

THESIS TOPIC

Impact of macroeconomic variables on housing prices and predicting the housing prices dynamics using random forest analysis and machine learning technique

Abstract

The real estate industry is critical to determine the standard of living in the country and creates huge potential for investments which contributes towards the growth in the overall economic activities of a country. The macroeconomic indicators such as CPI (Consumer Price Index), stock rate, gold rate, exchange rate, and oil prices are key determinants of the economic performance of a country and impact the investment opportunities and affect the supply and demand of housing in a country. This study covers the impact of macroeconomic variables on the housing price index in the region of South Asia comprising Pakistan, India, and China. The study covers the influence of these factors under the assumptions of Arbitrage theory. The prediction will be deduced from the implementation of a random forest algorithm with other neural network algorithms. The comparative analysis of these models will determine the accuracy and predictive prowess of different algorithms and how it contributes to the enhancement of predictive performance.

Keywords

macroeconomic, gold rate, exchange rate, stock rate, oil prices, arbitrage, random forest, neural network, price index

CHAPTER 1

This chapter explains the house price index and its importance in the real estate market and covers the historical background of the house price indexes and different methodologies developed over a period. The theories will be covered which will guide the relationship of macroeconomic factors and their impacts on house prices. The reason for motivation to do this study will be briefly discussed and the problem statement and research objectives with corresponding research questions will be covered in this chapter. The significance of the study will be discussed in the last part of this chapter.

1.1 INTRODUCTION

The success of a nation's economy can be determined by how much housing it incorporates into its society. Housing is an integral component of ownership, family growth, and the creation of wealth opportunities in the country's economy. The housing industry association with suppliers, household expenditures, and linkage to the financial system enabled them to drive the country's economy (Harris et al., 2006). With the passage of time and technological advancements, the trend shifted to live in urban cities rather than rural areas due to advanced infrastructure and good employment opportunities therefore the demand for housing increased significantly. Salam et al. (2020) stated that people tend to move to urban cities where their lifestyles can be improved. In China, urbanization took a rapid increase in four big cities where the rate of urban population increased from approx. 18% in 1978 to 64% in 2020 (Huang et al., 2022). Due to increased demand housing prices skyrocketed as well, and repercussions were felt in all levels of countries from developed to developing countries and the reasons vary depending on their economic situation, availability of land, and trade surplus/deficit. With the new developments on the rise and the increasing population, the price factor spread to new sites consisting of all necessities of life causing a ripple effect (Meen et al., 1999).

People tend to move towards the best opportunity and would like to live in such neighborhoods which can empower their families, upscale their employment chances and upgrade their living standards. This is observed that areas where public transport (bus, train) stations opened or their stop situated, the housing prices increased in that surrounding as it benefits people to reach home early and efficiently. Martin et al. (2014) reported that the investors capitalized on the increased demand and subsequently raised prices. The prices are dependent on multiple factors which include firstly the physical condition of the house and the market situation. The advertising agencies promote specific areas as per their contractual agreements, but it will only influence the public if the area is situated near necessary infrastructure such as hospitals, schools, work, and markets. In our current global world, once a platform is created by some developed country then cheap labor is imported from less developed nations to streamline the economic activities. According to Q. Zhou et al. (2020), this leads to a higher population density, which in turn prompts well laden classes to fulfill the growing housing demands. This contributes to increased productivity and fosters economic activity.

Due to rising demand in real-estate, various house pricing indexes emerged over the years which focused on different input parameters as per the requirements of the market during that time. These indexes monitor the health and stability of the market and potentially identify growth and investment opportunities which are helpful for investors and policy makers. The House pricing index (HPI) is a statistical measure that is used to measure the changes in house pricing over time and acts as an integral tool to analyze trends in the housing market. It also provides information about residential properties whose prices are getting appreciated or depreciated over the years. These indexes are deduced by different methodologies which consider different factors such as residential property characteristics, transaction details, and market conditions of that region. The indexes are used by investors, policymakers, researchers, and real estate professionals which can determine either investment on certain properties will be profitable to them or not. As mentioned by Shiller et al. (1993), investors require knowledge of the estimated house prices to make informed decisions about property investment and securities purchases. Their involvement has played a crucial role in driving the nation's economy. Adetunji et al. (2022) stated that the prediction of house price variation instead of exact specific house value is considered more beckoning to real-world applications as pricing is linked with multiple factors such as location, population, and characteristics of house therefore prediction required other variables as well in addition to house price index.

Hill et al. (2018) noted that Eurostat, a statistical authority responsible for providing statistical data on European countries, suggested the use of hedonic methods for determining house pricing indexes. However, Eurostat did not specify a particular method to be employed. Consequently, there is variation in house price indexes across different countries, as each country adopts different techniques for their determination. In China, the National Bureau of Statistics publishes data on different cities which includes commercial and residential properties while in India Reserve Bank of India developed a housing price index in 2010 which covers ten big cities of India, and this is based on the cost of construction methodology which incorporates the influence of factors like land, construction cost, and market conditions. In Pakistan, the State Bank of Pakistan introduced a housing price index in 2013 consisting of eight big cities of the country, and it's based on a hedonic model which includes the property attributes such as size, location, etc.

Macroeconomic indicators play a pivotal role in shaping the financial sector of a country, as they provide crucial insights into the overall state of the economy and influence the feasibility of potential enhancements. According to Hilbers et al. (2001), among the various indicators, unexpected fluctuations in housing prices have the potential to significantly impact the stability of the financial sector. Various macroeconomic indicators, such as the inflation rate, exchange rate, consumer price index, and others, have a direct bearing on the purchasing power of individuals, thereby affecting the demand for housing and subsequently impacting the real estate industry of the country. Research conducted by Wong et al. (2019) has revealed the presence of a long-run connection between macroeconomic indicators and housing prices in Malaysian housing market. According to their findings, the interest rate and exchange rate exhibited a negative impact on housing prices, implying that fluctuations in these variables can potentially influence the affordability of housing and deter prospective buyers. While the inflation rate and population showed a positive relationship with macroeconomic indicators,

indicating that change in indicators will influence a change in house price index. The acknowledgement of these macroeconomic indicators on housing prices, stakeholders can deduce potential influence on the financial sector and devise strategies to enable stability and economic growth.

In today's world, advanced machine learning methods are widely used, and researchers are reaping benefits of these approaches. The major advantage is their ability to manage structured and unstructured data and can efficiently extract meaningful insights from these datasets. The various aspects of housing can be determined by the machine learning algorithms and unknown hidden patterns can be identified to deduce further information (Yang et al., 2021). The conventional methods of house prediction often rely on simplistic models that may have constraints in capturing complex relationships in the data. However, with hybrid approach, the accuracy of house prediction can be significantly improved. The hybrid of algorithms shows substantial improvements in the accuracy of house prediction than conventional approaches (Gu et al., 2011)

Considering the existing research on the impact of macroeconomic variables on the house price index in China, India, and Pakistan, conducting new research aims to go deeper into the subject by determining the benefits and limitations of previous approaches. The focus of this research will be to comprehensively analyse the housing price index in these three countries and examine how macroeconomic factors influence it in the backdrop of unique dynamics in each country's real estate market considering their distinctive economic, political, social, and cultural contexts. This involves assessing essential macroeconomic indicators such as exchange rates, stock rates, gold rates, international oil prices, and consumer price index, and their effects on housing prices.

Overall, this research will be crucial for advancing knowledge in the field and providing valuable insights for various stakeholders involved in the housing sector to do well informed investment decisions

1.2 PROBLEM STATEMENT

Housing is an integral component of society which fulfills the basic human needs as well as act as potential investment opportunity for investors, policymakers, and entrepreneurs to reap benefits from real estate market. Therefore, accurate prediction of price is critical in doing well informed decisions. The housing prices changed due to change in economic activities in that region, Williams et al. (2016) examined the impact of economic policy on housing prices where with increase in interest rate, housing prices decreased. The analysis of impact of macroeconomic indicators are necessary to determine the expected variation in house pricing. Mohan et al. (2019) examined the impact of five macroeconomic indicators such as interest rate, unemployment rate, consumer price index, oil rate and Dow jones industrial average on housing prices. The prior studies incorporated various macroeconomic indicators to determine the impact on housing prices, but none managed to capture the impact of selected macroeconomic indicators under the assumptions of Arbitrage theory.

“It is pertinent that housing prices are always impacted by different microeconomic variables however the impact of macroeconomic variables such as gold price, oil price, exchange rate, interest rate and inflation rate towards house pricing has not been discussed. Moreover, the predictability of house price index within the framework of Arbitrage theory assumptions involving the ML techniques is still unaddressed”

1.3 RESEARCH QUESTION

The primary aim of this research is to examine the correlation between macroeconomic indicators and housing price indexes within the framework of Arbitrage theory. Considering this objective, the research will concentrate on addressing the following questions

1. Is there any impact of macroeconomic variables on housing prices under the Arbitrage theory
2. What is the predictability/strength of labor cost on the stock market
3. What is the policy implication for the housing market based on the results of this research

1.4 RESEARCH OBJECTIVES

In pursuit of our research question, this study aims to fulfill the following additional objectives

1. To assess the impact of macroeconomic variables on housing prices under the assumptions of Arbitrage theory
2. To predict the impact of gold rates on housing prices
3. To evaluate the strength of predictability of housing prices through macroeconomic variables using random forest analysis and neural networks
4. To compare the predictability and accuracy of both models

1.5 SIGNIFICANCE OF STUDY

The study's significance is more evident when considering study's impact on multiple levels: industrial, academic, and societal. The real estate market plays a significant role in the economy of Pakistan, and even during the recent COVID outbreak, the government implemented specific amnesty schemes to bolster this industry. These strategic measures aimed to benefit investors, individuals, and the overall economy by driving economic activity, creating wealth, and enhancing productivity.

1.5.1 Informed Policy Decisions

Policymakers recognize the importance of analyzing the relationship between macroeconomic indicators and housing prices. Housing market has significant implications for economic growth, financial stability, and social welfare. By identifying the key factors that influence housing prices, policymakers can develop targeted campaigns to promote housing affordability and resolve issues associated with housing market to ensure sustainable economic development. Iacoviello et al. (2010) stated that monetary policy shocks have a greater impact on housing prices compared to consumer prices. Therefore, policymakers analyze the potential

impact of policy changes on housing prices and able to make informed decisions regarding housing sector.

1.5.2 Assisting Investors

Investors in the real estate sector rely on accurate predictions of housing prices to make informed investment decisions. By incorporating machine learning techniques, the study provides a valuable insight for forecasting housing prices. Accurate predictions enable investors to determine market trends, identify potential investment opportunities, and evaluate risk profit ratio. According to a study conducted by Hong et al. (2020), it is affirmed that housing prices have a direct impact on investor sentiment. Therefore, investors can determine optimal entry and exit points in the market, identify under appreciated or overvalued properties, and make informed decisions regarding real estate investment opportunities.

1.5.3 Financial Risk Management

A comprehensive understanding of the relationship between macroeconomic indicators and housing prices can provide valuable insights for devising effective risk management strategies. The banks and other financial institutions can analyze the findings to assess the potential risks associated with housing market fluctuations. In an emerging economy, Ayberk et al. (2022) stated that variations in house prices exerts an impact on bank's lending decisions and strategies. This evidence demonstrates the significant benefits that relevant studies can offer to financial institutions and address any potential issues and lending practices changed as per prevailing situation.

1.5.4 Analysis under the Assumption of Arbitrage Theory

The inclusion of the Arbitrage Theory assumption in the study adds unique dimension to the analysis of housing markets. By exploring whether the theory holds true in the context of housing prices, the study analyzes the two of its major components i.e., market efficiency and price differentials. If the assumption is validated, it reinforces the notion that market forces quickly adjust housing prices to reflect relevant macroeconomic indicators, thereby enhancing market efficiency.

1.5.5 Machine Learning Approaches

This study utilized machine learning methods to improve predictive accuracy of results and analyze the strength and influence of relationship between independent and dependent variables.

The study holds substantial significance as it focused into the relationship between macroeconomic indicators and the housing price index, utilizing machine learning techniques such as random forest and other renowned machine learning methods for accurate housing price predictions. Additionally, the study examines the feasibility of Arbitrage Theory within the housing market context. The practical implications of this study extend to policymakers, investors, financial institutions, and the overall economy. The insights obtained from the study contribute to informed decision-making, enhanced risk management strategies, economic stability, and determining the optimal predictive modeling methodologies. Moreover, the

validation of the Arbitrage Theory assumption sheds light on market efficiency and provides valuable insights into price dynamics within the housing market.

1.6 THESIS STRUCTURE

The thesis document contains 5 chapters, introduction, literature review, methodology, estimation & modelling, analysis & results. The brief description of each chapter mentioned below

Chapter 1: This chapter provides a historical overview of the house price index and sets the foundation for addressing the problem statement, defining research objectives, and highlighting the significance of this study.

Chapter 2: This chapter explains the literature review in perspective of various approaches being used to calculate house price index and discuss the impact of macroeconomic variables on target variable. Further theoretical framework and underlying theories discussed.

Chapter 3: This chapter introduces the methodology employed in this research, along with a discussion on the data utilized. It also incorporates various machine learning approaches, techniques for determining feature importance, and the basis for predicting the house price index

Chapter 4: This chapter highlights the results of techniques being used followed by detailed analysis. This shows the importance of each variable and determine its impact on house price index of all countries.

Chapter 5: In this chapter, the results will be analyzed, and the findings will be explained to comprehend the significance and impact of each variable in predicting the house price index of three countries. The hypothesis will be tested against the obtained results and concluding remarks as well as the limitations of the study will be provided.

1.7 CHAPTER SUMMARY

This chapter discussed the reasons behind mass movement of population to urban setting. Further it described the house price index, purpose of its existence and its utilization to predict house prices. This also described the importance of machine learning approaches being utilized to study the impact of macroeconomic variables on housing prices in these countries. The chapter established the research gap and outlined the problem statement, research objectives, and research questions. The significance of this study explained within the framework of Arbitrage theory assumptions. Finally, the chapter concluded by providing an overview of the thesis structure for this research.

CHAPTER 2

In this chapter, the literature review is presented, providing a thorough discussion of the macroeconomic factors and how they influence the house price index. Moreover, it searches through the historical context of different indexes that have been utilized and currently employed approaches, offering a comparative analysis of these methods. Furthermore, this chapter broadly examines each explanatory variable and its effect on housing prices, evaluating its findings by existing research studies and references.

2. LITERATURE REVIEW

2.1 Comparative Analysis of methods

The methodologies which are being used to predict house price indexes are hedonic regression, repeat sales regression, and cost of construction methods. According to Court et al. (1939), the application of hedonic regression models originated when Andrew Court examined the changes in car prices over time, considering the car's specific characteristics. The model is based on Lancaster theory of consumer demand based in 1967 which states that customer take the benefit not from the actual content of item but from the features it possesses e.g., If consumer wants to buy a car, then it prefers specific type of acceleration, interior-exterior characteristics, safety, and security etc. therefore consumer is balancing the investment on car with cars features it possess. Similarly, to determine the implications of air pollution on housing prices, Ridker et al. (1967) utilized the hedonic method approach on housing in 1967 (Hill et al., 2013). Later, Sherwin Rosen presented a theory of hedonic pricing in 1974 which states a model that determines the features and internal-external factors which impact the price of item in the market (Xiao et al., 2017). The hedonic models deduced that price of item can be determined by the inclusion of features of that item to do the impact analysis. To determine the weight of each feature of item, an econometric equation developed

$$P = \alpha + \sum_{i=1}^K \beta_i * C_i + \varepsilon$$

Here P represents log price of item, C consists of features of item, β represents regression coefficient and ε is error term.

There are several factors to check prior implementation of hedonic method approach. The functional relationship between item and its attributes needs to be determined to check the linearity/non-linearity of hedonic price function (Lisi et al., 2013). The hedonic model incorporates item features therefore selection of variable is very important. The features such as size of house, type of house, number of rooms and floors, structural features, location, and distance to main city center etc. The list can include much more variables and selection should be based mainly on knowledge of dependent economic theories and research requirements (Amemiya 1980). There are several methods to resolve issues of multicollinearity in which

explanatory variables are highly correlated and heteroscedasticity where spread of data points across regression line are unequal (Xiao et al., 2017).

The repeat sales method based on determining the transaction prices of traded properties at least more than one time in given dataset. This method focused on the recent two consecutive transactions and not on property characteristics. These methods generally more suitable for large dataset where detailed data descriptions are non-existent. In 1963, Baily et al. (1963) stated BMN method which assigned equal weights irrespective of time interval between two transactions e.g., if any property sold after 10 years and in comparison, properties sold after 4 years, both will have same impact on index level. To enable fairness, in 1987-1989 Case and Shiller states that in case there will be more time between two consecutive transactions then with the increased time, the variance of log price will be higher as well, so they proposed to assign less weight if time between transactions is higher (Hill et al., 2022)

There are multiple flaws identified over the time related to repeat sales method. To implement this model, it is essential to have information on multiple transactions of the same property, which means that single-sale transactions are not included in the analysis. During the application of the repeat sales method, Case et al. (1989) stated that it is necessary to exclude single-sale transactions from the analysis. In the given dataset, these single-sale transactions comprised approximately 96% of the data, resulting in a significant portion of the dataset being rejected. Secondly this is a general phenomenon that if same property sold many times, then a bias emerged in buyers mind that such property may have some flaws or issues due to which no buyer kept this property for long also called lemon bias. The repeat-sales method encounters an additional challenge related to revisions of the entire index when new periods are included. Several studies have demonstrated that these revisions can be significant in magnitude (Hoesli et al., 1997)

The hedonic models usually generate more precise results in comparison to repeat sales regression especially if the dataset is small, another aspect is that repeat sales methods are likely to suffer from revisions as it determines valuation of house over time by checking various sale prices of same item which impact the precision and accuracy of models (Van de Minne et al., 2020). Repeat sales method is widely used in USA but it also captures more irrelevant features whereas hedonic regression required house pricing and house characteristics with their current condition. The availability of all this information is challenging and in the absence of proper information, hedonic model accuracy will be affected.

2.2 House Price Dynamics

Numerous approaches have been devised to create precise housing price indices across various countries. In 1987 Case-Shiller index was developed and considered among quite known and followed indexes in the United States. This is based on repeat sales methodology and its objective is to track price changes of the same selected properties over time. This method was quite useful for regional analysis but lack in incorporating smaller markets and includes properties having multiple sales transactions. In 1991 FHFA House Price Index was developed which resolved the issue of the Case-Shiller index and incorporates a wide range of properties including single homes to apartments. This model used data from Fannie Mae and Freddie Mac

mortgages (Truong et al., 2020), and it excludes properties that didn't finance through them creating a lapse in this index. In 2005 Eurostat House Price Index was developed which is designed for countries of Europe. This index covers the EU members and utilizes national statistical data to ensure consistency. As this index is utilized for multiple countries so the possibility of accurate historical data of all countries is difficult to ensure and consistency can be distorted (Mick Silver et al., 2012). Another aspect is of data quality variation across different states and countries will also affect accuracy. CoreLogic Home Price Index was published in 2006 and its based-on Case-Shiller methodology but as its data availability based on a larger number of metropolitan areas and property types therefore it can cover lot more areas in comparison to Case-Shiller index. The downside is that it has limited historical data in comparison to Case-Shiller index. In 2016 UK House Price Index was developed which covered all the countries residing in the United Kingdom, this index uses data from multiple sources such as national statistical offices, registers of Scotland, etc. Therefore, this index can cover all markets. The index was enabled in 2016 therefore there could be issues with the historical data of some regions which can affect the accuracy.

2.3 Effects of Real estate investment

Macroeconomic indicators have a significant impact on the housing price index across various types of investments. In residential real estate, changes in indicators like stock rates and exchange rates can affect the supply and demand of housing, consequently influencing individuals' purchasing power. Additionally, economic indicators such as interest rates can impact consumer behavior, affecting the demand for newly planned real estate projects like houses, apartments, and commercial properties. The profitability of these projects is directly influenced by interest rates, which can result in price increases if additional funds need to be recovered. Real estate investment trusts (REITs) are also influenced by economic indicators, as they are affected by overall market conditions and investor decisions. Fang et al. (2016) examined the relationship between real estate investment trust and macroeconomic indicators in China, Japan and Singapore and study findings concluded that there is significant evidence of existence of causal relationship between them. Lastly, mortgage loans are impacted by macroeconomic indicators, particularly interest rates. When interest rates rise, mortgage-backed properties become less attractive to investors, impacting the housing market's supply and demand dynamics. Conversely, when interest rates decrease, it leads to price appreciation, establishing an inverse relationship.

