Economics Influence on Housing price Index within Southeast Asian Countries

Steven Qin
20819760
s45qin@uwaterloo.ca

Tiankai Jiang
20834939
t57jiang@uwaterloo.ca

November 25, 2020

Group 17

1 Introduction

Talking about 2020, one of the ineluctable topics would be the impact of Coivd-19 and how to post-reaction would impact the financial market. It was unbelievable to see the U.S. stock market recover from the crisis in March. One of the most extreme examples would be Nasdaq where tech companies are significantly weighted. In hardly three months, it was not only recovered but also went beyond the post-high from before the pandemic. Base on CNBC's report, one of the largest no-commission online trading platforms Robinhood reports approximately 13 million active users. As we can say, more people have begun to establish attention to the Financial world since the market crash. Besides the stock market, the housing market has also increased. But it is challenging to explain such an abnormal increase base on fundamental knowledge because the world economy today remains far from being recovered due to the universal pandemic. We can recognize a higher than usual unemployment rate. One of the most objective indicators is the HPI, housing price index. In this project, we aim to explain the HPI using macroeconomic factors and perform trend analysis using a linear regression model.

2 Literature Review

This section reviews literature that focused on the relationship between housing price and macroeconomic factors. Some of the literature concentrated on macroeconomic factors, while others emphasized various analyzing techniques. Most of the papers concluded with several variables that they found could be the key factors that impact the housing price. The results are shown in the table 2.

Author(s) (year)	Macroeconomic Factors
Emiris (2016)	Mortgage loans, interest rates, GDP, inflation rate
Gasparėnienė et al. (2016)	Availability of bank loans, interest rate, inflation rate, GDP
Gupta et al. (2011)	10 variables in Iacoviello and Neri (2010), 110 additional variables from Boivin et al. (2009)
Pan and Wang (2013)	Personal income growth rate, income and labor force growth, bank instability
Goodhart and Hofmann (2008)	Shocks on GDP, CPI, interest rate
Beltratti and Morana (2010)	GDP, private consumption and investment, CPI inflation, short-and long-term interest rates
Tripathi (2019)	Price-to-income ratio, price-to-rent ratio, urbanization, per-capita GDP, rent, inflation, GDP growth rate, broad money, real exchange rate, percentage share of employment in services
Bouchouicha and Ftiti (2012)	Inflation rate, short-and long-term interest rates, employment rate, money supply
Renigier-Biłozor and Wiśniewski (2012)	Unemployment rate, population growth, consumption expenditure, household consumption expenditure and housing expenses

Table 1: Macroeconomic factors that have the impact on housing price level.

Emiris (2016) proposed a joint model to address the relationship between residential property price and several macro-economy indicators (e.g. mortgage loans, interest rates, the GDP, and inflation rate), aiming to forecast both the housing price and the economy in Belgium. Their approach is to

use a dynamic factor model estimated with maximum likelihood and the EM algorithm, in which they assume that HPI and other macroeconomic variables comove strongly. They assume HPI can be modeled as a sum of two components: a "common component" that is driven by an unobserved factor such as business cycle, and an "idiosyncratic component" that is uncorrelated with the common component. Their result shows that the prediction over 2008q1-2009q4 was relatively accurate during 2008 while it underestimated quarterly growth rates in 2009. The authors state that at this stage, it is still a reduced-form exercise, and future works should focus on the discovery of shocks and their impact on the HPI forecast.

Gasparėnienė et al. (2016) discussed the macroeconomic factors and house price level in Lithuania. The authors first summarized factors that could influence the house price from 14 previous papers, including GDP, employment rate, interest rate, and construction price, etc. Then, linear regression and methods for correlation were performed, accessing the impact of those factors on the average price of two-room apartments in Lithuania from 2008 to 2015. The result shows that the availability of bank loans and interest rate account most for the house price level, explain the variation of the house price by 79.03% and 49.23%, the inflation rate explains the price by 39.35%, and GDP has the most insignificant impact amount these four variables in Lithuania.

Plakandaras et al. (2014) stated that the issue among most house price forecasting models is that almost all time-series are aggregations of rather volatile elements, which introducing great error and making it difficult to generate an accurate model. In the paper, a three-step forecasting methodology was proposed to address this issue. First, a new method called Ensemble Empirical Mode Decomposition (EEMD) was used to smooth the original series data. Then, desired variables were selected using the Elastic Net approach. Finally, a Support Vector Regression model was constructed for forecasting. Using HPI data from 1989 to 2012 in the US, the result shows that their model outperforms all competing models in both in-sample and out-of-sample forecasting, and the model can make early accurate detection of the 2006-2009 price downturn.

Gupta et al. (2011) compared several time series models for HPI forecasting including dynamic stochastic general equilibrium (DSGE) model and the variations of vector autoregressive (VAR) models, such as Bayesian VAR (BVAR) models, Bayesian factor augmented VAR (BFAVAR), and small- and large-scale BVAR (SBVAR and LBVAR) models. 0 to 120 macro-economy factors are being used in different models to predict the HPI in the US. Using the period of 1976q1 to 2000q4 as the in-sample period and 2001q1 to 2005q2 as the out-of-sample horizon, they found that each model

performs better in different periods and different conditions. They conjecture that the utilization of fundamental economic variables may improve the forecasting performance over models that do not use such data. But the gains do not prove statistically significant. Therefore, additional research is required.

Inspired by the subprime mortgage crisis developed in 2007 and 2008, which was triggered by the 2005 housing bubble burst, Pan and Wang (2013) investigate the relationship between the bank stability and the house prices in the US. Factors such as non-performing loans, equity ratio, cost-income ratio, and return on assets were selected. A threshold model was applied in the experiment. Their result shows that the personal income growth rate is considered as the threshold variable which interacts with the house price indicators in the threshold models. And the house price increase with rising demand due to income and labor force growth. The result also indicated a negative correlation between bank stability and house price.

Goodhart and Hofmann (2008) assessed the linkages between money, credit, house prices, and economic activity in industrialized countries between 1973q1 and 2006q4. The analysis was performed on 17 industrialized countries based on a fixed-effects panel vector autoregressive (VAR) model. The result found that money growth influences house prices and credit, credit influences money, and house prices and house prices influence both credit and money. Also, it is found that shocks on GDP, CPI, and interest rate have a significant effect on house prices, indicating that the linkages between those factors are all bidirectional. However, this result also indicates that the choice of predictors in regression analysis and predictions should be carefully selected since all factors are correlated.

