Dean&Francis

ISSN 2959-6157

Causal Inference Analysis: Relationship Between Interest Rate and Housing Price

Zhaoyang Liu

International Leadership of Texas, Texas, 75044, United States

Corresponding author: liuandy284@gmail.com

Abstract:

In this research, the researchers are going to talk about the causal inference of interest rates on housing prices using a combination of causal inference techniques and predictive modeling. The researcher will introduce the basic introduction of causal inference, and four different rules-adjustment formulas, backdoor adjustment, front door adjustment, and collision situation- to identify the relationships between different variables, and based on these relationships, make a causal graph with causal direction. By employing decision-making trees and Vector Autoregression (VAR) models to predict the impact of interest rate changes on housing prices. This study offers insights into how fluctuations in interest rates influence housing markets, enhancing the predictive power and accuracy of macroeconomic decisions. The researcher collected data from the last ten years to make predictions and figure out the overall tendency of the housing market. Comparing the real data and the predicted data, finally figured out the importance of interest rates and exchange rates

Keywords: Decision Tree; Vector Autoregression Model; Housing prices.

1. Introduction

Causal inference is a branch of mathematics. It is a part of statistics, which deals with how to identify the relationships between different variables. In the real world, a lot of the policies and decisions rely on the understanding of causal inference. Because of that, causal inference can be applied in social science, medicine, and economy, and bring a deep influence in such fields [1].

In the basic statistic method, like regression analysis. It allows researchers to reveal correlations between variables, but it is really hard to support the causal interpretation. The method of causal inference can support this disadvantage, through constructing and proving the causal model, help people to get access to a deep understanding of the production process of data. During the research process, the random comparing test (RCTs) usually be recognized as the most reliable way, but based on the theory, economy, and application problem, RCTs cannot be used. Thus, recently, researchers developed a lot of methods which to observe and analyze the data, such as propensity score matching, instrumental variables, and synthetic control methods [2].

However, these means also need to face many challenges, like how to solve confounding variables. Accompany with the improvement of calculation ability, and the reliability of data, causal inference becomes a new method. In this case, based on this knowledge, the variables correlate, but this does not mean they have causal effects. For instance, the researcher collects data for monthly ice cream sales and monthly shark attacks around the United States each year and finds out that these two variables are highly correlated. But did these two variables have causal effects? Actually, these variables do not have any causal relations, the reason they have high correlation is that some other factors, such as season and temperature. The role of causal inference is to identify the causal effects between these variables. In this essay, the researcher will focus on the basic concept, main method and application of causal inference.

2. Methods and Theory

2.1 Background

The researcher can use adjustment formulas, backdoor adjustment, front-door adjustment, and collision situations to analyze the overall situation. These four methods need to be used under different conditions., the researcher will use the VAR model to make predictions or estimates for the model. Before the researcher illustrates these four rules, the author needs to be familiar with some basic relationships: Chain, fork, and collision [3].

Firstly, the chain is a relationship in which every variable affects each other in a linear order. For instance, there are three variables A, B, and C. For these three variables, A will cause B, and B will cause C. Based on this relationship, people can call that a chain relationship. The second relationship is the fork. Just like the name fork, the variables in this relationship will affect each other like a fork. For example, the variable A and C are both caused by the variable B. The last relationship is a collision. In collision conditions, A and C can both cause C. In this case, A and C will be mutually against.

2.2 Do operator and Adjustment formula

What is do operator? In fact, the operator focuses on canceling out all arrows and eliminating the interference of a fake path. Take this picture shown in Fig. 1 for example, there is a fake path between X and Y. The route of this fake path is U to X and U to Y, and people usually called the backdoor path. On the other hand, if the variable U is unknown, then the process needs to go through the variable Z. In this case, this concept called this front-door adjustment [4].