The influence of macroeconomic indicators on these investments can differ based on factors such as location, property type, market conditions, and the specific characteristics of the investments. Additionally, other factors, such as government policies, and investor sentiment, can also influence the relationship between macroeconomic indicators and the housing price index in these investment categories. Housing markets are influenced by a combination of regional, national, and global factors, which can result in diverse trends and dynamics therefore it is important to consider the specific context of every region and the prevailing economic conditions when analyzing the relationship between macroeconomic indicators and the house price index.

2.4 Macroeconomic Indicators Dynamics

This study aimed to assess the impact of multiple macroeconomic variables on house price index. These explanatory variables are gold prices, stock price, exchange rate, international oil rate and interest rate of China, India, and Pakistan

2.4.1 Gold Price Dynamics

The investors are constantly seeking investment opportunities that can enhance their risk-adjusted returns and provide portfolio diversification. Throughout history, gold has been widely regarded as one of the key investment options for achieving these objectives. Its unique properties, such as being a tangible asset with unique value and a long-standing store of wealth, have made it an attractive choice for investors across different time periods. Gold's ability to retain value and potentially do a protection against economic uncertainties and inflation has contributed to its persistent appeal as a major investment avenue. The impact of the macroeconomic indicator such as gold rate, on the house price index of countries can vary and is influenced by several factors. The specific characteristics of each country's housing market will contribute to the uniqueness of the relationship between gold rates and house prices.

Sindhu et al. (2013) examined This research paper primarily centers around the examination of several factors, namely the exchange rate between the US dollar and the Indian rupee, crude oil prices, repo rate, and inflation rate, and their respective connections to gold prices. The main emphasis of this paper is to explore the relationships between these factors and the prices of gold. One of the key findings of this study is the presence of an inverse relationship between the US dollar and gold prices. As the value of the US dollar fluctuates, it has an impact on the prices of gold in the opposite direction. Furthermore, the study reveals that crude oil prices also play a role in influencing gold prices. Changes in crude oil prices can have an impact on the movement of gold prices, although the nature of this relationship may vary. Additionally, the research highlights the interdependency between gold prices and repo rates. The repo rate, which is the interest rate at which the central bank lends money to commercial banks, can affect the prices of gold. Changes in repo rates can lead to corresponding adjustments in gold prices. Lastly, the study demonstrates a dependent and positive correlation between gold prices and inflation rates. As inflation rates increase, gold prices tend to rise as well. This presents evidence that there is impact of gold on inflation rate and consequently have indirect impact on real estate as well. In case of high inflation, investors avoid investing money on real estate and as a result real estate market impacted significantly.

Mensi et al. (2022) examined the price movements in one market, such as REITs, can influence and impact the prices of oil and gold, and vice versa. The aim is to understand the dynamics of these spillovers during periods of both high and low fluctuations in the financial markets. Based on MS-VAR (Markov-switching Vector Autoregressive) model, the interconnections between real estate investment trust (REIT), oil, and gold markets are examined. The analysis and evidence suggest that the REIT markets are influenced by both oil and gold markets, with the

strength of the impact varying depending on the severity of fluctuation in market. The impact is more significant once fluctuation severity is high as compared to low fluctuations,

KARAGOZ et al. (2021) focused on investigating the macroeconomic factors that influence the fluctuations in house prices at a detailed level in the Aegean region (Turkey). The Aegean region has experienced rapid population growth in recent years due to industrialization and significant immigration activity. The study considers various variables to determine their impact on the House Price Index, which serves as the dependent variable. The variables considered include the price of gold, the stock market index, the exchange rate between the Turkish Lira and Euro, the interest rate on housing loans, the consumer confidence index for the housing sector, and the regional consumer price index.

By utilizing ARDL (Autoregressive Distributed Lag) bound test, there is a prominent relationship exists between the variables for all studied regions. Based on the estimated results, the analysis reveals that house prices in the İzmir region are influenced by two key factors, namely gold prices, and the general price level. In the Aydın sub-region, the study identifies a more diverse set of variables that significantly impact housing prices, including gold prices, exchange rates, interest rates, population, and the general price level. On the other hand, in the Manisa sub-region, the analysis highlights three main variables that play a crucial role in shaping housing prices: exchange rates, interest rates, and the general price level.

Based on ARDL bound test, Ali et al. (2022) showed that oil price shocks and real income have a direct impact on housing market prices, both in the short and long run. Additionally, the results indicate that fluctuations in gold prices have a negative impact on housing prices, persisting in both the short and long run. As gold prices rise, there is a corresponding decrease in housing prices. This can be attributed to various factors, such as the preferences and behavior of investors and buyers in response to fluctuations in gold prices, as well as the specific characteristics and trends of China's housing market.

Hepşen et al. (2012) examined the potential existence of a lasting association between macroeconomic indicators and the property price index in Dubai. To achieve this objective, the study analyses monthly data over a duration of eight years 2003-2010. The cointegration analyses are employed as a methodology, the findings reveal the presence of relationship between gold prices and Dubai residential property price index. Salisu et al. (2020) explained that assets such as stocks, gold, and the variation in their prices affect inflation in the country which affects the buying power and affects housing prices.

These research papers explain the relationship between the macroeconomic variable i.e., gold and housing prices and provide significant evidence of presence of relationship between them however the house price index predictability is still unaddressed under the assumptions of Arbitrage Theory and machine learning techniques.

2.4.2 Oil Prices Dynamics

Mensi et al. (2013) explained a substantial volatility spillover between oil prices and S&P 500 index. The understanding of volatility spillover is critical for investors, policy makers and risk managers as it helps them in determining the risks associated with market fluctuations and change their investment strategies and focus on implementing mitigation measures. Based on VAR GARCH model, this research paper determines the highest correlation exists between S&P 500 index and WTI index (International Oil prices index), and with gold index as well concluded the fluctuation in these variables have direct impact on stock market index and they have indirect impact on investors behavior and market sentiment.

Similarly, Hoesli and Reka (2013) also shared a presence of volatility transfer between REIT (Real Estate Investment Trusts) and stock market of USA. This highlights that when oil and gold prices experience significant fluctuations, it creates a ripple effect. These fluctuations can directly influence the stock market, affecting investor sentiment and as a result, the real estate market, which is closely linked to the stock market, is also subject to these effects.

By employing a comprehensive Bloomberg REIT index as their unit of analysis, this research delves into the detailed relationship between Real Estate Investment Trusts (REITs) and various market factors including oil prices, stock markets, bond markets, interest rates, and inflation. Their findings suggest that REIT returns showed a positive response to expectations of high oil prices, surpassing the impacts witnessed in the stock and bond markets. This suggests that REITs are more sensitive or responsive to changes in oil prices than traditional stock and bond investments.

Nazlioglu et al. (2016) aimed to establish a framework that can effectively analyze and quantify the causal connections between oil prices and different categories of REITs. By accounting for smooth structural changes, the model offers an enhanced understanding of how fluctuations in oil prices influence the performance and behavior of various REIT types. By utilizing Granger causality-type tests, which analyze the predictive ability of past oil price movements on future changes in REIT returns, this study establishes a basis for understanding the influential connections between oil prices and the REIT sectors. The results emphasize the notable influence of oil price fluctuations on the performance and behavior of these various types of REITs specifically in the healthcare, hotel, warehouse/industrial, residential, and retail sectors.

Killins et al. (2017) examined and analyzed the relationship between these housing price indices and various measures of demand and oil supply shocks within the global crude oil market in perspective of Canadian and USA market. Based on SVAR (Structural Vector Autoregressive) model, this paper determines the influence of oil-specific fluctuations, particularly related to precautionary demand, it is noteworthy that these fluctuations have a more noticeable and significant impact on housing prices within the Canadian market compared to the housing market in the United States. This finding suggests that the oil trading status of a country, whether it is a net importer or net exporter, plays a crucial role in determining the magnitude of impact on housing prices.

Magnusson et al. (2019) explored the existence of a connection between oil prices and the occurrence of turning points in house prices. By doing a comprehensive analysis, this research focused on the intricate relationship between oil prices and both the increasing trends and critical junctures within the housing market. The findings presented in this study are derived from quarterly data encompassing the period from 1990 to 2018, specifically focusing on the Scandinavian countries. The linear regression model was utilized and incorporate macroeconomic indicators such as oil rate and logit model used to analyze the relationship between oil prices and house prices. The results of the analysis indicate a positive correlation between house price inflation and oil prices, leading to the conclusion that there is substantial evidence supporting the notion that fluctuations in oil prices significantly impact the real estate market.

These research papers investigate the connection between macroeconomic variables, specifically international oil rate, and housing prices. They present compelling evidence supporting the existence of a relationship between these variables. However, one aspect that remains unexplored is the predictability of the house price index within the framework of Arbitrage Theory and machine learning techniques.

2.4.3 Inflation Rate Dynamics

The impact of the inflation rate on the house price index is not uniform across all circumstances and may vary based on specific economic conditions, policies, and market dynamics. The research and data analysis plays a crucial role in gaining valuable insights into the precise quantitative relationship between inflation and housing prices. This aid in comprehending the magnitude and direction of this impact, offering a more comprehensive understanding of how inflation affects the fluctuations in the housing market.

Korkmaz et al. (2019) examined the potential impact of inflationary impact on the housing price index (HPI) in the context of Turkey. For the purposes of this study, comprehensive data from 26 regions of Turkey spanning the period between 2010 and 2019 has been employed. The study incorporates not only the house price index, but also consumer price index. The findings of the study, as determined by the Konya Causality test, suggested that to maintain stability in the real estate market, it is crucial to effectively manage and control the inflation rate. The results indicate a significant relationship between inflation and the real estate market, indicating that fluctuations in inflation can have an impact on the stability of housing prices.

In a study conducted by Kuang et al. (2015), the focus lies on examining the connection between inflation rates and housing prices specifically in the urban areas of China. The study encompasses a comprehensive dataset comprising information from 35 major cities across China. The data utilized for analysis covers a substantial timeframe ranging from 1996 to 2010. Utilizing a four-sector general equilibrium framework that encompasses firms, constructors, customers, and the central bank, this study examines the evidence of the relationship between inflation and housing prices. By considering the interactions and interdependencies among these key actors, the study aims to shed light on the dynamics between inflation and housing

prices. The findings of the study provide compelling evidence of a robust positive relationship between inflation and housing prices. These results suggest that as inflation increases, housing prices tend to exhibit a noticeable upward trend.

Zeitz et al. (2008) established a link between previous research conducted on a smaller dataset of houses and the findings obtained from analyzing a much larger sample of 136,000 houses located in the Jacksonville area. The study primarily focuses on utilizing time series data to investigate the relationship between inflation and the selling prices of houses over a specific period. By employing quantile regression analysis, this study examines whether inflation affects low-priced and high-priced houses similarly or if there are differential impacts. The study concludes that inflation exerts an equivalent impact on houses across different price ranges, suggesting a uniform relationship between inflation and housing prices irrespective of the price category.

TITMAN et al. (1982) developed an equilibrium model whose purpose is to examine the variation in inflation rate and its corresponding impact on housing sector. The model used in this study aims to estimate and quantify various aspects related to housing consumption. It specifically focuses on determining the quantity of housing services consumed by individuals, as well as identifying those who own or rent their homes. Additionally, the model seeks to assess the extent of rental property ownership among individuals. To achieve this, the model incorporates individual characteristics, such as demographics or socioeconomic factors, as well as factors like anticipated inflation rates and tax rates. By considering these variables, the model further examines how rental rates and purchase prices of housing are influenced.

Based on findings of the study, it demonstrated that due to the lack of indexing in the U.S. tax system for inflation rate changes, an increase in the expected inflation rate leads to a rise in real housing prices while causing a decline in real rental rates. This implies that the purchasing power of housing increases as inflation expectations rise. However, it is important to consider the elasticity of housing supply. If the supply of housing can respond to changes in demand, the consumption of housing services will also increase. In conclusion, the findings suggest that the interaction between inflation expectations, housing prices, rental rates can significantly impact housing consumption patterns and affordability.

Habanabakize et al. (2022) aimed to examine the impacts of inflation (measured by the consumer price index), interest rates, rental prices, and political risk on housing prices. The analysis is conducted using a comprehensive dataset comprising monthly data from January 2002 to December 2019. The study employs a combination of statistical techniques, including the VAR (Vector Autoregressive) model, the Johansen test for cointegration, VECM (Vector Error Correction Model), and Granger causality test. The findings of the study reveal that inflation, interest rates, rental prices, and political risk significantly influence housing prices in both the short run and the long run. Based on these findings, the researchers suggest that maintaining stability in the inflation rate, interest rates, and political situation is crucial to improve the housing market and achieve stability in real estate market.

Melnychenko et al. (2022) examined the interdependence between inflation rate and real estate market by analyzing the data of Poland real estate market. The study focused on analyzing the quarterly fluctuations in the average price of 1 square meter of housing across Poland's eight largest cities during the period spanning from 2009 to 2021. The analysis utilized a panel model based on regression analysis method, where variations in the price of one square meter of housing were examined as a function of changes in corresponding inflation. To further optimize the model and ensure consistency piecwise linear model was used in study and findings of study concluded that with the increase in inflation, the housing prices in largely all major cities of Poland increased and in only once city the impact was not prominent.

These research papers focused into the correlation between inflation rate, a macroeconomic variable, and housing prices, and providing substantial evidence in support of their interdependence. However, an area that has yet to be thoroughly examined is the predictability of the house price index. This aspect, which remains unaddressed within the assumptions of Arbitrage Theory and machine learning techniques, presents a avenue for further exploration and investigation.

2.4.4 Exchange Rate Dynamics

The exchange rate represents the value of one currency in terms of another currency. It is determined by various factors such as supply and demand dynamics, interest rate differentials, economic conditions, and market sentiment. The influence of the exchange rate on the house price index is not a straightforward relationship, as it can be affected by and interplay with various other economic factors. These factors include but are not limited to interest rates, inflation levels, and government policies. The complex dynamics and interactions among these variables necessitate a comprehensive analysis to fully understand the combined effects on the housing market.

Yang et al. (2012) aimed to investigate the connection between the real effective exchange rate of the Chinese currency (RMB) and housing prices. Based on the VAR (Vector Autoregressive) test, this study utilizes data with range of 2007-2010. The concluded findings vary for two scenarios: impact in long run and impact in short run. The model indicates a high accuracy in capturing the relationship between the variables under analysis. The linear fit of the model explains the influence and correlation between the two variables.

In the short run, an increase in housing prices leads to a decrease in the value of the RMB exchange rate, indicating a inverse relationship. However, in the long run, the impact of real estate prices on the RMB exchange rate turns positive, suggesting a direct relationship.

Bahmani-Oskooee et al. (2018) examined the relationship and assess the impact either fluctuations in exchange rate impact the housing prices in OECD countries or not. This study incorporates the quarterly data from 1994 to 2016 to do this analysis. Based on Granger causality approach this has identified that in half of samples housing prices impacting the exchange rate while in other cases exchange rate is impacting the housing prices.

The study conducted by Yamaka et al. (2022), focused on exploring the interrelationship between exchange rates and the real estate market. The primary objective is to analyze this

connection during periods of economic expansion and contraction. Using Granger causality tests, the researchers discovered compelling evidence of a significant relationship between exchange rates and housing prices. This relationship was found to be more pronounced during periods of economic expansion compared to periods of economic contraction. The consistent nature of this relationship was observed over an extended period, except for two specific time periods: 1998-1999, marked by a significant financial crisis in Asia, and 2008-2009, which witnessed one of the most severe financial crises in the history of the United States. The findings of this study shed light on the dynamic association between exchange rates and the real estate market, emphasizing their interdependence.

Marsel et al. (2022) focused on investigating the influence of various macroeconomic indicators, including the inflation rate, exchange rate, GDP, interest rate, and real estate, on the Indonesian stock exchange indexes. The study period spans from 2011 to 2019, and a total of 36 samples were collected using the saturated sampling method.

Using the multiple regression method, the researchers analyzed the data and derived significant findings. Specifically, they observed a notable negative relationship between the exchange rate and the real estate indexes listed on the Indonesian stock exchange. This indicates that changes in the exchange rate have an impact on the performance of the real estate sector within the stock market.

Njoroge et al. (2019) investigated the impact of the exchange rate on the residential property market in Kenya. To achieve this, the researchers collected and analyzed secondary data from the first quarter of 2005 to the fourth quarter of 2018 using a secondary data collection sheet. To obtain the most accurate and reliable results, the study employed various statistical tests and diagnostic tests. Specifically, a Vector Error Correction Model (VECM) was utilized to examine the short-run relationship between the exchange rate and the performance of residential properties, while an Auto-Regressive Distributed Lag (ARDL) model was employed to explore the long-run relationship.

The findings of the study revealed that the exchange rate had a negative impact on the performance of residential properties in Kenya in the short run. However, in the long run, the exchange rate was found to have a positive effect on the performance of residential properties.

The findings of these papers provide strong evidence supporting the existence of a connection between explanatory and dependent variable. However, there is still a research gap when it comes to understanding the predictability of the house pricing using Arbitrage Theory and machine learning techniques. This aspect requires further investigation to better comprehend how these advanced methods can contribute to forecasting the behavior of the house price index.

2.4.5 Stock Price Dynamics

The stock rate, often measured by reputable stock market indices such as the S&P 500, reflects the overall performance and fluctuations of publicly traded stocks. It is influenced by factors such as investor sentiment, corporate standing in market, economic indicators, and global market conditions. The remarkable progress and expansion of Asia's economy have resulted in a notable appreciation of both the stock and real estate markets within the Asian region. This upward trend has captured the attention and interest of investors worldwide, who recognize the potential for significant returns and growth in these markets.

Lin et al. (2011) examined the relationships between stock rate and real estate market and specifically targeted Asian countries such as Taiwan, China, South Korea, Japan, Hong Kong, and Singapore. Due to expansion of Asian market in recent years, the investors keen to diversify their portfolios and aim for maximum returns. The findings of this study indicates that there is interdependence exists between stock rate and real estate in Japan while limited interdependency exists between them in other Asian countries such as Taiwan, China and Hong Kong. For Singapore and South Korea, it provides potential chance for diversification of portfolios and investment avenues for keen investors.

Another interesting scenario emerged from this analysis is that sometimes stock market leads the real estate market in some countries while in other countries real estate market takes the precedence therefore this implies that there exists a relationship and dependency between stock rate and real estate market.

Similarly, Yousaf et al. (2020) analyzed the dependency existence between stock rate and real estate market of Pakistan. The data collected from four big cities of Pakistan i.e., Karachi, Rawalpindi, Lahore, Islamabad. The data of plot price index and house price index of these cities taken to find the evidence of this relationship. Based on Johansen cointegration test, this showed that the stock markets are linked with the real estate market. Furthermore, the study reveals that in the long run, there is a causal relationship from most house markets to the stock market, indicating that changes in the housing market have an influence on the stock market over a period. On the other hand, in the short run, the study finds evidence of a causal relationship from most plot markets to the stock market. This implies that in the short term, fluctuations in the plot market tend to impact the stock market. Overall, these findings suggest that the real estate market, both in terms of houses and plots, exhibits a leading role in driving changes in the stock market in Pakistan.

Apergis et al. (2011) investigated the interdependency of real estate market and stock market. This study utilizes data of USA and UK starting from 1985 to 2006 and analyzed both markets. Based on cointegration analysis and error correctness techniques to ensure consistency, the findings of the study reveal a strong degree of integration between these two markets, indicating that they are closely interconnected and influence each other's behavior.

In the study conducted by Ali et al. (2017), the researchers explored the causal dependency between the stock market and real estate market across various European countries. The analysis is based on a dataset comprising monthly data from 22 European countries, spanning the period from 2007 to 2010. The main objective of this study is to examine the presence of a relationship between these two markets, both in the short run and the long run. Using the

dynamic Ordinary Least Squares (OLS) approach, the researchers uncover interesting findings. They observed a positive relationship between the stock market and real estate market in fifteen of the examined countries. However, in five countries, a negative relationship is observed. These results suggest that there is indeed a causal relationship between these variables, but the direction of influence differs depending on the specific country and the time frame analyzed.