Beltratti and Morana (2010) aimed to explore the linkages between general macroeconomic conditions and the housing market and whether there is a feedback effect of housing price shocks on the real economy. A Factor-Augmented Vector Autoregressive (F-VAR) model is used. The paper discussed the effect of eleven macroeconomic variables, including GDP, private consumption and investment, CPI inflation, short-and long-term interest rates, etc., on the housing price. The result shows that for G-7 countries, global macroeconomic shocks account for about 40% in determining common house price fluctuations; productivity shocks are more important than demand shocks in determining house prices and the real estate itself accounts for about 20% of the house price fluctuation.

While most previous research focused on the house price in one country, Tripathi (2019) investigated the appropriate macroeconomic determinants of real house prices for 43 countries in the world from 1970 to 2017. The authors performed an OLS regression on the housing price. The result indi-

cates that the price-to-income ratio, price-to-rent ratio, urbanization, per-capita GDP, rent, inflation, GDP growth rate, broad money, and real exchange rate influence positively on housing prices. The percentage share of employment in services influence negatively on real house prices. Finally, there is not enough evidence to show the real interest rate affects the housing price. The target countries are separated into four groups based on their income, just like what we did in this paper. The author also found a gap between countries with different income levels, suggesting that the gap was mainly caused by urbanization and different regulation of housing prices in those countries.

Bouchouicha and Ftiti (2012) used dynamic coherence function (DCF), which was proposed by Priestley and Tong (1973), to analyze the real estate time series data of the UK and the US market. And they chose inflation rate, short-and long-term interest rates, employment rate, and money supply as the macroeconomic data. They found some synchronization between the UK and the US real estate markets in their long-run co-movements with the long-term interest rate, inflation, and employment rate. However, they also found desynchronization between the two countries in terms of economic growth, the money supply, and the short-term interest rate, which provided suggestions when selecting the independent variables in regression analysis, especially when comparing different countries' real estate data within the same time frame.

Finally, Renigier-Biłozor and Wiśniewski (2012) presented Artificial Neural Networks (ANN) model to select macroeconomic factors that could affect real estates. They chose 14 macroeconomic factors (e.g. GDP, unemployment rate, CPI...) as alternative independent variables and the house price index in Poland and Italy as dependent variables. Although ANN will not be used in this paper, it is usually provided with high accuracy and its result could be referred to as we choosing independent variables. They found for the developing country, i.e. Poland, unemployment rate, and population growth are the key factors that influence the house price. Whereas for a developed country, i.e. Italy, consumption expenditure, household consumption expenditure, and housing expenses have a major effect on the house price.

3 Methods

In this section, we will first talk about the candidate independent variables, their relationship with annual HPI and why do we select them. Then, the mathematical model will be proposed. Finally, since there are too many candidate variables, a reliable method will be proposed to filter those

variables.

3.1 Variables

To obtain the relationship between HPI and macroeconomic, one of the initial steps is to break down macroeconomic into smaller sectors that can be quantified. The public API we used for this project represents macroeconomics with more than thirty different sectors. Such sectors include GDP and Economic Growth, Consumption and Investment, Money, Labor Market, etc. Some other sectors contain relative information over the international index and ESG. Further, a sector contains multiple related indicators.

The essential part of construct this linear regression model is to, from top-down principal selection, reduce the number of potential sectors then designate principals to perform further calculations. To satisfy one of the important conditions by linear regression, we only select one variable from each sector to reduce the internal correlation in the set of independent variables. In the beginning, we chose 20 independent variables from 17 sectors, such as the rate of change of real GDP, the Unemployment rate, Government spending as percent of GDP, Personal income tax rate, External debt, percent of Gross National Income, etc. The tax rate factors were excluded because we observed a great chance for the tax rate to remain constant during an eleven consecutive year period. Also, because we have to ensure the data is complete for all four-country, and we had to exclude some variables accordingly. Subsequently, we added another fifteen variables to try to increase the accuracy we get from the regression analysis. Eventually, twenty-three qualified variables were chosen to perform the linger regression analysis. All of the remaining factors have to be complete and accurate as we attempted to prevent from filling missing data using addition calculation.

Moreover, we split the data into two groups, objective and supportive. These independent variables must also either directly reflect the domestic economic characteristics or indirectly relates to HPI. For example, the rate of change of real GDP is an objective variable because it would reveal the difference in the domestic economic status. On the other hand, the percent urban population would be supportive because it must correspond to a variable from the objective group before display the meaning relates to HPI. From this experiment, we observed that HPI relates to most of the independent variables. Whereas, a small portion of the macroeconomics factors, such as change in Consumer Price Index (inflation) and stock market capitalization as percent of GDP, only have a correlation coefficient around zero with HPI. After excluding these variables, there are still too many

variables that would make the model extremely complicated. Hence, we set a higher benchmark and only keep variables that prove to be strongly correlated with HPI, either positively or negatively. A few qualified candidates are GDP growth rate, unemployment rate, household debt to GDP, etc.

3.2 Model Description

There are three methods mentioned in section 2 to analyze and predict annual HPI data, namely, linear regression, time series analysis, and artificial neural network. We chose to use linear regression, like the one that was used in Gasparėnienė et al. (2016). The model is shown as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \epsilon$$

in which y represents the annual HPI and each x could be a candidate variable. The βs are the values we desire in the regression.

The reason we chose linear regression instead of the other two methods is that those two methods require a lot more data. Generally, all papers that did not use linear regression used data from the 1970s to 2010s. However, we only have 11 years of data on annual HPI, which is insufficient to perform those analyses. Also, as mentioned in section 1, we want to build a unified model for all four countries, which means we want to find the same independent variables for all four countries. Implementing a Vector Autoregressive model is too complicated for that.

3.3 Variables Selection

We have mentioned in the previous section that there are lots of independent variables that may influence the house price index. It is obvious that we cannot use all of them for regression. Therefore, the first step is to select the proper independent variables. We have the following constraints:

1. It is not possible to fit all four countries in the same linear function with the same coefficients, however, we should at least use the same independent variables for four regressions. Otherwise, the results are just four separate models without comparability.

2. When selecting the independent variables for regression, any selected variable should be strongly correlated with HPI.

3. When selecting the independent variables for regression, all selected variables should not be strongly correlated with each other.

Our strategy is(Shetye, 2019):

1. We construct the Pearson Correlation Table of all variables for all four countries. And remove a variable if, in any tables, the correlation between that variable and the annual HPI is less or equal to 0.5.

