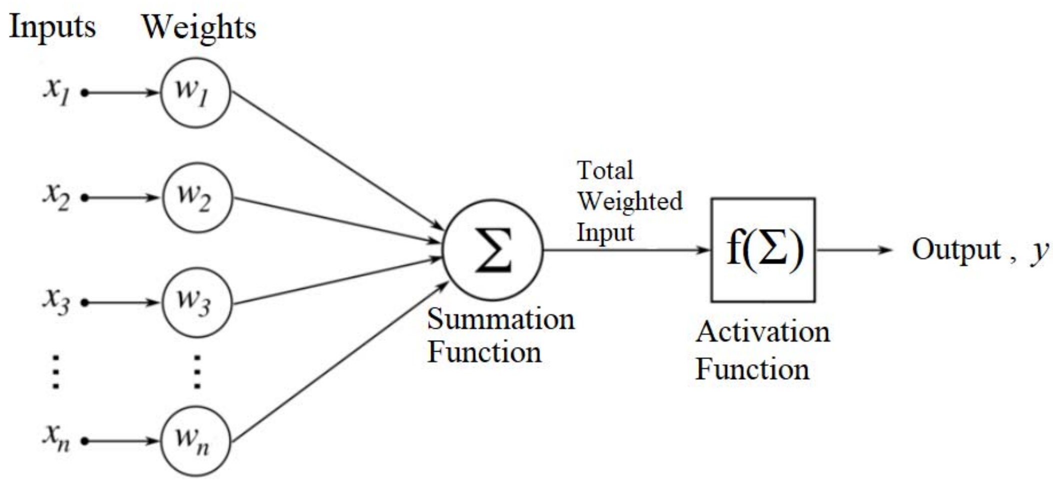

Figure 5.2. Structure of an Artificial Neuron

The other neuron will receive this output (signal) from this neuron and similarly from other neurons in the network and process the information in its cell body, produce output, and pass it to other neurons. This complete process of communication between the neurons in a neural network is demonstrated in Figure 5.3 in a small network consisting of three neurons.

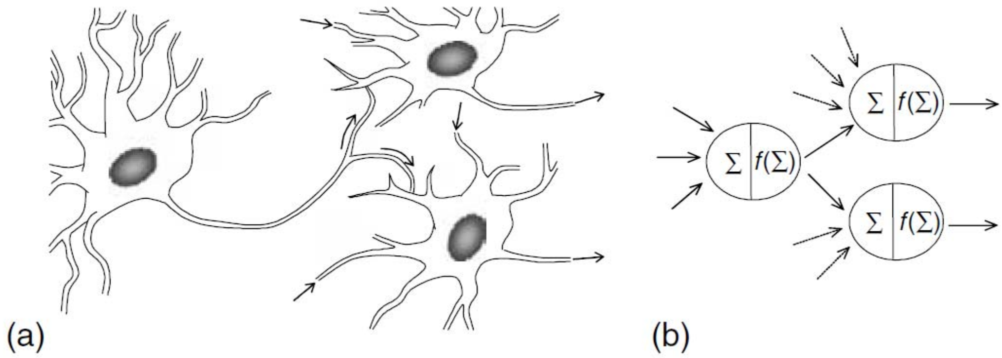

Figure 5.3. Communication between Neurons: (a) Communication of Three Biological Neurons, (b) Neural Network Model

5.1.4 Summation Function

The summation function calculates the net input by summing all inputs received by the artificial neuron multiplied by their weights. Some examples of the summation functions are listed in Table 5.1. There are no certain criteria for the selection of the summation function. It changes depending on the model generated.

Table 5.1. Summation Functions (adapted from Çelik (2018))

Name of the Function	Equation	Explanations
Weighted Total	$NET = \sum x_i . w_i$	Inputs and weight values are multiplied. The calculated values are added to each other.
Multiplication	$NET = \prod x_i . w_i$	Inputs and weight values are multiplied. The calculated values are multiplied.
Maximum	$NET = Max(x_i . w_i)$	Inputs and weight values are multiplied. The highest calculated value is taken.
Minimum	$NET = Min(x_i . w_i)$	Inputs and weight values are multiplied. The lowest calculated value is taken.
Incremental Total	$NET_k = NET_{k-1} + \sum x_i . w_i$	The weighted total is calculated. The previous weighted total is calculated.

5.1.5 Activation Function

As mentioned before, the basic operating system of an artificial neuron involves the summation of weighted inputs and producing and output by applying an activation function. It is also known as the transfer function. Generally, the activation functions are used in all layers of a neural network. The activation functions can be divided into two parts (1) linear activation functions and (2) non-linear activation functions. The commonly used activation functions are shown in Figure 5.4.

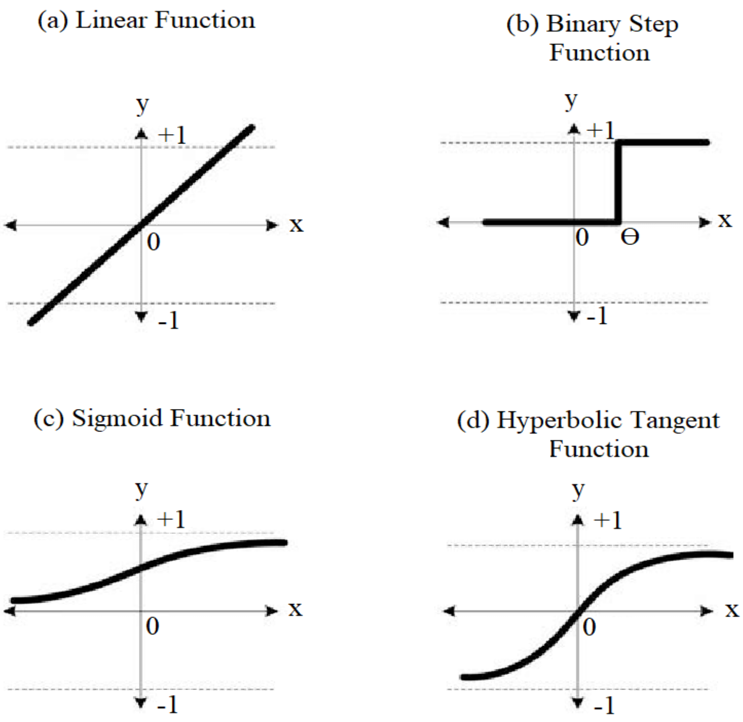

Figure 5.4. Common Activation Functions

5.1.5.1 Linear (Identity) Function

Linear or identity function is is the simplest activation function. It is given by the equation:

$$y = f(x) = x \qquad \text{for all values of x.} \tag{5.1}$$

Therefore, the output of the function will not be confined to a constant range. It takes all values from minus to plus infinity.

5.1.5.2 Binary Step Function

The binary step function is also known as the threshold function or Heaviside function. Single-layer networks usually use a step function to transform the net input to a binary output unit that may take only values of 0 or 1 or a bipolar output unit that may take only values of -1 or 1. The binary step function is defined by the equation:

$$y = f(x) = 1 \quad \text{if } x \geq \Theta$$

$$y = f(x) = 0 \quad \text{if } x < \Theta \tag{5.2}$$

where Θ represents the threshold value.

5.1.5.3 Sigmoid Function

The sigmoid function, also called binary sigmoid or logistic sigmoid function, is the most commonly used form of an activation function that is used to construct ANN models. It looks like an S-shaped curve that confines the output into a range between 0 and 1. The sigmoid function is defined by the equation:

$$y = f(x) = \frac{1}{1+e^{-ax}} \tag{5.3}$$

where a is the slope parameter or steepness parameter of the sigmoid function.

5.1.5.4 Hyperbolic Tangent Function

Different from the sigmoid function, the hyperbolic tangent function, also called a bipolar sigmoid function, confines the output into a range between -1 and 1. It is defined by the equation:

$$y = f(x) = \frac{1-e^{-2x}}{1+e^{-2x}} = \frac{2}{1+e^{-2x}} - 1 \tag{5.4}$$

There exist many other types of activation functions, including the Sine function, ReLU function, etc. Also, the derivatives or differentials of these functions may also be used in some cases.

