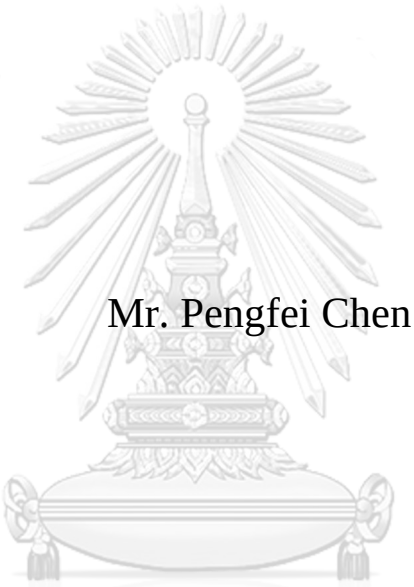
Forecasting the Thailand housing price indexA case study on condo HPI during the COVID-19

Mr. Pengfei Chen



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

An Independent Study Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Business and Managerial Economics
Field of Study of Business and Managerial Economics
FACULTY OF ECONOMICS
Chulalongkorn University
Academic Year 2020
Copyright of Chulalongkorn University

การพยากรณ์ดัชนีราคาที่อยู่อาศัยของประเทศไทยกรณีศึกษาคอนโด HPI ในช่วง COVID-19

นายเผิงเฟย เฉิน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต
สาขาวิชาเศรษฐศาสตร์ธุรกิจและการจัดการ สาขาวิชาเศรษฐศาสตร์ธุรกิจและการจัดการ
คณะเศรษฐศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2563
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Independent Study Title	Forecasting the Thailand housing price indexA case study on condo HPI during the COVID-19
By	Mr. Pengfei Chen
Field of Study	Business and Managerial Economics
Thesis Advisor	Assistant Professor PACHARASUT SUJARITTANONTA, Ph.D.

Accepted by the FACULTY OF ECONOMICS, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Arts

INDEPENDENT STUDY COMMITTEE

..... Chairman
(Assistant Professor RATIDANAI HOONSAWAT, Ph.D.)

..... Advisor
(Assistant Professor PACHARASUT SUJARITTANONTA, Ph.D.)

..... Examiner
(Associate Professor Chalaiporn Amonvatana, Ph.D.)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

เผิงเฟย เฉิน : การพยากรณ์ดัชนีราคาที่อยู่อาศัยของประเทศไทยกรณีศึกษาคอนโด HPI ในช่วง COVID-19. (Forecasting the Thailand housing price indexA case study on condo HPI during the COVID-19) อ.ที่ปรึกษาหลัก : ผศ. ดร.พัชรสุทธิ์ สุจริตตานนท์

HPI (ดัชนีราคาบ้าน) เป็นการคำนวณหามูลค่าการเปลี่ยนแปลงของราคาบ้านที่ขายให้กับครัวเรือน และวัดค่าจากเปอร์เซ็นต์การเปลี่ยนแปลงโดยมีดัชนีอยู่ที่ 100 ซึ่งถือว่าเป็นเกณฑ์มาตรฐาน กล่าวอีกนัยหนึ่งคือ HPI สะท้อนให้เห็นถึงความผันผวนของตลาดที่อยู่อาศัย นอกจากนี้การลงทุนในอสังหาริมทรัพย์ยังเป็นทางเลือกที่ดีสำหรับของผู้คนที่หลีกเลี่ยงความเสี่ยงต่างๆ ดังนั้น เมื่อเกิดการระบาดของ COVID-19 ทั่วโลก ในระยะสั้นการคาดการณ์การลอยตัวของ HPI อาจช่วยให้ตลาดที่อยู่อาศัย สามารถคาดการณ์สำหรับสิ่งที่จะเป็นไปได้ในอนาคต ทั้งในด้านรายบุคคลและด้านนโยบาย ในบทความนี้มีการจำลองแบบจำลอง 5 ประเภท บนพื้นฐานสมมุติฐานที่ว่า ก่อนประเทศไทยจะเปิดอีกครั้ง ด้วยข้อมูลตั้งแต่ 1 กรกฎาคม 2020 และคาดการณ์ HPI ในไตรมาสที่สามตามมา ผลการศึกษาพบว่าจากแบบจำลองทุกแบบจะถูกเปรียบเทียบโดยค่าความผิดพลาด และการทดลองแสดงให้เห็นถึงแบบจำลองการถดถอยพหุคูณตามความแม่นยำมีประสิทธิภาพค่อนข้างดีกว่าแบบจำลองอื่น ๆ อีกสี่แบบ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา	เศรษฐศาสตร์ธุรกิจและการจัดการ	ลายมือชื่อนิสิต .....
ปีการศึกษา	2563	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6284113729 : MAJOR BUSINESS AND MANAGERIAL ECONOMICS
KEYWOR D: house price index, forecasting

Pengfei Chen : Forecasting the Thailand housing price indexA case study on condo HPI during the COVID-19. Advisor: Asst. Prof. PACHARASUT SUJARITTANONTA, Ph.D.

HPI (house price index) measures the price development of houses sold to households. And it measures the price as a percentage change from some specific start date, which is treated as the benchmark with the index at 100. In another word, HPI reflects the housing market fluctuation in some respects. Moreover, real estate is a good type of investment for people for avoiding various risks. Hence, under the global rampant of COVID-19, forecasting the possible short-term floating of HPI would offer some foresight to the residential market both on individual and policy sides. In this paper, five kinds of models were constructed with the data before Thailand reopened (1st Jul 2020) and forecasting for the followed the third quarter HPI. Lastly, the capacity of every model will be compared by an error matrix. And the experiments demonstrated multiple regression model, based on accuracy, performs relatively better than the other four models.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study:	Business and Managerial Economics	Student's Signature .....
Academic Year:	2020	Advisor's Signature .....

ACKNOWLEDGEMENTS

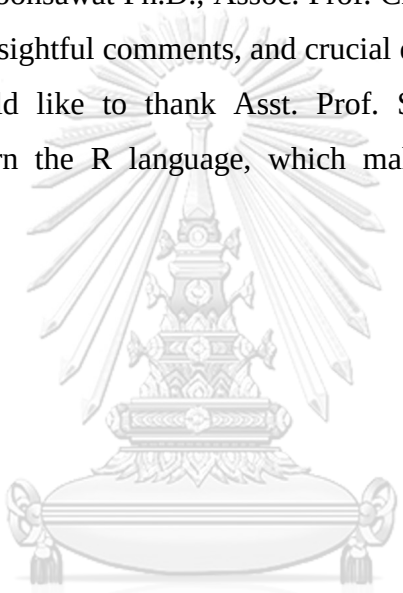
I would like to appreciate to all those who gave me the support to complete this Individual Study.

I would like to give my sincerer thanks to my advisor, Asst. Prof. Pacharasut Sujarittanonta Ph.D., for his patience, support, recommendation, feedback for improvement, and guidance from started this research until it is completed.

Besides my advisor, I would like to thank the rest of my thesis committee: Asst. Prof. Ratidanai Hoonsawat Ph.D., Assoc. Prof. Chalaiporn Amonvatana, Ph.D. for their encouragement, insightful comments, and crucial questions.

Lastly, I would like to thank Asst. Prof. San Sampattavanijia, Ph.D.. He encouraged me to learn the R language, which makes my IS could be completed smoothly.

Pengfei Chen



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

TABLE OF CONTENTS

Page

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

LIST OF TABLES

Page

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

LIST OF FIGURES

Page

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter 1

Introduction

1.1 Reason for the paper

In the first three quarters of 2020, the pandemic of COVID-19 causes a universal “stress reaction” in all of the social and economic sections. Especially, the general restriction on social distance brought a tough era for the most countries.

As the covid-19 worldwide pandemic, all the sectors of the economy are facing downward pressure. Under such a global recession, real estate is might be a good type of investment for people for avoiding risks. Real estate is the world’s single largest asset class, accounting for 60% of all global assets on some estimates. (1Schrimpf, et al. 2020) So, in America and some European countries house prices were increasing significantly during the pandemic. “According to unofficial series – which are timelier though less accurate than government data - America’s house prices are up 5% on a year ago. Germany’s are 11% higher. Britain’s hit an all-time high, in normal terms, in August.” (The economists, 2020) Different regions encountered various monetary and fiscal trajectories to treat the damages by this disease. Then if we focus on Thailand, people did relatively well, based on all measures, in the first round of pandemic. Which made consumers more confident for economic recovery. Since that, how would the Thai residential market fluctuate?

HPI measures the price development of houses sold to households. And it measures the price as a percentage change from some specific start date, which is treated as benchmark with index at 100. In another word, HPI reflects the housing market fluctuation in some respects. So, I would try to construct the different models to predict the HPI in the following parts. Supplementary, as the policy announced by the Thai government, about that foreigners cannot buy land in Thailand, only condominium units and apartments. The changes in the condominium market would be magnified by the environmental effects to a more significant extent. Hence, I merely focus on the condominium market for my forecasting model in this paper.

And in the following parts, the scenario of the experiment is simply scribed, which offers the reasons of choice of variables and modeling dataset’s division.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 The timeline of the pandemic in Thailand

The outbreak of COVID-19 was first identified in Wuhan, China, in December 2019. The World Health Organization declared the outbreak a Public Health Emergency of International Concern on 30 January, and a pandemic on 11 March (World Health Organization 2020). The COVID-19 outbreak is on its way to becoming a permanent part of human life and activities as it has been announced by WHO in recent times. Such kind of highly infectious disease does not just bring health concerns to the human beings, the words "new normal", which seems to outline the framework of comprehensively rethinking, is upcoming. In Thailand, the first confirmed COVID-19 case was reported on January 13, 2020. A state of emergency, instituted on March 26, had been extended to end-October first and prolonged until December.

The reopening began on August 1 allowing entry to certain non-Thai visitors, including medical tourists, filming crews, Thailand Elite card members, foreigners who have work permits, foreigners married to Thai nationals, and foreigners studying at educational institutions. (International Monetary Fund 2020)

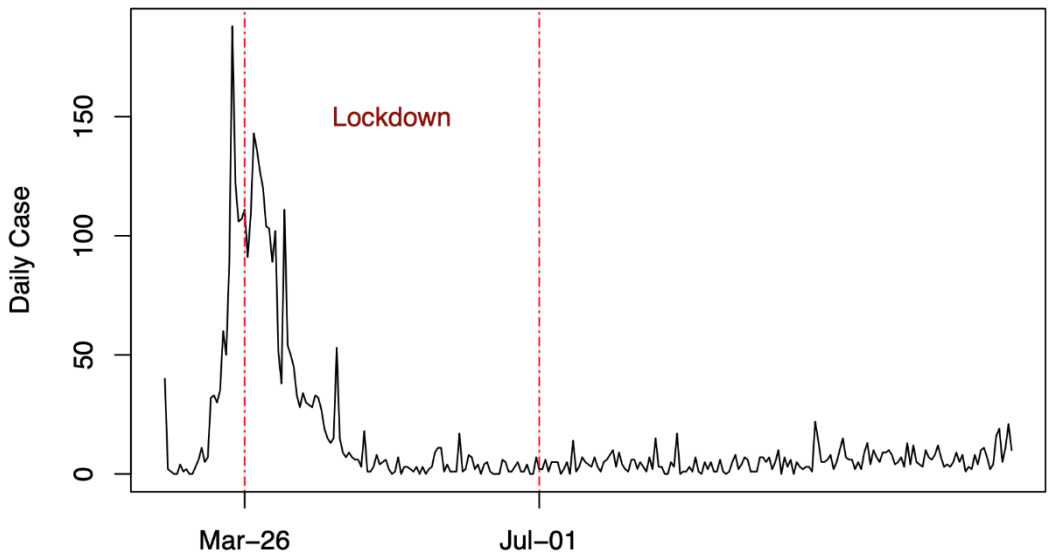


Figure 1: Daily increased infection case (data from Wikipedia, end at 1st Dec)

1.3 **Background of Thai economy**

1.3.1 GDP

In the span of less than one generation, Thailand is moving from a low-income to an upper-income country. Thailand has made noticeable progress in social and economic development during the last decades. Even if it was facing the Asia Financial Crisis and Subprime Global Financial Crisis in 1997 and 2008 respectively. However, the coronavirus rampantly attacked all the inhabitants around the world. Thai economy just encountered a plummet decrease in the second quarter of 2020, with a 12.1% YoY drop. Somehow, the main reason for this plummet could be explained by a countrywide lockdown for the prevention of Covid-19. Also, the worldwide pandemic makes most economists predicted some pessimistic expectations for the future, at least before the popularization of the vaccine.