The study finds that in the long run, the stock market and real estate market has bidirectional causality. On the other hand, in the short run, the stock market continues to lead the real estate market. These findings highlight the importance of considering both the short-term and long-term dynamics when analyzing the relationship between the stock market and real estate market in European countries.

In a study conducted by Batayneh et al. (2015), the researchers investigated the presence of a causal relationship between the stock market and housing prices in Saudi Arabia. The yearly data of Saudi Arabia collected starting from 1985 till 2012. By employing Granger causality and VAR (Vector Autoregression) tests, the analysis revealed that both the stock rate and economic growth have a significant impact on determining housing prices. The findings of the study indicated the existence of a causal relationship between the stock rate and housing prices, with the stock market playing a more prominent role in driving fluctuations in housing prices compared to economic growth.

These research papers present compelling findings that establish a strong relationship between macroeconomic variables, specifically the stock rate, and housing prices. However, there remains a gap in current research regarding the predictability using Arbitrage Theory and machine learning techniques.

2.4.6 Interest Rate Dynamics

The impact of the macroeconomic indicator, specifically the interest rate, on the housing price index refers to how variation in interest rates influence the overall movement of housing prices. The interest rate plays a significant role in the housing market as it affects the cost of borrowing money for purchasing a home.

When interest rates are low, it becomes more affordable for individuals to apply for loans to finance their home purchases. This increased affordability leads to higher demand for housing, which, in turn, can drive up housing prices. Lower interest rates also make it more attractive for investors to enter the housing market, further helped in price increase.

On the other hand, when interest rates rise, borrowing becomes more expensive, reducing the affordability of housing. This can impact demand and slow down the pace of housing price increases. Higher interest rates can also lead to a decrease in economic and investment activity in the housing market, as the cost of borrowing becomes less favorable.

Chen et al. (2022) indicated that the deposit interest rate has a substantial impact on housing prices and the overall activity in the real estate market. Furthermore, it is observed that an elevated level of deposit interest rate leads to a more effective regulation of house prices. Additionally, this higher interest rate serves as a deterrent against market bubbles, effectively

curbing excessive and unsustainable price growth. In summary, the study highlights the crucial role played by the deposit interest rate in shaping housing prices, market activity, and the prevention of anticipated bubbles in market.

Similarly, the evidence of this relationship validated from Vietnam. In Vietnam, Ho chi minh city is one of big economic powerhouse of country and lot of people tend to live in this city to drive economic activity. The purchase of house is difficult due to persistent housing prices fluctuations.

Bui et al. (2020) examined the impact of interest rate on housing prices in the city. The quarterly data of city collected from 2009 till 2018. The aim of this study to analyze the impact in both short and long run. Based on ARDL (Auto Regressive Distributed Lag) approach, the findings showed that variation in housing prices in city significantly caused by interest rate in both short and long run. Additionally, its specified that there is a positive impact on housing prices by interest rates in short run while impact changed to negative when analyzed in long run, providing evidence that there is interdependency between these variables.

The research conducted by Shi et al. (2014) focused on investigating the real estate market in New Zealand and aims to examine the relationship between interest rates and housing prices. The study utilized a dataset spanning from 1999 to 2009, during which the fluctuations in interest rates were influenced by various economic factors. Through their analysis, the researchers discovered a significant positive correlation between interest rates and housing prices. This implies that as interest rates increase, there is a corresponding upward movement in housing prices. The findings suggest that changes in interest rates play a crucial role in shaping the dynamics of the housing market in New Zealand.

Lin et al. (2021) investigated the housing market in Hong Kong and analyze the price fluctuations in real estate market. The data of price variations collected starting from 1981 to 2001. The influence of interest rates on housing prices exhibits a distinct pattern, particularly in the period before 1997 when inflationary situation was prevalent. During this time, interest rates were found to have a significant positive impact on housing prices.

In the study conducted by Zhang et al. (2017), the focus was on analyzing the relationship between real interest rates and housing prices in China using quarterly data. The researchers employed a Vector Error Correction (VEC) model to investigate this relationship and draw conclusions. The findings of the analysis revealed that there is a negative impact of interest rates on housing prices in the long run. This suggests that as interest rates increase, housing prices tend to decrease. However, in the short run, there were some indications of a positive impact of interest rates on housing prices. This means that in the short term, an increase in interest rates may lead to an initial rise in housing prices.

The VEC model allowed for the examination of both the long-term equilibrium relationship and the short-term dynamics between real interest rates and housing prices. It provided insights into the adjustments that occur in the housing market over time in response to changes in interest rates.

Above mentioned research papers provided enough evidence that interest rate has relationship with housing prices and any fluctuation in interest rate cause fluctuation in house price index as well. This relationship should be studied under the assumptions of Arbitrage theory to explore predictability further.

The provided research papers examine the relationship between macroeconomic variables, such as gold prices, oil rates, inflation rate, interest rate, exchange rate and housing prices. They highlight the influence of these factors on the housing market and provide insights into the dynamics of their interconnections. They highlight the importance of managing and controlling these macroeconomic factors to maintain stability & economic activity in the real estate market. These studies highlight the interdependencies and causal relationships between these variables, both in the short run and the long run. The studies employ various methodologies including VEC models, ARDL bound tests, VAR GARCH models, and SVAR models to analyze the relationships and quantify their impacts.

The research conducted in Asian countries such as Taiwan, China, South Korea, Japan, Hong Kong, and Singapore suggests that there is interdependence between the stock rate and real estate market in Japan, while limited interdependency exists in other Asian countries. These findings provide potential opportunities for portfolio diversification and investment avenues for keen investors. Similarly, studies conducted in Pakistan, European countries, Saudi Arabia, Vietnam, New Zealand, and China indicate that there is a significant relationship between the stock rate, interest rate, and housing prices in these regions. In some cases, the stock market leads the real estate market, while in others, the real estate market takes precedence. Additionally, the impact of interest rates on housing prices can be both positive and negative, depending on the short-term or long-term dynamics.

Predicting house prices is a complex task that involves analyzing various economic indicators, market dynamics, and policy factors. Arbitrage Theory, which is based on the principle of eliminating risk-free profit opportunities, can potentially be applied to forecast housing price movements by considering factors such as interest rates, inflation, and market inefficiencies. On the other hand, machine learning techniques offer a data-driven approach to predict house prices. These techniques involve training models on historical data and utilizing algorithms to identify patterns and relationships between variables. By considering factors such as inflation, exchange rates, interest rates, housing market indicators, and socioeconomic factors, machine learning models can learn to make predictions about future house price movements. Combining Arbitrage Theory and machine learning techniques could potentially enhance the accuracy and robustness of house price predictions. However, there remains a gap in current research regarding the predictability of housing prices using Arbitrage Theory and machine learning techniques. Further exploration of these approaches could provide valuable insights into the predictability and forecasting of housing price movements based on macroeconomic variables.

2.5 Underlying Theories

There are several theories which can explain the trend of population migrating to urban cities from rural areas and how prices soared due to high demand of housing once this population reach urban destinations.

The Supply and Demand Theory explained the relationship between price of a product and people willingness to buy or sell it. With the increase in price of product, supply increases while demand decreased and vice versa. This theory explained the high prices of housing in urban cities. As stated earlier that people opt for better quality of life and better growth opportunities for better lifestyle for their family so percentage of migrating people increased and their demand for housing also increased. Due to limited land in business hubs of cities the housing prices are way more than the outskirts of cities owing to low demand.

Arbitrage Theory: This theory explained that how market participants exploit price variations to make risk free profits. There are certain key elements of this theory

- Markets are efficient and price will adjust to reflect all available information
- Price variation occurred due to multiple factors and create opportunities for arbitrage
- Prices will revert quickly to their equilibrium levels

In the context of housing prices, arbitrage theory suggests that changes in macroeconomic indicators can create opportunities for investors or market participants to capitalize on price differentials between housing markets or properties.

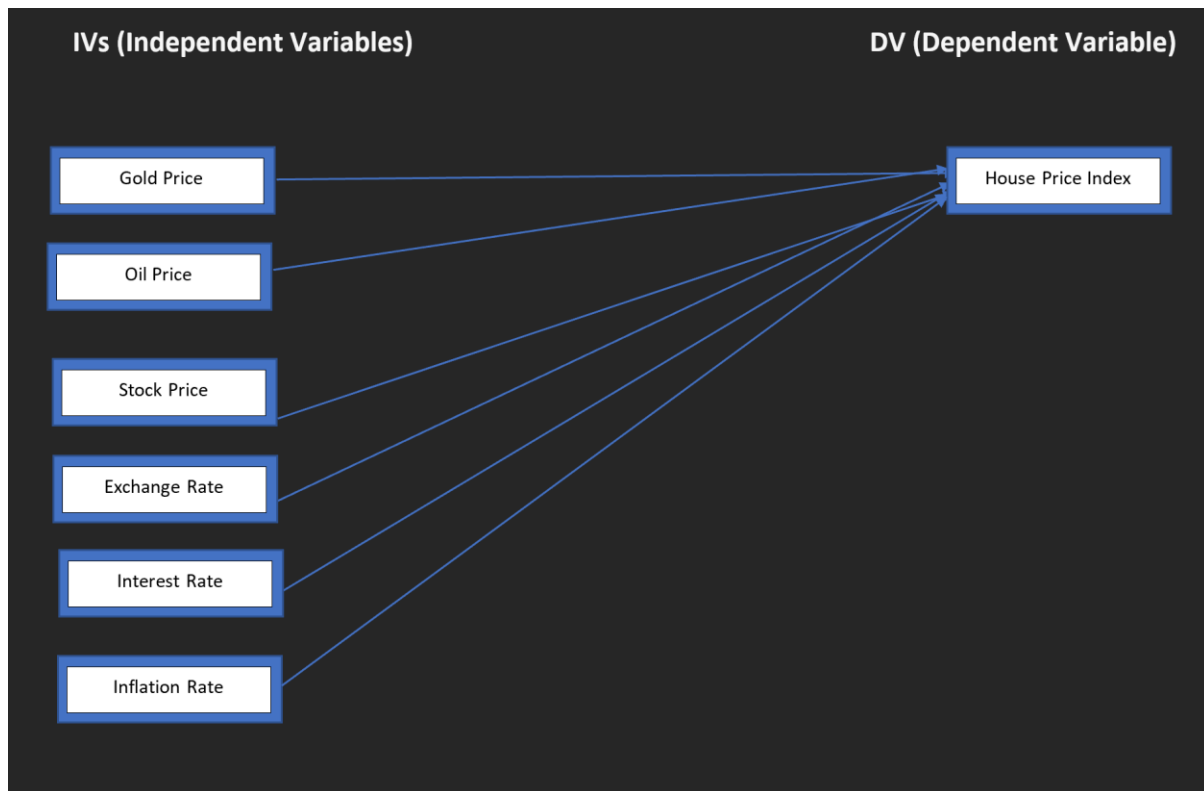
For example, if interest rates are expected to decrease, arbitrage theory suggests that housing prices may increase due to increased affordability and demand. In this scenario, investors may seek to purchase properties in anticipation of future price appreciation.

Arbitrage theory also considers other macroeconomic indicators that can influence housing prices, such as inflation rates, exchange, and interest rate. High inflation rates may impact the purchasing power of potential homebuyers, leading to fall in housing prices.

It's important to note that arbitrage theory relies on the assumption of market efficiency, meaning that prices in the housing market should accurately reflect all available information. However, in practice, markets may not always be perfectly efficient.

2.6 Theoretical Framework

The primary emphasis in this study is on the predictability aspect, and therefore, the study will not be specifically examining or incorporating mediator or moderator variables within the scope of our thesis. Our research objective is centered on developing predictive models and analyzing the factors directly influencing the phenomenon under investigation, rather than exploring the potential mediating or moderating factors that may be involved.



The theoretical framework states the independent and dependent variables and explains how each independent variable is explaining the variation and affecting dependent variable. In this model, the independent variables are exchange rate, gold rate, stock rate, oil prices, inflation rate and interest rate while dependent variable is house price index.

2.7 Hypothesis

The hypotheses will be structured to explore the specific impacts of selected macroeconomic indicators, such as gold prices, oil prices, stock rate, exchange rate and interest rate on housing prices.

H1: The change in oil price has effect on the change in housing prices

H2: The change in interest rate has effect on the change in housing prices

H3: The change in stock price has effect on the change in housing prices

H4: The change in gold price has effect on the change in housing prices

H5: The change in inflation rate has effect on the change in housing prices

H6: The change in exchange rate has effect on the change in housing prices

2.8 Chapter Summary

This chapter discussed in detail about different approaches being used to calculate house price index and comparative analysis of currently used house price indexes around the world. This

examined the supporting literature thoroughly regarding impact of macroeconomic indicators on housing prices and found conclusive evidence of existence of relationship between independent and dependent variable. The two theories discussed on basis of which literature review validated

ANALYZE2GETHER.COM

CHAPTER 3: METHODOLOGY

This chapter focused on research methodology, dataset collection and preprocessing approaches. The removal of multicollinearity discussed, and machine learning algorithms thoroughly explained in terms of actual implementation and enhancements in performance with utilization of hyperparameters. The measures which will be utilized to measure predictive prowess also incorporated in this chapter.

3.1 Data collection

The data of macroeconomic indicators (gold price, oil price, exchange rate, stock price, inflation rate, interest rate) of three countries China, India and Pakistan collected from multiple resources and aligned to consolidate into primary data. Managing the availability of data for all selected macroeconomic indicators in three different countries was challenging, so data had to be extracted from multiple sources. The data sources mentioned below

- Thegoldbaleconomy.com
- Investing.com
- Zameen.com
- rateinflation.com
- fxempire.com
- finance.gov.pk

The duration of collected data alongside its frequency and uniqueness of data mentioned below

- Duration: May 2013 to April 2023
- Frequency: Monthly
- Data Files: Individual excel files (Pakistan, China, India)
- Countries: China, India, and Pakistan
- Unique dataset: Exchange rate, Stock price, Inflation rate, Interest rate
- Shared dataset: International oil prices, international gold price

The dataset of oil prices and gold prices are shared between three countries. For oil prices, Brent oil considered as standard for setting international oil prices therefore historical data of brent oil prices considered for this research. Similarly regrading gold prices, LBMA considered as standard for international gold prices and historical data being utilized for this study.

The other macroeconomic indicators such as exchange rate, stock price, inflation rate and interest rate of each country differentiated from each other, and their unique values incorporated for China, India and Pakistan.

3.2 Sample Structure

This study includes the dataset of three south Asian countries, China, India, and Pakistan. The aim of this study is to determine the causal relationship between macroeconomic indicators and house price index of these countries. Kumar et al. (2003) also examined the power cycle of

these countries at regional and global politics stage amid uncertainties and instabilities in these countries.

China is amongst the leading economic powerhouse and experienced a significant and robust growth over a past decade or so. It has maintained a trade surplus economy and have a growing middle class and remarkable infrastructure.

India is seventh largest economy in the world but in comparison to China it has steady growth and moving at moderate pace. The country is facing trade deficit, but deficit is on continuous decline in recent years.

Pakistan is facing crucial crisis regarding economic activities in the country and have severe trade deficit and external debt issues. The macroeconomic indicators are severely impacted by the risk of default at hand.

While China, India, and Pakistan are geographically close, they have distinct cultural differences. However, India and Pakistan share some cultural similarities due to their historical connections. Both countries also face similar issues. In Pakistan, the real estate industry remains robust despite the country's challenging economic situation. Hence, this study focuses on these culturally unique yet similar countries with the potential to become economic powerhouses in the future. The aim is to analyze the impact of macroeconomic indicators on house price indexes in these countries using relevant datasets.

3.3 Data Preprocessing

The data of different sources merged into one consolidated primary excel sheet and loaded to perform the preprocessing and analysis parts. In dataset there are some instances where data was null and was not present in source as well. e.g., The house price index of countries extracted from source on quarterly basis as availability of daily was limited to recent years only. Therefore, percentage change of values between one quarter to another quarter calculated to fill the gap between two quarters.

$$\% \text{Change Q} = (\text{Present Quarter Value} - \text{Previous Quarter Value} / \text{Previous Quarter Value}) / 3$$

The above formula returns the outcome of % change between two quarters. Now to fill values in empty fields between two quarters following formula utilized

$$\text{Previously available Quarter value} + (\text{Previously available Quarter value} * \% \text{Change Q})$$

The outcome of this statement filled the blanks in the data and this approach utilized in whole sheet for the house price indexes. Prior to this approach another approach utilized to fill the gaps between filled cells which were forward filling and backward filling approach. In this approach, if current value is null then previous value inserted into current empty cell making it forward filling approach while in backward filling approach, post current value copied and inserted into current value to fill the gap. The limitation of this approach is that once same value filled in series then any calculation which includes the combination of previous and current value remained void as desired outcome couldn't retrieve further impacting the analysis of study.

3.4 Multicollinearity

The dataset contains multicollinearity when there is significant relationship exists between explanatory variables which impacts the predictive accuracy of the model. The focus of study is to explain the variation in target variable but with this anomaly, the impact of each independent variable on dependent variable cannot be calculated with accuracy. This caused standard errors to increase, and unreliable results processed. Shrestha et al. (2020) stated that multicollinearity impacts the explanatory variables in such a way that even significant variable became insignificant to the model. The other major impact is of overfitting where model gave good results on training set but once model exposed to testing or new dataset then model gave sharply different result which caused huge difference in actual and predicted values.

In this study, the same issue of multicollinearity was prevalent in the dataset and all value of variables have VIF (Variance Inflation Factor) greater than 10 showing high level of collinearity between independent variables. In dataset, amongst independent variables, interest rate and inflation rate have data in percentage whereas other indicators have their own measure of units.

Features	VIF
Exchange	139.07
Oil	23.73
Gold	100.35
Stock	31.48
Inflation	38.95
Interest	6.21

To develop same scale for all explanatory variables and remove multicollinearity from dataset, natural logarithm (LN) applied on each variable

Gold Price

The explanatory variable of gold price has high value of 100.35 which shows high level of collinearity between independent variables. To remove collinearity the impact of change in independent variable on change in house price index calculated. The natural logarithm applied on division of current gold price value and previous gold price value which is mentioned below

$$\Delta \text{ Change in Gold Price} = \text{LN} (\text{Current Month Gold Price} / \text{Previous Month Gold Price})$$

Oil Price

The explanatory variable of oil price has high value of 23.73 which breached the minimum of limit value 10, showing existence of collinearity between independent variables. To remove collinearity the impact of change in independent variable on change in house price index

calculated. The natural logarithm applied on division of current oil price value and previous oil price value which is mentioned below

$$\Delta \text{ Change in Oil Price} = \text{LN} (\text{Current Month Oil Price} / \text{Previous Month Oil Price})$$

Stock Price

The explanatory variable of stock price has high value of 31.8 which shows high level of collinearity between independent variables. To remove collinearity the impact of change in independent variable on change in house price index calculated. The natural logarithm applied on division of current stock price value and previous stock price value which is mentioned below

$$\Delta \text{ Change in Stock Price} = \text{LN} (\text{Current Month Stock Price} / \text{Previous Month Stock Price})$$

Exchange Rate

The explanatory variable of exchange rate has exuberant high value of 139.07 which shows high level of collinearity between independent variables. To remove collinearity the impact of change in independent variable on change in house price index calculated. The natural logarithm applied on division of current exchange rate and previous exchange rate value which is mentioned below

$$\Delta \text{ Change in Exchange Rate} = \text{LN} (\text{Current Month Exchange Rate} / \text{Previous Month Exchange Rate})$$

House Price Index (HPI)

The house price index variable is target variable of this study and to enable percentage change on all variables and align them on same scale, the natural logarithm applied on division of current month house price index value and previous month house price index value which

$$\Delta \text{ Change in House Price Index} = \text{LN} (\text{Current Month HPI} / \text{Previous Month HPI})$$

With the use of natural algorithm on independent (stock price, oil price, gold price, exchange rate) and dependent variable (house price index), the VIF value significantly decreased, and it became less than 10 for all variables. The following results calculated post implementation of natural logarithm

Features	Before - VIF	After - VIF
Exchange	139.07	1.28
Oil	23.73	1.168
Gold	100.35	1.029
Stock	31.48	1.269

Inflation	38.95	2.58
Interest	6.21	2.43

Now the multicollinearity between independent variables removed and this study can determine the individual variable impacts on target variables.