5.1.6 Artificial Neural Network Models

The structure of the neural network model includes three types of layers, namely input, output, and hidden layers. All neurons in the layers are connected to the neurons in the next layers. The connection type of neurons is significant in the classification of neural networks. The input layer consists of the neurons that are responsible for receiving information from the other neurons or from the environment. The direction of information flow in neural networks starts from the input layer, passes through the hidden layer (or layers), and ends in the output layer. The neurons in the hidden layer are responsible for collecting information and transferring it to the neurons in the output layer.

5.1.7 Single Layer Neural Networks (Perceptrons)

The neural networks consisting of only input and output layers are called single-layer neural networks or perceptrons (Rosenblatt, 1958). The perceptrons have a very limited ability to solve complex non-linear computational problems. For perceptrons, the threshold value is an important parameter determined by the user and plays an important role in generating the outputs.

5.1.8 Multi-Layer Neural Networks

In multi-layer neural networks, one or more hidden layers exist. In Figure 5.5, the structure of a multi-layer neural network having one hidden layer is demonstrated. The hidden layers give an ability to the neural network model to solve complex non-linear computational problems. These types of neural networks generally use the backpropagation or multilayer network Adaline (Madaline) algorithms for learning. In these algorithms, an iterative process is used to minimize the difference between actual output and desired output. The model keeps updating until the error is minimized.

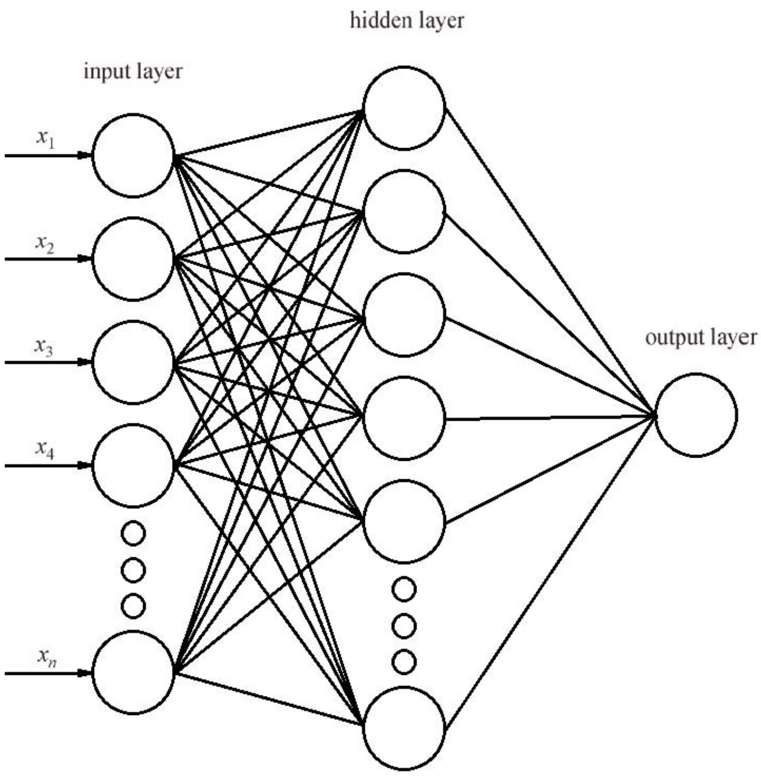

Figure 5.5. Simplified Structure of an Artificial Neural Network

5.1.9 Backpropagation and Feed-Forward Neural Networks

A feed-forward neural network is a type of neural network in which the connections between the neurons are fed forward. Connections that form cycles are not allowed in feed-forward neural networks. The inputs flow from the input layer to the hidden layer and from the hidden layer to the output layer. The backpropagation algorithm is a supervised learning method for multi-layer neural networks. The learning rule is to model the given function by updating the internal weights of the inputs to produce the expected or desired output. The backpropagation algorithm requires the usage of activation functions that are differentiable.

5.2 Application of the ANN Model to HPI Forecasting

By using the selected nine input variables, a backpropagated artificial neural network (ANN) model is utilized to predict the HPI of Turkey for the time range of January 2010 to December 2021. The model is developed with a gradient descent algorithm using online learning (one-by-one sampling update). As the programming language, Matlab is used in the model's development process. The model is developed using the training set that consists of 120 monthly observation values and validated using the validation set that includes randomly split 24 monthly observation values. The details of the developed ANN model are listed below:

- The developed ANN model consists of nine input parameters and one output parameter.
- A single hidden layer exists in the structure of the ANN model.
- The developed ANN model contains nine neurons in a single hidden layer.
- The inputs of the model are normalized to the interval [-1,+1] using all data, including 144 observation values for all parameters.
- No output normalization is applied.
- The activation function in the hidden layer is selected as the sigmoid function
- The activation function in the output layer is selected as a linear y=x function.
- A total number of 10,000 epochs are used.
- The error function is selected as 0.5 x [$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$].
- The weights are randomly initialized on the interval [-1,+1].
- The learning rate is selected as 0.005.

In order to check the ANN model’s ability on unseen data and eliminate the possible problems such as overfitting, memorizing, and selection bias, the cross-validation method is applied to the time-series data. According to this, the time-series data set is split into six randomly generated sets (folds) for training and validation. One fold of the six folds has remained for the validation set, and the other five folds are used as the training set. This method is known as leave-one-out cross-validation. In order to validate the consistency of the training set, the five folds in the training set of the model are used to carry out further training and testing. The flow of the cross-validation algorithm is demonstrated in Figure 5.6. A total of 5 runs are carried out for cross-validation, and in each round, four folds are used for training the model, and one fold is used for testing the model.

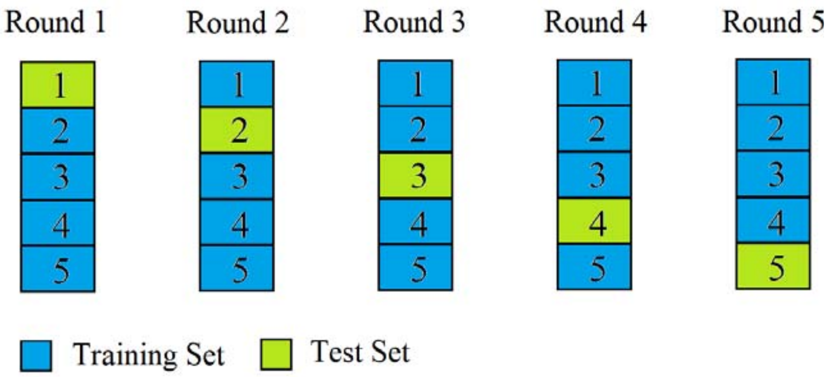

Figure 5.6. The Flow of Cross-Validation Algorithm

Table 5.2. shows the calculated MAE, MAPE, and RMSE values of the developed backpropagated ANN models for training and test sets of each round of cross-validation. In each round, an ANN model is generated using four folds of data, including 96 samples, and the predictive accuracy of the models is evaluated using the remaining one fold of data, including 24 samples, as the test set. The minimum and maximum values of the MAE are calculated as 0.17 and 0.22 for the training sets and 1.38 and 2.11 for the test sets, respectively. The minimum and maximum values of the MAPE are calculated as 0.25 and 0.33 for the training sets and 1.64 and 1.97 for the test sets, respectively. The minimum and maximum values of the RMSE are calculated as 0.24 and 0.35 for the training sets and 2.20 and 3.79 for the test set, respectively. This implies the consistency and useability of the training data set for the development of the final ANN model for predicting the HPI.

Table 5.2. Measures of Accuracy for Cross-Validation

A. Round 1

CV-Round 1	**# of Samples**	**MAE**	**MAPE**	**RMSE**
Training	96	0.17	0.25	0.24
Test	24	1.88	1.97	2.75

B. Round 2

CV-Round 2	**# of Samples**	**MAE**	**MAPE**	**RMSE**
Training	96	0.21	0.30	0.31
Test	24	1.79	1.66	2.95

C. Round 3

CV-Round 3	**# of Samples**	**MAE**	**MAPE**	**RMSE**
Training	96	0.22	0.33	0.35
Test	24	1.38	1.64	2.20

D. Round 4

CV-Round 4	**# of Samples**	**MAE**	**MAPE**	**RMSE**
Training	96	0.18	0.27	0.26
Test	24	1.92	1.95	3.09

E. Round 5

CV-Round 5	**# of Samples**	**MAE**	**MAPE**	**RMSE**
Training	96	0.20	0.28	0.32
Test	24	2.11	1.84	3.79

After cross-validation of the training set, the final ANN model is developed. The actual HPI values and the predicted HPI values obtained using the developed ANN model are demonstrated in the line graph given in Figure 5.7.