2. After the first step, all variables left are strongly correlated with annual HPI. We construct the Pearson Correlation Table again, this time, for all independent variables left. And we only pick one from a group, if that group of variables is strongly correlated with each other (here we pick r=0.7).

We use China as an example for step 1 (Figure 1). Here actually we only use the annual HPI row.

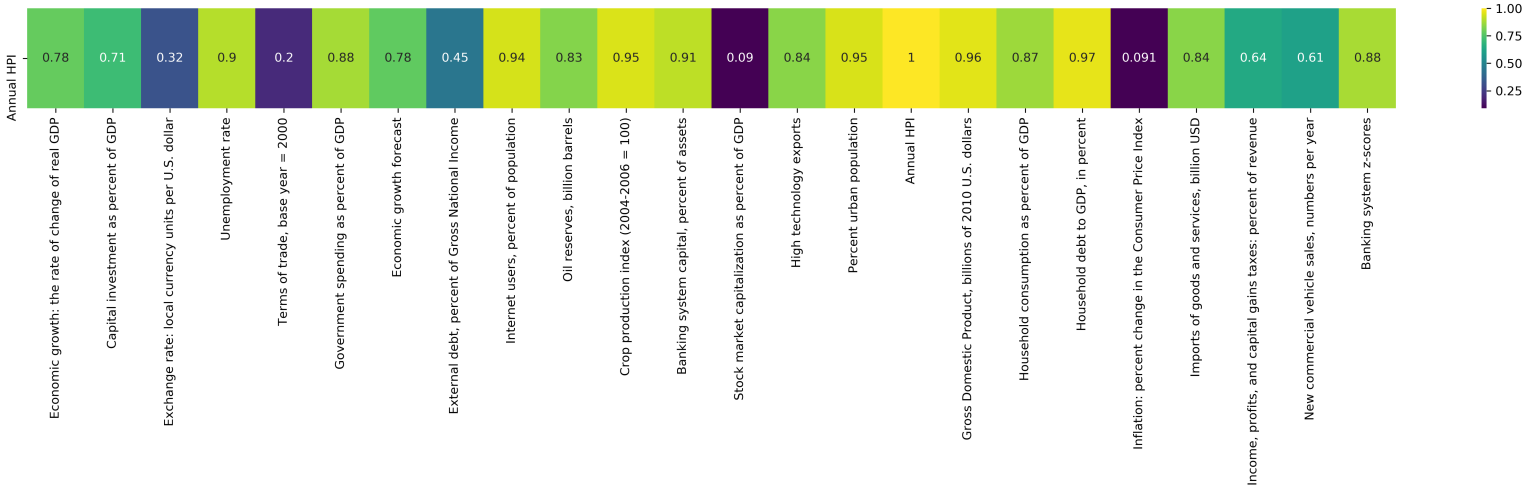


Figure 1: Pearson Correlation Table in Step 1

After doing this step for all four countries, we plot the correlation table for all variables left (Figure 2).

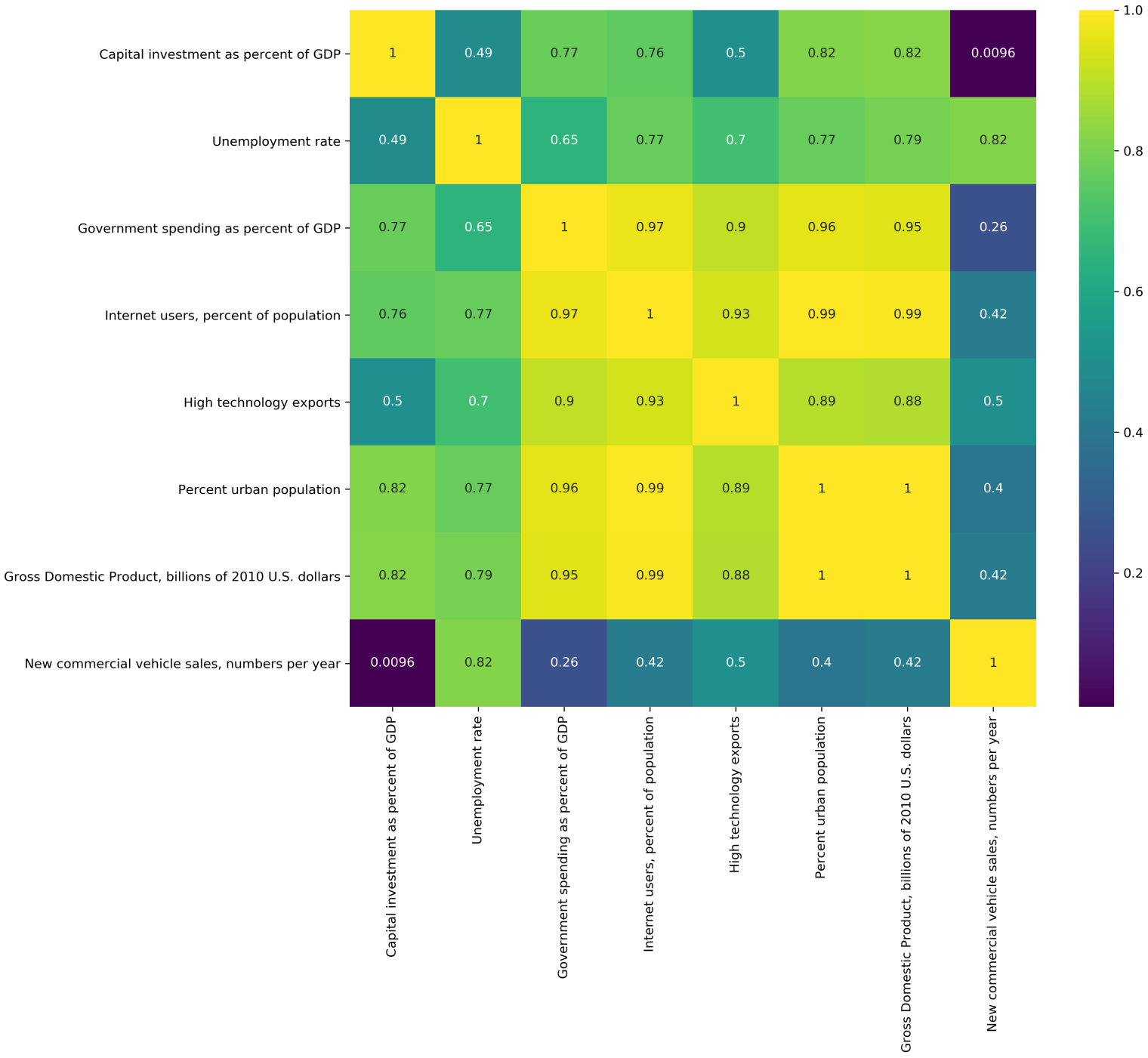


Figure 2: Pearson Correlation Table in Step 2

Lots of independent variables are strongly correlated, which does not satisfy the assumption of linear regression. Based on the results mentioned in the literature reviews, here we pick only the

Unemployment rate and Government spending as percent of GDP (row 2 and row 3).