3.5 Analysis & Estimation Techniques

This study emphasized on determining the causal relationship between explanatory variables and target variable and assess the impact of each macroeconomic indicator on housing price index of three countries. The analysis and estimation approaches refer to set of methods being utilized to analyze the dataset thoroughly, utilized hyperparameters to enhance predictive performance and apply estimation measures to perform comparative analysis of algorithms. This study employed two machine learning models namely Random Forest and XGBoost to perform required predictions on selected dataset.

3.5.1 Random Forest

In implementation phase, Random Forest will be used which is widely used supervised machine learning algorithm and most preferable to solve classification and regression problems. This algorithm comprises a subset of decision trees, and the more trees we include, the more accurate our predictions of the dependent variable become. Random Forest utilizes an averaging method across the data subsets and combines multiple classifiers to address complex problems. It builds upon the implementation of decision trees and operates as an ensemble method, leveraging the combined predictions of various decision trees to produce more precise results. It is important for Random Forest to perform well that the variables involved have low correlation with each other. The major advantage of using this algorithm is to cater overfitting issue where model gave accurate results in training but failed to perform in testing data. Another advantage of this model that it can capture non-linear and complex relationships better than others in most cases. Song et al. (2021) employed four machine learning approaches to determine the pressure sore and findings concluded that random forest performed better amongst all employed methods.

Feature Importance

The purpose of this study is to determine the relationship between explanatory and target variables. The random forest method provides a way to perform the feature importance approach and identify the key features of the model who are contributing more to the prediction of target variable. The feature importance can be calculated by evaluating decrease in impurity in each node of decision tree. This also involved the node probability which can be determined by total count of samples that reach the node and divided by total number of samples. The higher the outcome of this calculation, more important that feature will be to the model.

Model based on Gini impurity to determine the feature importance

$$\text{Gini Impurity} = \sum_{i=1}^C f_i(1 - f_i)$$

where C represents count of unique labels
 f_i represents the frequency of a label

$$\text{RFfi sub}(i) = \sum j \in \text{all trees} * \text{normfi sub}(ij) / T$$

Where T represents total count of trees
 $\text{RFfi sub}(i)$ represents importance of feature calculated from all trees
 $\text{normfi sub}(i,j)$ represents normalized feature importance in the tree

Hyperparameters

To further increase the performance of this algorithm, the hyperparameters can be further analyzed and values can be changed as per requirement of model. The following values needs to be changed to improve the prediction prowess. The commonly used hyperparameters which were utilized in this study as well mentioned below

- **N_estimators:** This includes the total number of decision trees to be included in the model. The increase in number of `n_estimators` have direct relationship with performance of the model but in case the number of trees included are of large count then it increases the computational complexity of the model and consequently impact the performance of the model.
- **Min_samples_split:** This involves the minimum samples required to do the split at node level. With the increase in its value the overfitting issue in model can be resolved.
- **Min_samples_leaf:** This involves the minimum samples required to be available on leaf node. This also helped in resolving overfitting issues of the model.
- **Max_features:** This involves the maximum number of features to be considered. The lower value helped in reducing correlation amongst trees whereas the higher value can increase the computational complexity.
- **Max_depth:** This involves the maximum depth allowed for each tree in the model. In case depth is large then complexity will be increased, and it can lead to overfitting issues.
- **Bootstrap:** This involves data sampling with replacement and consequently helped in enabling randomness in the model.

The optimal values of tuning of these hyperparameters vary as per dataset and acceptable tradeoffs.

3.5.2 XGBoost

XGBoost also called Extreme Gradient Boost is an extension to gradient boost decision tree to enhance performance and accuracy of model. This is based on decision tree algorithms and

focuses on ensemble learning which is to enhance accuracy of prediction by ensemble of relatively less accurate models. This has three main components

- **Loss Function:** This determines the performance of model in terms of accuracy with existing data and this will determine the difference between observed and predicted values
- **Weak Learner:** This is used for classification of data with high error rate and weak accuracy
- **Additive Model:** This is used for addition of models on incremental basis and reach towards final model

This model used enhanced regularized model to resolve overfitting issues. It can effectively determine the complex relationship between independent and target variable with better accuracy. In case of small dataset, these models prone to overfitting when model is not properly regularized. Ibrahim et al. (2021) employed multiple models such as SVM, XGBoost and neural networks to predict correct groundwater level in Malaysia. The results showed that XGBoost performed significantly better than neural networks and SVM (support vector machines) in various situations.

Feature Importance

The feature importance in XGBoost algorithm is based on gain concept which is responsible for the improvement in objective function of the model. The gain in XGBoost represents the relative importance of a feature in the model. It is calculated by considering the contribution of each feature for each tree in the model. If a feature has a higher gain value compared to another feature, it indicates that the former is more significant in generating predictions. Essentially, the gain helps us understand the importance of different features relative to each other within the XGBoost model. The ranking of most impacting features calculated in the code of this study, and it's attached in Annex.

Hyperparameters

To make the XGBoost algorithm perform better, it is important to analyze and tweak the hyperparameters according to the specific needs of the model. Changing the values of the following hyperparameters can help improve the accuracy of predictions. These hyperparameters are commonly used and few of them were also utilized in this study.

- **N_estimators:** This determines the number of trees to build in the model. With the increase in number of trees the performance will be increased but at same time training time may also increase.
- **Learning_rate:** This act as learning curve for the model. Usually, smaller value of learning rate is preferred to maintain stable learning over the time. Though this requires more iterations to align.
- **Max_depth:** This is commonly used hyperparameter to define the maximum depth of the tree. More the higher value the more it has chance to capture complex relationship,

but trade-off will be issue of overfitting therefore optimal value needed to incorporate to avoid this issue.

- Subsample: This included the fractions of samples to be utilized in each iteration of boosting. The value of under 1 considered optimal as it helped in randomness and avoidance of overfitting issues.
- Colsample_bytree: This one manages the fraction of features to be utilized for each tree. This also introduces randomness via selection of subset of features and helped in reducing correlation issues.
- Reg_alpha, Re_lambda: These are responsible for enabling regularization (L1, L2) in the model and their optimum values can be beneficial to avoid overfitting issues.
- Min_child_weight: This includes the minimum sum of weight required in child node during tree development which consequently helped in sustaining model stability.
- Gamma: This involves the minimum loss reduction needed for a split to occur. The higher values can prevent overfitting issues

The hyperparameters and feature importance techniques implemented in the model. To avoid overfitting and generalization issues, there is a need to set optimal values of hyperparameters to get precise and desired results.

3.5.3 Prediction Measures

These are some of machine learning algorithms which are relevant to house prediction problem solving. The usage and selection of algorithms will be decided once best attributed selected for feature selection. Once method selected and implemented with the steps mentioned above, the interpretation will be the key in identifying the accuracy of results. To measure the accuracy of results following metrics used

Mean Squared Error (MSE)

This is calculated as mean of squares of difference between actual value and the predicted values

$$\sum_{i=0}^n \frac{(y_i - y_i^{\wedge})^2}{n}$$

The variables used in above equation explained below

y_i	=	actual	value	where	i	is	iterative	observed	value
$y_i^{\wedge}$	=		predicted		value		of	ith	observation
n	=				number		of		observations

Mean Squared Error is useful for comparing different models and assessing their performance. However, this is to be noted that the values of MSE are in squared units, which might not be as easy to interpret as MAE. Additionally, due to squared units, MSE can be influenced by large errors or outliers more than MAE.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is determined by finding the absolute difference between each predicted value and its corresponding actual value. These absolute differences are added together for all the samples and then divided by the total number of samples to obtain the average.

This is calculated as mean of absolute value of difference between actual and predicted value

$$\sum_{i=0}^n \frac{|y_i - \hat{y}_i|}{n}$$

The variables used in above equation explained below

y_i = actual value where i is iterative observed value
 $\hat{y}_i$ = predicted value of i th observation
 n = number of observations

This method considers errors equally irrespective of size of error; therefore, this method is suitable when all errors are considered equally important and there is no significant difference in the importance of estimations. The lower value indicates better model performance, as it indicates small difference between actual and predicted values.

Root Mean Squared Error (RMSE)

The root mean squared error is calculated by taking the square root of mean squared error

$$\text{RMSE} = \sqrt{\text{MSE}}$$

RMSE is also sensitive to outliers or larger errors like MSE due to squaring operation. The lower value of RMSE indicates better accuracy and overall performance of the model. RMSE operates in the unit of target variable which enables it to better measure to interpret results in comparison to MSE. According to Chai et al. (2014), it is stated that RMSE (Root Mean Squared Error) outperforms MAE (Mean Absolute Error) when the error distribution follows a Gaussian or normal distribution.

Coefficient of determination(R^2)

This is the commonly used approach to evaluate the performance of regression model in fitting the data. This method explains the variability of singular factor caused by its association with another factor. The value resides between 0 and 1 and called as goodness fit. The higher the value the more fit that model is, and variability can be explained. The low r square value is more likely the worse result for predictive values. R square alone cannot be trusted for determining the reliability of model as other measures needs to be considered for the accuracy of model.

$R \text{ square} = \text{Sum of squares regression} / \text{Total Sum of squares}$

The various performance measures are being used to evaluate the model, and the results will be presented in the next chapter. It is important to acknowledge that each measure has its own advantages and limitations. Therefore, it is crucial to select and utilize the measures based on the specific requirements of the study, taking into consideration their acceptable limitations.

3.6 Chapter Summary

This chapter discusses two algorithms: Random Forest and XGBoost, which are used in the study. The enhancements of the model are performed by tuning the hyperparameters of these algorithms and determining key features through feature importance analysis. Additionally, performance measure methods are discussed and employed to interpret the results. The next chapter will focus on the discussion and interpretation of these results.

Chapter 4 Estimation and predictive modeling

In this chapter, the focus is on showing the outcomes of model iterations, specifically in selected countries. The results deduced from training and testing datasets will be presented and

comparative analysis of performance metrics also included. The hyperparameters of both algorithms will be tuned to enhance the accuracy and testing results will be presented in perspective of feature importance

4.1 Model development

The objective of this study is to understand how economic factors influence house price index. This is crucial to identify the key features that have a significant influence on the house price index.

For data analysis, measures were taken to address multicollinearity, which refers to high correlation amongst selected features. This was done to ensure influence of each variable on the house price index and impact could be individually assessed.

The variables such as stock price, exchange rate, gold price, oil price, and the house price index values were transformed using the natural logarithm. This transformation was applied to enable same scale across all variables. While variables such as inflation rate and interest rate were already expressed as percentages and did not require any further transformation.

By making these modifications to the data, the study aims to establish a model that captures how changes in the independent variables correspond to changes in the house price index. This will enable a better understanding of the individual impact and contribution of each variable in influencing the housing market.

$$\Delta \text{ House Price Index} = \Delta (\text{Gold Price} + \text{Oil Price} + \text{Stock Price} + \text{Exchange Rate} + \text{Inflation Rate} + \text{Interest Rate})$$

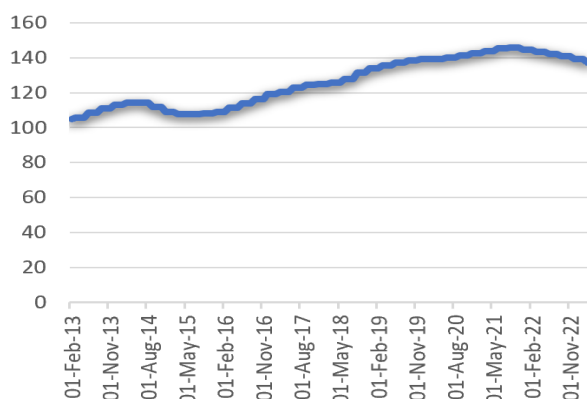
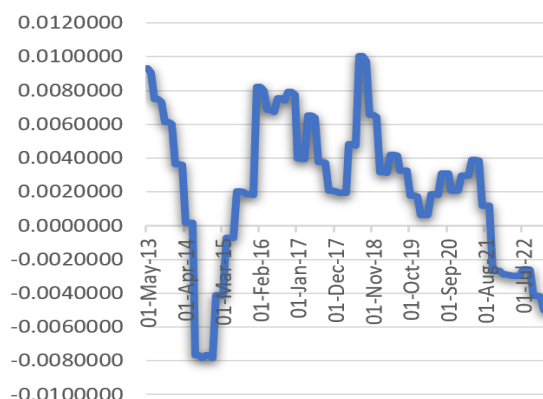
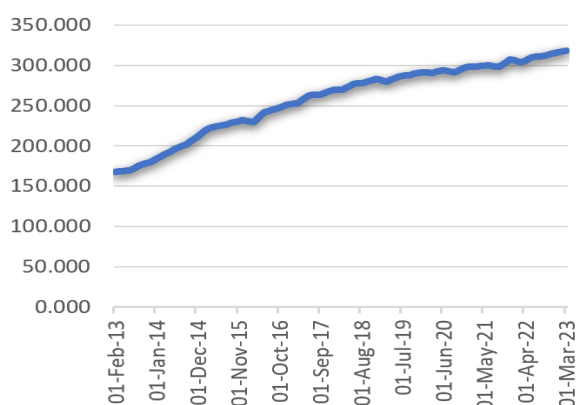
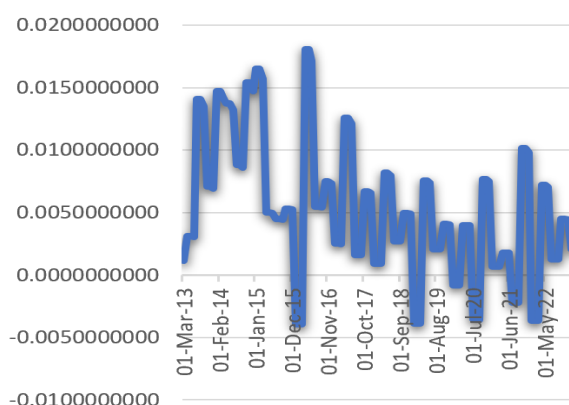
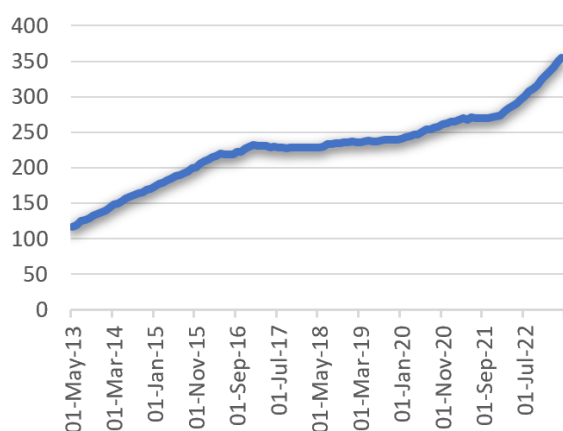
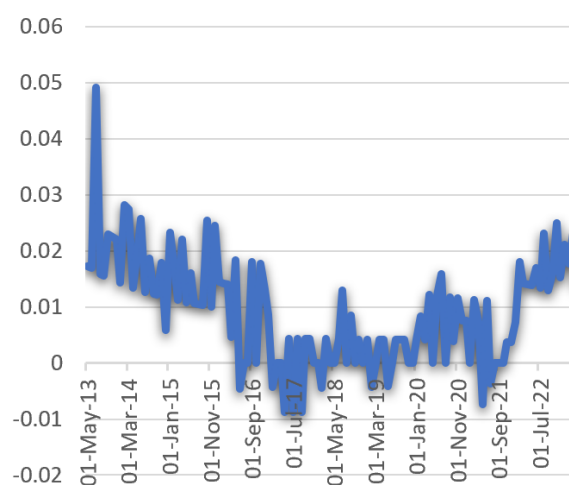
4.1 Resources used for implementation

The code implementation of this study performed in python and google collab was used to execute results which incorporates all necessary libraries. The dataset was split into training and testing with distribution of 80% assigned to training dataset whereas 20% assigned to testing dataset.

4.2 House Price Index Dynamics Country wise

The house price index of each country; China, India and Pakistan extracted and plotted to analyze the performance of target variable over the years. The actual values of house price index of each country as well as percentage change values visualized against duration of dataset to analyze the general trend of target variable

China

China HPI - Actual Values **Δ Change in China HPI****India****India HPI - Actual values** **Δ Change in India HPI****Pakistan****Pakistan HPI - Actual Values** **Δ Change in Pakistan HPI**

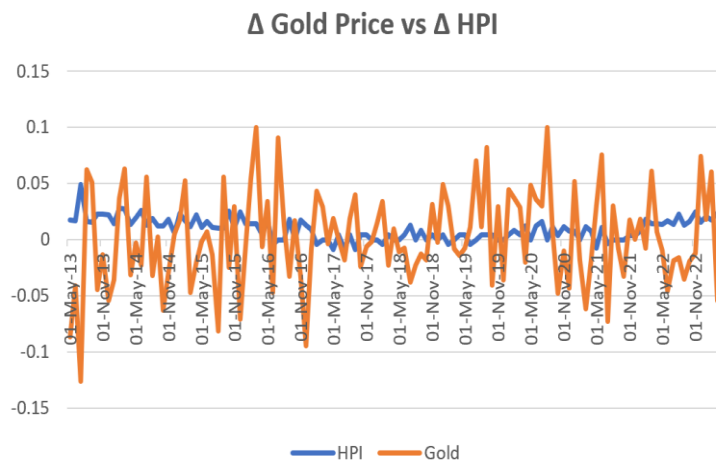
The visualizations depict the actual values of the House Price Index (HPI) for each country, as well as the changes in house price indexes over time.

4.3 Impact of Δ macroeconomic indicators with Δ house price indexes

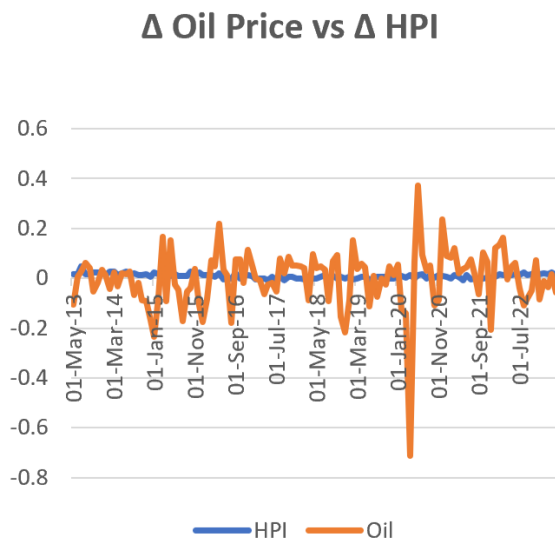
The impact on change in housing price index can be evaluated by assessing the change in macroeconomic indicators. The comparison of HPI against each independent variable will be visualized and this will be followed for each country China, India, and Pakistan