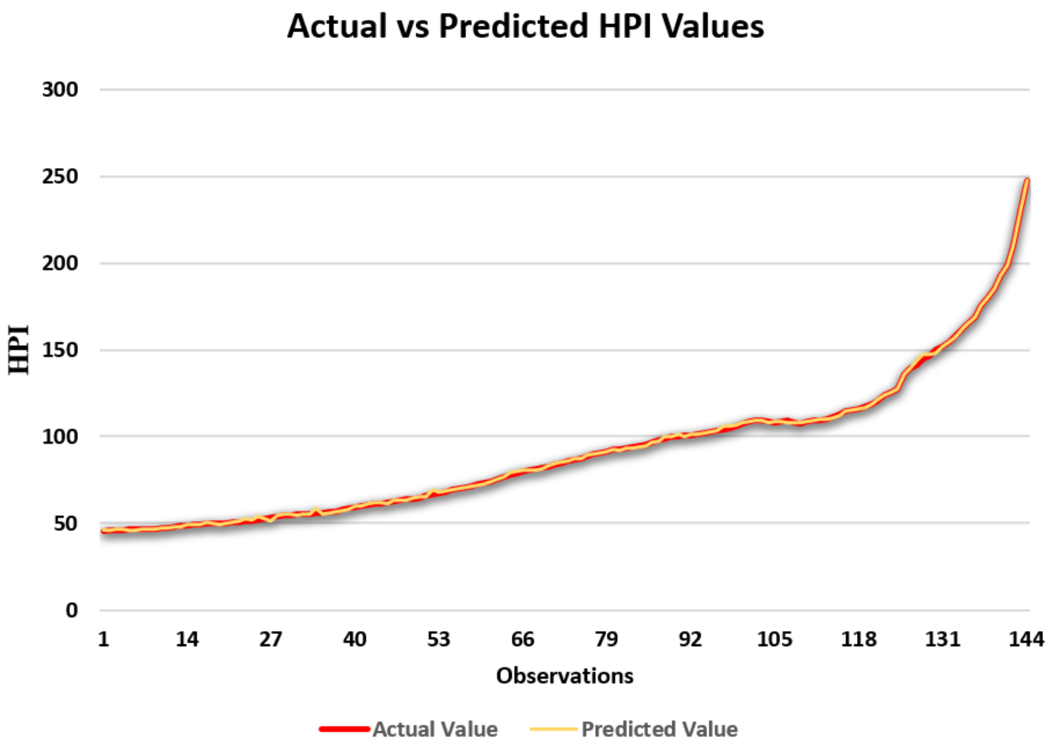

Figure 5.7. ANN Model – Actual vs. Predicted HPI Values

The structure of the developed ANN model consists of nine inputs and nine neurons in a single hidden layer. The weight coefficients that are input to the hidden layer of the model are shown in Table 5.3. X1, X2, …, and X9 represent the input parameters of the model. Whereas N1, N2, …, and N9 represent the neurons in the hidden layer.

The weight coefficient is an indicator of how significant a change in the input will have an influence on the output. A change in an input having a low weight coefficient will have a negligible effect on the output. However, a change in an input having a larger weight coefficient will have a significant effect on the output. By looking at the weight coefficients of the inputs in Table 5.3., it can be observed that X3 (Consumer Price Index) is the most significant input parameter for the prediction of the HPI. The weight coefficients of the neurons that are hidden in the output layer are given in Table 5.4.

Table 5.3. Weight Coefficients of the ANN Model (Input to Hidden Layer)

	N1	N2	N3	N4	N5	N6	N7	N8	N9
X1	-0.59471	0.40436	1.09763	-4.26910	-1.79831	5.21926	0.46058	1.25233	0.76678
X2	-4.40375	-0.07844	-0.71280	-3.20351	-0.10627	-3.70212	2.39580	0.96633	1.05310
X3	6.74590	6.71780	17.92036	-3.51655	-1.32653	5.96138	6.03671	6.29593	17.32605
X4	1.42315	0.11589	0.71282	-1.32735	-0.25710	1.08276	-0.42284	-0.54115	-1.67935
X5	2.24793	1.32703	-1.53840	-1.24773	0.12896	3.35882	-0.50154	0.59095	1.04030
X6	0.98978	0.69137	-1.50376	4.11296	-0.06562	-0.14061	-0.77816	0.53174	0.26530
X7	-0.79236	0.44710	0.59242	-6.78408	-0.49161	0.08780	2.23629	-0.15824	-1.00552
X8	0.81037	1.78170	-0.29810	-2.36323	0.40209	1.59341	0.43728	3.54939	-1.46652
X9	0.02195	0.14809	-1.18050	2.46871	-1.60298	0.93218	-0.46679	0.25381	0.86723
C	-6.73934	4.29407	14.15633	6.67507	5.81280	-5.72026	-1.42111	-0.66592	11.06614

Table 5.4. Weight Coefficients of the ANN Model (Hidden to Output Layer)

	Weight Coefficient
N1	35.95419
N2	23.15622
N3	17.36758
N4	14.05568
N5	15.91676
N6	37.90786
N7	34.90771
N8	36.48366
N9	21.92375
Const.	15.36029

To compare the actual and predicted values, and evaluate the predictive performance of the developed model, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) values are calculated for training and validation sets as measures of forecast accuracy.

Table 5.5. shows the calculated MAE, MAPE, and RMSE values of the developed backpropagated ANN model for training and validation sets. MAE, MAPE, and RMSE values are calculated as 0.35, 0.48, and 0.46 for the training set. On the other hand, for the validation set, MAE, MAPE, and RMSE values are obtained as 1.12, 1.58, and 1.50, respectively.

Table 5.5. Measures of Accuracy for the Backpropagated ANN Model

	Training	Validation
MAE	0.35	1.20
MAPE	0.48	1.58
RMSE	0.46	1.50

5.3 Comparison with Previous Studies

In previous studies, Radzi et al. (2012) predicted the house prices index of Malesia using a backpropagated ANN model. They obtained a MAPE value of 8% to validate the predictive performance of their model. Li and Chu (2017) introduced backpropagated ANN models for predicting two different house price indices in Taiwan. For evaluating the predictive accuracy of these two models, they obtained MAPE values of 1.71 and 2.06 and RMSE values of 2.05 and 2.58. When these values are compared with the backpropagated ANN model developed in this thesis, it can be observed that the model outperforms the previously developed ANN models for predicting the house price index.

Moreover, the results of all three measures of accuracy indicate that the predictive performance of the developed backpropagated ANN model is considerably better than the developed OLR regression model, as will be discussed in Chapter 6.

CHAPTER 6

DISCUSSION OF FINDINGS

In the scope of this thesis, two independent models are developed for predicting the house price index (HPI) of Turkey using time series data of nine explanatory variables. The first model is the multiple linear regression (MLR) model, which is developed using OLS regression, and the other model is the backpropagated artificial neural networks (ANN) model developed with a gradient descent algorithm using online learning.

The results of the developed MLR and ANN models revealed that the input parameter X3 (Consumer Price Index) is the most significant variable in the determination of HPI movements in Turkey. This supports the previous studies claiming that there is a strong relationship between the consumer price index and house price movements (Abelson et al., 2005; Aliefendioğlu et al., 2022; Çetin, 2021; Dougherty & Order, 2016; Hacıevliyagil et al., 2022; McGibany & Nourzad, 2004; Milunovich, 2020; Lai et al., 2013; Sağlam & Abidinoğlu, 2020; Wang et al., 2021, Wei & Cao, 2017). This finding is not surprising since the monthly change in the consumer price index represents the rate of inflation in Turkey. According to the results of the developed models, the second most significant parameter in the determination of HPI movements in Turkey is X6 which stands for the unemployment rate, which is an important economic parameter for all economies in the world. This finding is in line with the findings of the previous studies suggesting that the house price movements can be determined significantly by using the unemployment rate (Abelson et al., 2005; Aderibigbe & Chi, 2018 Ciarlone, 2015; Hacıevliyagil et al., 2022; Hanewald & Sherris, 2013; Kestel & Yilmaz, 2020; Özgüler et al., 2022; Plakandaras et al., 2015; Rapach & Strauss, 2007; Rapach & Strauss, 2009; Wei & Cao, 2017). The results of the models also revealed that the third most significant parameter in the determination of HPI movements in Turkey is X7 which stands for the Brent Oil Price, which is an important global economic parameter for all countries. The effect of oil prices on house price movements is also studied by Padilla (2005), Beltratti and Moorana (2010), and Plakandras et al. (2015). Their results were similar to the results obtained in this thesis, indicating that there is an important relationship between the change in global oil prices and house price movements.