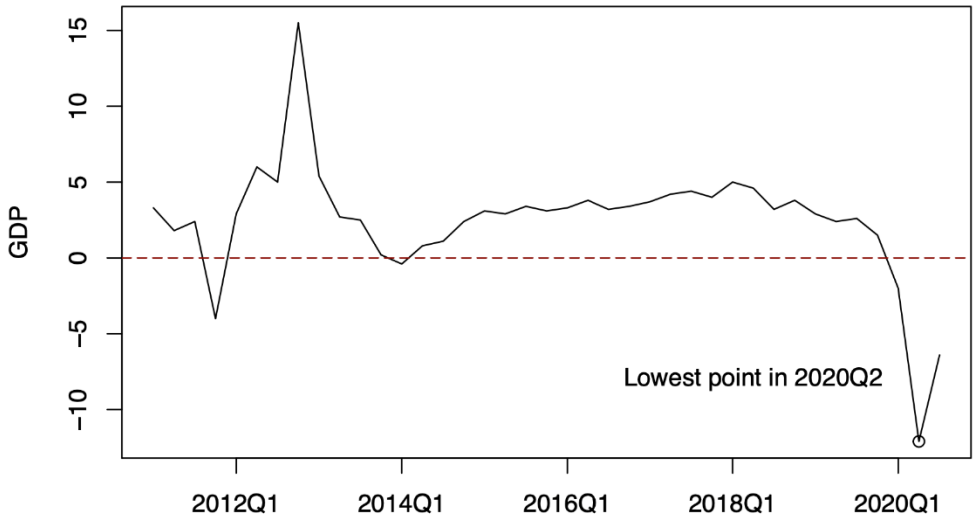


Figure 2: Thailand GPD quarterly YoY data (data from Bank of Thailand)

1.3.2 Tourism industry

Thailand, a country which is famous for its excellent natural travel resource all around the world. The tourism industry continued to expand in 2016 and remained an important driver of growth. Revenues from such sector made up 12.3% of GDP. In which, a large share is coming from European and Chinese tourists. Some of them are also a relatively large group that contributed to the condominium demand market.

1.3.3 Condominium market

There were over 175,000 house and condominium units worth more than 765 billion baht left unsold at the and 2019 from Real Estate Information Center (REIC). That is a large amount of oversupply for the real estate market. More interesting, Chinese buyers were contributed to 70% of demand in 2018 and 2019. However, the pandemic hindered foreign investors come to Thailand. For local investors, banks look more into their ability to repay the mortgage and prolonged the time for the approval process, recent years. In another hand, the loss of tourists and the hurtle of flight transportation as the pandemic, which made all sectors of the economy in Thailand face a decline, such as unemployment rate increase, manufactory industries had to be closed for the prevention of Covid-19, etc. In specific terms, Condominium prices rose by 4.4% (4.6% in real terms) during the year to Q1 2020, a sharp improvement from last year's 1% growth. However, quarter-on-quarter, condo prices dropped 1% (-0.3% in real terms) during the second quarter. (globalpropertyguide.com, 2020, Jun) As the cause, in the second quarter, Thailand encountered the overall lockdown. In such a scenario, we need to detect deeper for the patterns trace. Especially, take the pandemic measurements into account.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.4 Expected Benefits

The procedure of this experiment tries to use constructed model to find out the house price index pattern, which further to be able to predict the short-term HPI in a reasonable manner. I hope to make more insight into the data via such a paper, especially in such a data hegemony era.

And the analysis of this experiment gives some trace to understand the mechanism of society and theories.

Chapter 2

Literature Review

The COVID-19 pandemic had caused and is leading to a dramatic decrease in consumption in all segments of demand and supply sides. It is as the matter of course that following with a larger reduction in prices and a decline in workers' per capita income. The next step, a rise in the unemployment rate had swept most of the nations in the first half of this year, which further depressed consumption. And then, the drop in private income reduces aggregate demand and nominal house prices. Hence, the economic environment changed a lot in all dimensions, for example, the propensity of consumption, expectations for the future, confidence for the economy reinvigoration, assets allocation, etc. Of course, all of these reflects on the real estate sector will show a different trace in different countries.

Vary kind of methods have been booming out for nominal price forecasting and analysis purpose.

In the paper of Ewing et al (2007), they utilized the autoregressive-moving average model (ARMA) as the basic framework for their analysis. To estimates changes in the local housing price index (HPI) as a function of several control variables as well as dichotomous variables that correspond to the tornadoes and hurricanes. Moreover, they introduce a dummy variable, which indicates 1 is in the period of the tornado or hurricane disasters, 0 is otherwise.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Plakandaras, et al were motivated by looking for an early warning system for forecasting sudden house price drops with direct policy implications. They proposed a novel hybrid forecasting methodology that combines the Ensemble Empirical Mode Decomposition (EEMD) from the field of signal processing with the Support Vector Regression (SVR) methodology that originates from machine learning. And they used the MAPE and DS as the accuracy measurements, which detect both absolute quantity error and direction error.

A more-straight method, Strömberg, Hedman, and Broberg (2011) forecasted the house price index in Stockholm County in 2011-2014 by constructing a multiple regression model that included four influential macroeconomic variables. They are

real disposable income, real debt, number of the house per capita, and unemployment rate.

Besley and Mueller (2012) exploit data on the pattern of violence both on space and time span to estimate the impact of the peace process in Northern Ireland on house prices. After established a negative correlation between killings and house prices, they estimate the parameters of a Markov switching model with conflict and peace as latent states, and this model is used to estimate the size of the peace dividend, as captured in house price changes.

More similar for my topic, Del Giudice et al (2020) constructed a multiple regression model, in which the dependent variable is average housing price (AHP), to analyze the impact of the COVID-19 on the real estate market in the Italy region. And in the regressors aspect, the independent variables are YEAR (the year indication), AHI (the average household income), UNEMP (the number of unemployment), PI (the per capita income); IMI (real estate market index depending on the number of housing transactions and the number of housing stock offered for sale), JEX (the number of judicial foreclosures or real estate judicial execution).

And for another fashion type of methodology, Kauko, Hooimeijer, and Hakfoort (2002) examined neural network modeling with an application to the housing market in Finland.

Park and Bae (2015) point out “The analysis of the housing market and housing price valuation literature indicates two principal research trends: the use of the hedonic-based regression approach and artificial intelligence techniques for developing housing price prediction models.” However, the accuracy of the hedonic-based regression varies with potential limitations relating to model assumptions and estimations, and especially on the selection of independent variables. In contrast, the machine learning approach is better performance at this point, but worse in the interpretation of the parameters.

More recently, Norouzi et al (2020) developed an experiment to explore the impacts of COVID-19 on oil and electricity demand in China, which through a comparison between the regression model and neural network model. The process begins with environmental scanning to determine the possible effective driving

variables, then the main variables are being selected, and the data is gathered for them, and then the data is processed and analyzed through two different methods.

As all above, many classical methods are introduced by the works of literature for the analysis of the structural correlations, which make it difficult to choose. However, as the limitation of my capacity. In this paper, I will choose the housing price index as my dependent variable and via constructing univariate and multivariate ARIMA, multiple regression, elastic regression, neural network such five kinds three directions of models to looking for the information beneath the data.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Chapter 3

Methodology

In the previous literature review, I have mentioned that some abundant methods and models have been structured for forecasting AHP (average house price) and HPI. Further, most of the studies are concentrated on the structural correlations, which specifically restricted space, time, and motivation. For this paper, I put all the sights on investigating the Thailand condominium market under the impact of the COVID-19 pandemic. Because of the limitation of the writer, some of the classic methodologies are not available for this paper. Therefore, this experiment proposes multi-model forecasting and comparison of results procedure.

3.1 Data description

Normally, for house price forecasting, two categories of data are widely used: 1. Properties of house, like, the number of rooms, location, number of hospitals around, number of schools around, etc. 2. Economic dataset, like interest rate, unemployment rate, crime rate, etc. Further, if the focus on the HPI forecasting, the latter is chosen more. And there is another popular direction is treating HPI as time series and concentrating on the information on its own.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

In this paper, HPI is treated as the dependent variable. It measures the price development of houses sold to households. And the measures the price as a percentage change from some specific start date, which is treated as the benchmark with the index at 100. In Thailand, the House Price Index refers to a hedonic regression method with the three-month moving average. The index is calculated from 17 commercial banks' mortgage loans in Bangkok and vicinities (Bangkok, Samut Prakan, Nonthaburi, Pathum Thani, Nakhon Pathom, and Samut Sakhon).

For independent variables, I firstly choose consumers price index (general), unemployment rate, policy rate, the exchange rate (to USD, EUR, RMB respectively) on the predecessors' foundation. And considering about situation of the location of my study, I add the consumers' confidence index andforeign tourist numbers into

account. Lastly, the infection condition and lockdown are both treated as dummy control variables.

Back to the data volume, normally, more is better. It offers more information underlying the phenomenon. However, as the economic indices changed a lot because of the pandemic, which makes big volume data set decreases the accuracy of the analysis, and too small a volume data set to increase the uncertainty because of lack of enough information. Hence that, after the check of raw data trends, I decided to construct models with monthly data of Jan 2011 to Jun 2020 and try to forecast with models using the data of Jul to Sep 2020. After the 2008 Subprime Global Crisis, the trend of data is relatively stable. Especially notice: there are three months of unemployment rate missed because of lockdown, which are April May and June respectively. They will be replaced by the number of July, which I assumed the lockdown risen the unemployment rate to July level directly.

Table 1*:Variables description (Monthly data from Jan 2011- Jun 2020)*

Variable	***Description***	***Text Sign***	***Measerment***
Dependent Variable	House Price Index	HPI	Percentage
Independent Variables	Time	t	Number(series)
	Consumers Price Index	CPI	Number
	Consumers Confident Index	CCI	Number
	Unemployment Rate	UN	Number
	Policy Rate	INT	Percentage
	Exchange Rate	USD	Number
		EUR	Number
		RMB	Number
	Tourist Number	NUM	Number (in thousand)
Control	Infection Condition	INF	Dummy (0|no, 1|yes)
	Lockdown	LD	Dummy (0|no, 1|yes)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Table 2*: Basic information of variables (length: 117)*

	HPI	CPI	CCI	UN	INT	US	EUR	RMB	NUM
Min.	112.4	91.93	17.00	0.3896	0.50	29.07	33.42	4.278	0
1st Qu.	131.3	99.07	33.60	0.6314	1.50	30.95	37.45	4.733	1896
Median	159.7	100.36	37.50	0.7459	1.75	32.01	38.92	4.978	2420
Mean	154.9	99.77	36.48	0.7904	1.923	32.33	39.09	4.975	2377
3rd Qu.	177.6	101.44	40.40	0.8864	2.50	33.25	40.71	5.213	2990
Max	197.3	103.31	52.00	1.5315	3.50	36.16	44.87	5.648	3947

HPI, UN, INT, USD, EUR, RMB, NUM sourced from Bank of Thailand; CCI from National Statistical Office; and CPI from Bureau of Trade and Economic Indices

Note: ***The more detail information about raw data is plotted in Appendix A***

3.2 Model selection

Forecasting of various economic indices is not just deepened the broad boundary of the science of economic itself, but more important on offering some foresight about this complex world to better government policies, planning, and decision making.