PAKISTAN

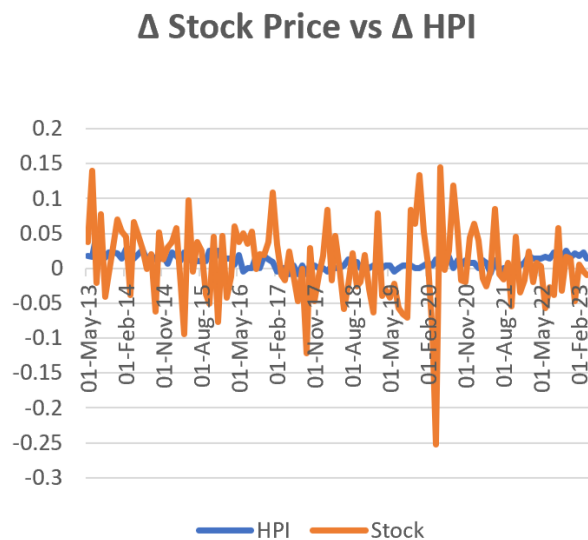
Δ Gold Price $\rightarrow$ Δ HPI

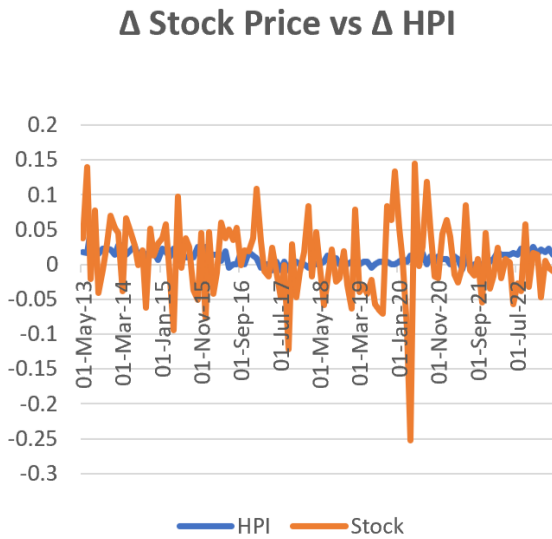
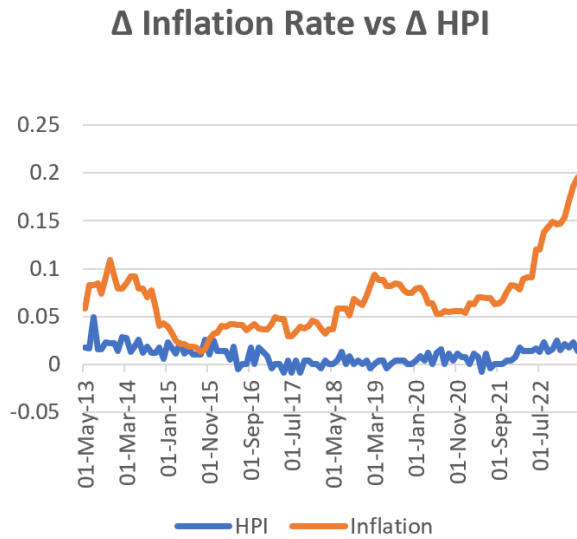
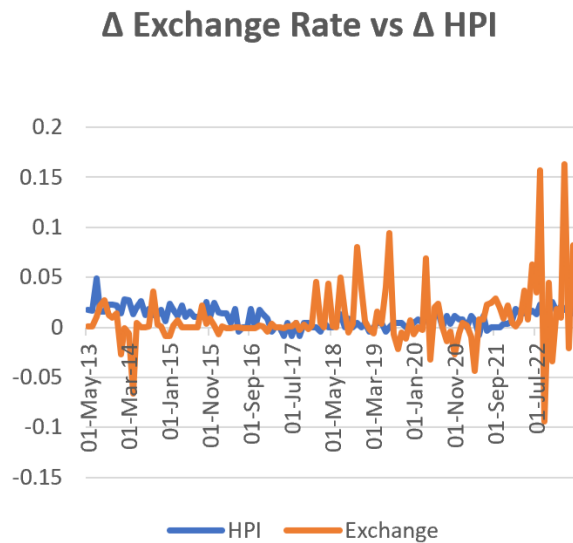
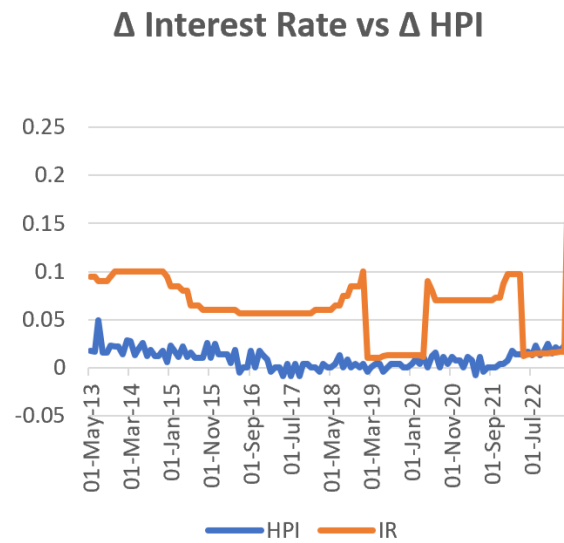


Δ Oil Price $\rightarrow$ Δ HPI

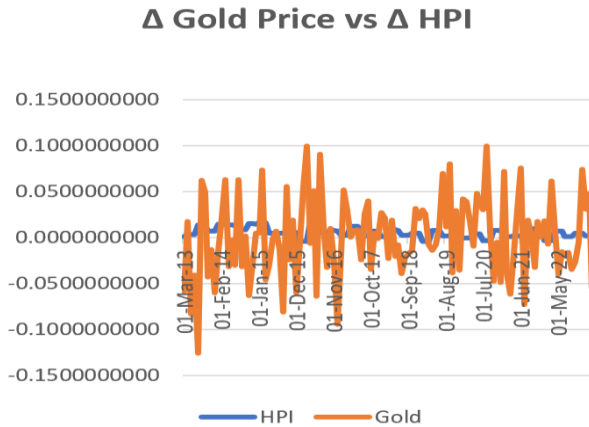
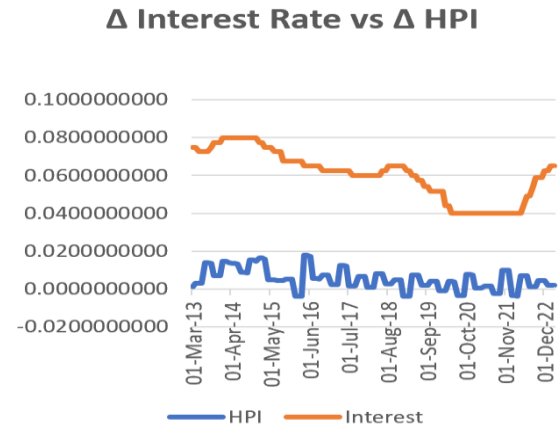
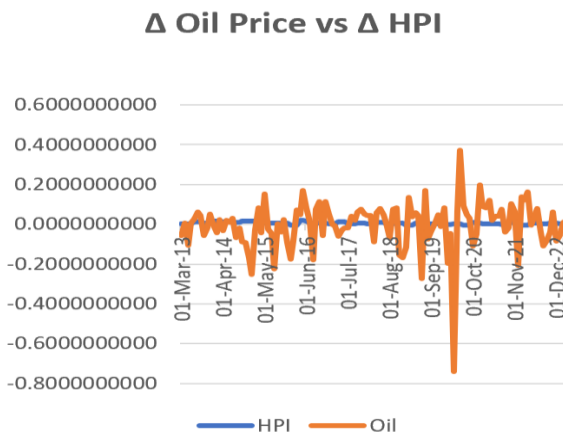
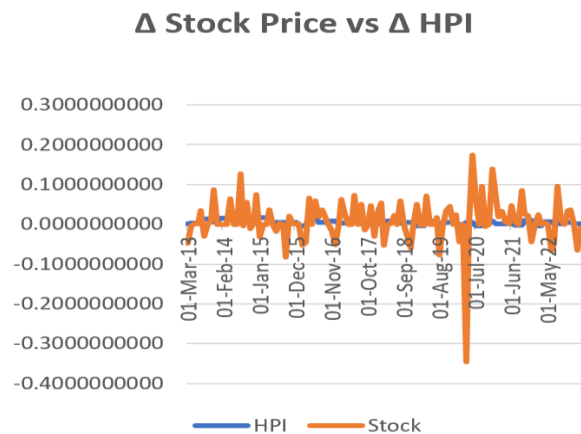
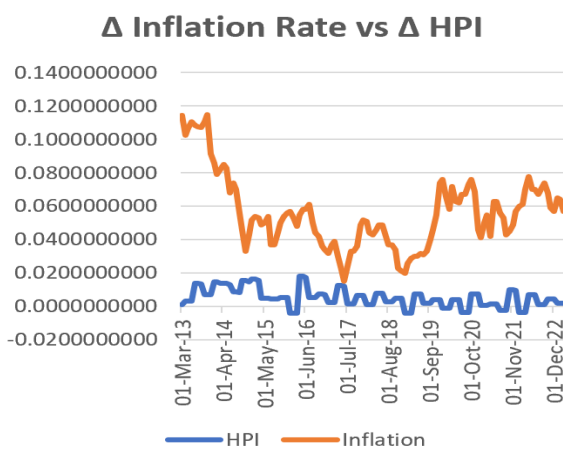
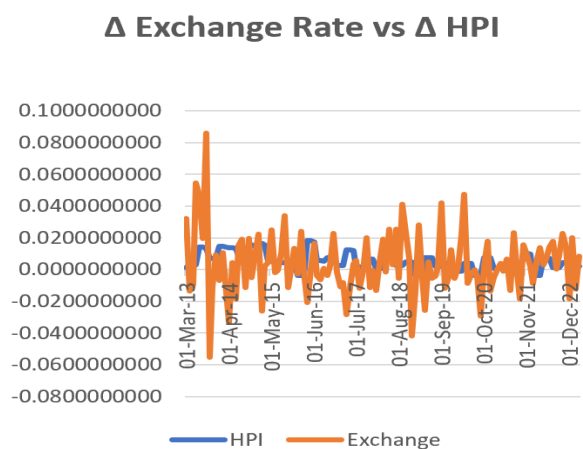


Δ Stock Price $\rightarrow$ Δ HPI



Δ Stock Price $\rightarrow$ Δ HPI **Δ Inflation Rate $\rightarrow$ Δ HPI** **Δ Exchange Rate $\rightarrow$ Δ HPI** **Δ Interest Rate $\rightarrow$ Δ HPI**

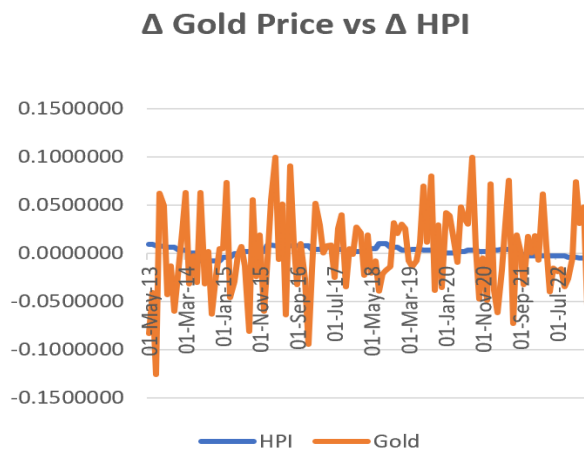
The visualization of the impact of changes in macroeconomic indicators on the house price index (HPI) of Pakistan reveals that there are occasions when these explanatory variables project a similar effect on the HPI. However, in most cases, the substantial fluctuations in the independent variables do not conform to an equivalent impact on the HPI.

INDIA **Δ Gold Price -> Δ HPI** **Δ Interest Rate -> Δ HPI** **Δ Oil Price -> Δ HPI** **Δ Stock Price -> Δ HPI** **Δ Inflation Rate -> Δ HPI** **Δ Exchange Rate -> Δ HPI**

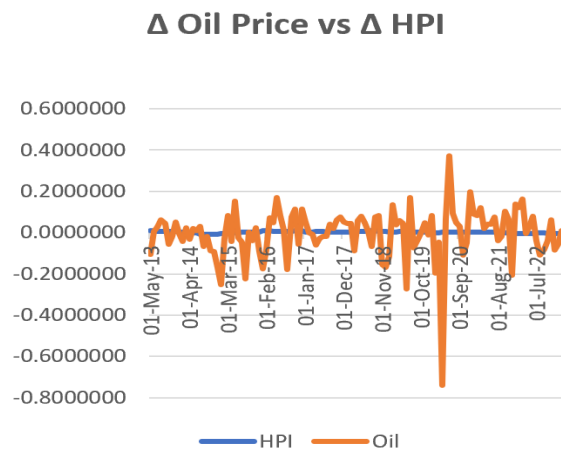
The visualization of the impact of changes in macroeconomic indicators on the house price index (HPI) of India reveals that there are instances when these explanatory variables such as stock price exhibit a similar magnitude of effect on the HPI. However, in most cases, the substantial fluctuations in the independent variables do not produce an equivalent impact on the HPI. Further Interest rate didn't impact the change in HPI of India.

CHINA

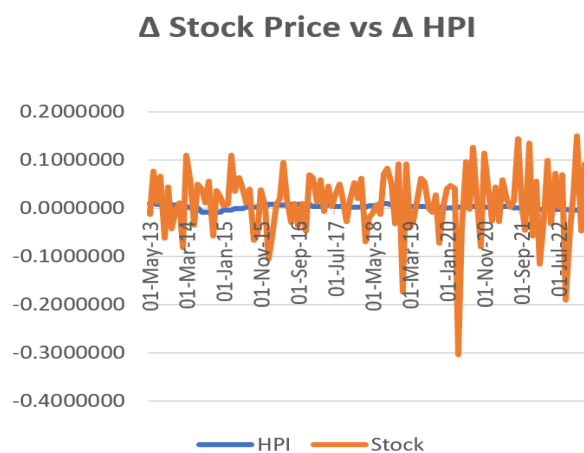
Δ Gold Price $\rightarrow$ Δ HPI



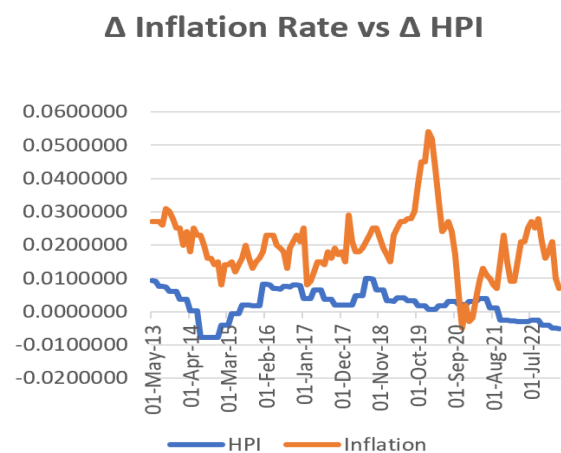
Δ Oil Price $\rightarrow$ Δ HPI

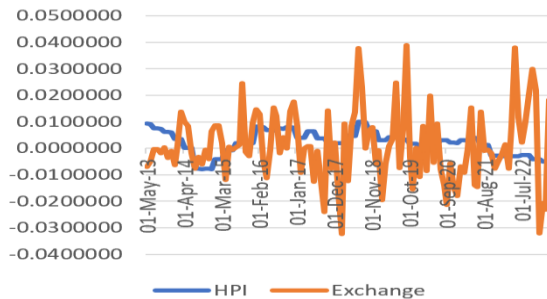
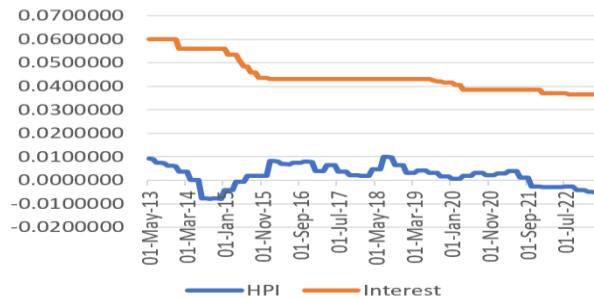


Δ Stock Price $\rightarrow$ Δ HPI



Δ Inflation Rate $\rightarrow$ Δ HPI



Δ Exchange Rate $\rightarrow \Delta$ HPI **Δ Exchange Rate vs Δ HPI** **Δ Interest Rate $\rightarrow \Delta$ HPI** **Δ Interest Rate vs Δ HPI**

Upon examining the visualizations of the relationship between changes in macroeconomic indicators and the house price index (HPI) of China, it becomes apparent that there are limited cases where these explanatory variables exhibit a parallel impact on the HPI. According to the visualizations, variations in the interest rate and inflation rate fail to adequately explain the fluctuations observed in the HPI.

4.4 Random Forest

The random forest algorithm utilized to determine the impact of explanatory variables on house price indexes of these countries. The visualization showcases the outcomes of model training and testing, highlights the most significant features, and provides a comparative analysis of the actual and predicted house price indexes across these countries.

PAKISTAN

Hyperparameters utilized to further tune the results and improved the testing and training results. The tuning of hyperparameters on Pakistan dataset set as below

Hyperparamters Tuning	
n_estimators	500
min_Samples_split	15
min_samples_leaf	5
max_features	log2
max_depth	30

Model Training and Testing Results

The performance measures metrics utilized to observe the performance of model in training and testing. Following is the outcome of results obtained for Pakistan dataset

Country	Model Training			Model Testing		
	MAE	MSE	RMSE	MAE	MSE	RMSE
PAKISTAN	0.0023847	0.0000095	0.00307857	0.0104409	0.00014094	0.011872

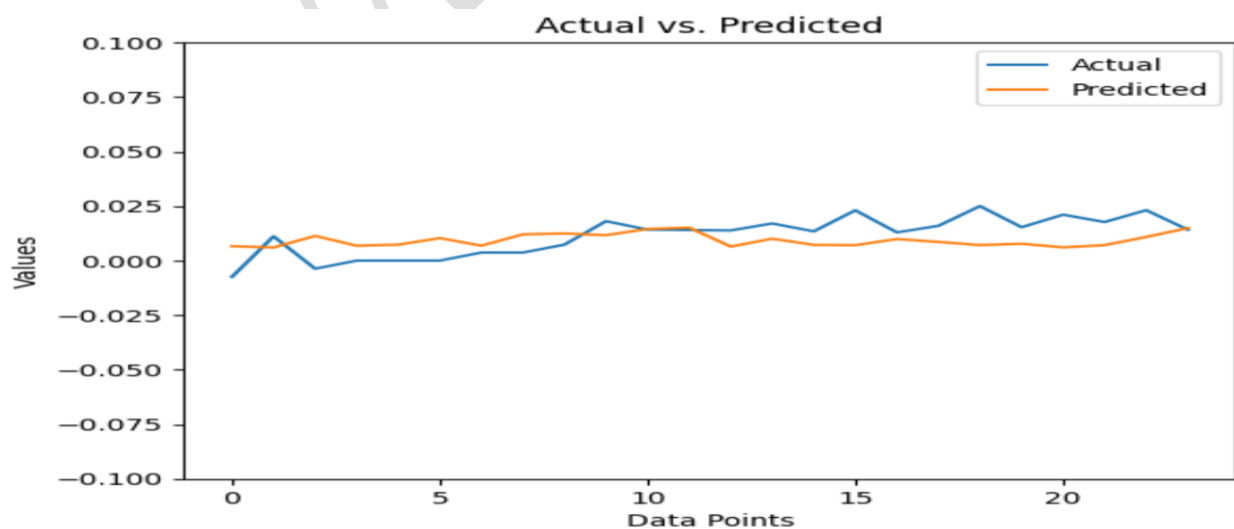
As per results, its observed that the model achieved low error rate between training and testing phases and results are considerably close to actual values.

The difference in results of training and testing phase achieved post-employment of hyperparameters. The error rate decreased across all values and percentage difference achieved mentioned below alongside absolute difference in values

PAKISTAN	Model Training			Model Testing		
	MAE	MSE	RMSE	MAE	MSE	RMSE
Before Hyperparameters	0.0023847	0.0000095	0.00307857	0.0104409	0.00014094	0.011872
Post Hyperparameters	0.0050207	0.000044	0.00663493	0.0080883	0.0000887	0.009423
Difference	-0.002636	-0.0000345	-0.0035564	0.0023526	0.00005224	0.002448
% Difference Achieved	-0.263609	-0.0034543	-0.355636	0.2352579	0.00522376	0.244841

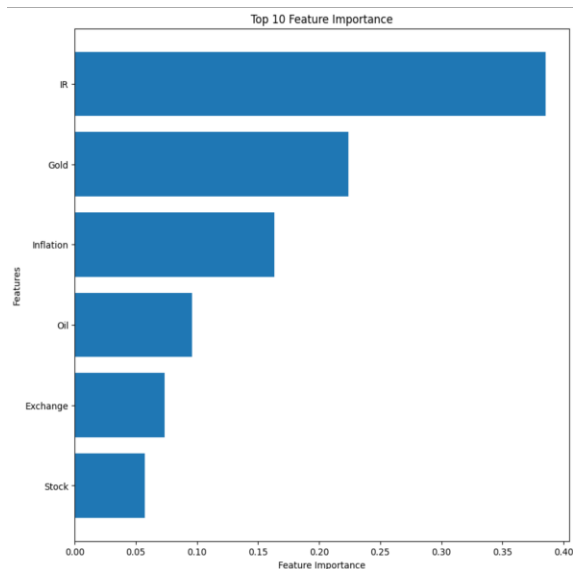
Actual HPI vs Predicted HPI

Below is the graph to show the representation of actual and predicted values of target variable which is house price index (HPI). The results seem to be close to each other citing low error rate



Most important features

By utilization of feature importance implementation in random forest algorithm, the most important features by ranking which are contributing the most change in HPI of Pakistan are mentioned



This is to be mentioned that in prior results before tuning hyperparameters the visualization representing relationship between change in interest rate on change in HPI didn't reflect much significant contribution.