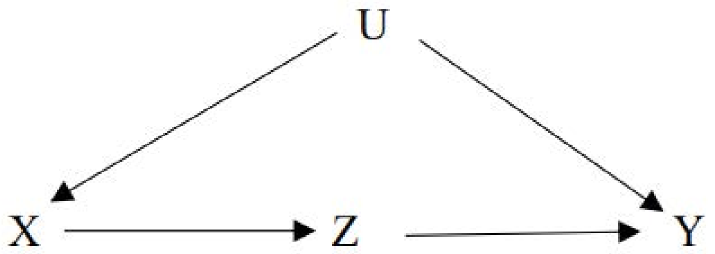

Fig. 1 Illustrate a basic causal graph

In fact, an idealized method of estimating the effect of an interest rate would be to simply consider two scenarios. On the one hand, administer the change in interest rate (do(X=1)) to the entire population and observe how many recovers. On the other hand, administer the change in interest rate to zero (do(X=0)) and observe how many recovers. In these scenarios, the equation will look like this:

$$P(Y \mid do(X=1)) - P(Y \mid do(X=0)) \tag{1}$$

This is known as the “Causal Effect Difference” or “Average Causal Effect”. By controlling the variable X, the author can measure the difference between two different outcomes. The causal graph will look like this shown in Fig. 2.

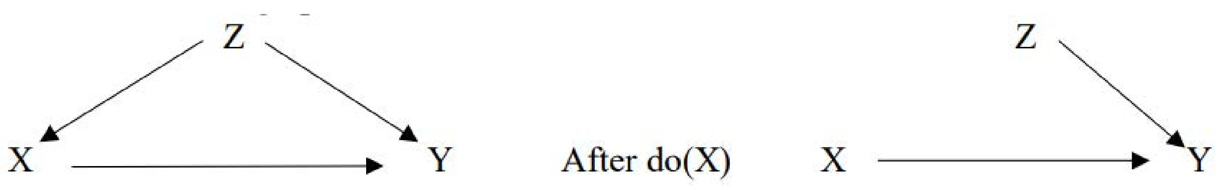

Fig. 2 Illustrate the function of do operator

Under this condition, the use of the ACE expression will like this:

ISSN 2959-6157

$$P(y \mid do(x)) = \sum_z P(y \mid x,z)P(z) \tag{2}$$

This is adjustment formula.

2.3 Backdoor and Front-door Adjustment

Backdoor adjustment is a mathematical strategy that can be used in causal inference. It includes identifying the real causal relationships between different variables and illustrating casual direction maps which represent these causal relationships. The confounding variable is the third variable that affects the assumed cause and the assumed effect. It has an impact on both treatment and outcomes and can lead to biased estimates of treatment effectiveness. By identifying confounding factors, researchers can effectively adjust for these confounding variables when controlling and blocking all backdoor paths. Finally, in observational studies where randomization is not feasible, backdoor adjustments are critical to obtaining unbiased estimates, allowing researchers to understand the true effect of an intervention by considering confounding effects (see Fig. 3).

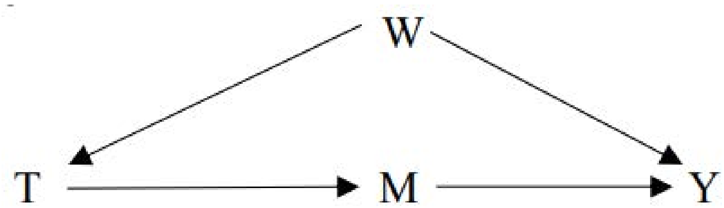

Fig. 3 illustrates a causal relationship. W = unknow policies, T = interest rates, M = exchange rates, Y = housing prices

For front-door adjustment, it is a kind of adjustment strategy when there is no backdoor path between two different variables. In this case, the researcher can choose a mediator to create a front-door path. Now based on the graph below, the author wants to analyze the causal effects of interest rates on housing prices. This process needs to pass through a mediator—exchange rates. (Because the policy is unknown, and cannot be observed, so the author needs to use front-door adjustment) To be more specific, the researcher can divide this event into three parts: First, the researcher needs to identify the causal effects of interest rates on exchange rates, just like the following equation [5]

$$P(m \mid do(t)) = P(m \mid t) \tag{3}$$

“Do” interest rates mean there is no backdoor path from interest rate to exchange rate because housing price is a collider.