The actual observation values and predicted values of HPI obtained using the MLR and ANN models are illustrated in Figure 6.1. As it can be seen from Figure 6.1., the closeness of fit to the actual value is much better in the ANN model.

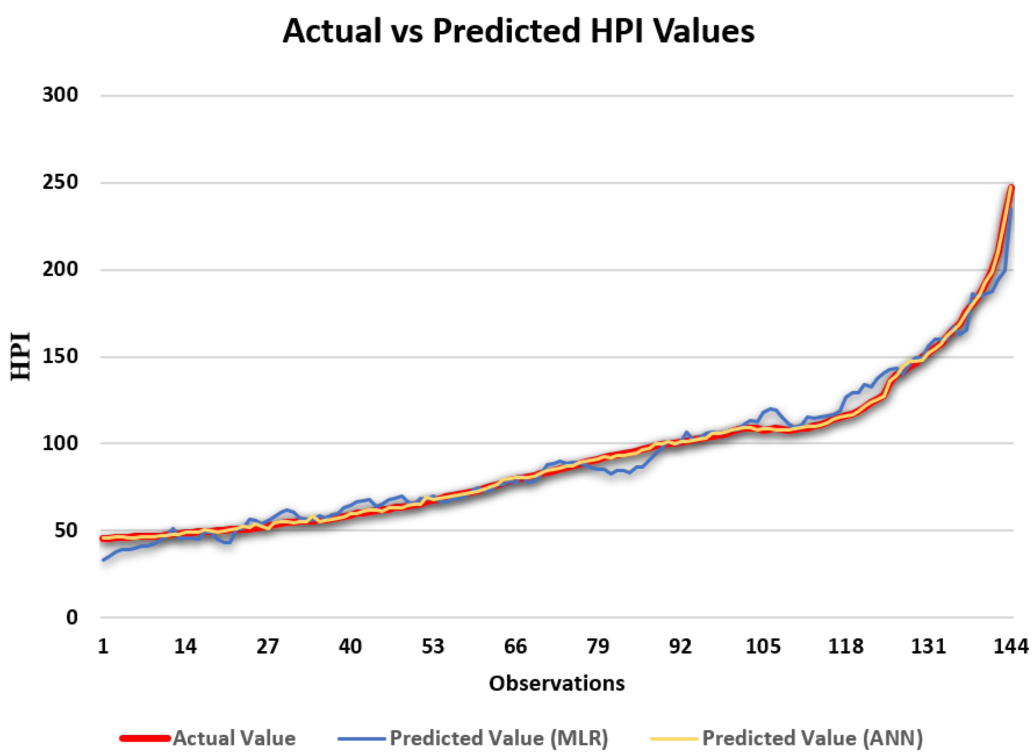

Figure 6.1. Actual and Predicted Values of HPI

In order to evaluate the predictive performance of the models, the predicted values generated using the validation set of the data are compared. As can be seen from Table 6.1., MAE, MAPE, and RMSE values of the MLR model are 3.10, 4.20, and 4.23, respectively. Whereas MAE, MAPE, and RMSE values of the ANN model are 1.20, 1.58, and 1.50, respectively. The results of all three measures of accuracy indicate that the predictive performance of the developed ANN model is considerably better than the developed OLR regression model. This is not unexpected, given that the ANN model, unlike the MLR model, can handle nonlinearities in the data. When the MAPE values are considered as the measure of accuracy, the predictive performance of the ANN model developed in this thesis also outperforms the previous ANN models introduced by Radzi et al. (2012) and Li & Chu (2017) to predict the house price indices of the eastern countries using macroeconomic parameters.

Table 6.1. Measures of Accuracy for the Developed Models (Validation Set)

	MLR Model	**ANN Model**
MAE	3.10	1.20
MAPE	4.20	1.58
RMSE	4.23	1.50

CHAPTER 7

CONCLUSIONS

In the last few years, Turkey has faced an astronomic amount of inflation. In parallel with this, in March 2022, the house price index (HPI) of Turkey increased by 110% annually, which is more than doubled (CBRT, 2022). The extreme volatility in the housing prices increases the uncertainty in the residential construction industry and makes it more difficult for the industry practitioners to make revenue estimations of the projects in the early stages. Correspondingly, the primary objective of this thesis is to investigate the macroeconomic determinants behind the house price movements in Turkey. For this purpose, the time period between January 2010 and December 2021 is selected as the study period, and two empirical models are utilized to predict the monthly movements of the house price index of Turkey. The first model is the multiple linear regression (MLR) model developed using OLS regression, and the other model is the backpropagated artificial neural networks (ANN) model developed with a gradient descent algorithm using online learning. Due to the MLR model's ease of implementation, it is chosen to be employed in this study. The ANN model, on the other hand, is chosen since it produces forecasts with a high level of accuracy utilizing time-series data.

After analyzing the collected time series data of 25 candidate input parameters, the models are generated by using the selected nine input parameters. The selected nine input parameters include the number of housing permits (NHP), consumer confidence index (CCI), consumer price index (CPI), industrial production index (IPI), the employment rate (ER), unemployment rate (UR), Brent oil price (BOP), gold price (GP), and the volatility index (VIX).

Overall, the findings of this study show that among the nine input parameters employed, the consumer price index (CPI) is the most significant macroeconomic parameter in determining the house price index (HPI) movements in Turkey. This parameter is followed by two more significant parameters, unemployment rate (UR) and Brent oil price (BOP), respectively. When the predictive performance of the utilized models is compared, the backpropagated artificial neural networks (ANN) model outperforms the multiple linear regression (MLR) model. This is not unexpected because the ANN model, unlike the MLR model, is able to handle the nonlinearities in the data. Additionally, the ANN model built in this thesis shows superior predictive accuracy when compared to earlier ANN models developed to estimate the housing price index of different nations.

The findings of this thesis contribute to theory by filling the gap in the literature on macroeconomic determinants of house prices in Turkey by identifying the significant parameters. Moreover, the performance of prediction methods and strategies used to develop models may help researchers from other countries to develop similar models. Besides, this thesis contributes to practice by introducing two empirical models that can predict the house price movements in Turkey effectively with high predictive accuracy. These MLR and ANN models can be used by industry practitioners and policy makers in the residential construction industry to constantly monitor the house price movements in the market.

As a future work, it would be interesting to evaluate the predictive accuracy of the developed MLR and ANN models by predicting the monthly values of the house price index of Turkey using the forthcoming years' time series data and comparing them with the actual values. It would also be interesting to develop additional empirical models on the same data, such as VEC, SVR, ARDL, etc., and compare their results. Furthermore, the development of MLR and ANN models on time series data obtained for G20 countries having emerging economies similar to Turkey, such as Argentina, Brazil, India, Indonesia, Mexico, Saudi Arabia, and South Africa, and a comparison of findings would be a fruitful research topic.

REFERENCES

Abelson, P., Joyeux, R., Milunovich, G., & Chung, D. (2005). Explaining house prices in Australia: 1970–2003. *Economic record*, *81*, S96-S103.

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive science*, *9*(1), 147-169.

Adams, Z., & Füss, R. (2010). Macroeconomic determinants of international housing markets. *Journal of Housing Economics*, *19*(1), 38-50.

Aderibigbe, T., & Chi, H. (2018). Investigation of Florida housing prices using predictive time series model. In *Proceedings of the Practice and Experience on Advanced Research Computing* (pp. 1-4).

Afşar, A. (2018). Türkiye'de konut fiyatlarını belirleyici ekonomik faktörlerin analizi. *Sosyal, Beşeri ve İdari Bilimler'de Akademik Araştırmalar.5*(1), 129-146.

Agnello, L., & Schuknecht, L. (2011). Booms and busts in housing markets: Determinants and implications. *Journal of Housing Economics*, *20*(3), 171-190.

Aliefendioğlu, Y., Tanrivermis, H., & Salami, M. A. (2021). House price index (HPI) and Covid-19 pandemic shocks: evidence from Turkey and Kazakhstan. *International Journal of Housing Markets and Analysis*.