INDIA

Hyperparameters utilized to further tune the results and improved the testing and training results. The tuning of hyperparameters on India dataset set as below

Hyperparameters Tuning	
n_estimators	900
min_Samples_split	15
min_samples_leaf	1
max_features	log2
max_depth	70

Model Training and Testing

The performance measures metrics utilized to observe the performance of model in training and testing. Following is the outcome of results obtained for India dataset

Country	Model Training			Model Testing		
	MAE	MSE	RMSE	MAE	MSE	RMSE
India	0.002916	0.0000129	0.003603	0.00407	0.0000236	0.004867

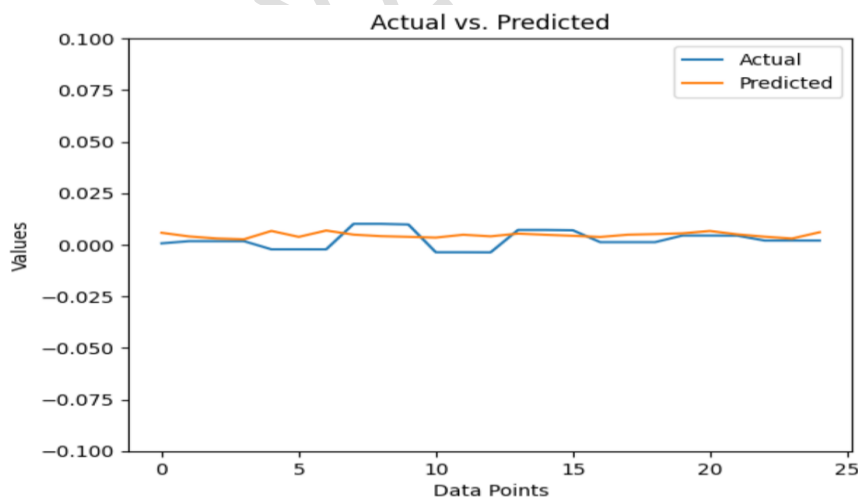
As per results, its observed that the model achieved low error rate between training and testing phases and results are considerably close to actual values.

The difference in results of training and testing phase achieved post-employment of hyperparameters. The error rate decreased for MSE and RMSE whereas for MAE error rate increased by 0.13%. The percentage difference achieved mentioned below alongside absolute difference in values

INDIA	Model Training			Model Testing		
	MAE	MSE	RMSE	MAE	MSE	RMSE
Before Hyperparameters	0.0015318	0.0000038	0.0019728	0.0040239	0.0000241	0.004916
Post Hyperparameters	0.002916	0.0000129	0.003603	0.0040702	0.0000236	0.004867
Difference	-0.001384	0.0000008	-0.00163	-0.000046	0.0000005	0.0000486
% Difference Achieved	-0.13842	0.00008	-0.16302	-0.0046	0.00005	0.00486

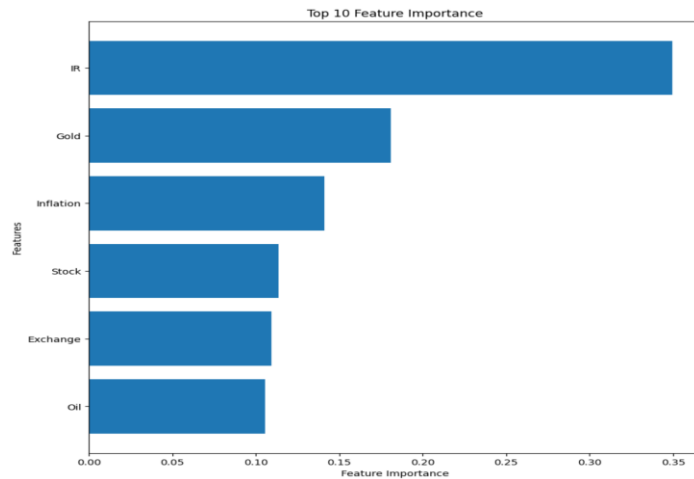
Actual HPI vs Predicted HPI

Below is the graph to show the representation of actual and predicted values of target variable which is house price index (HPI). The results seem to be close to each other citing low error rate



Most important features

By utilization of feature importance implementation in random forest algorithm, the most important features by ranking which are contributing the most change in HPI of India are mentioned



This is to be mentioned that in prior results before tuning hyperparameters the visualization representing relationship between change in interest rate on change in HPI didn't reflect much significant contribution.

CHINA

Hyperparameters utilized to further tune the results and improved the testing and training results. The tuning of hyperparameters on China dataset set as below

Hyperparameters Tuning	
n_estimators	700
min_Samples_split	5
min_samples_leaf	1
max_features	log2
max_depth	70

Model Training and Testing Results

The performance measures metrics utilized to observe the performance of model in training and testing. Following is the outcome of results obtained for China dataset

Country	Model Training			Model Testing		
	MAE	MSE	RMSE	MAE	MSE	RMSE
China	0.001086	0.0000022	0.0014835	0.00543	0.0000361	0.006011

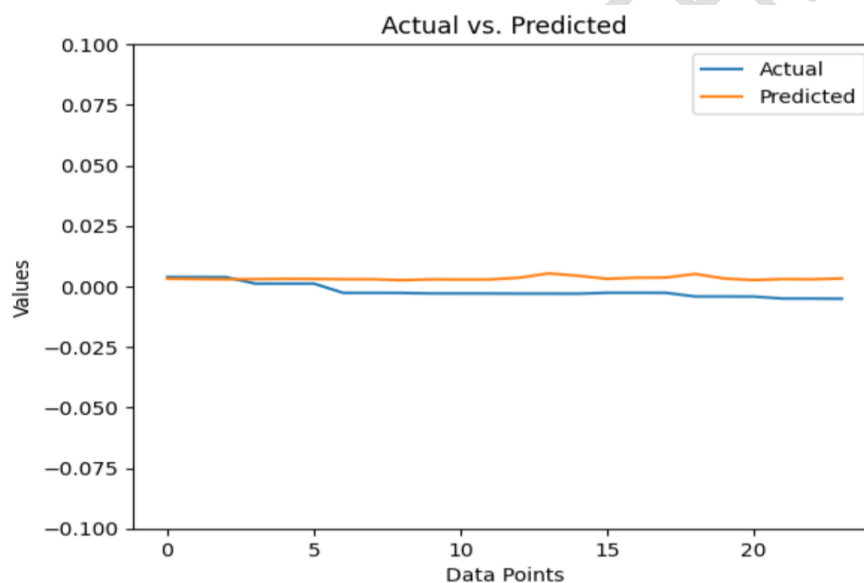
As per results, its observed that the model achieved low error rate between training and testing phases and results are considerably close to actual values.

The difference in results of training and testing phase achieved post-employment of hyperparameters. The error rate decreased in MSE while slight increase observed in MAE and RMSE

CHINA	Model Training			Model Testing		
	MAE	MSE	RMSE	MAE	MSE	RMSE
Before Hyperparameters	0.000719	0.0000009	0.000959	0.00503	0.0000304	0.005517
Post Hyperparameters	0.001086	0.0000022	0.0014835	0.00543	0.0000361	0.006011
Difference	-0.00036	0.0000008	-0.000524	-0.000046	0.0000005	-0.000494
% Difference Achieved	-0.0367	0.00008	-0.05245	-0.0046	0.00005	-0.0494

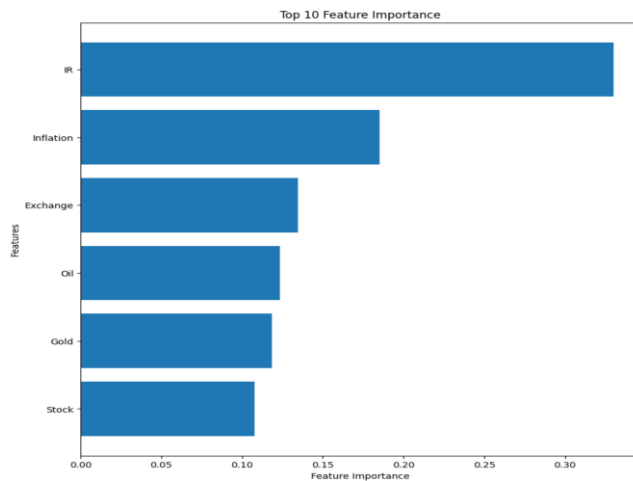
Actual HPI vs Predicted HPI

Below is the graph to show the representation of actual and predicted values of target variable which is house price index (HPI). The results seem to be close to each other citing low error rate



Most important features

By utilization of feature importance implementation in random forest algorithm, the most important features by ranking which are contributing the most change in HPI of China are mentioned



This is to be mentioned that in prior results before tuning hyperparameters the visualization representing relationship between change in interest rate on change in HPI didn't reflect much significant contribution.

4.5 XGBoost

The XGBoost algorithm was employed to analyze the influence of explanatory variables on house price indexes in multiple countries. The results of model training and testing were visualized to illustrate the outcomes. The visualization highlights the most important features and offers a comparative analysis of the actual and predicted house price indexes across the countries.

PAKISTAN

The hyperparameters were adjusted to optimize the results and enhance the performance of training and testing on the dataset specific to Pakistan. The tuning of hyperparameters aimed to improve the accuracy and reliability of the model's predictions for the Pakistani dataset.

Hyperparameters Tuning

n_estimators	500
min_child_weight	5
max_depth	30

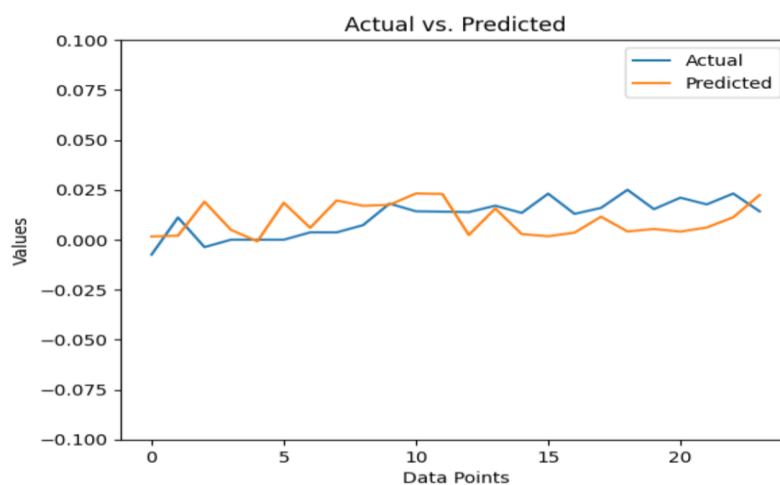
Model Training and Testing Results

The difference in results of training and testing phase achieved post-employment of hyperparameters. The error rate decreased in MSE while slight increase observed in MAE and RMSE, the percentage difference achieved mentioned below alongside absolute difference in values

PAKISTAN	XGBoost (Hyperparameters)		
	MAE	MSE	RMSE
Training	0.0004597	0.0000005	0.000713
Testing	0.0103968	0.0001473	0.012137
Difference	-0.009937	0.0000008	-0.01142
% Difference Achieved	-0.99371	0.00008	-1.1424

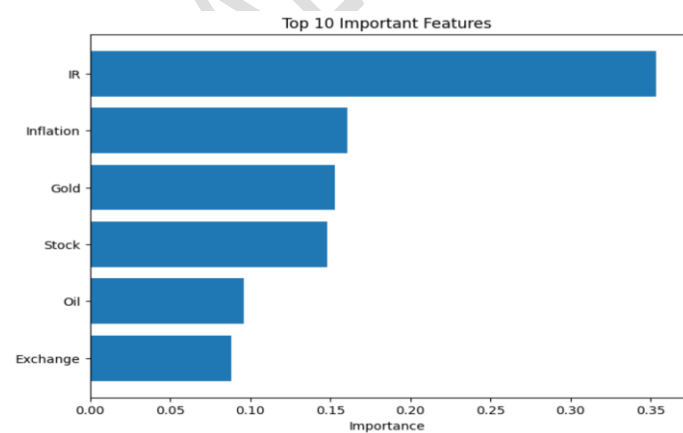
Actual HPI vs Predicted HPI

Below is the graph to show the representation of actual and predicted values of target variable which is house price index (HPI). The results seem to be close to each other citing low error rate



Most important features

By utilization of feature importance implementation in random forest algorithm, the most important features by ranking which are contributing the most change in HPI of Pakistan are mentioned



This is to be mentioned that in prior results before tuning hyperparameters the visualization representing relationship between change in interest rate on change in HPI didn't reflect much significant contribution.

INDIA

The hyperparameters were adjusted to optimize the results and enhance the performance of training and testing on the dataset specific to India. The tuning of hyperparameters aimed to improve the accuracy and reliability of the model's predictions for the Indian dataset.

Hyperparamters Tuning	
n_estimators	500
min_child_weight	15
max_depth	30

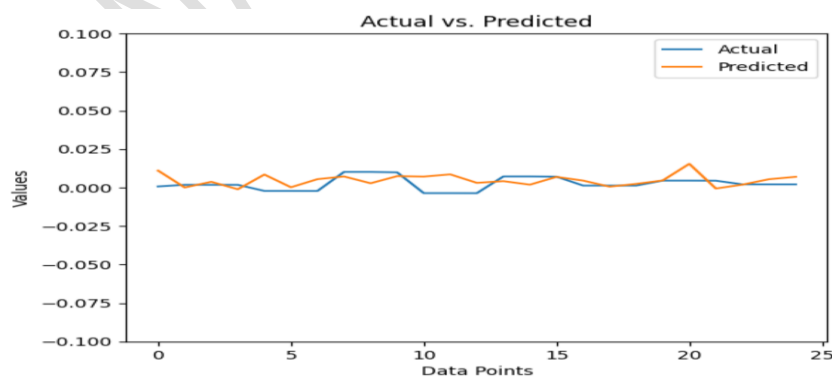
Model Training and Testing Results

The difference in results of training and testing phase achieved post-employment of hyperparameters. The error rate decreased in MSE while slight increase observed in MAE and RMSE, the percentage difference achieved mentioned below alongside absolute difference in values

INDIA	XGBoost (Hyperparameters)		
	MAE	MSE	RMSE
Training	0.0011698	0.0000022	0.001501
Testing	0.0047157	0.0000361	0.006016
Difference	-0.003546	0.0000008	-0.00452
% Difference Achieved	-0.35459	0.00008	-0.45156

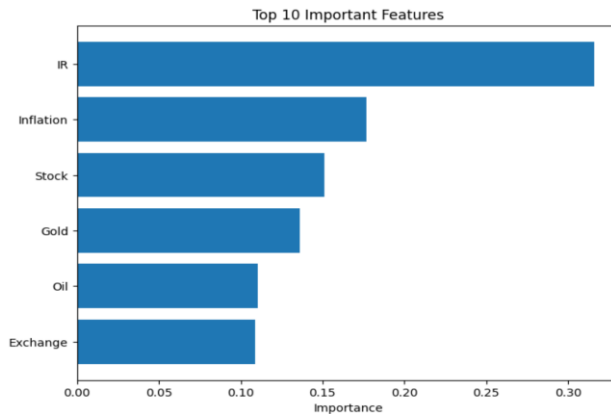
Actual HPI vs Predicted HPI

Below is the graph to show the representation of actual and predicted values of target variable which is house price index (HPI). The results seem to be close to each other citing low error rate



Most important features

By utilization of feature importance implementation in random forest algorithm, the most important features by ranking which are contributing the most change in HPI of India are mentioned



This is to be mentioned that in prior results before tuning hyperparameters the visualization representing relationship between change in interest rate on change in HPI didn't reflect much significant contribution.

CHINA

The hyperparameters were adjusted to optimize the results and enhance the performance of training and testing on the dataset specific to China. The tuning of hyperparameters aimed to improve the accuracy and reliability of the model's predictions for the dataset.

Hyperparamters Tuning	
n_estimators	500
min_child_weight	5
max_depth	30

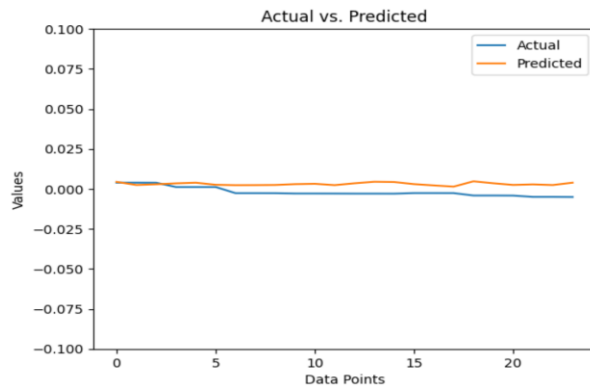
Model Training and Testing Results

The difference in results of training and testing phase achieved post-employment of hyperparameters. The error rate decreased in MSE while slight increase observed in MAE and RMSE, the percentage difference achieved mentioned below alongside absolute difference in values

CHINA	XGBoost (Hyperparameters)		
	MAE	MSE	RMSE
Training	0.0005835	0.00000061	0.000784
Testing	0.0051842	0.0000329	0.005736
Difference	-0.004601	0.0000008	-0.00495
% Difference Achieved	-0.46007	0.00008	-0.49524

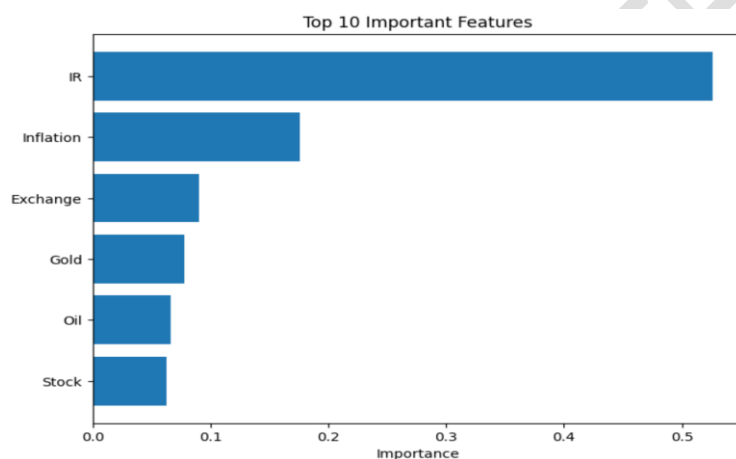
Actual HPI vs Predicted HPI

Below is the graph to show the representation of actual and predicted values of target variable which is house price index (HPI). The results seem to be close to each other citing low error rate



Most important features

By utilization of feature importance implementation in random forest algorithm, the most important features by ranking which are contributing the most change in HPI are mentioned



This is to be mentioned that in prior results before tuning hyperparameters the visualization representing relationship between change in interest rate on change in HPI didn't reflect much significant contribution.

4.6 Comparative Analysis of Models

The comparative analysis of random forest and xgboost involving results of testing data mentioned below

Prediction Accuracy	Random Forest (testing)			XGBoost (testing)		
	MAE	MSE	RMSE	MAE	MSE	RMSE
Pakistan	0.0080883	0.000089	0.009423	0.0103968	0.0001473	0.012137
India	0.0040702	0.000024	0.004867	0.0047157	0.000036	0.006016
China	0.00543	0.000036	0.006011	0.0051842	0.000033	0.005736

As per results observed for both models, it seems that for every country different model will be preferable as error rate is changing per country dataset. In India dataset, XGBoost performed better than random forest and gave less error rate. In Pakistan case, random forest performed better while in China scenario, MAE and RMSE performed better in random forest while MSE performed better in XGboost.

4.7 Comparative Analysis of Models

The comparative analysis of topmost contributing variables retrieved by random forest and XGBoost mentioned below

TOP 5 FEATURES	Random Forests (Testing)					XG Boost (Testing)				
IVs	1	2	3	4	5	1	2	3	4	5
Pakistan	Interest	Gold	Inflation	Oil	Exchange	Interest	Inflation	Gold	Stock	Oil
India	Interest	Gold	Inflation	Stock	Exchange	Interest	Inflation	Stock	Gold	Oil
China	Interest	Inflation	Exchange	Oil	Gold	Interest	Inflation	Exchange	Gold	Oil

As per observed results this can be seen that both algorithms rightly predict interest rate as top contributor in change in HPI for all countries. For Pakistan, top 3 are interest rate, gold price and inflation rate for both algorithms whereas similar results from both models can be seen in case of China, where top 3 are interest rate, inflation rate and exchange rate. In case of India, gold is second key contributor as per random forest whereas inflation rate considered corresponding key contributor from XGBoost results. The inflation and stock rate comes third and fourth for random forest whereas same variables come on second and third with respect to XGBoost results.