Second, identify the causal effects of the exchange rate on housing prices $P(y \mid do(m))$. In fact, there is a backdoor path from exchange rate to interest rate to policy to housing prices. Fortunately, the researcher can block that backdoor path by condition on the variable interest rate. In this case, the author can use a backdoor path here, and write the equation:

$$P(y \mid do(m)) = \sum_t P(y \mid m,t)P(t) \tag{4}$$

Finally, the author needs to add these two outcomes together, and finally calculate the probability of interest rate on housing prices. Based on what talked about in the previous section, the researcher can write the final equation like this [6]

$$P(y \mid do(t)) = \sum_m P(m \mid do(t))P(y \mid do(m)) \#(5)$$

The researcher will use three variables A, B, and C to provide further explanation. The variables A and B, are independent of each other, but both influence a third variable C. Here, C is the collider. The paths might look like A to C and B to C. For instance, suppose you are studying the relationship between exercise (A) and diet (B) on heart health. Both exercise and diet can influence the body mass index (BMI, C). In this case, BMI is a collider. If you control for BMI in the analysis (e.g., by only looking at people with a certain BMI), you might find a correlation between exercise and diet, even if there is no direct causal relationship between them. This correlation is purely due to conditioning on the collider C. In other situations, in the economic field, economic growth is a common outcome influenced by multiple factors. Interest rates, inflation rates, government policies, and the global economic environment can all influence economic growth. In this case, economic growth is a collider for all the arrows coming from these factors.

2.4 Decision Tree and VAR Model

The decision tree is a graph indicator for decisions and the outcomes of these decisions, it includes consequences, resource cost, and application. It usually is used in decision analysis and statistics. Based on the data that the user inputs in this model, the decision tree can predict the outcomes and construct a model to illustrate these outcomes. The combination of decision trees is by nodes and branches, starting through the question and then showing the internal nodes that represent subsequent decisions or tests on attributes, and ending with leaf nodes that signific the final outcomes [7]. In each internal node, based on the specific standard, such as Gini impurity or information gain, spread the data, forming branches, leading to other outcomes. The route from the outcome to the leaf nodes shows that this model can be used in classification or regression models. The decision tree is easy to explain and observe, which leads to the popularity of this model for practitioners and stakeholders. It can be applied in a lot of fields, like categorization, regression, and decision analysis, helping to make a good decision from different scenarios and events. Although it offers some benefits, for instance, simplicity, minimal data preprocessing, and the ability to handle both numerical and categorical data, it also will occur in overfitting situations, extremely in some complicated trees. The structure of the decision tree can be affected by some small changes, which shows this model is very sensitive. Overall, the decision tree is a powerful tool for making decisions and prediction modeling, and provides a comprehensive understanding of the relationships between different variables.

A Vector Autoregression (VAR) model is a statistical model used to capture the linear interdependencies among multiple time series data, widely utilized in econometrics and finance for analyzing and forecasting systems where multiple variables influence each other over time. Unlike univariate models that focus on a single time series, VAR models analyze multiple time series simultaneously, allowing for the examination of how each variable affects the others. It is then a linear time-series model designed to capture the joint dynamics of multiple time series with e, normally distributed $N(0,\Sigma)$. The model consists of a system of equations, with each equation representing one of the variables in terms of its own lagged values and the lagged values of the other variables, and the parameters are typically estimated using methods like Ordinary Least Squares (OLS) [8]. VAR models are commonly applied in economic forecasting for predicting indicators such as GDP, inflation, and interest rates, and they assist in assessing the impact of monetary or fiscal policy changes on an economy, as well as analyzing relationships between different financial assets. While they offer flexibility and do not require strong assumptions about the underlying data distribution, VAR models can risk overfitting due to the inclusion of many variables and lags, and their results can be difficult to interpret because of the complexity of interactions among multiple variables. Regarding the error terms (innovation) of each endogenous variable which could be correlated and constitute a disadvantage then it is better to create new error terms that are orthogonal (structural) with diagonalized variance-covariance matrix of the error terms, this new error term Ɛt (normally distributed) therefore avoid the crossed. Overall, VAR models are powerful tools for understanding and forecasting the behavior of interconnected time series, providing valuable insights into dynamic systems.