In classical econometrics, a regression model is always a straightforward choice for most starters. The parameters illustrate some extent of a clear relationship between regressors and regressand. But under the hypothesis of the BLUE, properties of raw data, like collinearity, skedasticity, normality, causality, etc, all of these might bring biases and lead to systematic errors. So, it is especially necessarily needing a comprehensive fundamental understanding about properties of variables.

In the other direction, in recent years, as the rapid development of utilization of machine learning algorithm on data science. Make the purposes, that just looking for patterns under the data and interpret the outputs, are much easier for practice usage. Maybe, some critics suspect this process of data mining, which ignoring any supported theories. But their efficiency has been proving by more empirical evidence.

I have no preference for any of these two directions, as they both show their tremendous power in all aspects of the economic fields. For such reason, I choose five kinds of models to complete my goal in this paper from both two directions.

a) Uni-variate ARIMA

Definitely, the HPI is a time series variable (see the plot in Appendix A). The ARIMA model has a proven good fit performance in such kind of data. The idea of this method is that past values in the time series have information about the current state.

b) Multi-variate ARIMA

Compare with the previous one, the multivative method treats the other variables like a matrix and also offers the information for the predictand.

c) Multiple regression

Classic method, a regressive model is constructed to detect the relationships between HPI and the other variables. As the assumption of efficient market theory, I lagged all independent variables one period, assume that the HPI_t is reflecting the information of independent variables in time *t-1*.

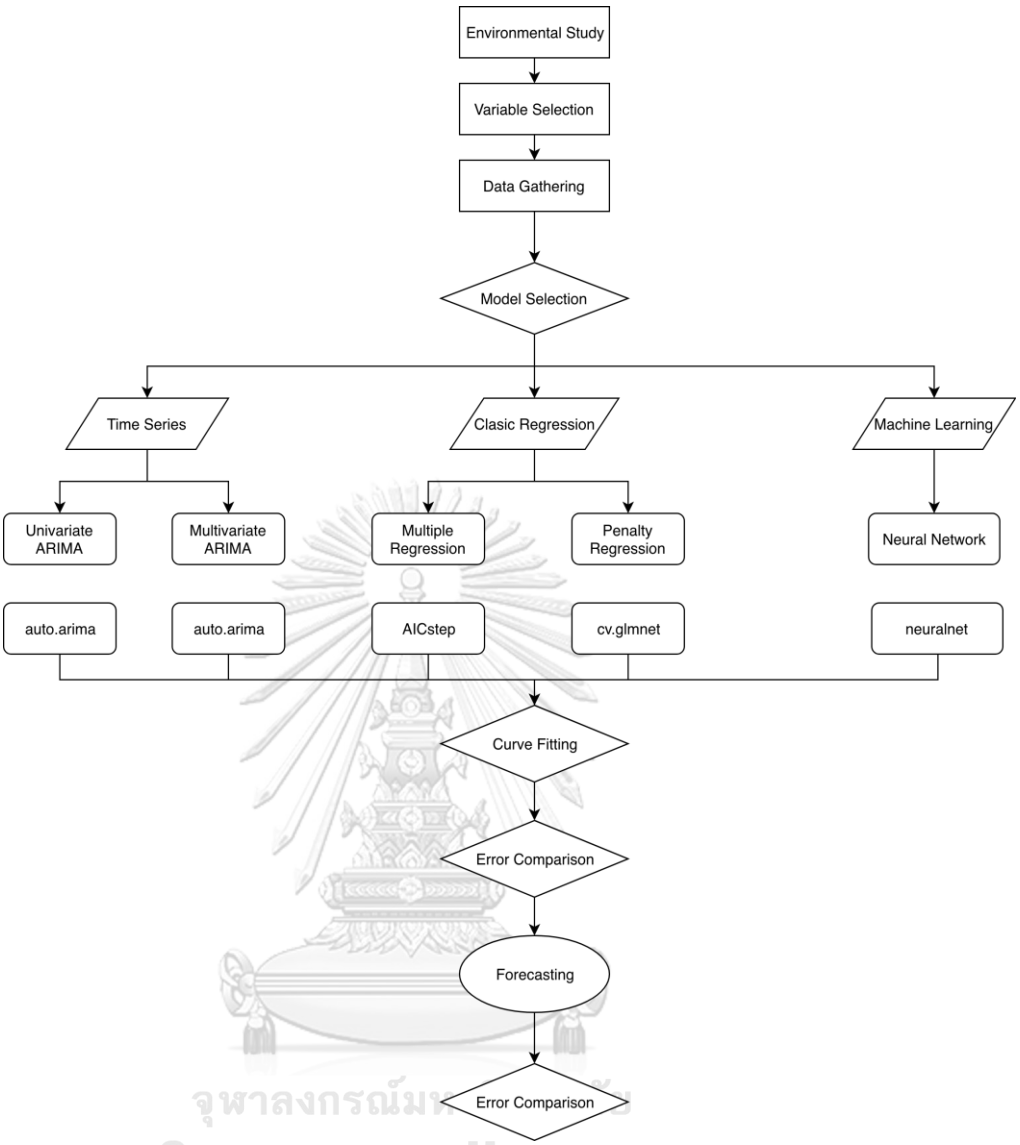
d) Elastic net regression

Considering the out-of-sample forecasting, the elastic net regression is structured for the purpose of the penalty to the sensitivity of *Xs* to *Y*.

e) Neural network

Widely used machine learning algorithm. Although dependent and independent variables are not changed, the network and nodes are not representing the predefined relationship as regressive methods. But the relation is defined by the algorithm.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 3: Flow plot

3.3 Accuracy judgement

Take it a step further, the most important thing for forecasting and prediction is accuracy. The main metrics for accuracy judgment in this paper are DS (directional symmetry), and MAPE (mean absolute percentage error).

So, for each model. The following procedure will be progressed:

Step1: Adjust data for the specific model

Step2: Divide data into modeling set and forecasting set (modeling set: Jan 2011-Jun 2020, forecasting set: Jul-Sep 2020)

Step3: Fit the modeling set with the mechanism of different model

Step4: Construct a matrix of different aspect error metrics with cross-validation

Step5: Predict the HPI with forecasting set data

Step6: Compare the forecasting data and real data

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter 4

Result

In the previous parts, the main logic chain has been simply narrated. Five models will be specifically processed in this part, in the time series, the regression, and the machine learning, respectively. Emphasize again, my most concern is forecasting, which means the model's capacity of predicting out-of-sample data should be significant. Hence, the method of cross-validation will be combined with each model to enhance out-of-sample forecasting.

Cross-validation (CV) is a statistical method used to estimate the skill of machine learning models. The logic is straightforward, randomly splitting the train set into k-fold and use *k-1* folds as a sub-data set to fit the model and forecasting the remaining fold, then compare the prediction with the real data, repeatedly until traversing all k folds. At last, return the evaluation scores and the model with the optimal mean skill. It could reduce the biases caused by sample limitation and improve the out-of-sample prediction capacity. Considering the data length and the time series property, the 10-fold CV will be processed.

4.1 Univariate ARIMA model

ARIMA is short for Autoregression Integrated Moving Average. Under the philosophy "let the data speak for themselves, which is widely used in the field of time series forecasting. It bases on the idea that the past values of time series could predict the future value. The index of housing price is obviously a time series, and the main pattern of the data is plotted like below:

Figure 4*: HPI of known data (Jan-2011~Sep-2020)*

Furthermore, when fitting the ARIMA model, two main kinds of forms, which are univariate ARIMA and multivariate ARIMA, are widely used in the practice. In 4.1 and 4.2 they will be constructed in each.

In a univariate manner first, an ARIMA (p, d, q) model follows the Box-Jenkins (BJ) Methodology. The method consists of four steps:

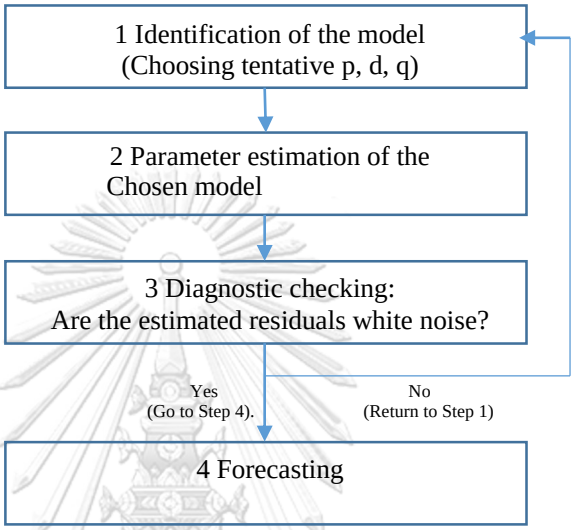


Figure 5*: BJ methodology flow plot*

Figure 4 is the trend of known HPI, which from January 2011 to Sep 2020. Obviously, for a time series, the stationarity analysis is the first step of identification. A stationary time series is the properties of $\{y_t\}$, in the period n, the distribution of it, $y_t, \ldots\ldots, y_{t+n}$, does not depend on t. Thus, time series with the trend, or with seasonality, are not stationary — both these two will affect the value of the time series at different points. In the contrast, a white noise series is stationary,$x_t = e_t, e_t \sim N(0, \sigma^2)$, which does not matter when you observe it, it almost looks the same at any point. If decomposing this series into the different part, could get,

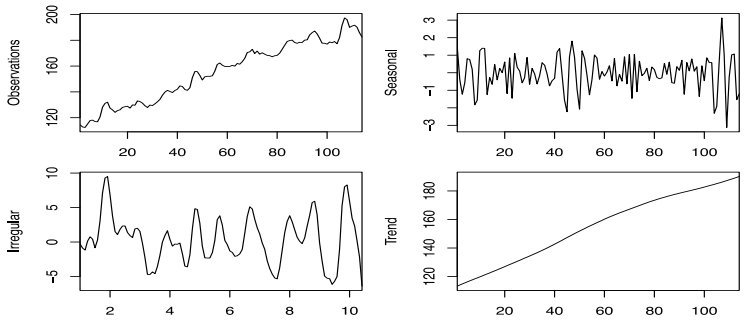


Figure 6*: Decomposing of HPI series*

From figure 6, the main trend of HPI is clearly shown. Since the main direction of the model is there. Then, for the purpose of test the stationarity of a time series, ACF (Auto-Correlation Function) plot and PACF (Partial Auto-Correlation Function) plot are widely used as a kind of intuitional judge method. If the plot is not so clear, a Dickey-Fuller unit root test will be processed. However, in this specific data set, it is clear that the property of non-stationary (Appendix B, part 1, figure 15). And the following steps is according complex test and calculation to choose p, q, and n, even the seasonality parameters. Thank for the development of the computer calculation, it is easy to conduct this process with a simple code. Especially, I use R to help me to get the model.