Arkes, J. (2019). *Regression analysis: A practical introduction.* Routledge.

Badurlar Öner, İ. (2008). Türkiye'de Konut Fiyatları ile Makro Ekonomik Değişkenler Arasındaki İlişkinin Araştırılması. *Anadolu üniversitesi sosyal bilimler dergisi*, *8*(1), 223-238.

Banking Regulation and Supervision Agency (BRSA). (n.d.). Monthly Banking Sector Data. Retrieved June 2, 2022, from https://www.bddk.org.tr/BultenAylik/en

Bayır, M. (2019). Türkiye'de Para Politikasının Konut Fiyatlarına Etkisi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *14*(3), 657-674.

Bayır, M., Güvenoğlu, H. & Kutlu, Ş. Ş. (2019). Konut Fiyatlarının Belirleyicileri Üzerine Ampirik Bir Analiz. *In II. International Conference on Empirical Economics and Social Science* (pp. 760-774).

Beltratti, A., & Morana, C. (2010). International house prices and macroeconomic fluctuations. *Journal of Banking & Finance*, *34*(3), 533-545.

Canbay, Ş., & Mercan, D. (2020). Türkiye'de Konut Fiyatları, Büyüme Ve Makroekonomik Değişkenler Arasındaki İlişkinin Ekonometrik Analizi. *Yönetim ve Ekonomi Araştırmaları Dergisi*, *18*(1), 176-200.

Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, *21*(3), 77-88.

Case, K. E., & Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics*, *18*(3), 253-273.

Case, K. E., & Mayer, C. J. (1996). Housing price dynamics within a metropolitan area. *Regional Science and Urban Economics*, *26*(3-4), 387-407.

Çelik, Ö. (2018). A research on machine learning methods and its applications. *Journal of Educational Technology and Online Learning*, *1*(3), 25-40.

Çetin, A. C. (2021). Türkiye'de Konut Fiyatlarına Etki Eden Faktörlerin Analizi. *Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi*, *5*(1), 1-30.

Central Bank of The Republıc of Turkey (CBRT). (n.d.). Electronic Data Delivery System. Retrieved June 2, 2022, from https://www.tcmb.gov.tr/wps/wcm/connect/en/tcmb+en

Ciarlone, A. (2015). House price cycles in emerging economies. *Studies in Economics and Finance*.

Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example, 5th edn.* John Wiley & Sons.

Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of regression analysis*. John Wiley & Sons.

Coskun, Y., Seven, U., Ertugrul, H. M., & Alp, A. (2020). Housing price Dynamics and bubble risk: the case of Turkey. *Housing Studies*, *35*(1), 50-86.

Dilber, İ., & Sertkaya, Y. (2016). 2008 Finansal Krizi Sonrası Türkiye'de Konut Fiyatlarının Belirleyicilerine Yönelik Analiz. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, *4*(1), 11-29.

Dougherty, A., & Van Order, R. (1982). Inflation, housing costs, and the consumer price index. *The American Economic Review*, *72*(1), 154-164.

Fausett, L. V., & Elwasif, W. (1994, June). Predicting performance from test scores using backpropagation and counterpropagation. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)* (Vol. 5, pp. 3398-3402). IEEE.

Gok, I. Y., & Keceli, A. (2015). Determinants of house prices in Turkey: comparative analysis of development regions. In *The 2015 WEI International Academic Conference Proceedings* (pp. 110-122).

Gupta, S., & Gupta, A. (2019). Dealing with noise problem in machine learning data-sets: A systematic review. *Procedia Computer Science*, *161*, 466-474.

Hacıevliyagil, N., Drachal, K., & Eksi, I. H. (2022). Predicting House Prices Using DMA Method: Evidence from Turkey. *Economies*, *10*(3), 64.

Hanewald, K., & Sherris, M. (2013). Postcode-Level House Price Models for Banking and Insurance Applications. *Economic Record*, *89*(286), 411-425.

Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, *4*, 60-78.

Helleiner, E. (2011). Understanding the 2007–2008 global financial crisis: Lessons for scholars of international political economy. *Annual review of political science*, *14*, 67-87.

Hepsen, A., & Bas, N. K. (2009). Housing market activity and macroeconomic variables: an analysis of Turkish dwelling market under new mortgage system. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, *38*(1), 38-46.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.

Hopfield, J. J. (1987). Learning algorithms and probability distributions in feed-forward and feed-back networks. *Proceedings of the national academy of sciences*, *84*(23), 8429-8433.

Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits: A model. *Science*, *233*(4764), 625-633.

Hopfield, J. J., & Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. *Biological cybernetics*, *52*(3), 141-152.

Investing.com. (n.d.). World Financial Markets. Retrieved June 2, 2022, from https://www.investing.com/markets/

Jud, G. D., & Winkler, D. T. (2002). The dynamics of metropolitan housing prices. *The journal of real estate research*, *23*(1/2), 29-46.

Kolcu, F., & Yamak, N. (2018). Gelir Ve Faiz Oranlarının Konut Fiyatları Üzerindeki Kısa Ve Uzun Dönem Etkileri. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 141-152.

Lai, J., Zhang, L., Duffield, C. F., & Aye, L. (2013). Engineering Reliability Analysis in Risk Management Framework: Development and Application in Infrastructure Project. *IAENG International Journal of Applied Mathematics*, *43*(4).

Li, L., & Chu, K. H. (2017, May). Prediction of real estate price variation based on economic parameters. In *2017 International Conference on Applied System Innovation (ICASI)* (pp. 87-90). IEEE.

Lin, W. Y., & Tsai, I. C. (2019). Trader differences in shanghai's a-share and b-share markets: Effects on interaction with the shanghai housing market. *Journal of Asian Economics*, *64*, 101-128.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133.

McDonald, J. F., & McMillen, D. P. (2010). *Urban economics and real estate: theory and policy*. John Wiley & Sons.

McGibany, J. M., & Nourzad, F. (2004). Do lower mortgage rates mean higher housing prices?. *Applied Economics*, *36*(4), 305-313.

Meidani, A. A. N., Zabihi, M., & Ashena, M. (2011). House prices, economic output, and inflation interactions in Iran. *Research in Applied Economics*, *3*(1), 1.

Mikhed, V., & Zemčík, P. (2009). Do house prices reflect fundamentals? Aggregate and panel data evidence. *Journal of Housing Economics*, *18*(2), 140-149.

Milunovich, G. (2020). Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, *39*(7), 1098-1118.

Minsky, M., & Papert, S. (1969). *Perceptrons.* M.I.T. Press.

Neter, J., Wasserman, W., & Whitmore, G. A. (1993). *Applied statistics* (pp. 531-649). Boston: Allyn and Bacon.

Nyakabawo, W., Miller, S. M., Balcilar, M., Das, S., & Gupta, R. (2015). Temporal causality between house prices and output in the US: A bootstrap rolling-window approach. *The North American Journal of Economics and Finance*, *33*, 55-73.

Padilla, M. A. (2005). *The effects of oil prices and other economic indicators on housing prices in Calgary, Canada* (Doctoral dissertation, Massachusetts Institute of Technology).

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications*, *42*(6), 2928-2934.

Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the US real house price index. *Economic Modelling*, *45*, 259-267.

Radzi, M. S. M., Muthuveerappan, C., Kamarudin, N. & Mohammad, I. S. (2012). Forecasting the House Price Index Using Artificial Neural Networks. *International Journal of Real Estate Studies*, *7*(1), 43-48.

Rapach, D. E., & Strauss, J. K. (2007). Forecasting real housing price growth in the eighth district states. *Federal Reserve Bank of St. Louis. Regional Economic Development*, *3*(2), 33-42.

Rapach, D. E., & Strauss, J. K. (2009). Differences in housing price forecastability across US states. *International Journal of Forecasting*, *25*(2), 351-372.

Raykar, V. C., & Saha, A. (2015). Data Split Strategies for Evolving Predictive Models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 3-19). Springer, Cham.

Rochester, N., Holland, J., Haibt, L., & Duda, W. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on information Theory*, *2*(3), 80-93.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Sağlam, C., & Abdioğlu, Z. (2020). Türkiye'de Tüketici Fiyatları ile Hedonik Konut Fiyatları Arasındaki İlişki: Panel Veri Analizi. *Journal of Yaşar University*, *15*(57), 117-128.

Samarasinghe, S. (2016). Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications.