3. Result

The author identifies several factors that may will affect the result: Interest rate, inflation rate, economic growth, capital flows, market expectations, government policies, and global economic environment. Observing these factors and using the four different strategies mentioned in the last section to make a causal graph shown in Fig. 4. This causal graph gives people a comprehensive understanding of the entire market.

ISSN 2959-6157

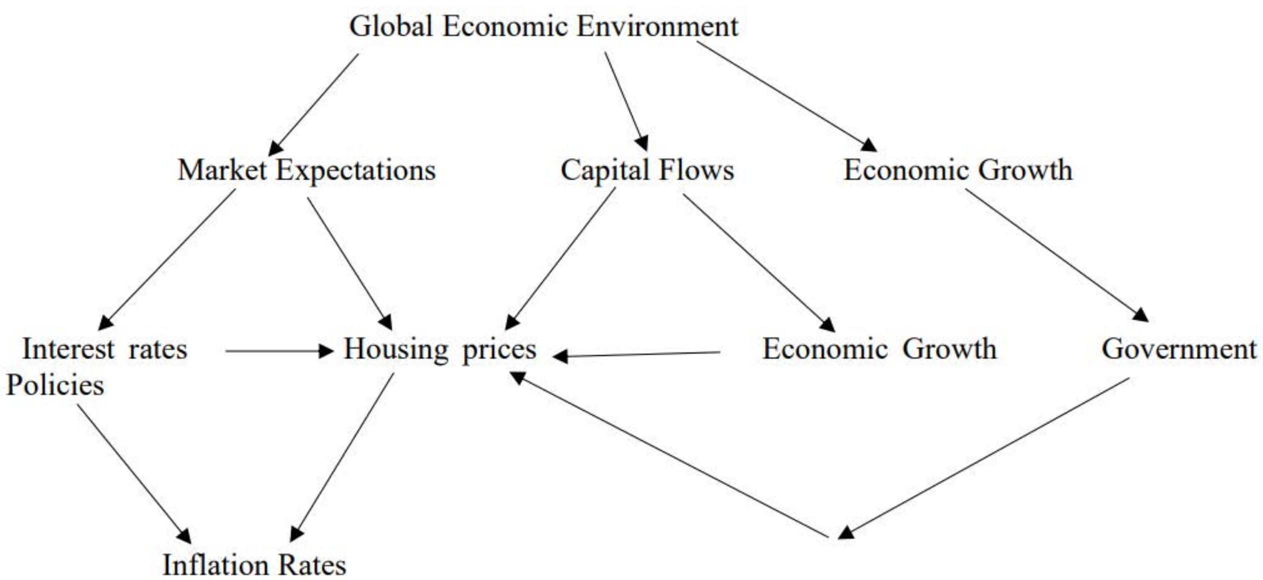

Fig. 4 Illustrate the causal relationship in the housing market

3.1 Decision tree for Predicting Exchange Rate

The researcher finds the interest rates, the exchange rate of the Chinese yuan to the US dollar, and the housing index. By using this tool, the researcher can observe the outcome through the deep of the tree and the distribution of this model. From the following graph shown in Fig. 5, the decision tree shows a complicated model, there are a lot of branches for this tree, which means numerous situations need to be considered. Also, because of this deep tree system, to ensure the accuracy of the outcome. The researcher used the regression model, which can increase the stability of this model.