4 Data

We chose annual HPI as our dependent variable and based on the previous research, we chose 23 candidate variables for regression analysis. The relationship between those candidate variables and the annual HPI has been discussed in section 3. The appendix contains a description of each variable, its type, how it was constructed, and the data source. The table below shows the mean, standard deviation, minimum, and maximum of each variable. We group the result by countries.

Variable	mean	std	min	max
Economic growth: the rate of change of real GDP	0.673	2.353	-5.420	4.190
Capital investment as percent of GDP	23.025	1.130	21.300	24.320
Exchange rate: local currency units per U.S. do...	100.538	13.793	79.790	121.040
Unemployment rate	3.692	0.999	2.290	5.100
Terms of trade, base year = 2000	65.403	5.525	58.440	74.370
Government spending as percent of GDP	19.893	0.262	19.490	20.250
Economic growth forecast	0.672	2.353	-5.420	4.190
Internet users, percent of population	85.934	6.391	78.000	93.180
Oil reserves, billion barrels	0.040	0.000	0.040	0.040
Crop production index (2004-2006 = 100)	91.250	4.264	81.800	96.700
Banking system capital, percent of assets	5.270	0.412	4.300	5.630
Stock market capitalization as percent of GDP	89.781	26.454	54.010	127.860
High technology exports	112398.783	12738.820	98537.260	133518.230
Percent urban population	91.205	0.478	89.990	91.700
Gross Domestic Product, billions of 2010 U.S. d...	5908.509	231.440	5470.800	6210.700
Household consumption as percent of GDP	57.198	1.461	55.339	58.960
Household debt to GDP, in percent	53.427	17.807	0.000	62.500
Inflation: percent change in the Consumer Price...	0.300	1.079	-1.400	2.800
Imports of goods and services, billion USD	852.938	119.272	626.200	998.180
Income, profits, and capital gains taxes: perce...	47.820	2.358	42.110	50.960
New commercial vehicle sales, numbers per year	809939.636	65941.198	701188.000	894125.000
Banking system z-scores	15.349	2.031	10.330	18.013

Table 2: Mean, Standard deviation, Min and Maximum of each variable (Japan)

Variable	mean	std	min	max
Economic growth: the rate of change of real GDP	7.849	1.413	6.110	10.640
Capital investment as percent of GDP	44.761	1.685	42.630	46.660
Exchange rate: local currency units per U.S. do...	6.534	0.277	6.140	6.910
Unemployment rate	4.525	0.131	4.280	4.720
Terms of trade, base year = 2000	85.492	5.045	79.050	92.530
Government spending as percent of GDP	15.854	0.736	14.590	16.919
Economic growth forecast	7.820	1.404	6.110	10.560
External debt, percent of Gross National Income	13.532	2.100	8.920	16.950
Internet users, percent of population	47.054	10.472	28.900	62.712
Oil reserves, billion barrels	22.919	3.194	16.000	25.930
Crop production index (2004-2006 = 100)	137.225	12.943	116.800	156.683
Banking system capital, percent of assets	7.448	1.282	5.600	9.341
Stock market capitalization as percent of GDP	58.013	12.249	41.260	74.020
High technology exports	606382.805	114654.551	359273.920	759591.883
Percent urban population	54.210	4.120	47.880	60.310
Gross Domestic Product, billions of 2010 U.S. d...	8414.664	1982.378	5502.000	11537.200
Household consumption as percent of GDP	36.858	1.790	34.330	39.433
Household debt to GDP, in percent	37.790	10.745	23.500	53.793
Inflation: percent change in the Consumer Price...	2.291	1.511	-0.700	5.600
Imports of goods and services, billion USD	1980.514	437.674	1042.530	2548.880
Income, profits, and capital gains taxes: perce...	22.371	3.175	18.864	28.270
New commercial vehicle sales, numbers per year	3933418.000	359135.486	3313479.000	4370795.000
Banking system z-scores	20.852	2.269	17.130	24.116

Table 3: Mean, Standard deviation, Min and Maximum of each variable (China)

Variable	mean	std	min	max
Economic growth: the rate of change of real GDP	5.950	1.759	1.45	7.33
Capital investment as percent of GDP	22.235	3.120	17.43	27.15
Exchange rate: local currency units per U.S. do...	46.639	3.682	42.23	52.66
Unemployment rate	3.135	0.598	2.15	3.86
Terms of trade, base year = 2000	79.358	3.932	73.52	84.20
Government spending as percent of GDP	10.863	0.913	9.71	12.53
Economic growth forecast	5.951	1.760	1.45	7.34
External debt, percent of Gross National Income	23.135	3.074	20.13	28.49
Internet users, percent of population	39.152	15.417	9.00	60.05
Oil reserves, billion barrels	0.140	0.000	0.14	0.14
Crop production index (2004-2006 = 100)	114.112	3.567	109.60	119.30
Banking system capital, percent of assets	10.213	0.719	9.50	11.70
Stock market capitalization as percent of GDP	76.018	10.980	49.03	88.41
High technology exports	33703.015	283.599	33502.48	33903.55
Percent urban population	46.128	0.625	45.33	47.15
Gross Domestic Product, billions of 2010 U.S. d...	269.055	55.812	194.10	360.90
Household consumption as percent of GDP	72.208	0.859	70.19	73.21
Household debt to GDP, in percent	0.000	0.000	0.00	0.00
Inflation: percent change in the Consumer Price...	3.136	1.363	0.70	5.20
Imports of goods and services, billion USD	98.596	31.768	54.01	151.72
Income, profits, and capital gains taxes: perce...	40.729	1.716	36.53	42.08
New commercial vehicle sales, numbers per year	142465.909	38579.063	86333.00	226384.00
Banking system z-scores	18.112	1.127	16.19	20.71

Table 4: Mean, Standard deviation, Min and Maximum of each variable (Philippines)