The model was fitted with *auto.arima( )* function is (the residual check plot in Appendix B, Part 1):

Table 3*: Result of optimal univariate model with auto.arima() function*

Series: modeling set					
ARIMA (4,1,0) with drift					
Coefficients:					
	AR1	AR2	AR3	AR4	drift
	0.4576	-0.0139	-0.4911	0.1972	0.6272
s.e.	0.0949	0.0895	0.0902	0.0989	0.2211
Sigma^2 estimated as 4.122		Log likelihood = -238.3			
AIC = 488.6	AICc = 489.39	BIC = 504.97			

After check the residuals, this model's residuals show a white noise property.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2 Multivariate ARIMA

In 4.1, the univariate ARIMA is fitted to the modeling data set, which in the logic of letting the data talking itself, but may not include the information from other relevant places. Like in this case, or the other variables have some impact on HPI to a different extent in the theory. If they could explain some of the historical variations then more accurate forecasting would be got. Actually, the idea is regression allowing an autocorrelation in the error term. Again, the *auto.arima()* function in R will be used, the only difference with the univariate model is all the other variables are combined into a matrix *xreg* and brought into the function.

Table 4 *: Result of optimal multivariate model with auto.arima() function*

Series: modeling set

Regression with *ARIMA (5,0,0)* errors

Coefficients:

	AR1	AR2	AR3	AR4	AR5	Intercept	t	CPI	CCI
	1.2129	-0.4165	-0.4852	0.6225	-0.3083	180.8017	0.7666	-0.9230	-0.0297
s.e.	0.0992	0.1416	0.1288	0.1456	0.1051	40.1160	0.0382	0.4019	0.0704
	EUR	RMB	INF	NUM					
	-6.3920	28.0478	-0.5031	0.3652					
s.e.	8.9277	9.1559	1.6188	0.5967					

Sigma^2 estimated as 3.569 Log likelihood=-228.54

AIC = 485.08 AICc = 489.32 BIC = 523.39

4.3 Regression model

As the classic method to analyze the under the relationship between regressant with regressors, it shows some extent of proven performance in a straightforward manner and excellent interpretation properties. However, all of this convenient method is constructed on the fundamental assumptions. Like I mentioned in the previous parts, the variables I choose for predicting the HPI are all in some respect of understanding the objective connections, maybe somehow under hypothesis. Then, a direct process of multiple regression model with all variables is constructed. The result is seemed wonderful, a good t-test for the significant level of coefficients, high adjusted R square, fine goodness of fit ANOVA. Unfortunately, this model is nonsense for practice as the reason for time series, which violates the assumption of the regression model. Just as the figure 14 in Appendix A, highly colinear between some variable, and the DW test tell an autocorrelation for the u terms. However, if just for forecasting purposes multicollinearity may not be too problematic. Since, in this situation, we are only interested in predicted values but not parameters.

For the further choice of the independent variables, I use the *lm()* and *step()* function in R to run the data under the standard of lowest AIC.

The final result as:

$$HPI_t = \beta_0 + \beta_1 t + \beta_2 CPI_{t-1} + \beta_3 CCI_{t-1} + \beta_4 lnEUR_{t-1} + \beta_5 lnRMB_{t-1} + \beta_6 INF_{t-1} + \beta_7 NUM_{t-1} + \mu$$

Table 5*: Result of multiple linear regression model by step() function*

	Estimate	***Std. Error***	***t value***	***Pr (>|t|)***		***VIF***
(Intercept)	151.20853	32.75700	4.616	1.11e-05	***	
t	0.73963	0.03055	24.207	<2e-16	***	10.5035
CPI	-0.53721	0.36217	-1.483	0.140989		9.1180
CCI	-0.29725	0.08405	-3.537	0.000604	***	4.3751
EUR	-12.33625	8.46646	-1.457	0.148081		3.8801
RMB	38.70758	5.98981	6.462	3.30e-09	***	1.5926
INF	5.14270	2.35304	2.186	0.031070	*	2.4760
NUM	3.11473	0.74635	4.173	6.21e-05	***	3.1185

Residual standard error: 3.269 on 105 degrees of freedom
Multiple R-squared: 0.9818 Adjusted R-squared: 0.9806
F-statistic: 810.8 on 7 and 105 df p-value: < 2.2e-16

*(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)*
Note: The residual checking plot in the Appendix B part 2

The final model residuals show somehow not bad properties. And the VIF test shows all the variables are under 10.

Back to the coefficient signs, all seem reasonable. When consumers were becoming not so confident about the consumption market, the wealth would flow into the more stable market such as real estate, which pushes the demand curve to the right and the house price increase. Interestingly, in this model, the pandemic encounters positive relation with the housing price index, which showed the same relation just like some western countries. Even the t-test is just significant in 5% level, is relative enough to explain to some extent. From it, the HPI moving the same direction because of the infection, nevertheless the data in the modeling set just contain half a year that have the infection cases, so, further cross validation and analysis are needed to test the model and make a conclusion.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.4 Elastic Net regression

To avoid overfitting problems, kinds of penalty regression methods are invented for the sight of people. $E(e) = Bias^2 + Variance + \sigma^2$, as the complexity of the model increase, the square of bias decrease, by contrast, the variance increase. So, the under logic is to add some penalties for the original models to make the dependent variable less sensitive to the independent variables, in another word, to minimize the variance in the cost of bringing in some bias.

According to this, to bring in some bias means to decrease the complexity of the model. More specifically, to decrease the number of predictors. There are 10 independent variables in my model, and some of them are correlated. As the analysis in part 4.3, the OLS parameter estimates have a large variance, thus making the model unreliable. For the purpose of optimizing the total error, to balance the bias-variance trade-off, which is suitable for utilizing the penalty regression.

Normally, ridge regression (L2) adds the penalty as the square term, lasso regression (L1) adds the penalty as the absolute value term.

$$L1 = \sum_{i=1}^{n} residual_i{}^2 + \lambda\left(\left|\hat{\beta}_1\right| + \cdots + \left|\hat{\beta}_m\right|\right)$$

$$L2 = \sum_{i=1}^{n} residual_i{}^2 + \lambda\left(\hat{\beta}_1{}^2 + \cdots + \hat{\beta}_m{}^2\right)$$

Just focus on the penalty parts, when lambda equals 0, $\hat{\beta}_{ridge}$ and $\hat{\beta}_{lasso}$ equal to $\hat{\beta}_{OLS}$, when lambda increase to ∞, $\hat{\beta}_{ridge}$ and $\hat{\beta}_{lasso}$ become to 0.

Because of different sensitivity of these two, to combine them with alpha into elastic net regression would more reliable on my case.

$$Penalty\ term = \lambda_1\left(\left|\hat{\beta}_1\right| + \cdots + \left|\hat{\beta}_m\right|\right) + \lambda_2\left(\hat{\beta}_1{}^2 + \cdots + \hat{\beta}_m{}^2\right)$$
$$= \lambda[\alpha Lasso + (1-\alpha)Ridge]$$

I run the data via *cv.glmnet( )* function, and the detailed codes for programming in the Appendix C table 10. The results of alpha and lambda in different levels are listed below:

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Table 6 *: Elastic net model choosing process in different alpha value*

FIT_name	***R_Square***	***MSE***	***s***
Alpha0	0.96853	0.03273	0.15521
Alpha0.1	0.98043	0.01932	0.01743
Alpha0.2	0.98080	0.01902	0.01835
Alpha0.3	0.98049	0.01944	0.02138
Alpha0.4	0.98059	0.01933	0.01931
Alpha0.5	0.98075	0.01914	0.01696
Alpha0.6	0.97986	0.02023	0.02250
Alpha0.7	0.98012	0.01982	0.01757
Alpha0.8	0.98053	0.01933	0.01401

Alpha0.9	0.97963	0.02037	0.01807
Alpha1	0.98065	0.01915	0.01121

s is the value of the penalty parameter lambda at which predictions are required. Default is the value s="lambda.1se" stored on the CV object.

When alpha equals 0.2 and lambda equals 0.01835, getting the best MSE and R square.

4.5 Neural networks

Neural networks, one of the most popular artificial algorithms, cover a broad of concepts and technologies. However, a lot of people call it a black box, as it is really hard to understand it and do some meaningful interpretation with its parameters just like the regression method. The ANNs imitate the structure of a nervous system. It composes several nodes and strings of connections which are call neurons and synapses respectively in the nervous system. In every node, there is a non-linear activation function (curved line). In each connection, there is a weight or bias term. When the algorithm gets the inputs, each input is multiplied by its weight then sum up plus the bias term. Then this result is sent into one node, processed by the activation function. After repeatedly in each node in the first layer, the results will be transmitted to the next layer if there is. Finally, the last layer's data are weighted and plus bias term and output the fitted data. Under the chain rule, this form of artificial neural networks could even approximate any continuous function in a precise manner. Thank you for this reason, I chose this algorithm as one of my models.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

The main function in R is *neuralnet( ).*

Table 7*: The error term of neural networks chosen in different number of nodes*

	First Layer							
Second Layer	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***
2	0.04067	0.06641	0.05708	0.05230	0.06221	0.07052	0.12052	0.07319
3	0.07267	0.09697	0.03276	0.05219	0.04263	0.05779	0.06214	0.09876
4	0.07406	0.10307	0.05721	0.07730	0.08035	0.08884	0.06034	0.03362
5	0.06782	0.07074	0.06176	0.05287	0.04122	0.05067	0.08986	0.05379

Note: In this paper I choose two layers to construct the neural models, and the detailed codes of for loop is showed in Appendix B part 4.

After the procedure of loop. The optimal ANN model is with two layers as c (10,3).

4.6 Accuracy and Forecasting

Until now the all five models have been structured. The following mission is to compare their capacity of forecasting.

Firstly, the five models fit the data like:

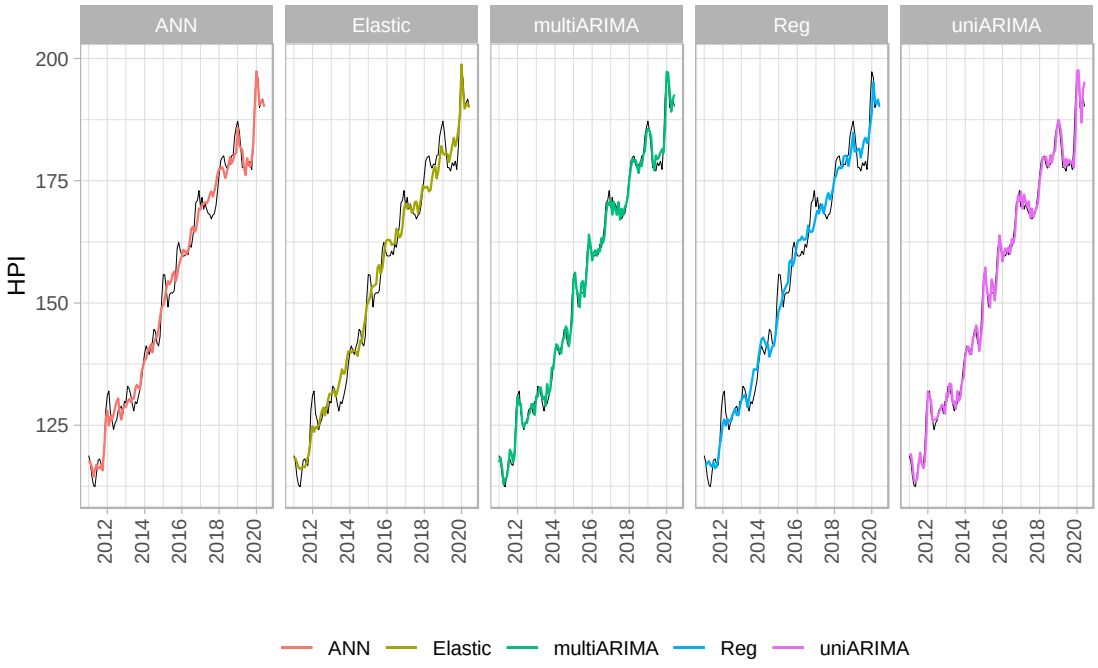


Figure 7: The quality of fit for five model.