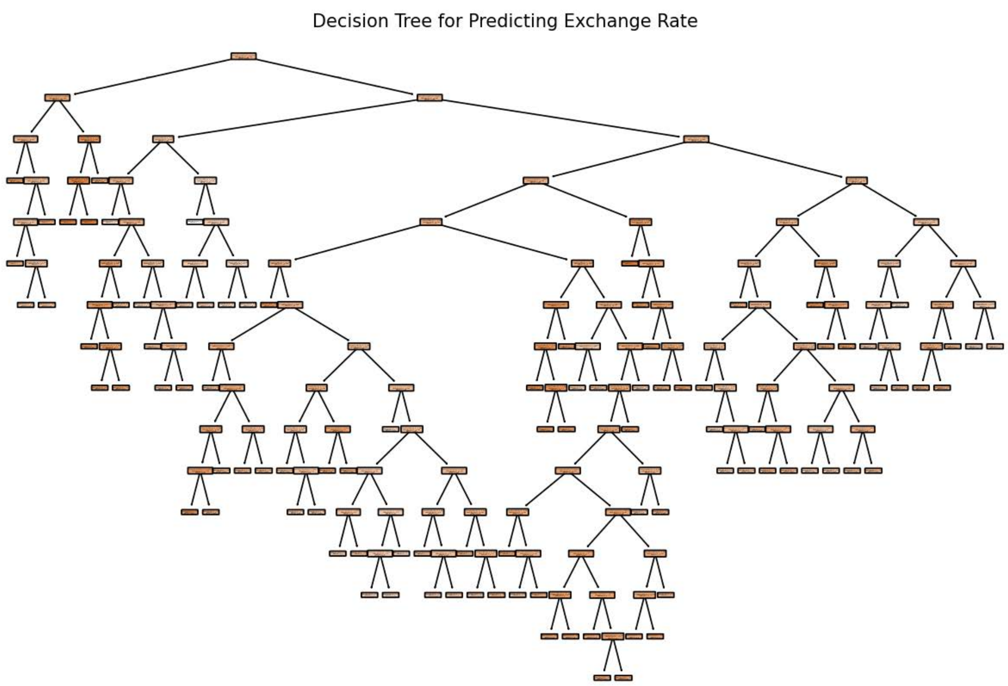

Fig. 5 The result of decision tree

The decision tree analysis reveals that both the Interest Rate and the House Price Index contribute almost equally to predicting the Exchange Rate (USD/CNY), with the Interest Rate having a slightly higher importance (about 51%) compared to the House Price Index (about 49%).

3.2 Compare with actual and predicted by using the VAR model

The data that the author used is from the last ten years in China. The time interval of each data is a month. The author uses the VAR model to make the predictions and find out the fluctuations. In this case, making the compare and to see the overall tendency of the entire market. The VAR model's prediction for the exchange rate (USD/CNY) is plotted against the actual values, see Fig. 6. The mean squared error (MSE) of the prediction is approximately 0.0182, which indicates the average squared difference between the actual and predicted values. There is a certain error between the prediction of the VAR model and the actual exchange rate, but the overall trend is still relatively close, indicating that the fluctuation of interest rate and housing price can indeed reflect the fluctuation of the exchange rate to a certain extent.