4.8 Chapter Summary

In this chapter, the impact of changes in macroeconomic indicators on house price indexes for all countries was presented. Two algorithms, random forest and XGBoost, were utilized to analyze and predict the outcomes. The results of model testing and training were extracted and presented. Furthermore, the top key contributors to the target variable were identified, and the results were showcased. The graphs were used to visualize the relationship between actual and predicted values of HPI. Lastly, a comparative analysis of the extracted important features and testing results achieved from random forest and XGBoost was conducted and presented.

Chapter 5 Discussion and Conclusion

The purpose of this chapter is to discuss and assess the model's findings to support the hypothesis and move closer to the conclusion. To determine its influence and contribution to the change in house price indexes, this study also compares the error rate with the absence of a variable included in the discussion. The chapter will conclude by discussing the results of the available research.

5.1 Evaluation of Hypothesis

This study formulated six hypotheses regarding change in macroeconomic indicators with corresponding impact on change in house price indexes of Pakistan, China, and India. The model will be analyzed and evaluated on basis of feature importance of both algorithms and correlation to validate the feature importance outcome. The partial plots will be used where correlation factor is low to determine the impact and reason of non-significant values.

This study consist dataset of three countries therefore hypothesis validation will be done by country wise

Hypothesis validation in case of China

In case of China, the outcome by random forest algorithm observed that interest rate, inflation and exchange rate are top 3 contributors to change in house price index whereas oil, gold and stock prices are least contributed towards change in target variable. This is further validated by determining the correlation between macroeconomic indicators and house price index.

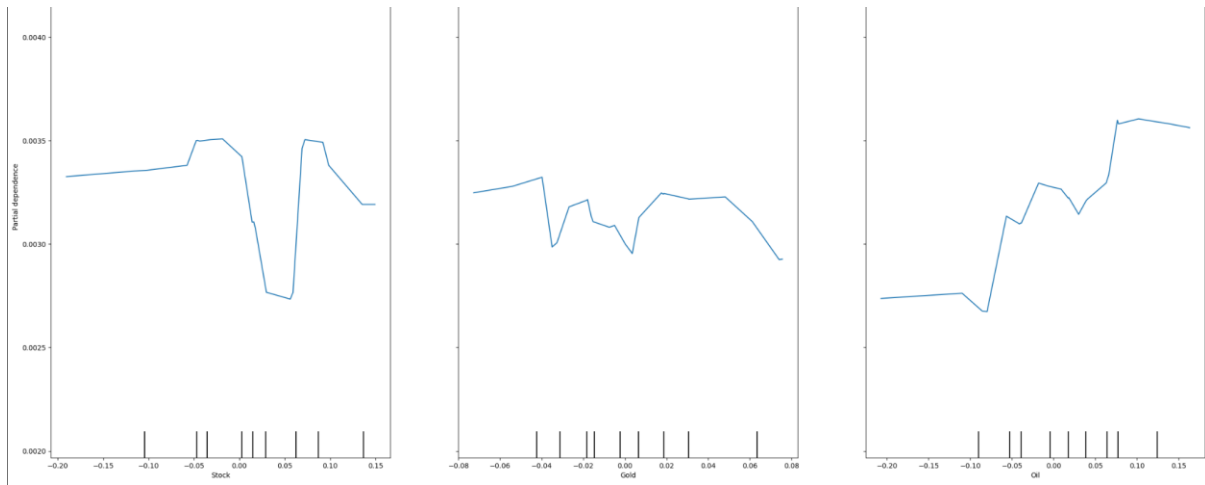
Correlation Results

Correlation - China	
Stock	-0.013095
Gold	-0.009548
Oil	0.077516
Exchange	0.076158
Inflation	0.145463
Interest Rate	0.061106

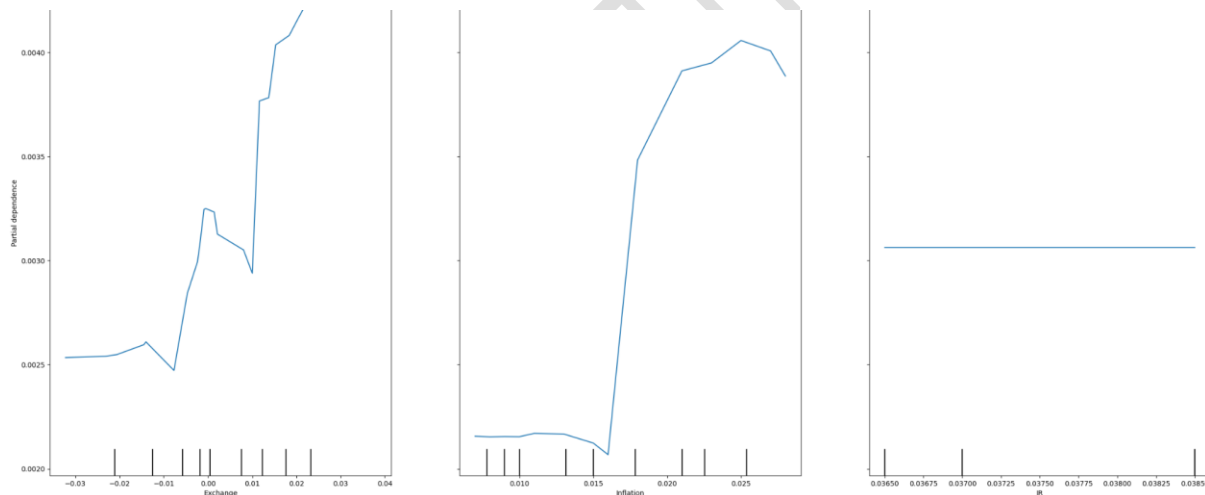
As per above results, stock and gold have weak negative relationship and with the increase in stock and gold, the house price index in China decreases slightly. The close to zero value showed a weak association with target variable. The oil, exchange and inflation showed a moderately positive relation and indicated the existence of association between independent and target variable. The Interest rate has weak relationship which conflicts with the feature importance results determined by random forest and xgboost. Therefore, partial plots utilized to further check the associations and accept or reject the hypothesis. The inflation and exchange positive association was validated as they were in top 3 contributors observed by both algorithms.

Partial Plots

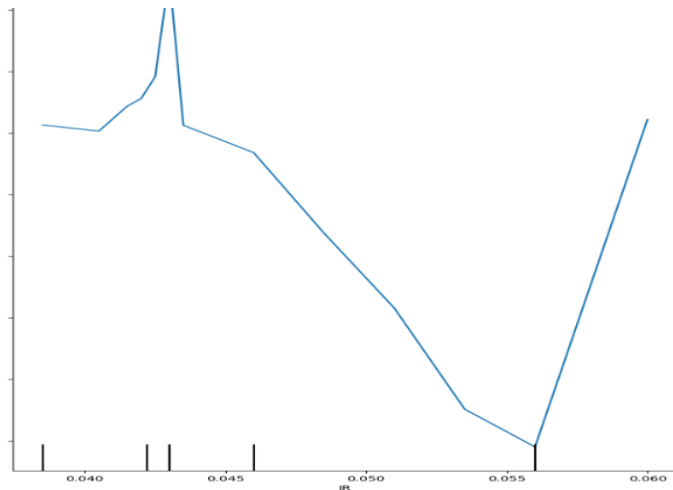
The partial plots utilized to check the impact of single feature on target variable and assess the direction of relationship. The following graphs generated to show partial plots



This can be seen from partial plots as well that stock, oil and gold which are considered as low contributors are showing change in HPI with change in each of these variables, but change is not much significant. The interest rate, inflation and exchange plots mentioned below



These plots showed that there is considerable change in HPI with change in exchange rate and inflation rate further validating the feature importance ranking of these variables. The interest rate which was top contributor showed straight line indicating no association between them. This is further checked with training predicted data to assess the reason of this straight line. As per checked there is significant variation exists between change in interest rate and change in HPI.



As time series data is being used, the splitting technique employed is designed to ensure that the predictive testing phase includes values that may not have a significant impact on the Home Price Index (HPI). The interest rate of China generally doesn't change much and remain constant which further validates the straight-line result.

As per above evaluations, this can be observed that all variables have association with target variable and there is evidence of change in HPI impacted by change occurred in macroeconomic indicators though stock and oil have weak association.

Hypothesis validation in case of India

In case of India, the outcome by random forest algorithm observed that interest rate, gold and inflation are top 3 contributors to change in house price index whereas stock, exchange and oil prices are least contributed towards change in target variable. This is further validated by determining the correlation between macroeconomic indicators and house price index.

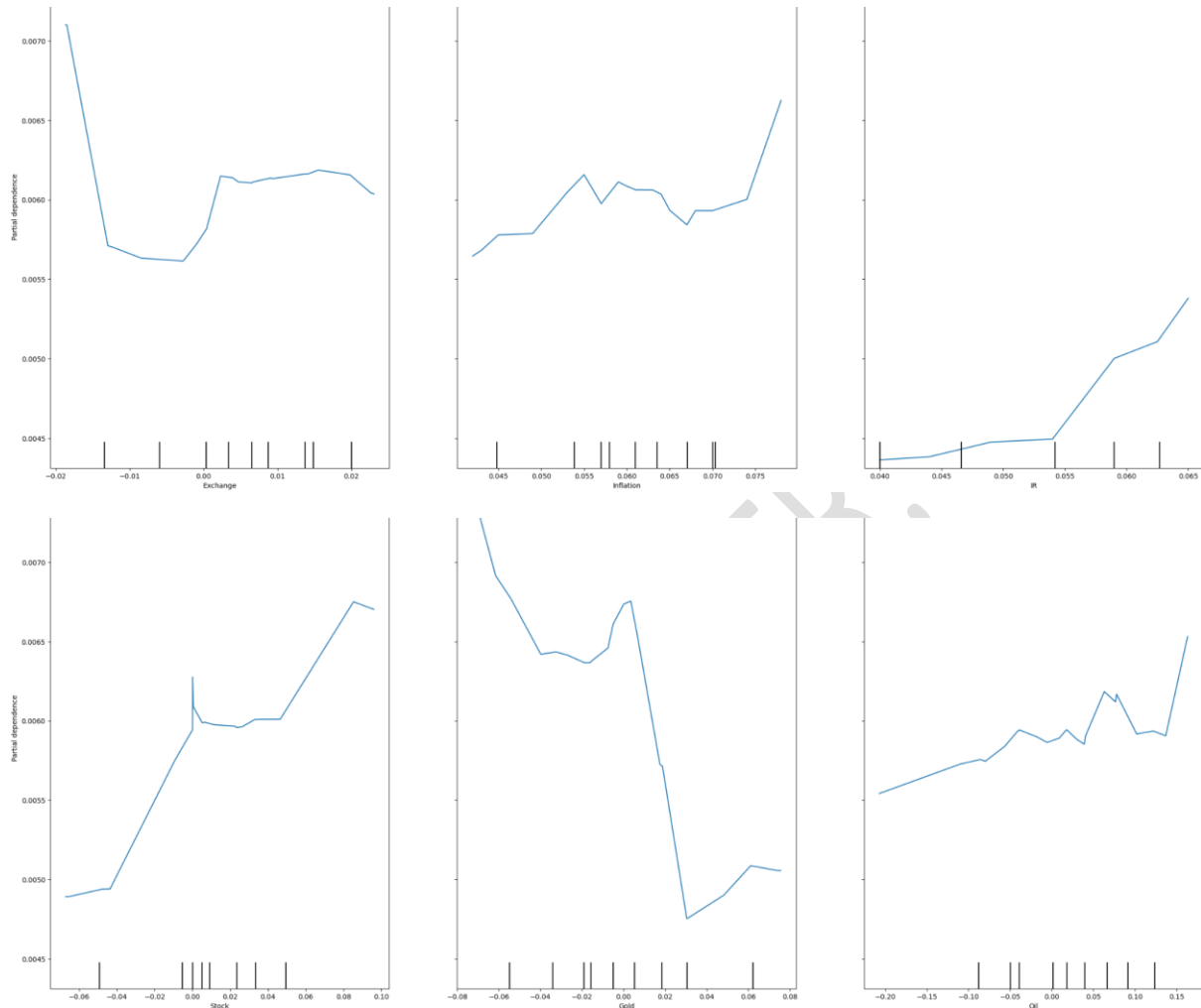
Correlation Results

Correlation - India	
Stock	0.04098
Gold	-0.19467
Oil	-0.056888
Exchange	0.010661
Inflation	0.134044
Interest Rate	0.497154

As per results observed, interest rate, inflation and gold has moderately positive association with target variable and the stock, exchange and oil has weak relationship validating the results of feature outcome extracted prior.

Partial Plots

The partial plots utilized to check the impact of single feature on target variable and assess the direction of relationship. The following graphs generated to show partial plots



As per results obtained, this can be observed that all macroeconomic indicators showed change in HPI with change in their values however as per predicted by feature importance and correlation, the interest rate, gold, and inflation is showing more fluctuations in HPI in comparison though there is clear evidence of strong association in interest rate, inflation and gold and weak association in stock, exchange, and oil.

Hypothesis validation in case of Pakistan

In case of Pakistan, the outcome by random forest algorithm observed that interest rate, gold and inflation are top 3 contributors to change in house price index whereas oil, exchange and stock prices are least contributed towards change in target variable. This is further validated by determining the correlation between macroeconomic indicators and house price index.

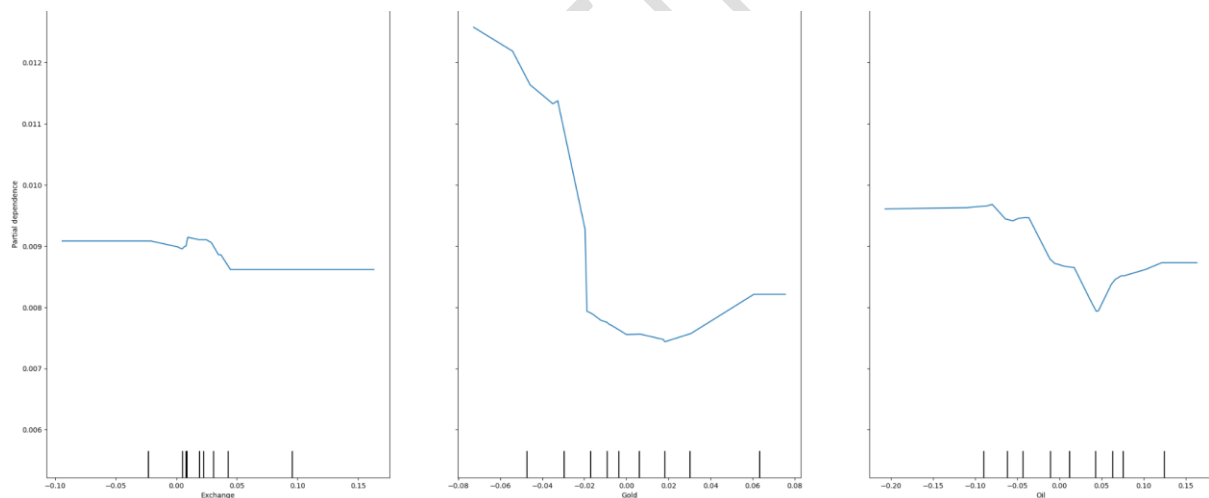
Correlation Results

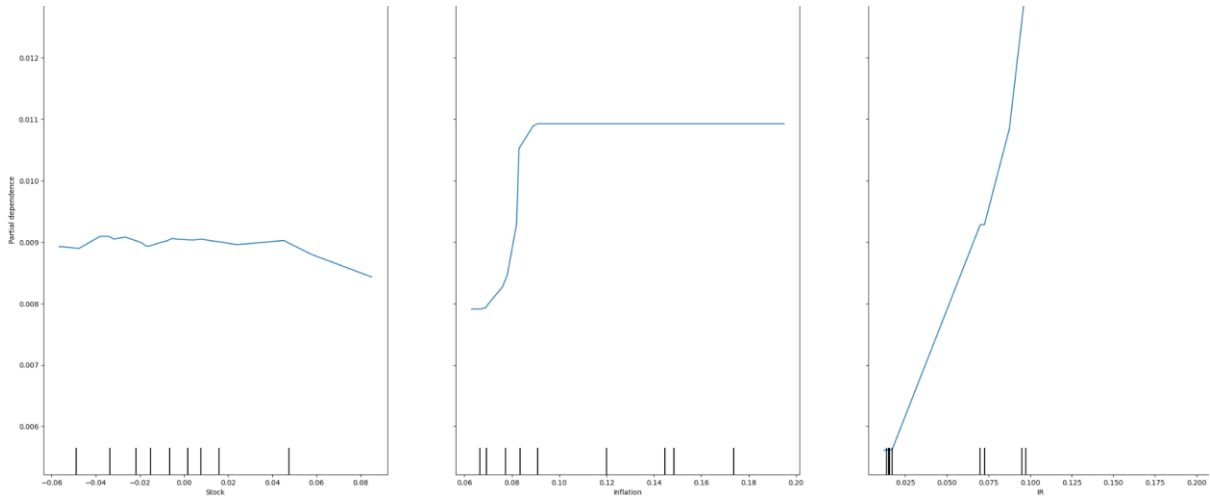
Correlation - Pakistan	
Stock	-0.042466
Gold	-0.254761
Oil	-0.06129
Exchange	0.042906
Inflation	0.28629
Interest Rate	0.231846

As per above results this can be seen that stock, oil and exchange rate have weak association with HPI and interest rate, inflation and gold prices have moderate association with HPI. This validated the outcome of feature importance determined earlier.

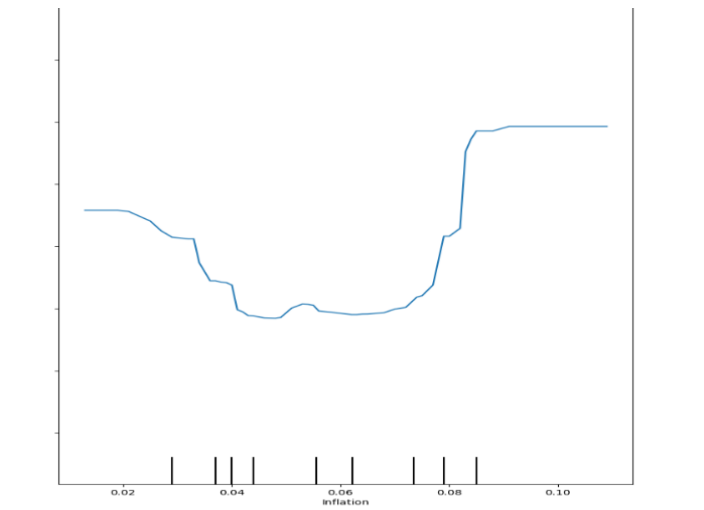
Partial Plots

The partial plots utilized to check the impact of single feature on target variable and assess the direction of relationship. The following graphs generated to show partial plots





As per above results, stock and exchange rate have insignificant impact on HPI whereas interest rate and gold are causing significant variations in HPI. As inflation rate is amongst top 3 contributors as per feature importance and correlation therefore partial plot generated for predicted training dataset to determine the impact



The above graph validated the impact of inflation rate on HPI and validated the results of feature importance and correlation. In case of Pakistan all variables have some association with change in HPI however exchange and stock have weak association with change in HPI.

5.2 Comparative Analysis of Countries

The selection of these countries for this study provided interesting insights due to huge population and diverse and emerging markets. Amongst them China observed a rapid growth in their economy and India is also moderately growing their economy whereas Pakistan is in crisis comparatively therefore it is interesting to analyze the contrast and people preferences in housing market. These countries have influential power in the region as well therefore combination of these countries enabled us to determine the relationship of economic indicators on HPI.

In case of Pakistan and India, stock prices have weak association with change in HPI, due to growing economies, people tend to move towards long term investments and desire potential returns.

5.3 Conclusion

The study aimed to determine the impact between changes in macroeconomic indicators on change in house price index of China, India, and Pakistan. The time series data of these countries analyzed with the predictive prowess of random forest and xgboost algorithms. These algorithms further tuned by configuring hyperparameters settings and ranking of feature importance generated by both algorithms. The performance metrics such as MAE, RMSE and MSE were utilized to analyze predictive accuracy. The focus of study lies on examining six hypotheses regarding the impact of macroeconomic indicators (stock prices, oil prices, gold prices, exchange rates, inflation rate and interest rate) on HPI changes. These hypotheses were further authenticated by analytical methods such as feature importance by random forest and xgboost, correlation identification between features and HPI, and partial plots generation to identify the individual feature contribution in influencing the change in HPI.