Variable	mean	std	min	max
Economic growth: the rate of change of real GDP	6.845	1.266	5.02	8.500
Capital investment as percent of GDP	34.689	4.111	30.17	40.220
Exchange rate: local currency units per U.S. do...	59.015	9.084	45.73	70.420
Unemployment rate	5.543	0.121	5.33	5.670
Terms of trade, base year = 2000	97.596	6.501	89.99	108.190
Government spending as percent of GDP	10.847	0.483	10.30	11.770
Economic growth forecast	7.113	1.644	4.18	10.260
External debt, percent of Gross National Income	20.443	1.949	17.52	23.300
Internet users, percent of population	17.412	8.497	5.12	32.000
Oil reserves, billion barrels	5.235	0.539	4.42	5.680
Crop production index (2004-2006 = 100)	140.275	13.405	112.80	159.317
Banking system capital, percent of assets	7.108	0.205	6.70	7.390
Stock market capitalization as percent of GDP	76.156	13.504	55.25	97.390
High technology exports	15466.125	3058.267	10778.73	20273.090
Percent urban population	32.443	1.287	30.59	34.470
Gross Domestic Product, billions of 2010 U.S. d...	2197.400	483.493	1544.40	2964.000
Household consumption as percent of GDP	57.820	1.755	54.72	60.240
Household debt to GDP, in percent	9.841	0.838	8.90	11.300
Inflation: percent change in the Consumer Price...	7.664	3.010	2.50	12.000
Imports of goods and services, billion USD	524.751	84.219	347.18	639.010
Income, profits, and capital gains taxes: perce...	44.837	2.997	41.25	50.270
New commercial vehicle sales, numbers per year	730269.091	148067.419	449391.00	1005380.000
Banking system z-scores	16.813	0.470	15.96	17.560

Table 5: Mean, Standard deviation, Min and Maximum of each variable (India)

5 Results

5.1 Normality of the Dependent Variable

Shapiro-Wilks test was used to check normality of the dependent variable.

H_0: The variable is normally distributed.

H_a: The variable is not normally distributed.

For China, India, Japan, Philippines, the p-values are **0.4830, 0.4636, 0.2070, 0.1855** respectively. All four values are greater than $\alpha = 0.05$. Therefore, we failed to reject H_0 and concluded that the dependent variable is normally distributed.

Kolmolgorov-Smirnoff test was also used to check normality of the dependent variable.

$$D = \max_{1 \leq i \leq N} (F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i))$$

H_0: The variable is normally distributed.

H_a: The variable is not normally distributed.

For China, India, Japan, Philippines, the result are **0.1873, 0.1224, 0.1338, 0.1536** respectively. According to the One-Sample Kolmogorov-Smirnov Table (Appendix B), when $n = 11$ and $\alpha = 0.05$, the critical value $D_{n,a} = 0.3912$. All four results are less than the critical value. Therefore, we failed to reject the H_0. The dependent variable is normally distributed.

5.2 Mean Comparison of the Dependent Variable

One-tail two sample independent t-test was performed on all four countries. Therefore, there are 6 comparisons in total.

Let μ_1 denotes the average HPI of the country in the first column, μ_2 denotes the average HPI of the country in the second column.

$H_0 : \mu_1 \geq \mu_2$

$H_a : \mu_1 < \mu_2$

$df = 11 + 11 - 2 = 20$, the critical value for an one-tail t-test when $\alpha = 0.05$ is $t_{crit} = 1.725$.

The results are shown in table 6.

Country A	Country B	t_{obj}	t_{crit}	result
India	Philippines	1.395607763	0	0
Philippines	China	2.481775986	0	1
China	Japan	2.196236113	0	1
Philippines	Japan	3.070888374	0	1
India	China	3.929769778	0	1
India	Japan	4.449066711	0	1

Table 6: Comparison Results

For the comparison between India and Philippines, $t_{obj} = 1.3956 < t_{crit}$. We failed to reject H_0 and conclude that there is not enough evidence to show that the average HPI of the Philippines is greater than that of India. For all the other five comparisons, $t_{obj} > t_{crit}$, we reject H_0 and conclude that the HPI of the country in the second column is indeed greater than that in the first column.

5.3 Regression and Residuals Analysis

The regression results are summarized in the table 7.

For all four countries, the Prob(F-statistic) is close to zero, indicating that the overall regression is meaningful. All countries have a Durbin-Watson between 1 and 2, implying that the regression results are reliable from the interpretation side of this metric. According to R-squared, the unemployment rate and government spending are accounts for 80 to 90 percent of the variation in annual HPI.

Based on the result, we can get the regression equations for the four countries, in which $\hat{y}$ is the annual HPI, x_1 stands for the unemployment rate, and x_2 stands for Government spending as a percent of GDP.

Japan: $\hat{y} = -4.856x_1 - 5.4x_2 + 229.38$

China: $\hat{y} = -61.183x_1 + 9.599x_2 + 238.571$

Philippines: $\hat{y} = -47.361x_1 + 35.132x_2 - 72.193$

India: $\hat{y} = -607.825x_1 - 81.276x_2 + 4451.38$

For the Philippines, the P-value of the constant is 0.774, which is abnormal. That's because of the unemployment rate and government spending are strongly correlated in the Philippines. However, as we are finding a universal model for all four countries, we have to make some compromise here.

	Japan	China	Philippines	India
Model:	OLS	OLS	OLS	OLS
Dependent Variable:	Annual HPI	Annual HPI	Annual HPI	Annual HPI
No. Observations:	11	11	11	11
Df Model:	2	2	2	2
Df Residuals:	8	8	8	8
R-squared:	0.965	0.957	0.932	0.851
Adj. R-squared:	0.956	0.946	0.914	0.814
AIC:	35.5778	59.7291	97.2204	109.2955
BIC:	36.7715	60.9228	98.4141	110.4891
Log-Likelihood:	-14.789	-26.865	-45.610	-51.648
F-statistic:	110.0	88.39	54.40	22.82
Prob (F-statistic):	1.51e-06	3.51e-06	2.20e-05	0.000495
Scale:	1.1847	10.645	321.61	963.98
const_Coef.	229.38	238.571	-72.1933	4451.38
const_Std.Err.	26.1221	69.3646	243.306	649.735
const_t	8.78105	3.43938	-0.296718	6.85107
const_P>|t|	2.22013e-05	0.0088319	0.774239	0.000130848
UnemploymentRate_Coef.	-4.85583	-61.1828	-47.3606	-607.825
UnemploymentRate_Std.Err.	0.34461	10.3284	24.161	90.384
UnemploymentRate_t	-14.0908	-5.92375	-1.9602	-6.72492
UnemploymentRate_P>|t|	6.251e-07	0.000352204	0.0856311	0.000148837
GovSpending_Coef.	-5.39993	9.59926	35.1322	-81.2761
GovSpending_Std.Err.	1.31389	1.84297	15.8131	22.7249
GovSpending_t	-4.10989	5.20857	2.22171	-3.57653
GovSpending_P>|t|	0.00339113	0.000813845	0.0570298	0.00722465
Omnibus:	0.532	4.142	0.539	0.354
Prob(Omnibus):	0.766	0.126	0.764	0.838
Skew:	-0.192	-0.954	0.401	0.241
Kurtosis:	1.985	3.319	2.385	2.642
Durbin-Watson:	1.453	1.995	1.853	1.195
Jarque-Bera (JB):	0.541	1.714	0.469	0.165
Prob(JB):	0.763	0.425	0.791	0.921
Condition No.:	1615	1177	515	856

Table 7: Regression Results

We performed k-s test on the residuals, results are shown in table 8.