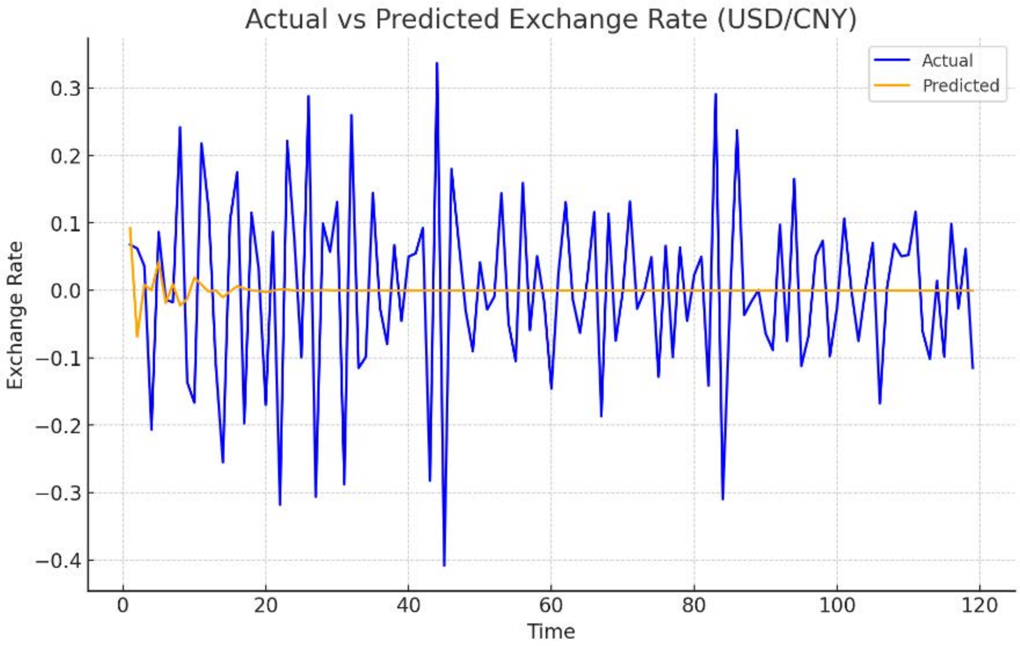

Fig. 6 The result of VAR model

4. Conclusion

In this research, the researchers talk about the causal effects of interest rates on housing prices through four concepts and two mathematic tools. Creating a causal graph through interest rates, economic growth, inflation rates, government policies, and market expectations. This study indicates that interest rates have more influence than exchange rates by using the decision tree. The use of a decision tree and VAR model is very important, First, the use of a decision tree provides the importance of interest rates and exchange rates on the housing market, Also, it illustrates the effect of other key factors. Next, the application of the VAR model is used to conduct a dynamic analysis of interest rates, housing prices, and exchange rates. Under this condition, compare the prediction tendency and real tendency. Finally, figure out that the interest rates have certain effect on housing prices. This paper provides an in-depth insight into the impact of interest rate policy on housing prices and has important practical significance for policymaking and economic forecasting in the real estate market.

References

[1] Njoku, Obinna. Decision Trees and Their Application for Classification and Regression Problems. MSU Graduate Theses, 2019.

[2] Nicolas-Domenic Reiter, Jonas Wahl, Andreas Gerhardus, Jakob Runge. Causal Inference on Process Graphs, Part II: Causal Structure and Effect Identification. 2024, arXiv.2406.17422.

[3] Richard Anderson, Dennis L. Hoffman, Robert H. Rasche. A Vector Error-Correction Forecasting Model of the U.S. Economy. Journal of Macroeconomics, 2002, 24(4):569–598.

[4] Keke Gai, Zijun Wang. Decision Tree-Based Federated

ISSN 2959-6157

Learning: A Survey. Blockchains 2024, 2(1): 40-60.
[5] Konstantin Kashin, Adam N. Glynn. Front-Door Versus Back-Door Adjustment With Unmeasured Confounding: Bias Formulas for Front-Door and Hybrid Adjustments With Application to a Job Training Program. Journal of the American Statistical Association, 2018, 113(523): 1040–49.
[6] Walsh, Vincent. Time: The Back-Door of Perception. Trends in Cognitive Sciences, 2003, 7(8): 335–38.
[7] Marcel Wienöbst, Benito Van der Zander, Maciej Liskiewicz. Linear-Time Algorithms for Front-Door Adjustment in Causal Graphs. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(18): 20577-20584.
[8] Ruska Ventsislavova Krasteva. Do illegal Migration Affect tourism Arrivals in Greece? A Vector Autoregression Analysis. Revista de Gestão Social e Ambienta, 2024, l18(7): e8093.