Clearly, all five model shows some good on fitting to the original data. The more quantitative evidence should be processed.

In the classic regression, we judge the quality of a model with R square, t test of coefficients' significant and F test for goodness of fit, etc. The similar logic, normally, in the forecasting model, we compare different kinds of defined distance from central value. Such as, ME (mean error), RMSE (root mean squared error), MAE (mean absolute error), MPE (mean percentage error), MAPE (mean absolute percentage error), DS (directional symmetry). I will process the last two as my judgement standard in this paper.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

$$DS = \frac{100}{n}\sum_{i=2}^{n} d_i \ , where\ d_i = \begin{cases} 1 & if\ (y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) > 0 \\ 0 & otherwise \end{cases}$$

From the comparison of the fitted values with real data. Mostly the same as the plot shows. The multi ARIMA and neural network perform a good quality on fitting in MAPE and DS respectively. As the previous part mentioned different kinds of ideas could fit a squiggle line to match the data in the various mechanism. However, I need to emphasize again, the core purpose of my focusing is forecasting capacity. If the model fits the data very well, is it somehow kind of overfit? Thus, possess the weak power of prediction. Hence, I use the random sample cross validation process to test the performances of each model. The Arima model is directly tested by the *tsCV( )* function with a 10 continuous periods forecasting. For the other three, the modeling set was randomly sampled 90% of them as CV training data, and the remaining 10% as CV testing data. Using the CV training data fit the chosen model and then forecasting the CV testing part. Calculate the MAPE and DS respectively. Loop this 10 times, and get the average value of each. The result as below:

Table 8*: The comparison of accuracy*

Model	***Modeling Accuracy***		***Out-of-sample Accuracy***	
	MAPE(%)	***DS(%)***	***MAPE(%)***	***DS(%)***
Uni ARIMA	1.0143	59.29	2.6019	64.21
Multi ARIMA	0.9382	56.64	1.7674	61.00
Regression	1.6243	58.04	1.6626	86.24
Elastic Net	1.7344	50.44	1.7135	75.23
Neural Network	1.0089	62.83	1.8442	76.15

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

In the cross-validation process, the regression model performance best. Although, five models express the different capacities in out-of-sample forecasting, the outcomes are all in receivable range.

So, I use all of them to forecasting the third quarter's HPI.

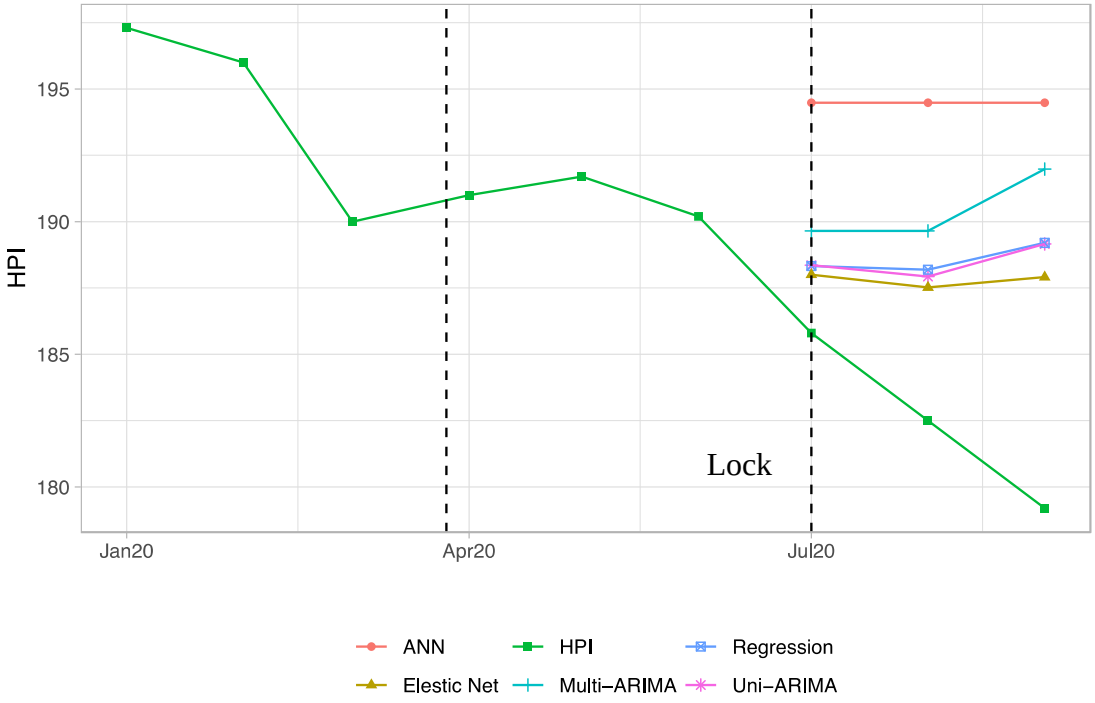


Figure 8: Forecasting of the third quarter's HPI

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Chapter 5

Discussion

5.1 Mechanism of this study

In the previous chapters, five kinds of models have been constructed, tested, and used to predict the HPIs from Jul to Sep respectively.

Like Pfefferbaum, B., & North, C. S. (2020) mentioned in their paper, “Public health emergencies affect the health, safety, and well-being of both individuals and communities. These effects may translate into a range of emotional reactions (such as distress or psychiatric conditions), unhealthy behaviors (such as excessive substance use), and non-compliance with public health directives (such as home confinement and vaccination) in people who contract the disease and in the general population.” Of course, it also affected the personal economic behavior in various aspects too. Except for all the economic environment changes which may be relatively easy to be observed and quantitated, the more complex thing is the mental situation of people which is hard to take a precise account of and then to reflect into the market analysis. Lockdown infringed on human freedoms, large and growing financial losses. In addition, all the messages, from various social media platforms, depressed mental health profoundly. Moreover, the emergency is a so-called emergency because of its suddenness. All of these might contribute to the unexplained changes in reality. Under this concern, the longer people would be the more getting used to that. So, I took inflection and lockdown as dummy variables into my modeling process, it could squiggle the fit curve to some extent, and give more accurate forecasting.

Back to the methodology, I gathered all the other quantitative data both with theory and my intuition as much as possible. Then analyze them, what would offer the information in a more detailed manner for the goal object. Following, I coded a 90% random sampling and 10% validation for checking the out-of-sample forecasting performance. I assume the difference is the unexplained part beyond of model.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2 Discussion of data

Firstly, about the quality of all variables, I processed importance analysis HPI ~ all the other variables.

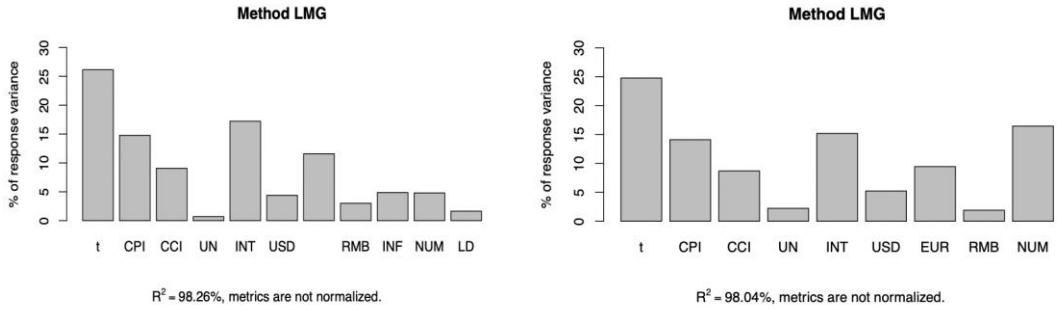


Figure 9: Relative importance for HPI (left: all data, right: before pandemic)

As the reason of time series, t is changed into a sequence as 1 to *n*. Put the t aside, the other variables show some changes in response to the variance when separately process with whole data set and before pandemic set. Especially, infection situation, number of tourists and unemployment rate.

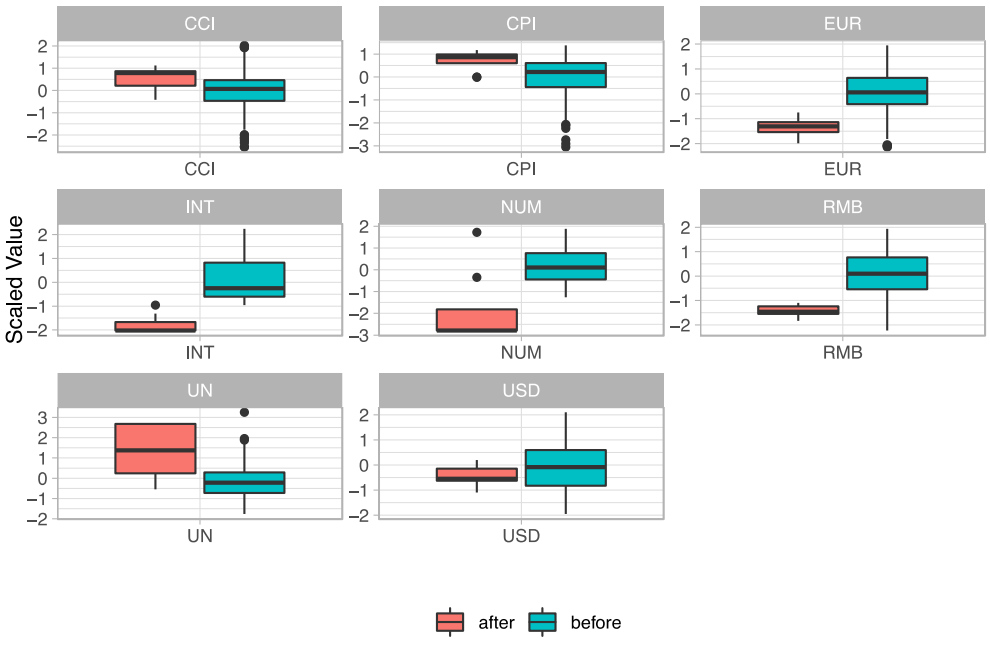


Figure 10: Box plot of independent variables in before and after pandemic

Further insight into variables' variations, obviously, the previous NUM and UN appear huge gaps (around three sigmas) before and after the pandemic. The

exchange rate to EURO and RMB also declined significantly. Of course, under such a tough situation, for the purpose of stimulating consumption, the interest rate decreased too.

5.3 Discussion of model

5.3.1 ARIMA

As the logic of the autoregression integrated moving average, this model digs deep into the data itself and explains the future with history. And as I used both the univariate ARIMA and multivariate ARIMA, all the information reflected in the model is mined from HPI history data, which means if there is no or less information about uncertainties in the historical modeling set the model would just forecasting under the previous rules. Hence, they perform worse in the time series cross validation.

5.3.2 Multiple regression

This model uses the OLS mechanism to estimate the coefficients of all variables and to check the significant levels. Thanks to the R language. By step method to choose the least AIC model with useful regressors. In my regression model, CPI CCI lnEUR lnRMB INF and NUM were chosen as independent variables. It takes account of INF and NUM which both fluctuate a lot before and after the pandemic. Thus, it performs relatively well in a random cross validation procedure.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.3.3 Elastic net regression

After adding some penalty terms on the regression model, making the elastic model less sensitive to the fluctuation of explanatory variables. Nevertheless, it sacrifices some precision also because of this merit. So, in both modeling and CV accuracy results, it performs mediocre.