Country	p-value
Japan	0.94095
China	0.16866
Philippines	0.00139
India	0.00135

Table 8: Residual K-S Test Results

The result shows that only the residual of Japan is normally distributed. In the regression result, Prob(Omnibus) is low for China, also indicating the result is not normally distributed. For the other 3 countries, the Prob (Omnibus) is relatively high(around 0.8) so the data is somewhat normal, but not altogether ideal.

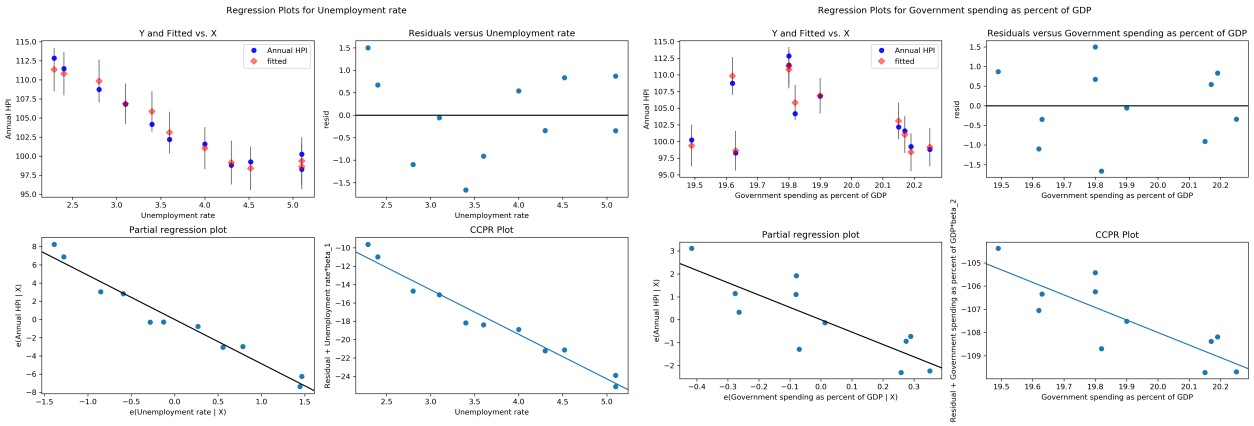
Breusch-Pagan test was used to check the heteroscedasticity of the regression.

H_0: Homoscedasticity is present.

H_a: Heteroscedasticity exists.

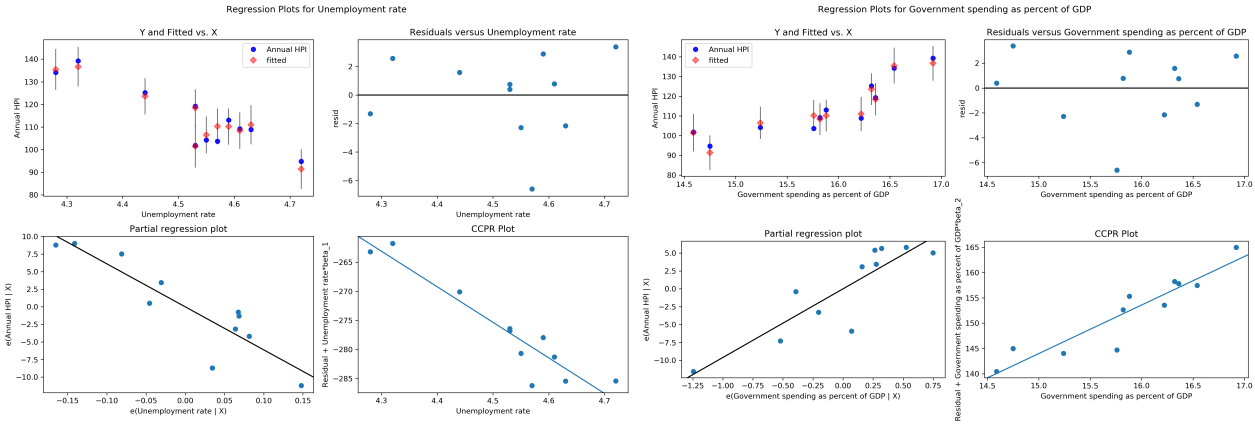
The p-values of the test are **0.7657, 0.0186, 0.2902, 0.0876** for China, India, Japan, Philippines respectively. Hence, for China, Japan and Philippines, we failed to reject the null hypothesis and we concluded that the models are not heteroskedastic. However, for India, the p-value is less than $\alpha = 0.05$, we reject H_0 and concluded that the model is heteroskedastic.

For this project, heteroskedasticity represents the level of volatility in the market. For example, there exists a correlation between the housing price and inflation, money supply. Although we may assume the real price of the HPI remains constant during the observation, the nominal price, which is reflected in HPI, continuous to grow. Hence, we choose to believe the volatility would increase along the HPI(Hayes, 2020; Nguyen, 2020).