The results of study indicated that stock and oil prices have generally weak influence on HPI changes in these countries. In contrast, inflation rate and interest rates have consistent positive relationship with HPI. The gold prices have consistently negative or inverse relationship with change in HPI in all countries. It is important to recognize the limitations of this study, including the availability and duration of data for all macroeconomic indicators across the countries examined. However, despite these limitations, the study has provided valuable insights into the impacts of economic variables on changes in the house price index (HPI). These findings offer valuable information regarding how factors such as stock prices, oil prices, gold prices, exchange rates, interest rates, and inflation rates can potentially influence the fluctuations in HPI. As per this study its observed that, the associations of features with HPI may be strong or weak but there is clear evidence that change in macroeconomic indicators influenced the change in house price index of China, India and Pakistan under the assumptions of Arbitrage theory.

Chapter Summary

This chapter focused on authentication of hypothesis and determining the weak and strong association between macroeconomic indicators and HPI. The hypothesis was then validated by outcome of correlation factor and further analyzed by partial plots to identify the impact of single feature on target variable. The purpose and conclusion of study with evidence observed discussed in the end.

References

- Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House Price Prediction using Random Forest Machine Learning Technique. In *Procedia Computer Science* (Vol. 199, pp. 806–813). Elsevier BV. <https://doi.org/10.1016/j.procs.2022.01.100>
- Ali, G., & Zaman, K. (2017). Do house prices influence stock prices? Empirical investigation from the panel of selected European Union countries. In *Economic Research-Ekonomska Istraživanja* (Vol. 30, Issue 1, pp. 1840–1849). Informa UK Limited. <https://doi.org/10.1080/1331677x.2017.1392882>
- Ali, M., Samour, A., Joof, F., & Tursoy, T. (2022). Oil prices and gold prices on housing market in China: novel findings from the bootstrap approach. In *International Journal of Housing Markets and Analysis*. Emerald. <https://doi.org/10.1108/ijhma-09-2022-0132>
- Amemiya, T. (1980). Selection of Regressors. *International Economic Review*, 21(2), 331. doi:10.2307/2526185
- Apergis, Nicholas & Lamprinidis, Lampros. (2011). More Evidence on the Relationship between the Stock and the Real Estate Market.
- Ayberk, İ., & Önder, Z. (2022). House prices and bank loan portfolios in an emerging market: The role of bank ownership. In *Economic Modelling* (Vol. 106, p. 105683). Elsevier BV. <https://doi.org/10.1016/j.econmod.2021.105683>
- Bahmani-Oskooee, M., & Wu, T.-P. (2018). Housing prices and real effective exchange rates in 18 OECD countries: A bootstrap multivariate panel Granger causality. In *Economic Analysis and Policy* (Vol. 60, pp. 119–126). Elsevier BV. <https://doi.org/10.1016/j.eap.2018.09.005>
- Batayneh, K. I., & Al-Malki, A. M. (2015). The Relationship between House Prices and Stock Prices in Saudi Arabia: An Empirical Analysis. In *International Journal of Economics and Finance* (Vol. 7, Issue 2). Canadian Center of Science and Education. <https://doi.org/10.5539/ijef.v7n2p156>
- Bui, T. N. (2020). IMPACTS OF INTEREST RATE ON HOUSING PRICES: EVIDENCE FROM HO CHI MINH CITY, VIETNAM. *International Transaction Journal of Engineering, Management*, 11A05A: 1-7. <https://doi.org/10.14456/ITJEMAST.2020.81>

Cellmer, R., Cichulska, A., & Belej, M. (2020). Spatial Analysis of Housing Prices and Market Activity with the Geographically Weighted Regression. *ISPRS International Journal of Geo-Information*, 9(6), 380. <https://doi.org/10.3390/ijgi9060380>

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. In *Geoscientific Model Development* (Vol. 7, Issue 3, pp. 1247–1250). Copernicus GmbH. <https://doi.org/10.5194/gmd-7-1247-2014>

Chen, C., Zhai, H., Wang, Z., Ma, S., Sun, J., Wu, C., & Zhang, Y. (2022). Experimental Research on the Impact of Interest Rate on Real Estate Market Transactions. In F. Tramontana (Ed.), *Discrete Dynamics in Nature and Society* (Vol. 2022, pp. 1–14). Hindawi Limited. <https://doi.org/10.1155/2022/9946703>

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939785>

Dr, Sindhu. (2013). A study on impact of select factors on the price of Gold. *IOSR Journal of Business and Management*. 8. 84-93. 10.9790/487X-0848493.

Du, M., Zhang, X., Wang, Y., Tao, L., & Li, H. (2020). An operationalizing model for measuring urban resilience on land expansion. *Habitat International*, 102, 102206. <https://doi.org/10.1016/j.habitatint.2020.102206>

Fang, H., Chang, T.-Y., Lee, Y.-H., & Chen, W.-J. (2016). The impact of macroeconomic factors on the real estate investment trust index return on Japan, Singapore and China. In *Investment Management and Financial Innovations* (Vol. 13, Issue 4, pp. 242–253). LLC CPC Business Perspectives. [https://doi.org/10.21511/imfi.13\(4-1\).2016.11](https://doi.org/10.21511/imfi.13(4-1).2016.11)

Garriga, C., Hedlund, A., Tang, Y., & Wang, P. (2021). Rural-urban migration and house prices in China. *Regional Science and Urban Economics*, 91, 103613. <https://doi.org/10.1016/j.regsciurbeco.2020.103613>

Gu, J., Zhu, M., & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems With Applications*, 38(4), 3383–3386. <https://doi.org/10.1016/j.eswa.2010.08.123>

Habanabakize, T., & Dickason, Z. (2022). Political risk and macroeconomic effect of housing prices in South Africa. In *Cogent Economics & Finance* (Vol. 10, Issue 1). Informa UK Limited. <https://doi.org/10.1080/23322039.2022.2054525>

Harris, R., & Arku, G. (2006). Housing and economic development: The evolution of an idea since 1945. In *Habitat International* (Vol. 30, Issue 4, pp. 1007–1017). Elsevier BV. <https://doi.org/10.1016/j.habitatint.2005.10.003>

Hepşen, A., & Vatansever, M. (2012). RELATIONSHIP BETWEEN RESIDENTIAL PROPERTY PRICE INDEX AND MACROECONOMIC INDICATORS IN DUBAI HOUSING MARKET / SANTYKIS TARP GYVENAMOJO NT INDEKSO IR MAKROEKONOMINIŲ RODIKLIŲ DUBAJAUS BŪSTO RINKOJE. In *International Journal of Strategic Property Management* (Vol. 16, Issue 1, pp. 71–84). Vilnius Gediminas Technical University. <https://doi.org/10.3846/1648715x.2011.602756>

Hilbers, P., Lei, Q., Zacho, L., 2001. Real Estate Market Developments and Financial Sector Soundness (WP/01/129; IMF Working Paper)

- Hill, R.J., 2013. Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys* 27 (5), 879–914. <https://doi.org/10.1111/j.1467-6419.2012.00731.x>.
- Hoesli, M., & Reka, K. (2011). Volatility Spillovers, Comovements and Contagion in Securitized Real Estate Markets. In *The Journal of Real Estate Finance and Economics* (Vol. 47, Issue 1, pp. 1–35). Springer Science and Business Media LLC. <https://doi.org/10.1007/s11146-011-9346-8>
- Hoesli, M., Giaccotto, C., & Favarger, P. (1997). In *The Journal of Real Estate Finance and Economics* (Vol. 15, Issue 1, pp. 93–109). Springer Science and Business Media LLC. <https://doi.org/10.1023/a:1007753406308>
- Hong, Y., & Li, Y. (2020). Housing prices and investor sentiment dynamics: Evidence from China using a wavelet approach. In *Finance Research Letters* (Vol. 35, p. 101300). Elsevier BV. <https://doi.org/10.1016/j.frl.2019.09.015>
- Huang, X., & Chen, M. (2022). Understanding the role of housing in rural migrants' intention to settle in cities: Evidence from China. *Habitat International*, 128, 102650. <https://doi.org/10.1016/j.habitatint.2022.102650>
- Iacoviello, M., & Neri, S. (2010). Housing market spillovers: Evidence from an estimated DSGE model. *American Economic Journal: Macroeconomics*, 2, 125-64
- Ibrahim Ahmed Osman, A., Najah Ahmed, A., Chow, M. F., Feng Huang, Y., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. In *Ain Shams Engineering Journal* (Vol. 12, Issue 2, pp. 1545–1556). Elsevier BV. <https://doi.org/10.1016/j.asej.2020.11.011>
- Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2021). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, 111, 104919. <https://doi.org/10.1016/j.landusepol.2020.104919>
- KARAGÖZ, K., & ÖZKUBAT, G. (2021). Impact of Macroeconomic Factors on Housing Prices: An Analysis for Aegean Region. In *Journal of Yaşar University* (Vol. 16, Issue 62, pp. 867–889). Journal of Yasar University. <https://doi.org/10.19168/jyasar.846027>
- Kau, J.B., Keenan, D.C., 1995. An overview of the option-theoretic pricing of mortgages. *J. Hous. Res.* 6, 217–244.
- KAUKO, T., HOOIMEIJER, P., & HAKFOORT, J. (2002). Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modelling. *Housing Studies*, 17(6), 875–894. <https://doi.org/10.1080/02673030215999>
- Kiely, T. J., & Bastian, N. D. (2019). The spatially conscious machine learning model. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 13(1), 31–49. <https://doi.org/10.1002/sam.11440>
- Killins, R. N., Egly, P. V., & Escobari, D. (2017). The impact of oil shocks on the housing market: Evidence from Canada and U.S. In *Journal of Economics and Business* (Vol. 93, pp. 15–28). Elsevier BV. <https://doi.org/10.1016/j.jeconbus.2017.07.002>
- Korkmaz, Ö. (2019). The relationship between housing prices and inflation rate in Turkey. In *International Journal of Housing Markets and Analysis* (Vol. 13, Issue 3, pp. 427–452). Emerald. <https://doi.org/10.1108/ijhma-05-2019-0051>

Kuang, Weida and Liu, Peng, (2015), Inflation and House Prices: Theory and Evidence from 35 Major Cities in China, *International Real Estate Review*, 18, issue 2, p. 217-240, <https://EconPapers.repec.org/RePEc:ire:issued:v:18:n:02:2015:p:217-240>.

Kumar, S. (2003). Power Cycle Analysis of India, China, and Pakistan in Regional and Global Politics. *International Political Science Review / Revue Internationale de Science Politique*, 24(1), 113–122. <http://www.jstor.org/stable/1601333>

Lin, C.-C., & Tsai, I.-C. (2021). THE SPECIAL EFFECT OF INTEREST RATE CUTS ON HOUSING PRICES. In *Journal of Business Economics and Management* (Vol. 22, Issue 3, pp. 776–798). Vilnius Gediminas Technical University. <https://doi.org/10.3846/jbem.2021.14576>

Lin, T. C., & Lin, Z.-H. (2011). Are stock and real estate markets integrated? An empirical study of six Asian economies. In *Pacific-Basin Finance Journal* (Vol. 19, Issue 5, pp. 571–585). Elsevier BV. <https://doi.org/10.1016/j.pacfin.2011.05.001>

Lisi, G., 2013. On the functional form of the hedonic price function: a matching-theoretic model and empirical evidence. *Int. Real Estate Rev.* 16 (2), 189–207.

Macroeconomic Determinants of House Prices in Malaysia. (2020). In *Jurnal Ekonomi Malaysia* (Vol. 54, Issue 1). Penerbit Universiti Kebangsaan Malaysia (UKM Press). <https://doi.org/10.17576/jem-2020-5401-11>

Magnusson, A., & Makdessi, L. (2019). Is there a relationship between oil prices and house price inflation? (Dissertation). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-44471>

Martin, E. W., & Shaheen, S. A. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. *Journal of Transport Geography*, 41, 315–324. <https://doi.org/10.1016/j.jtrangeo.2014.06.026>

Meen, G. (1999). Regional House Prices and the Ripple Effect: A New Interpretation. *Housing Studies*, 14(6), 733–753. <https://doi.org/10.1080/02673039982524>

Melnychenko, O., Osadcha, T., Kovalyov, A., & Matskul, V. (2022). Dependence of Housing Real Estate Prices on Inflation as One of the Most Important Factors: Poland's Case. In *Real Estate Management and Valuation* (Vol. 30, Issue 4, pp. 25–41). Walter de Gruyter GmbH. <https://doi.org/10.2478/remav-2022-0027>

Mensi, W., Beljid, M., Boubaker, A., & Managi, S. (2013). Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. In *Economic Modelling* (Vol. 32, pp. 15–22). Elsevier BV. <https://doi.org/10.1016/j.econmod.2013.01.023>

Mensi, W., Reboredo, J. C., Ugolini, A., & Vo, X. V. (2022). Switching connectedness between real estate investment trusts, oil, and gold markets. In *Finance Research Letters* (Vol. 49, p. 103112). Elsevier BV. <https://doi.org/10.1016/j.frl.2022.103112>

Mick Silver, 2012. "Why House Price Indexes Differ: Measurement and Analysis," IMF Working Papers 2012/125, International Monetary Fund.

Mohan, S., Hutson, A., MacDonald, I., & Lin, C. C. (2019). Impact of macroeconomic indicators on housing prices. In *International Journal of Housing Markets and Analysis* (Vol. 12, Issue 6, pp. 1055–1071). Emerald. <https://doi.org/10.1108/ijhma-09-2018-0070>

- Nazlioglu, S., Gormus, N. A., & Soytaş, U. (2016). Oil prices and real estate investment trusts (REITs): Gradual-shift causality and volatility transmission analysis. In *Energy Economics* (Vol. 60, pp. 168–175). Elsevier BV. <https://doi.org/10.1016/j.eneco.2016.09.009>
- Njoroge, C. G., Muturi, W., & Oluoch, O. (2019). EXCHANGE RATE AND PERFORMANCE OF THE RESIDENTIAL PROPERTY MARKET IN KENYA. In *International Journal of Finance & Banking Studies* (2147-4486) (Vol. 8, Issue 3, pp. 88–100). Center for Strategic Studies in Business and Finance SSBFNET. <https://doi.org/10.20525/jfbs.v8i3.507>
- Ridker, R.G., Henning, J.A., 1967. The determinants of residential property values with special reference to air pollution. *Rev. Econ. Stat.* 49 (2), 246–257. <http://www.jstor.org/stable/1928231>.
- Salam, S., Kalam Azad, Md. A., Salam, Md. A., & M. Islam, M. (2020). Rural-to-Urban Migration and Realization of Expected Better Life in Bangladesh: An Empirical Study in Rajshahi City Corporation. In *International Journal of Asian Social Science* (Vol. 10, Issue 10, pp. 612–622). <https://doi.org/10.18488/journal.1.2020.1010.612.622>
- Salisu, A. A., Raheem, I. D., & Ndako, U. B. (2020). The inflation hedging properties of gold, stocks and real estate: A comparative analysis. In *Resources Policy* (Vol. 66, p. 101605). Elsevier BV. <https://doi.org/10.1016/j.resourpol.2020.101605>
- Shi, S., Jou, J.-B., & Tripe, D. (2014). Can interest rates really control house prices? Effectiveness and implications for macroprudential policy. In *Journal of Banking & Finance* (Vol. 47, pp. 15–28). Elsevier BV. <https://doi.org/10.1016/j.jbankfin.2014.06.012>
- SHILLER, R. J. (1993). Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures. *The Journal of Finance*, 48(3), 911–931. <https://doi.org/10.1111/j.1540-6261.1993.tb04024.x>
- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. In *American Journal of Applied Mathematics and Statistics* (Vol. 8, Issue 2, pp. 39–42). Science and Education Publishing Co., Ltd. <https://doi.org/10.12691/ajams-8-2-1>
- Song, J., Gao, Y., Yin, P., Li, Y., Li, Y., Zhang, J., Su, Q., Fu, X., & Pi, H. (2021). The Random Forest Model Has the Best Accuracy Among the Four Pressure Ulcer Prediction Models Using Machine Learning Algorithms. In *Risk Management and Healthcare Policy: Vol. Volume 14* (pp. 1175–1187). Informa UK Limited. <https://doi.org/10.2147/rmhp.s297838>
- TITMAN, S. (1982). The Effects of Anticipated Inflation on Housing Market Equilibrium. In *The Journal of Finance* (Vol. 37, Issue 3, pp. 827–842). Wiley. <https://doi.org/10.1111/j.1540-6261.1982.tb02226.x>
- Tra, C. I. (2010). A discrete choice equilibrium approach to valuing large environmental changes. *Journal of Public Economics*, 94(1–2), 183–196. <https://doi.org/10.1016/j.jpubeco.2009.10.006>
- Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, 174, 433–442. <https://doi.org/10.1016/j.procs.2020.06.111>
- Tu, M., Li, W., Orfila, O., Li, Y., & Gruyer, D. (2021). Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. *Transportation Research Part D: Transport and Environment*, 93, 102776. <https://doi.org/10.1016/j.trd.2021.102776>

- Uzuner, G., & Adewale, A. A. (2019). Does asymmetric nexus exist between agricultural land and the housing market? Evidence from non-linear ARDL approach. *Environmental Science and Pollution Research*, 26(8), 7677–7687. <https://doi.org/10.1007/s11356-019-04211-4>
- Wang, X., Li, K., & Wu, J. (2020). House price index based on online listing information: The case of China. *Journal of Housing Economics*, 50, 101715. <https://doi.org/10.1016/j.jhe.2020.101715>
- Williams, John C., Measuring the Effects of Monetary Policy on House Prices and the Economy (October 2016). BIS Paper No. 88b, Available at SSRN: <https://ssrn.com/abstract=2861309>
- Yamaka, W., Liu, J., Li, M., Maneejuk, P., & Dinh, H. (2022). Analyzing the Causality and Dependence between Exchange Rate and Real Estate Prices in Boom-and-Bust Markets: Quantile Causality and DCC Copula GARCH Approaches. In *Axioms* (Vol. 11, Issue 3, p. 113). MDPI AG. <https://doi.org/10.3390/axioms11030113>
- Yang, L., & Zhiqiang, H. (2012). on Correlation between RMB Exchange Rate and Real Estate Price based on Financial Engineering. In *Systems Engineering Procedia* (Vol. 3, pp. 146–152). Elsevier BV. <https://doi.org/10.1016/j.sepro.2011.11.020>
- Yang, L., Liang, Y., Zhu, Q., & Chu, X. (2021). Machine learning for inference: using gradient boosting decision tree to assess non-linear effects of bus rapid transit on house prices. *Annals of GIS*, 27(3), 273–284. <https://doi.org/10.1080/19475683.2021.1906746>
- Yousaf, I., & Ali, S. (2020). Integration between real estate and stock markets: new evidence from Pakistan. In *International Journal of Housing Markets and Analysis* (Vol. 13, Issue 5, pp. 887–900). Emerald. <https://doi.org/10.1108/ijhma-01-2020-0001>
- Zhang, S., & He, M. (2017). An Empirical Study on the Influence of Price scissors of M1-M2 Growth on House Prices. In *Proceedings of the 2016 2nd International Conference on Economics, Management Engineering and Education Technology (ICEMEET 2016)*. 2016 2nd International Conference on Economics, Management Engineering and Education Technology (ICEMEET 2016). Atlantis Press. <https://doi.org/10.2991/icemeet-16.2017.5>
- Zhou, Q., Shao, Q., Zhang, X., & Chen, J. (2020). Do housing prices promote total factor productivity? Evidence from spatial panel data models in explaining the mediating role of population density. *Land Use Policy*, 91, 104410. <https://doi.org/10.1016/j.landusepol.2019.104410>
- Zhou, T., Clapp, J. M., & Lu-Andrews, R. (2021). Is the behavior of sellers with expected gains and losses relevant to cycles in house prices? *Journal of Housing Economics*, 52, 101750. <https://doi.org/10.1016/j.jhe.2021.101750>
- Zietz, J., Sirmans, G. S., & Smersh, G. (2008). The Impact of Inflation on Home Prices and the Valuation of Housing Characteristics Across the Price Distribution. In *Journal of Housing Research* (Vol. 17, Issue 2, pp. 119–137). Informa UK Limited. <https://doi.org/10.1080/10835547.2008.12091992>