Sari, R., Ewing, B. T., & Aydin, B. (2007). Macroeconomic variables and the housing market in Turkey. *Emerging markets finance and trade*, *43*(5), 5-19.

Smith, L. B., Rosen, K. T., & Fallis, G. (1988). Recent developments in economic models of housing markets. *Journal of economic literature*, *26*(1), 29-64.

Stevenson, S. (2000). A long-term analysis of regional housing markets and inflation. *Journal of Housing Economics*, *9*(1-2), 24-39.

Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, *84*(7), 1896-1900.

Tunc, C. (2020). The effect of credit supply on house prices: Evidence from Turkey. *Housing Policy Debate*, *30*(2), 228-242.

Turkish Statistical Institute (TURKSTAT). (n.d.). Data Portal for Statistics. Retrieved June 2, 2022, from https://data.tuik.gov.tr/

Varlık, N. (2020). Ekonomik büyümenin konut fiyatları üzerindeki asimetrik etkisi: NARDL uygulaması. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *13*(2), 352-367.

Wang, X., Gao, S., Zhou, S., Guo, Y., Duan, Y., & Wu, D. (2021). Prediction of House Price Index Based on Bagging Integrated WOA-SVR Model. *Mathematical Problems in Engineering*, *2021*.

Wei, Y., & Cao, Y. (2017). Forecasting house prices using dynamic model averaging approach: Evidence from China. *Economic Modelling*, *61*, 147-155.

Werbos, P. (1974). Beyond regression:" new tools for prediction and analysis in the behavioral sciences. *Ph. D. dissertation, Harvard University*.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Stanford Univ Ca Stanford Electronics Labs.

Yeap, G. P., & Lean, H. H. (2020). Nonlinear relationship between housing supply and house price in Malaysia. *International Journal of Strategic Property Management*, *24*(5), 313-322.

Yilmaz, B. & Kestel, A.S. (2020). Forecasting house prices in Turkey: GLM, VaR and time series approaches. *Journal of Business Economics and Finance*, *9*(4), 274-291.

Zhao, L., Mbachu, J., & Liu, Z. (2020). Modelling Residential Building Costs in New Zealand: A Time-Series Transfer Function Approach. *Mathematical Problems in Engineering*, *2020*.

Zheng, X., & Tian, H. (2018, November). House Price Forecast Based on Dynamic Model Averaging Model Combined With Web Search Index. In *2018 International Conference on Cloud Computing, Big Data and Blockchain (ICCBB)* (pp. 1-14). IEEE.

APPENDICES

A. Time Series Data Used in the Model

#	Month	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	Jan-10	45.40	34,373	84.32	174.07	56.84	40.70	12.10	71.20	1104.31	24.62
2	Feb-10	45.60	38,920	86.45	176.59	57.51	41.00	12.10	76.36	1090.25	19.50
3	Mar-10	46.00	55,904	89.08	177.62	68.08	41.30	11.70	80.37	1110.81	17.59
4	Apr-10	46.20	58,425	89.92	178.68	66.12	41.10	11.30	86.19	1151.68	22.05
5	May-10	46.50	55,622	91.21	178.04	69.05	41.20	11.10	73.00	1206.50	32.07
6	Jun-10	46.60	66,074	92.68	177.04	72.17	41.20	10.90	74.94	1233.38	34.54
7	Jul-10	46.70	59,607	92.71	176.19	72.69	41.20	11.00	77.50	1191.80	23.50
8	Aug-10	46.80	53,663	92.78	176.90	70.04	41.20	11.00	75.51	1220.13	26.05
9	Sep-10	46.90	53,238	95.21	179.07	67.70	41.20	10.90	80.77	1270.50	23.70
10	Oct-10	47.20	58,228	93.33	182.35	77.30	41.30	10.60	82.47	1336.94	21.20
11	Nov-10	47.50	56,662	95.47	182.40	70.03	41.70	10.00	86.02	1379.67	23.54
12	Dec-10	48.10	311,918	94.41	181.85	85.12	41.90	10.00	93.23	1382.42	17.75
13	Jan-11	48.30	33,812	94.45	182.60	70.72	42.50	9.70	98.97	1349.13	19.53
14	Feb-11	48.80	30,523	96.15	183.93	68.54	42.90	9.40	112.27	1376.25	18.35
15	Mar-11	49.00	44,752	95.93	184.70	79.03	42.90	9.30	116.94	1423.63	17.74
16	Apr-11	49.30	45,928	95.80	186.30	76.03	42.80	9.30	126.59	1455.08	14.75
17	May-11	49.60	57,574	95.27	190.81	79.55	42.70	9.20	117.18	1504.00	15.45
18	Jun-11	49.80	60,383	97.97	188.08	82.98	42.70	9.20	111.71	1530.37	16.52
19	Jul-11	50.00	56,357	97.02	187.31	81.88	43.20	9.10	115.93	1568.40	25.25
20	Aug-11	50.00	59,363	94.72	188.67	78.88	43.10	8.80	116.48	1757.70	31.62
21	Sep-11	50.60	41,147	95.72	190.09	80.91	43.10	8.70	105.42	1765.90	42.96
22	Oct-11	50.90	65,799	92.03	196.31	89.86	43.10	8.30	108.43	1678.40	29.96
23	Nov-11	51.10	48,819	92.68	199.70	80.14	42.80	8.50	111.22	1732.38	27.80
24	Dec-11	51.20	103,236	92.82	200.85	92.09	42.90	8.30	108.09	1646.40	23.40
25	Jan-12	51.70	40,532	92.70	201.98	74.98	42.70	8.40	110.26	1657.75	19.44
26	Feb-12	52.50	46,254	94.00	203.12	73.39	42.50	8.50	122.23	1736.50	18.43
27	Mar-12	53.00	62,250	93.30	203.96	83.77	42.80	8.40	123.41	1675.80	15.50
28	Apr-12	53.60	61,642	88.90	207.05	80.30	43.00	8.10	118.66	1650.62	17.15
29	May-12	54.30	95,224	91.80	206.61	86.81	43.20	8.10	103.86	1596.43	24.06
30	Jun-12	54.70	64,728	91.10	204.76	86.15	43.20	8.10	94.17	1588.25	17.08
31	Jul-12	55.10	63,584	91.90	204.29	85.43	43.00	8.00	105.93	1594.93	18.93
32	Aug-12	55.40	55,395	89.90	205.43	76.45	43.00	8.30	113.93	1630.15	17.47
33	Sep-12	55.50	55,158	89.10	207.55	87.10	43.20	8.30	111.36	1766.00	15.73
34	Oct-12	55.70	65,801	85.30	211.62	84.84	43.70	8.40	109.89	1742.95	18.60
35	Nov-12	56.00	69,047	89.90	212.42	91.38	43.90	8.40	110.84	1719.45	15.87
36	Dec-12	56.40	87,811	89.30	213.23	91.67	44.00	8.60	110.80	1676.68	18.02
37	Jan-13	56.90	44,453	91.80	216.74	78.18	44.10	8.60	115.55	1663.50	14.28
38	Feb-13	57.60	50,507	92.00	217.39	77.04	44.30	8.60	112.20	1631.00	15.51
39	Mar-13	58.40	67,282	92.20	218.83	85.94	44.10	8.60	108.46	1593.10	12.70
40	Apr-13	59.20	74,297	92.80	219.75	87.08	43.90	8.80	101.53	1495.12	13.52