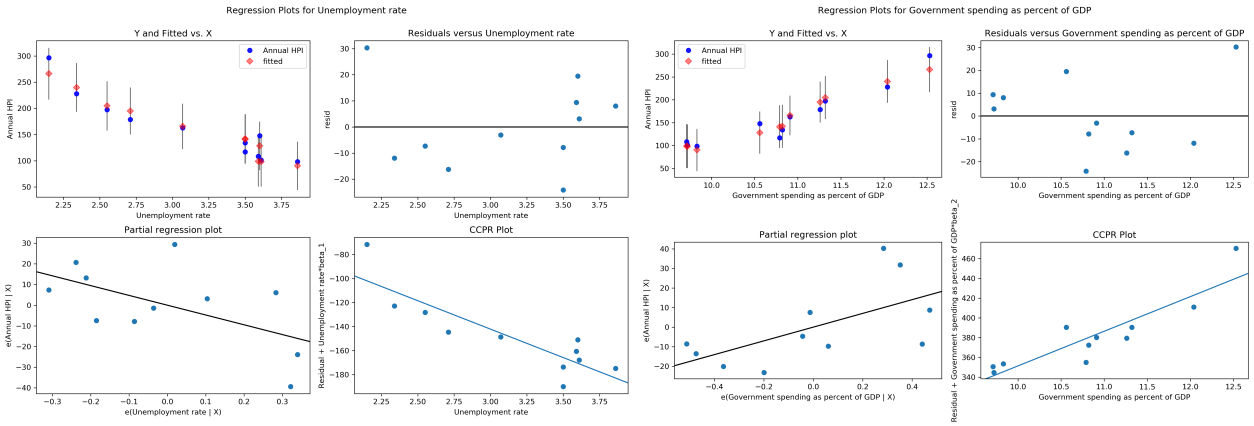
(a) Residual vs. Unemployment Rate (b) Residual vs. Government Spending as percent of GDP

Figure 3: Residual Plot (Japan)



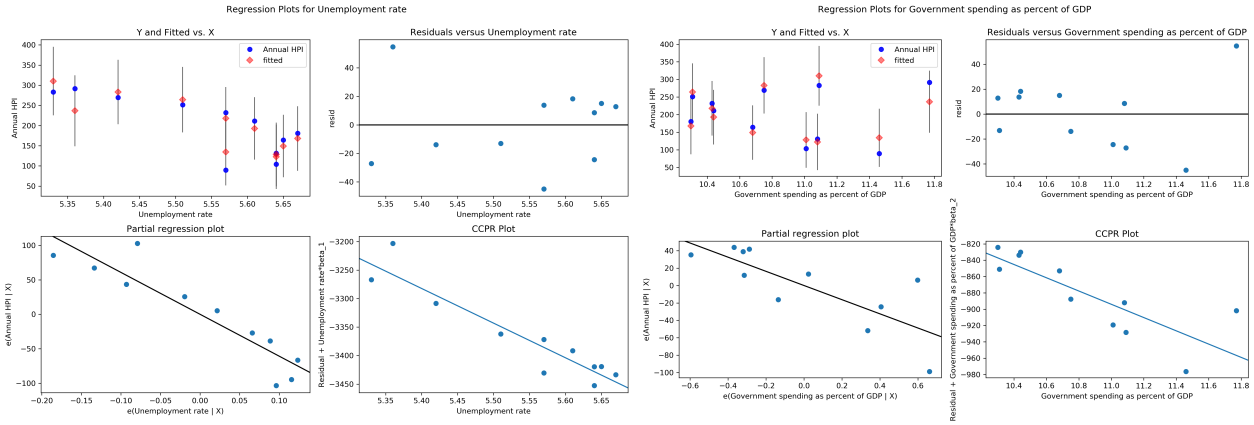
(a) Residual vs. Unemployment Rate (b) Residual vs. Government Spending as percent of GDP

Figure 4: Residual Plot (China)



(a) Residual vs. Unemployment Rate (b) Residual vs. Government Spending as percent of GDP

Figure 5: Residual Plot (Philippines)



(a) Residual vs. Unemployment Rate (b) Residual vs. Government Spending as percent of GDP

Figure 6: Residual Plot (India)

We also created residual plots for all four countries (Figure 3 to 6). For both unemployment rate and government spending as a percent of GDP, the residuals for all four countries randomly scattered around zero (the upper right plots), which is an indication that heteroscedasticity is not a problem with either predictor variable in the model(Zach, 2020).

6 Implication and Conclusion

Through this experiment, we successfully implement ways to represent HPI with independent variables selected from a pool of macroeconomic factors. However, many of the macroeconomic factors may represent overlapping information. For example, capital investment and household consumption can both shown as a percentage of GDP. Thus, it is essential to minimize such correlation when choosing independent variables to increase the accuracy of the linear regression. In this experiment, we find a universal linear model using the unemployment rate and government spending as a percent of GDP. Although the model we found certainly has great potential to improve, we have to make some compromise as we are looking for a universal model for all four countries. From time to time, the domestic situation of a country can not speak for another. But, we believe that we have successfully demonstrated the relationship between HPI and macroeconomics. Furthermore, a linear equation can express this relationship including the feature to predict future changes.

References

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

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *American Economic Review*, 99(1):350–84.

Bouchouicha, R. and Ftiti, Z. (2012). Real estate markets and the macroeconomy: A dynamic coherence framework. *Economic Modelling*, 29(5):1820 – 1829.

Emiris, M. (2016). A dynamic factor model for forecasting house prices in belgium. national bank of belgium working paper no. 313.

Gasparėnienė, L., Remeikienė, R., and Skuka, A. (2016). Assessment of the impact of macroeconomic factors on housing price level: Lithuanian case. *Intellectual Economics*, 10(2):122 – 127.

Goodhart, C. and Hofmann, B. (2008). House prices, money, credit, and the macroeconomy. *Oxford Review of Economic Policy*, 24(1):180–205.

Gupta, R., Kabundi, A., and Miller, S. M. (2011). Forecasting the us real house price index: Structural and non-structural models with and without fundamentals. *Economic Modelling*, 28(4):2013 – 2021.

Hayes, A. (2020). Heteroskedasticity-investopedia.

Iacoviello, M. and Neri, S. (2010). Housing market spillovers: Evidence from an estimated dsge model. *American Economic Journal: Macroeconomics*, 2(2):125–64.

Nguyen, J. (2020). Is there a correlation between inflation and home prices?-investopedia.

Pan, H. and Wang, C. (2013). House prices, bank instability, and economic growth: Evidence from the threshold model. *Journal of Banking and Finance*, 37(5):1720 – 1732.

Plakandaras, B., Gupta, R., Gogas, P., and Papadimitriou, T. (2014). Forecasting the u.s. real house price index. *SSRN Electronic Journal*.

Priestley, M. B. and Tong, H. (1973). On the analysis of bivariate non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 35(2):153–166.

Renigier-Biłozor, M. and Wiśniewski, R. (01 Dec. 2012). The impact of macroeconomic factors on residential property price indices in europe. *Folia Oeconomica Stetinensia*, 12(2):103 – 125.

Shetye, A. (2019). Feature selection with sklearn and pandas.