5.3.4 Neural network

Because of its capacity to squiggle fit line matching the real data very well, or say, over-fit. It shows a relatively good accuracy in the modeling set, which could approximately understand as it covering all information in the train data. When it facing an out of data forecasting, some of the gained information become surplus.

5.4 Analysis of predicted results

From the figure 8 is clear that the all of predictions greater than the real data. And the gaps are very significant in visual.

Table 9*: The real data and forecasting data*

	HPI	*Uni-ARIMA*	*Multi-ARIMA*	*Regression*	*Elestic Net*	*ANN*
Jul	185.80	188.36	189.65	188.33	188.00	194.48
Aug	182.50	187.93	189.65	188.19	187.52	194.48
Sep	179.20	189.16	191.98	189.20	187.91	194.48

Just like the discussion of the model. The regression model performs best in out-of-sample prediction. In the contrast, although, the neural network model fits the modeling data relatively well, it shows the worst capacity on forecasting in the third quarter. This kind of over-fit might stem from the large variance between the modeling and forecasting set, which distort the most sensitive model to the largest extent. Further data are needed to process a deeper comparison of the forecasting capacities of models.

5.5 The limitations of this study

Currently, this disease is still rampantly spreading around the human world. Especially in the writing time, the second wave is attacking most of the regions that entered into winter. Until the vaccine could be widely inoculated, might be, some countries started to get used to this new normal, regardless of daily life economic behaviors and mental status, etc. As the figure 11 below, if just focus on the trend of real data, it decreased at beginning of the pandemic and increased from starting of lockdown, then sharply drop from May.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

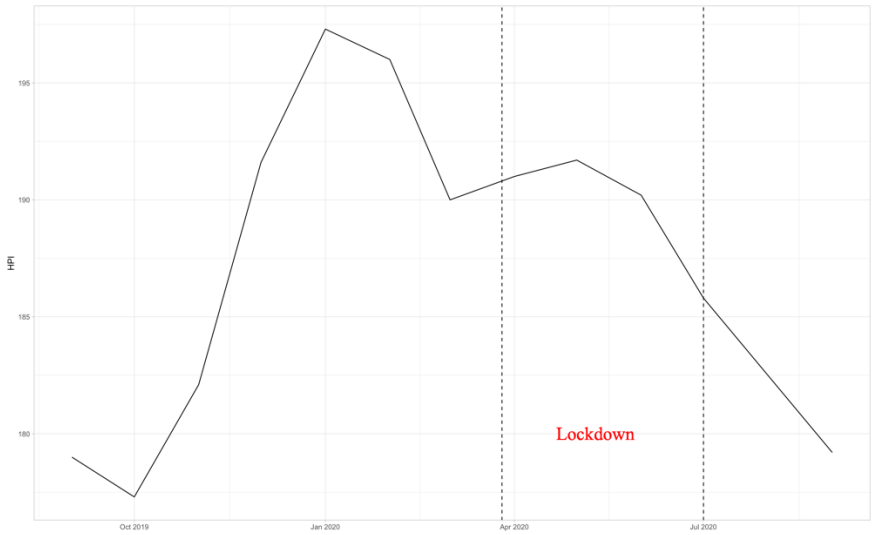


Figure 11: One-year trend (Sep 2019- Sep2020)

So, firstly, because of the unfinished status, there are full of uncertainties in all economic sectors. Which means the independent variables might change randomly to a huge extent that could not be explained by the historical data. Thus, the model, which is structured from the previous data, will become unreliable to forecast the future.

Secondly, in this study, infected cases and lockdown are simplified as two dummy variables. As the discussion before about the human mental health problem, these two variables are not just impact on people by a yes-or-no manner, but likely in a more subtle manner caused by specific level. Especially, the number of infected cases changed huge in the beginning of the pandemic in Thailand. In the result, the model will miss any detailed information in these parts.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Chapter 6

Recommendations

Training the model by different methods with historical data, which makes the models could explain the information as more as possible. Of course, as the mentioned in former chapters, various methods underlying different kinds of theories, thus best suit different types of data, possess their own advantages, and diversely perform in prediction. In the end, according to the five sources of quantitative results, the evidence point that if just following the economic data the HPI would higher than the real one, the meaning is pandemic caused some extent negative effect on the HPI in Thailand which could not be explained by the models trained with previous empirical evidence. Of course, the precondition of such a conclusion is based on the variables are reasonably chosen, the data are reliably collected and properly tidied, the model is suitably defined and tested, and so on.

Of course, as the limitation of my understanding in this field and since the situation of crisis is still ongoing, there is not complete data to compare and analyze the comprehensive effects on the condominium HPI in Thailand. However, as far, there are some recommendations from this study as follows.

- According to the importance analysis, except choosing the consuming price index and interest rate as regressors under the empirical experience, the consumer confident index also could explain the HPI variance relatively efficiently.
- Since the tourism industry is contributed to a considerable part of Thai GDP and foreign buyers account for a non-negligible term of condominium consuming. The effects of tourist numbers and the exchange rate of certain countries are especially significant. And this also brings some rethinking about weights of the tourism industry in Thailand, particularly in the situation like worldwide pandemic which tremendously damaged it.
- In the environmental study, the unemployment rate is frequently chosen as a explain variable and processed in models. Nevertheless, it

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

useless in the situation of the Thai region. As the unemployment rate fluctuates rather small, min at 0.3896 and max at 1.5315, in the span of ten years, and it is really in a stunning low level with mean at 0.7894.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Appendix A:
Visualization of the detail information about raw data