41	May-13	60.00	94,605	95.40	220.07	91.88	43.80	8.80	100.43	1409.85	16.30
42	Jun-13	60.60	76,511	94.40	221.75	91.22	43.60	8.80	102.49	1316.12	16.86
43	Jul-13	61.20	73,915	95.90	222.44	93.69	43.40	9.00	107.89	1279.81	13.45
44	Aug-13	61.50	58,349	94.90	222.21	77.59	43.40	9.00	115.97	1334.12	17.01
45	Sep-13	62.00	67,454	91.50	223.91	96.31	43.50	9.00	107.85	1348.93	16.60
46	Oct-13	62.50	56,599	93.70	227.94	87.73	43.50	9.10	107.53	1307.66	13.75
47	Nov-13	63.30	70,813	96.80	227.96	100.39	43.60	9.10	111.07	1275.75	13.70
48	Dec-13	63.60	101,353	94.40	229.01	100.92	43.80	9.00	109.95	1218.76	13.72
49	Jan-14	64.10	68,786	91.80	233.54	87.54	45.00	9.20	108.16	1249.35	18.41
50	Feb-14	64.80	73,925	89.50	234.54	83.36	44.80	9.50	108.98	1307.25	14.00
51	Mar-14	65.70	154,341	92.60	237.18	93.89	45.40	9.50	105.95	1337.75	13.88
52	Apr-14	66.80	73,104	97.40	240.37	92.89	45.40	9.70	108.63	1303.87	13.41
53	May-14	67.50	127,631	95.00	241.32	94.26	45.10	9.70	109.21	1281.20	11.40
54	Jun-14	68.20	86,434	93.20	242.07	95.71	45.80	9.70	111.03	1287.62	11.57
55	Jul-14	69.30	63,771	93.10	243.17	90.64	45.10	9.90	104.94	1314.06	16.95
56	Aug-14	70.20	85,583	93.20	243.40	89.34	44.60	10.20	101.12	1292.00	11.98
57	Sep-14	70.90	88,894	94.10	243.74	102.30	45.70	10.10	94.67	1232.75	16.31
58	Oct-14	71.50	47,633	91.70	248.37	92.38	45.20	10.40	84.17	1209.05	14.03
59	Nov-14	72.40	63,945	90.90	248.82	100.09	45.00	10.50	71.89	1177.50	13.33
60	Dec-14	73.10	93,256	90.40	247.72	108.50	45.50	10.30	55.27	1195.56	19.20
61	Jan-15	74.20	50,907	89.60	250.45	86.73	45.60	10.50	47.52	1244.45	20.97
62	Feb-15	75.40	57,137	89.10	252.24	84.18	45.20	10.70	61.89	1223.93	13.34
63	Mar-15	76.60	78,206	86.70	255.23	99.67	45.50	10.50	53.69	1176.65	15.29
64	Apr-15	77.70	77,419	87.50	259.39	99.26	45.60	10.50	63.90	1198.05	14.55
65	May-15	78.90	79,747	86.70	260.85	98.83	45.70	10.20	63.16	1199.59	13.84
66	Jun-15	79.90	87,949	89.30	259.51	104.51	46.20	10.20	60.31	1180.32	18.23
67	Jul-15	80.60	71,722	87.50	259.74	95.08	46.00	10.10	53.29	1127.85	12.12
68	Aug-15	81.10	74,386	84.80	260.78	101.57	45.90	10.10	47.97	1120.73	28.43
69	Sep-15	81.70	64,141	82.20	263.11	97.06	46.00	10.20	47.29	1122.83	24.50
70	Oct-15	82.70	61,810	86.30	267.20	108.69	46.40	10.20	48.00	1155.35	15.07
71	Nov-15	83.70	73,344	95.20	268.98	107.49	46.20	10.20	43.73	1077.38	16.13
72	Dec-15	84.50	116,659	93.60	269.54	116.93	46.10	10.40	36.61	1070.85	18.21
73	Jan-16	85.70	83,383	92.20	274.44	91.33	46.40	10.20	33.14	1100.91	20.20
74	Feb-16	86.60	85,372	90.00	274.38	95.81	46.00	10.10	35.92	1211.93	20.55
75	Mar-16	87.50	91,892	90.10	274.27	106.41	46.00	10.20	36.75	1254.46	13.95
76	Apr-16	88.60	85,372	91.30	276.42	102.69	46.50	10.10	45.64	1241.82	15.70
77	May-16	89.70	83,875	91.50	278.02	107.11	46.40	10.30	49.26	1256.33	14.19
78	Jun-16	90.50	82,085	91.60	279.33	107.37	46.20	10.70	48.05	1280.55	15.63
79	Jul-16	91.30	53,141	88.90	282.58	86.90	45.90	11.10	40.76	1336.80	11.87
80	Aug-16	92.30	76,824	94.70	281.76	105.49	46.20	11.10	47.94	1339.43	13.42
81	Sep-16	92.80	66,694	94.80	282.27	93.47	45.80	11.60	48.24	1325.01	13.29
82	Oct-16	93.50	78,270	95.30	286.33	113.01	46.30	11.50	46.20	1262.38	17.06
83	Nov-16	94.00	88,368	91.60	287.81	113.72	46.30	11.60	47.95	1234.48	13.33
84	Dec-16	94.80	127,449	87.40	292.54	117.58	46.10	12.10	54.96	1149.51	14.04
85	Jan-17	95.50	70,451	88.70	299.74	97.07	45.90	12.00	55.25	1187.90	11.99
86	Feb-17	96.70	68,839	88.00	302.17	96.25	46.20	11.90	53.36	1234.77	12.92
87	Mar-17	97.90	98,287	90.10	305.24	113.22	45.80	11.90	52.20	1230.22	12.37
88	Apr-17	98.70	98,167	92.00	309.23	110.16	46.40	11.60	49.46	1274.07	10.82

89	May-17	99.80	100,286	94.00	310.61	113.71	47.00	11.10	49.40	1244.08	10.41
90	Jun-17	100.30	230,164	92.30	309.78	105.11	46.50	11.00	47.08	1258.97	11.18
91	Jul-17	100.50	104,432	92.40	310.24	112.48	47.20	10.60	51.99	1239.85	10.26
92	Aug-17	100.80	98,921	93.00	311.85	113.69	47.30	10.50	52.69	1281.22	10.59
93	Sep-17	101.40	330,735	92.00	313.88	110.57	47.20	10.30	57.02	1311.75	9.51
94	Oct-17	102.20	66,382	89.90	320.40	125.87	47.50	10.30	61.35	1277.26	10.18
95	Nov-17	102.80	49,404	87.60	325.18	124.95	47.60	10.10	63.53	1281.58	11.28
96	Dec-17	103.40	81,690	88.20	327.41	130.17	47.60	9.90	66.73	1270.85	11.04
97	Jan-18	104.10	46,580	92.70	330.75	109.17	47.70	9.80	67.78	1333.01	13.54
98	Feb-18	104.80	42,894	93.40	333.17	105.33	47.40	10.00	66.08	1331.32	19.85
99	Mar-18	105.70	52,023	92.50	336.48	120.66	47.50	10.00	69.02	1324.69	19.97
100	Apr-18	107.10	57,835	91.70	342.78	114.91	47.60	10.20	75.92	1333.28	15.93
101	May-18	108.60	85,376	90.70	348.34	121.09	47.50	10.60	76.45	1306.38	15.43
102	Jun-18	109.10	76,147	90.60	357.44	107.16	47.30	10.70	77.44	1278.50	16.09
103	Jul-18	109.30	59,692	92.10	359.41	120.86	47.70	10.80	74.16	1237.43	12.83
104	Aug-18	108.70	37,042	88.20	367.66	100.84	47.30	11.20	76.94	1202.88	12.86
105	Sep-18	107.80	42,652	81.10	390.84	114.80	47.30	11.50	82.72	1196.70	12.12
106	Oct-18	108.50	41,079	78.80	401.27	119.82	47.00	11.60	74.84	1221.31	21.23
107	Nov-18	109.20	39,222	81.20	395.48	116.20	46.60	12.20	57.71	1221.42	18.07
108	Dec-18	108.00	85,406	80.10	393.88	117.22	46.40	12.70	50.57	1253.95	25.42
109	Jan-19	107.60	22,198	80.50	398.07	100.97	45.50	13.60	62.46	1286.73	16.57
110	Feb-19	108.60	20,030	79.20	398.71	100.09	45.70	13.80	65.03	1319.78	14.78
111	Mar-19	109.00	34,253	81.30	402.81	115.12	46.30	13.90	67.93	1303.78	13.71
112	Apr-19	109.30	16,977	83.60	409.63	113.58	45.60	13.80	72.19	1285.66	13.12
113	May-19	110.20	18,056	76.90	413.52	120.96	45.90	13.70	66.78	1282.23	18.71
114	Jun-19	111.00	9,622	79.80	413.63	96.97	45.80	13.70	67.52	1385.80	15.08
115	Jul-19	112.40	16,591	78.30	419.24	120.39	45.60	13.90	64.07	1414.08	16.12
116	Aug-19	114.40	19,202	79.10	422.84	99.12	45.30	14.00	61.04	1489.62	18.98
117	Sep-19	115.20	19,316	77.70	427.04	119.55	45.30	14.00	60.99	1504.65	16.24
118	Oct-19	116.00	21,296	78.50	435.59	123.14	45.50	13.50	59.30	1495.43	13.22
119	Nov-19	117.10	43,424	81.30	437.25	121.01	45.70	13.20	64.50	1472.89	12.62
120	Dec-19	118.80	77,310	80.70	440.50	128.88	45.60	13.40	67.77	1479.18	13.78
121	Jan-20	120.90	22,722	81.40	446.45	108.69	44.50	13.10	57.77	1561.69	18.84
122	Feb-20	123.70	24,549	79.60	448.02	111.35	44.40	12.60	51.31	1601.80	40.11
123	Mar-20	125.40	33,244	81.10	450.58	114.05	42.60	12.90	14.85	1621.58	53.54
124	Apr-20	127.50	29,470	78.10	454.43	78.22	40.60	13.60	18.11	1679.92	34.15
125	May-20	135.90	29,177	82.70	460.62	84.10	41.30	13.50	34.15	1724.10	27.51
126	Jun-20	139.50	78,764	82.70	465.84	114.09	42.30	13.30	41.64	1724.82	30.43
127	Jul-20	141.40	57,836	82.30	468.56	119.62	41.80	14.10	43.13	1847.27	24.46
128	Aug-20	144.40	37,072	79.40	472.61	115.15	42.90	12.70	45.22	1964.40	26.41
129	Sep-20	146.70	45,846	82.00	477.21	133.07	43.10	12.70	40.30	1921.06	26.37
130	Oct-20	149.90	46,801	81.90	487.38	134.95	43.00	13.10	36.33	1903.37	38.02
131	Nov-20	152.20	50,847	80.10	498.58	131.48	43.20	13.10	46.84	1878.86	20.57
132	Dec-20	154.90	97,227	80.10	504.81	144.69	42.90	12.90	51.22	1864.98	22.75
133	Jan-21	157.70	43,248	83.30	513.30	116.63	43.50	12.80	55.25	1872.32	33.09
134	Feb-21	161.70	44,019	84.50	517.96	117.74	43.60	13.20	65.86	1787.08	27.95
135	Mar-21	165.50	86,476	86.70	523.53	136.96	44.60	13.00	63.52	1717.01	19.40
136	Apr-21	168.80	61,479	80.20	532.32	129.19	44.40	13.80	67.73	1758.23	18.61