Tripathi, S. (2019). Macroeconomic Determinants of Housing Prices: A Cross Country Level Analysis. MPRA Paper 98089, University Library of Munich, Germany.

Zach (2020). How to Create a Residual plot in python.

Appendices

A Data Source and Description

All data in this paper is from the Global Economy.

Economic growth: the rate of change of real GDP: Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2010 U.S. dollars. Type: Ratio.

Capital investment as percent of GDP: Gross capital formation (formerly gross domestic investment) consists of outlays on additions to the fixed assets of the economy plus net changes in the level of inventories. Type: Ratio.

Exchange rate: local currency units per U.S. dollar: Official exchange rate refers to the exchange rate determined by national authorities or to the rate determined in the legally sanctioned exchange market. It is calculated as an annual average based on monthly averages (local currency units relative to the U.S. dollar). Type: Ratio.

Unemployment rate: Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Type: Ratio.

Terms of trade: Net barter terms of trade index is calculated as the percentage ratio of the export unit value indexes to the import unit value indexes, measured relative to the base year 2000. Type: Ratio.

Government spending as percent of GDP: General government final consumption expenditure (formerly general government consumption) includes all government current expenditures for purchases of goods and services (including compensation of employees). Type: Ratio.

Economic growth forecast: Year-on-year percent changes in constant price GDP. The base year is country-specific. Type: Ratio.

External debt, percent of Gross National Income: Total external debt stocks to gross national income. Total external debt is debt owed to nonresidents repayable in currency, goods, or services. Type: Ratio.

Internet users, percent of population: Internet users are individuals who have used the Internet (from any location) in the last 3 months. Type: Ratio.

Oil reserves, billion barrels: Proved reserves of crude oil are the estimated quantities of all liquids defined as crude oil, which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from reservoirs under existing economic and operating conditions. Type: Ratio.

Crop production index: Crop production index shows agricultural production for each year relative to the base period 2004-2006. It includes all crops except fodder crops. Type: Ratio.

Banking system capital, percent of assets: Ratio of bank capital and reserves to total assets. Capital and reserves include funds contributed by owners, retained earnings, general and special reserves, provisions, and valuation adjustments. Capital includes tier 1 capital, which is a common feature in all countries' banking systems, and total regulatory capital, which includes several specified types of subordinated debt instruments that need not be repaid if the funds are required to maintain minimum capital levels. Total assets include all nonfinancial and financial assets. Type: Ratio.

Stock market capitalization as percent of GDP: Market capitalization is the share price times the number of shares outstanding for listed domestic companies. Investment funds, unit trusts, and companies whose only business goal is to hold shares of other listed companies are excluded. Data are end of year values. Type: Ratio.

High technology exports: High-technology exports are products with high R&D intensity, such as aerospace, computers, pharmaceuticals, scientific instruments, and electrical machinery. Data are in current U.S. dollars. Type: Ratio.

Percent urban population: Urban population refers to people living in urban areas as defined by national statistical offices. Type: Ratio.

Gross Domestic Product: GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. Type: Ratio.

Household consumption as percent of GDP: Household final consumption expenditure is the market value of all goods and services, including durable products (such as cars, washing machines, and home computers), purchased by households. Type: Ratio.

Household debt to GDP, in percent: The total outstanding debt of households to banks and other financial institutions as percent of GDP. Type: Ratio.

Inflation: percent change in the Consumer Price Index: Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used. Type: Ratio.

Imports of goods and services, billion USD: Imports of goods and services represent the value of all goods and other market services received from the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. Type: Ratio.

Income, profits, and capital gains taxes: Taxes on income, profits, and capital gains are levied on the actual or presumptive net income of individuals, on the profits of corporations and enterprises, and on capital gains, whether realized or not, on land, securities, and other assets. Type: Ratio.

New commercial vehicle sales, numbers per year : The indicator estimates the number of new commercial vehicle registrations and sales that took place within a country in a year. Type: Ratio.

Banking system z-scores: The index captures the probability of default of a country's banking system. Z-score compares the buffer of a country's banking system (capitalization and returns) with the volatility of those returns. Type: Ratio.

B One-Sample Kolmogorov-Smirnov Table

n\α	0.001	0.01	0.02	0.05	0.1	0.15	0.2
1		0.99500	0.99000	0.97500	0.95000	0.92500	0.90000
2	0.97764	0.92930	0.90000	0.84189	0.77639	0.72614	0.68377
3	0.92063	0.82900	0.78456	0.70760	0.63604	0.59582	0.56481
4	0.85046	0.73421	0.68887	0.62394	0.56522	0.52476	0.49265
5	0.78137	0.66855	0.62718	0.56327	0.50945	0.47439	0.44697
6	0.72479	0.61660	0.57741	0.51926	0.46799	0.43526	0.41035
7	0.67930	0.57580	0.53844	0.48343	0.43607	0.40497	0.38145
8	0.64098	0.54180	0.50654	0.45427	0.40962	0.38062	0.35828
9	0.60846	0.51330	0.47960	0.43001	0.38746	0.36006	0.33907
10	0.58042	0.48895	0.45662	0.40925	0.36866	0.34250	0.32257
11	0.55588	0.46770	0.43670	0.39122	0.35242	0.32734	0.30826
12	0.53422	0.44905	0.41918	0.37543	0.33815	0.31408	0.29573
13	0.51490	0.43246	0.40362	0.36143	0.32548	0.30233	0.28466
14	0.49753	0.41760	0.38970	0.34890	0.31417	0.29181	0.27477
15	0.48182	0.40420	0.37713	0.33760	0.30397	0.28233	0.26585
16	0.46750	0.39200	0.36571	0.32733	0.29471	0.27372	0.25774
17	0.45440	0.38085	0.35528	0.31796	0.28627	0.26587	0.25035
18	0.44234	0.37063	0.34569	0.30936	0.27851	0.25867	0.24356
19	0.43119	0.36116	0.33685	0.30142	0.27135	0.25202	0.23731
20	0.42085	0.35240	0.32866	0.29407	0.26473	0.24587	0.23152
25	0.37843	0.31656	0.30349	0.26404	0.23767	0.22074	0.20786
30	0.34672	0.28988	0.27704	0.24170	0.21756	0.20207	0.19029
35	0.32187	0.26898	0.25649	0.22424	0.20184	0.18748	0.17655
40	0.30169	0.25188	0.23993	0.21017	0.18939	0.17610	0.16601
45	0.28482	0.23780	0.22621	0.19842	0.17881	0.16626	0.15673
50	0.27051	0.22585	0.21460	0.18845	0.16982	0.15790	0.14886