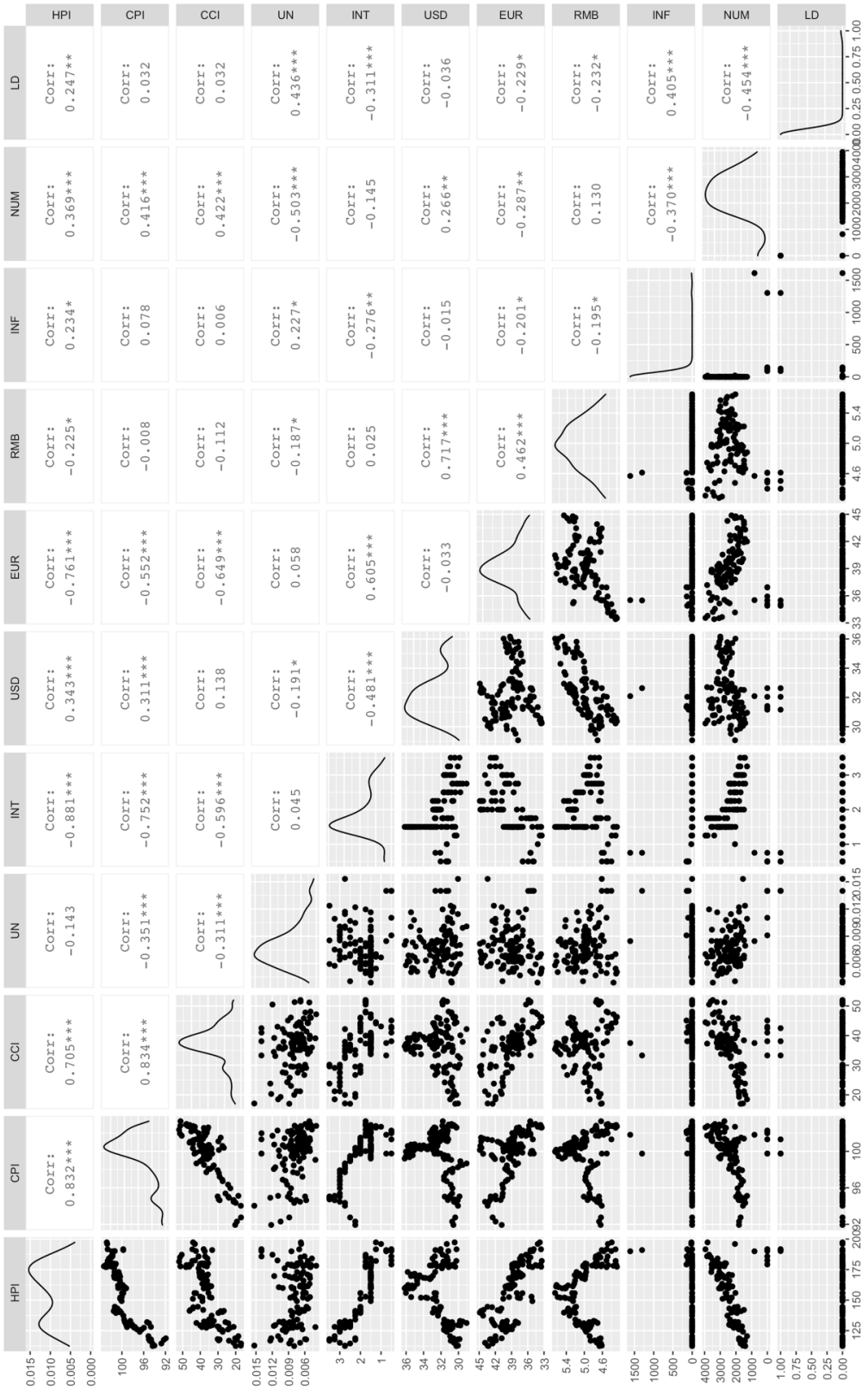


Figure 12: Scatter plot about raw data

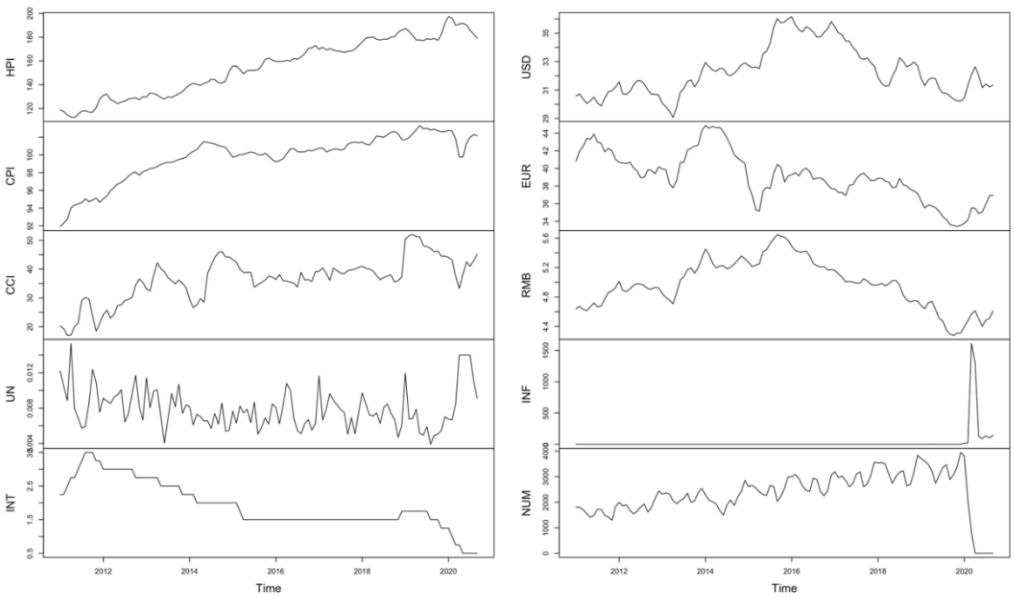


Figure 13: Variables vs Time

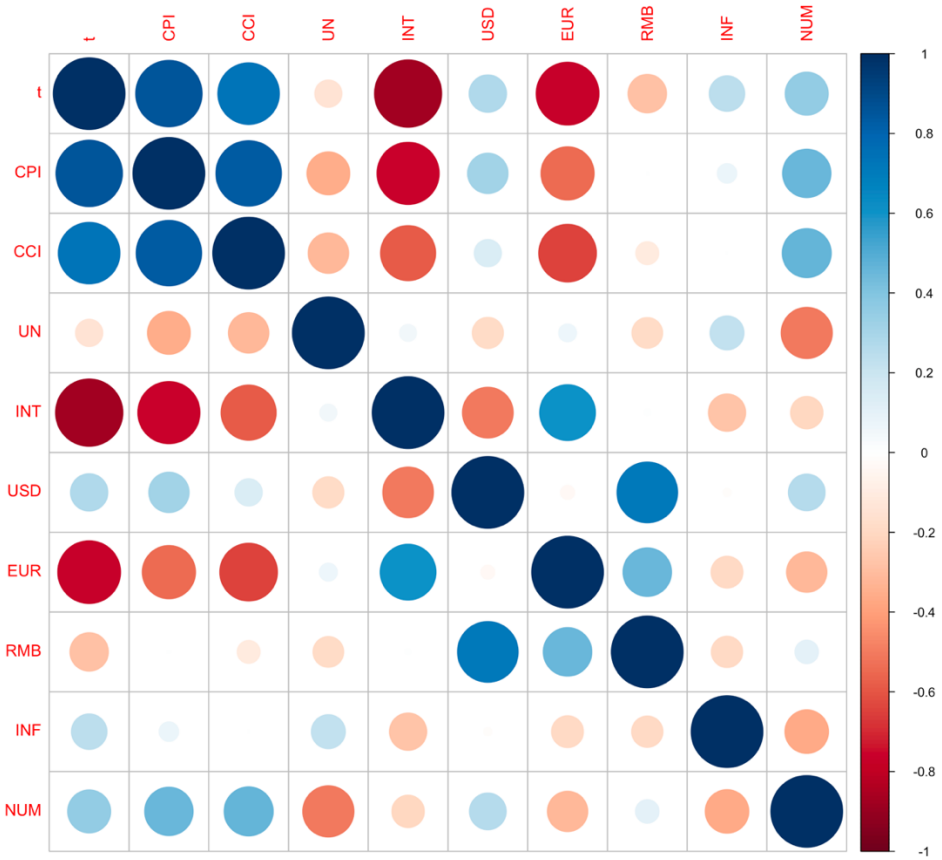


Figure 14: Correlation between independent variables (raw data)