137	May-21	175.40	37,358	77.30	537.05	117.24	44.20	13.30	69.36	1862.62	16.76
138	Jun-21	180.30	58,143	81.70	547.48	141.18	45.20	10.60	76.94	1832.85	15.83
139	Jul-21	185.40	47,325	79.50	557.36	117.81	45.40	11.20	77.72	1810.55	18.24
140	Aug-21	192.70	48,510	78.20	563.60	138.21	45.40	11.60	73.45	1778.57	16.48
141	Sep-21	198.90	53,424	79.70	570.66	144.87	46.00	11.40	77.81	1780.26	23.14
142	Oct-21	210.20	54,702	76.80	584.32	142.17	46.20	11.20	83.10	1777.80	16.26
143	Nov-21	229.20	57,995	71.10	604.84	150.65	46.40	11.30	70.86	1831.07	27.19
144	Dec-21	247.40	114,988	68.90	686.95	165.56	46.70	11.40	77.24	1793.21	17.22

B. Actual and Predicted Values of HPI

Observation #	Y (Actual)	Y (Predicted, MLR)	Y (Predicted, ANN)
1	45.40	33.03	45.96
2	45.60	35.02	46.15
3	46.00	38.24	46.43
4	46.20	39.18	46.59
5	46.50	39.45	46.21
6	46.60	39.71	46.22
7	46.70	41.40	46.38
8*	46.80	40.98	46.30
9*	46.90	42.44	46.64
10	47.20	44.61	47.20
11	47.50	46.72	47.38
12	48.10	51.54	48.02
13	48.30	45.82	47.68
14	48.80	45.88	48.96
15	49.00	45.89	49.42
16	49.30	45.47	49.38
17*	49.60	49.07	50.70
18	49.80	49.22	50.10
19	50.00	45.10	49.39
20	50.00	43.21	49.98
21	50.60	43.28	50.32
22	50.90	51.34	51.25
23*	51.10	51.32	52.70
24	51.20	56.64	51.83
25*	51.70	55.93	54.09
26	52.50	54.02	52.90
27*	53.00	55.85	51.21
28	53.60	57.91	54.64
29	54.30	60.45	54.98
30	54.70	62.25	55.36
31	55.10	60.66	54.81
32	55.40	57.15	55.22
33	55.50	56.88	55.55
34*	55.70	56.07	58.61
35	56.00	58.46	55.44
36	56.40	57.59	56.17
37	56.90	59.10	56.82
38*	57.60	59.89	57.32
39	58.40	62.94	57.93

40	59.20	64.77	59.82
41*	60.00	66.73	59.89
42	60.60	67.53	61.05
43	61.20	67.66	62.05
44	61.50	64.08	61.86
45	62.00	65.04	61.56
46*	62.50	67.84	63.47
47	63.30	68.77	63.42
48	63.60	70.21	63.22
49	64.10	66.49	64.68
50	64.80	65.08	65.11
51	65.70	68.29	65.48
52*	66.80	68.26	69.27
53	67.50	70.13	68.12
54	68.20	67.57	68.65
55*	69.30	66.69	69.57
56	70.20	68.30	70.26
57	70.90	68.49	70.81
58	71.50	70.27	71.38
59	72.40	72.81	71.63
60*	73.10	74.44	72.76
61	74.20	73.42	73.98
62	75.40	73.22	75.02
63	76.60	76.33	76.81
64	77.70	76.42	79.06
65	78.90	78.79	79.84
66	79.90	78.12	80.94
67	80.60	81.46	80.42
68	81.10	78.11	80.65
69	81.70	78.36	81.41
70	82.70	82.68	83.05
71	83.70	88.14	84.51
72	84.50	88.46	85.26
73	85.70	89.93	86.15
74	86.60	88.64	87.22
75	87.50	89.20	87.44
76	88.60	88.46	89.22
77	89.70	87.71	89.81
78	90.50	85.70	90.52
79	91.30	85.15	91.34
80	92.30	85.13	92.83
81	92.80	82.85	92.19
82	93.50	84.93	93.27
83	94.00	84.91	93.36

84	94.80	83.20	94.02
85	95.50	86.38	94.82
86	96.70	86.91	96.78
87*	97.90	90.44	97.04
88	98.70	94.09	99.65
89	99.80	97.98	100.11
90*	100.30	101.22	101.37
91	100.50	99.78	99.90
92	100.80	100.35	101.21
93	101.40	106.44	101.52
94	102.20	101.85	102.28
95	102.80	103.14	102.73
96	103.40	105.81	103.38
97*	104.10	106.93	105.84
98*	104.80	106.04	105.91
99	105.70	106.73	106.33
100	107.10	107.66	108.18
101	108.60	108.32	108.41
102	109.10	111.58	109.38
103	109.30	113.18	109.07
104	108.70	112.96	108.08
105	107.80	118.00	108.52
106	108.50	120.21	108.65
107	109.20	119.12	107.70
108	108.00	115.35	108.19
109	107.60	111.58	108.18
110	108.60	109.55	108.71
111*	109.00	110.76	109.40
112	109.30	115.22	109.73
113	110.20	114.44	110.31
114	111.00	115.07	110.71
115	112.40	116.16	111.95
116	114.40	116.81	114.32
117*	115.20	118.64	115.49
118	116.00	126.43	116.28
119	117.10	129.56	116.68
120	118.80	129.48	118.92
121*	120.90	133.95	121.41
122	123.70	132.56	123.70
123	125.40	137.59	125.44
124	127.50	140.45	127.50
125	135.90	142.64	136.00
126	139.50	143.68	139.40
127*	141.40	140.03	143.85

128*	144.40	145.16	147.13
129*	146.70	149.14	147.17
130*	149.90	149.08	147.77
131	152.20	155.89	152.23
132	154.90	160.00	154.87
133	157.70	159.90	157.66
134	161.70	160.13	161.84
135	165.50	167.18	165.54
136	168.80	162.73	168.88
137	175.40	165.64	175.48
138	180.30	186.10	180.30
139	185.40	184.93	185.41
140	192.70	186.23	192.73
141	198.90	187.74	198.92
142	210.20	194.37	210.21
143	229.20	199.10	229.22
144	247.40	235.01	247.38

* Observations included in the validation set.