Appendix B

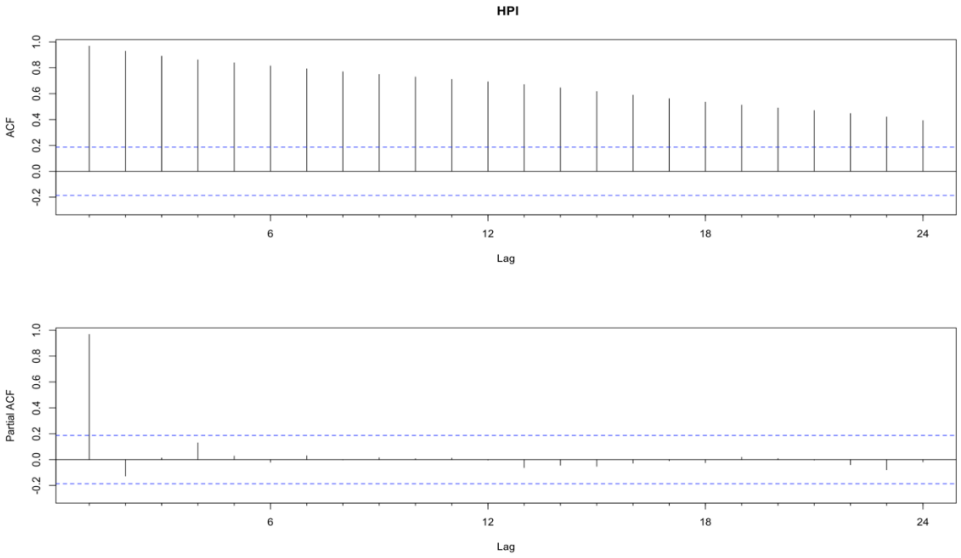


Figure 15: ACF & PACF plot for stationary judgement

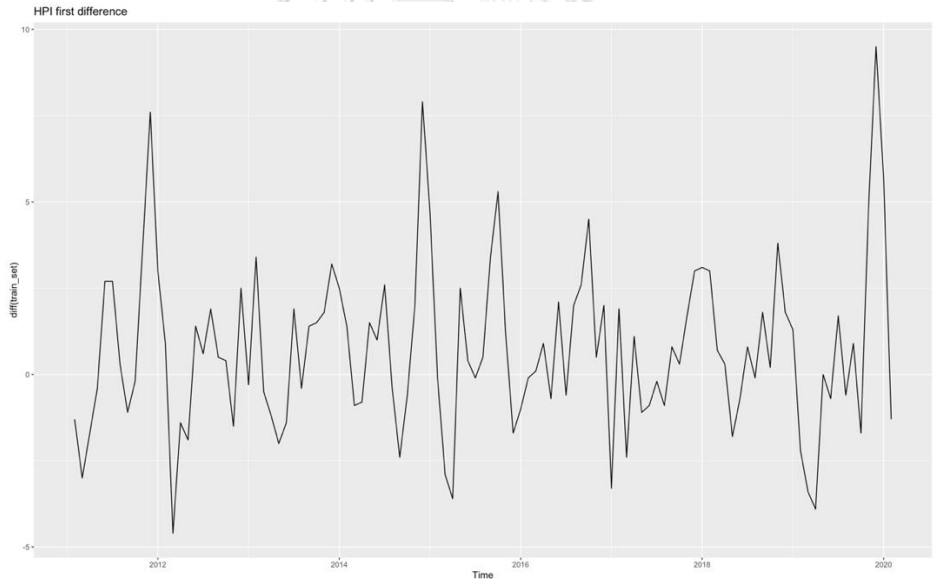


Figure 16: HPI with first order difference

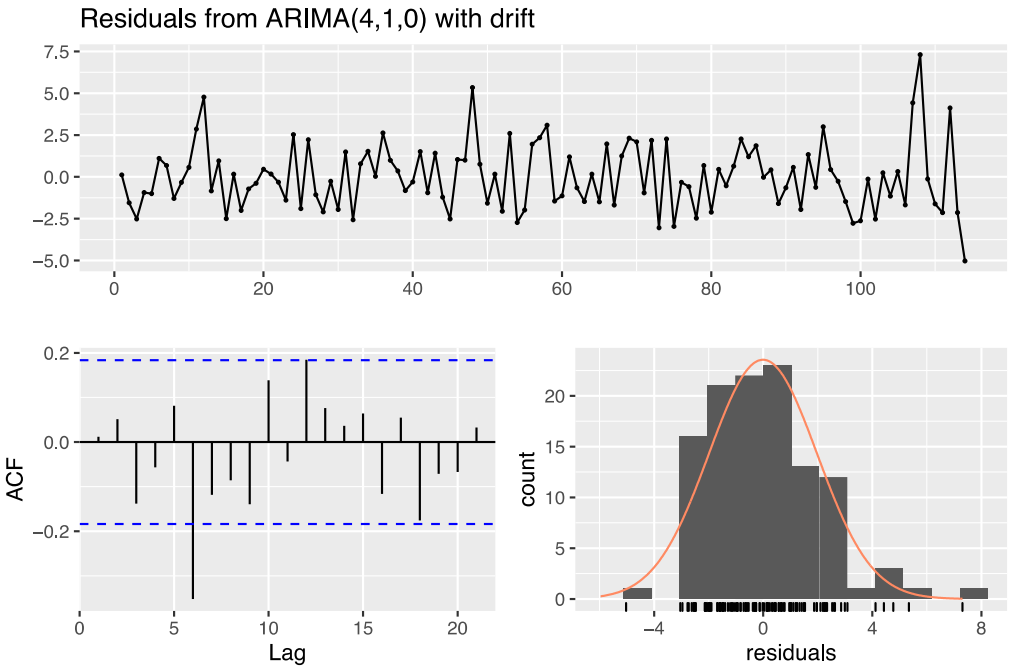


Figure 17: Residuals checking plot for univariate ARIMA model

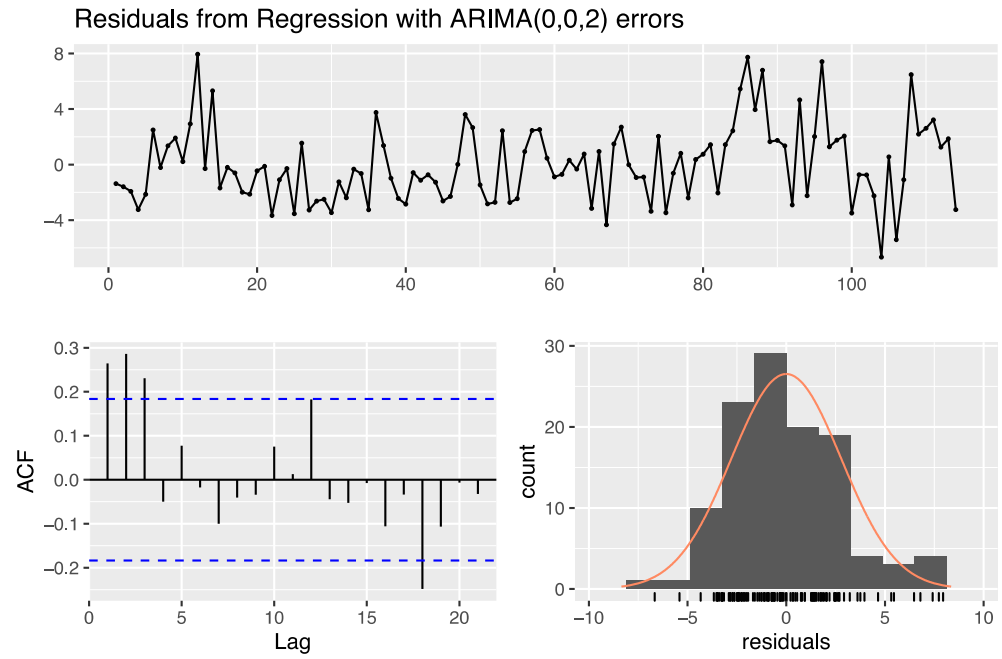


Figure 18: Residuals checking plot for multivariate ARIMA model

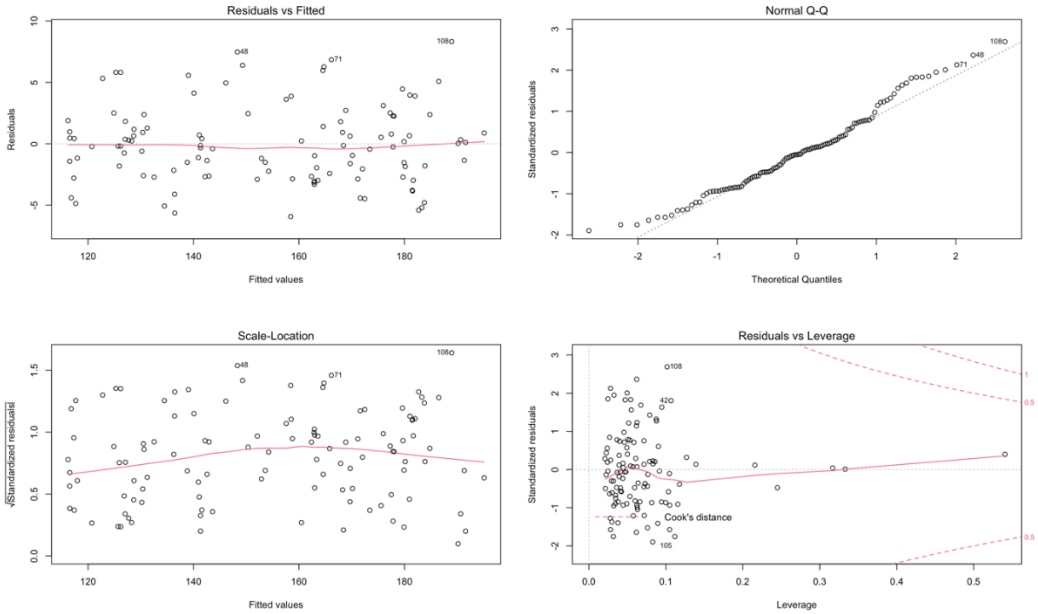


Figure 19: Residuals checking plot for multiple regression model

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Appendix C

Table 10: The choosing process of optimal elastic net model

```
library(caret)
library(tidyverse)
library(readxl)
library(glmnet)
library(forecast)
# Normalize the data
whole_ela <- whole %>% mutate(t=c(1:117),
                        UN=100*UN,
                        INF=ifelse(INF==0,0,1),
                        NUM=NUM/1000) %>% scale() %>% as.data.frame()
data_ela <- whole %>% select("HPI")%>% summarise(sd=sd(HPI), avg= mean(HPI))
sd <- data_ela$sd
avg <- data_ela$avg
denor_ela<- function(x){
  return(x*sd+avg)
}
m_ela <- whole_ela[1:114,]

# Seprate the x and y
x_m_ela <- m_ela[,-2] %>% as.matrix()
y_m_ela <- m_ela[,2] %>% as.matrix()

# Run the data in different of alpha value
set.seed(2020)
list_of_fits_ela <- list()
for(i in 0:10){
  fit_name <- paste0("Alpha", i/10)
  list_of_fits_ela[[fit_name]]<-cv.glmnet(x_m_ela, y_m_ela,type.measure = "mse",alpha=i/10,
                                  family="gaussian")

}
results_ela <- data.frame()
set.seed(2020)
for(i in 0:10){
  fit_name <- paste0("Alpha", i/10)
  predicted <- predict(list_of_fits_ela[[fit_name]],s=list_of_fits_ela[[fit_name]]$lambda.1se,newx=x_m_ela)
  mse<- mean((y_m_ela-predicted)^2)
  rsq <- cor(y_m_ela, predicted)^2
  s <- list_of_fits_ela[[fit_name]]$lambda.1se
  temp<- data.frame(Rsq=rsq,MSE=mse,s=s,fit_name=fit_name)
  results_ela <- rbind(results_ela, temp)
}
results_ela
write.csv(results_ela, "m_ela results.csv")
```

CHULALONGKORN UNIVERSITY

Table 11: The choosing process of optimal ANN model

```
library(forecast)
library(tidyverse)
library(readxl)
library(neuralnet)
# Tidy data
whole_ann <- whole%>% mutate(t=c(1:117),
                        UN=100*UN,
                        INF=ifelse(INF==0,0,1),
                        NUM=NUM/1000)
data_ann <- whole_ann %>% select("HPI")%>% summarise(max=max(HPI), min= min(HPI))
max_hpi <- data_ann$max
min_hpi <- data_ann$min
norm_ann <- function(x){
  (x-min(x))/(max(x)-min(x))
}
denor_ann<- function(x){
  x*(max_hpi-min_hpi)+min_hpi
}

data_scale_ann <- sapply(whole_ann,norm_ann)
m_ann <- data_scale_ann[1:114,]

# Model choosing
set.seed(2020)
list_of_fits_ann <- list()
for(i in 8:15) {
  for(j in 2:5){
    fit_name <- paste0("fit_nn", i,"-",j)
    list_of_fits_ann[[fit_name]] <- neuralnet(HPI~., data = m_ann, hidden = c(i,j))
  }
}

results_ann <- data.frame()
for(i in 8:15){
  for(j in 2:5){
    fit_name <- paste0("fit_nn", i,"-",j)
    error <- list_of_fits_ann[[fit_name]]$result.matrix[1,1]
    temp <- data.frame(Model=fit_name, Error=error)
    results_ann <- rbind(results_ann, temp)
  }
}
results_ann
write.csv(results_ann,"m_ann results.csv")
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Table 12: Random Cross Validation process

```
# out of sample comparison
# uni arima cv
fun_uar <- function(x,h,order){
  forecast(Arima(x, order=order), h=h)
}

fr_uar <- tsCV(m_uar, fun_uar, order=c(4,1,0), h=10)[,3]%>% as.matrix()
data_uar <- cbind(real,fr_uar) %>% as.data.frame()
aco_uar <- data_uar %>% mutate(pe=abs(fr_uar/real),si=ifelse(fr_uar>0,1,0)) %>%
  summarize(aco_uar=mean(pe,na.rm=T),dso_uar=mean(si,na.rm=T))

# multi arima cv
result_of_mar <- list()
for (i in 1:10) {
  n <- sample(1:104,1)
  test <- m_mar[n:(n+10),]
  xr <- as.matrix(xre[n:(n+10),])
  pr_mar <- forecast(fit_mar,xreg = xr)$mean
  temp <- data.frame(test=test,pre=pr_mar)
  result_of_mar[[i]] <- temp
}
r_mar <- result_of_mar
aco <- list()
do <- list()
for (i in 1:10){
  aco[[i]] <- accuracy(r_mar[[i]]$HPI,r_mar[[i]]$pre)
  data_mar <- r_mar[[i]] %>% as.data.frame()
  do[[i]] <- data_mar %>%mutate(pre=as.vector(pre)) %>%
    mutate(r=lag(HPI,1),reg=lag(pre,1))%>%
    mutate(dr=HPI-r, df=pre-reg,ds=ifelse(dr*df>0,1,0)) %>%
    summarize(ds_mar= mean(ds,na.rm=T))
}

# reg cv
set.seed(2020)
result_of_reg <- data.frame()
for (i in 1:10){
  index <- sample(1:nrow(m_reg), round(0.9*nrow(m_reg)))
  train_cv <- m_reg[index,]
  train_cv <- as.data.frame(train_cv)
  test_cv <- m_reg[-index,]
  test_cv <- as.data.frame(test_cv)
  fit_reg <- lm(HPI~t+CCI+EUR+RMB+INF+NUM, data = train_cv)
  pr_reg <- predict(fit_reg, test_cv)
  test <- test_cv[,2]
  temp <- data.frame(test=test,pre=pr_reg)
  result_of_reg <- rbind(result_of_reg, temp)
}
r_reg <- result_of_reg
aco_reg <- accuracy(r_reg$test,r_reg$pre)
do_reg <- r_reg %>% mutate(r=lag(test),reg=lag(pre)) %>%
  mutate(dr=test-r, df=pre-reg,ds=ifelse(dr*df>0,1,0)) %>%
  summarize(ds_reg= mean(ds,na.rm=T))

# ela cv
set.seed(2020)
result_of_ela <- data.frame()
for (i in 1:10){
  index <- sample(1:nrow(m_ela), round(0.9*nrow(m_ela)))
  train_cv <- m_ela[index,]
  test_cv<- m_ela[-index,]
  x_train_ela <- as.matrix(train_cv[,-2])
  y_train_ela <- as.matrix(train_cv[,2])
  x_test_ela <- as.matrix(test_cv[,-2])
  ela <- glmnet(x_train_ela,y_train_ela, family = "gaussian",alpha = 0.2)
  pr_ela <- predict(ela, newx=x_test_ela, s=0.01835)
  pr_ela <- sapply(pr_ela, denor_ela)
  test <- sapply(test_cv[,2],denor_ela)
  temp <- data.frame(test=test, pre= pr_ela)
  result_of_ela <- rbind(result_of_ela, temp)
```

```
}
r_ela<- result_of_ela
aco_ela <- accuracy(r_ela$test,r_ela$pre)
do_ela <- r_ela %>% mutate(r=lag(test),reg=lag(pre)) %>%
  mutate(dr=test-r, df=pre-reg,ds=ifelse(dr*df>0,1,0)) %>%
  summarize(ds_ela= mean(ds,na.rm=T))

# ann cv
set.seed(2020)
result_of_ann <- data.frame()
for (i in 1:10){
  index <- sample(1:nrow(m_ann), round(0.9*nrow(m_ann)))
  train_cv <- m_ann[index,]
  test_cv <- m_ann[-index,]
  nn <- neuralnet(HPI~., data = train_cv,hidden = c(10,3))
  pr_ann <- compute(nn, test_cv[,-2])$net.result
  pr_ann <- sapply(pr_ann, denor_ann)
  test <- sapply(test_cv[,2],denor_ann)
  temp <- data.frame(test=test, pre=pr_ann)
  result_of_ann <- rbind(result_of_ann, temp)
}
r_ann<- result_of_ann
aco_ann <- accuracy(r_ann$test,r_ann$pre)
do_ann <- r_ann %>% mutate(r=lag(test),reg=lag(pre)) %>%
  mutate(dr=test-r, df=pre-reg,ds=ifelse(dr*df>0,1,0)) %>%
  summarize(ds_ann= mean(ds,na.rm=T))
```


จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

REFERENCES

Besley, T., & Mueller, H. (2012). Estimating the Peace Dividend: The impact of violence on house prices in Northern Ireland. *American Economic Review*, *102*(2), 810-833.

Damodar, N., Gujarati ; Dawn, C., Porter. (2008). *Basic Econometrics*. McGraw-Hill Education.

Del Giudice, V., De Paola, P., & Del Giudice, F. P. (2020). COVID-19 infects real estate markets: Short and mid-run effects on housing prices in Campania region (Italy). *Social sciences*, *9*(7), 114.

DELMENDO, L. C. (June 30, 2020). *Property in Thailand brings good returns. So why isn't the market more vibrant?* https://www.globalpropertyguide.com/Asia/Thailand/Price-History

Ewing, B. T., Kruse, J. B., & Wang, Y. (2007). Local housing price index analysis in wind-disaster-prone areas. *Natural Hazards*, *40*(2), 463-483.

IMF. (2020, November 15, 2020). *Policy Responses to COVID19*. IMF. https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19#T

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.

Norouzi, N., de Rubens, G. Z., Choubanpishehzafar, S., & Enevoldsen, P. (2020). When pandemics impact economies and climate change: exploring the impacts of COVID-19 on oil and electricity demand in China. *Energy Research & Social Science*, *68*, 101654.

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

Pfefferbaum, B., & North, C. S. (2020). Mental health and the Covid-19 pandemic. *New England Journal of Medicine*, *383*(6), 510-512. https://www.nejm.org/doi/pdf/10.1056/NEJMp2008017?articleTools=true

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

Schrimpf, A., Shin, H. S., Tarashev, N., & Upper, C. (2020). Cross-border commercial real estate investment in Asia-Pacific1. *BIS Quarterly Review*, 15.

Strömberg, P., Hedman, M., & Broberg, M. (2011). Forecasting the House Price Index in Stockholm County 2011-2014: A multiple regression analysis of four influential macroeconomic variables.

Why, despite the coronavirus pandemic, house prices continue to rise. (2020). *The Economist*. https://www.economist.com/finance-and-economics/2020/09/30/why-despite-the-coronavirus-pandemic-house-prices-continue-to-rise

VITA

NAME Pengfei Chen

DATE OF BIRTH 15 Nov 1984

PLACE OF BIRTH China

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University