Title:

A Time Series Analysis of Housing Price Index, CPI, Unemployment Rate, and Interest Rate: Case Study of French Metropolitan Area

Author Name:

1. Amin Ilyas
2. Hélène Do
3. Logba Elvire Ursule Vera Yapi
4. Nico Benedikt Horstmann
5. Pritam Ritu Raj

Date:

21 December 2022

Table of Content

1. Introduction

Following Covid-19 pandemic and its economic aftereffects, household primary income in France fell sharply in 2020. However, this drop was offset by a decrease in salary tax deductions and an increase in social benefits. Thanks to these measures' household purchasing power remained stable on average in 2020 compared to previous recessions (1973, 1993 and 2009). Households although significantly reduced their consumption (−6.5% in terms of value compared to 2019), given the restrictions that were put in place concerning transport, culture, restaurants and so on. As a result, household savings increased considerably, reaching 21.4% of disposable income in 2020, compared to 15.1% the year before.

The current growing trend in the real estate market is explained by on the one hand this increase in savings and on the other hand the fact that the lockdowns made people think about moving to the suburbs to access outdoor space and to get a larger property in order to get more comfortable while working remotely.

We are wondering if the current market trend is giving a hint of future a housing bubble crash? We may expect home prices to fall substantially, as central bank will raise interest rates to bring inflation under control. Housing represents a good asset when it comes to inflation, because the home's value rises with inflation rate and because it is a leveraged asset.

However, inflation is tightly related to both interest rate and unemployment. As W. Phillips stated, inflation and unemployment have a stable and inverse relationship. Which means that higher inflation is associated with lower unemployment and vice versa. Plus, interest rate is the primary tool used by central banks to manage inflation. Thus, they tend to move in the same direction. That is why we choose to realize a Multivariate Time Series Analysis based on three variables coupled with the House Price Index (HPI) of France, Consumer Price Index (CPI), Interest rate and finally unemployment rate.

2. Data Source

There are four time series used in this project taken from multiple sources. The first index data, HPI, were obtained from the INSEE database found in its website at www.insee.fr. Another index data, unemployment rate, were also obtained from the same database. The selection of INSEE was due to the fact that all information published are produced by the official statistical authority (data and studies) and also is available free of charge and in compliance with statistical confidentiality, in publications, on computer storage media or on the websites of INSEE and ministerial statistical offices.

The next index data, CPI, were obtained from the FRED Economic database found in its website at www.fred.stlouisfed.org. FRED is an online database consisting of hundreds of thousands of economic data time series from scores of national, international, public, and private sources. FRED, created and maintained by the Research Department at the Federal Reserve Bank of St. Louis, goes far beyond simply providing data: It combines data with a powerful mix of tools that help the user understand, interact with, display, and disseminate the data. In essence, FRED helps users tell their data stories.

The last index data, interest rate, were provided from Webstat by Banque de France Eurosytème database found at www.webstat.banque-france.fr. The data are based on 10-year Benchmark Bond rate computed annually. The original frequency of the data is daily. This portal contains about 35 000 series of Banque de France and partner international organizations (of which ECB, Eurostat). The database allows users to search for statistical series, to compare and display them in different views: charts, tables and individually.

3. Selection of Parameters and Sampling

3.1. House Price Index (HPI)

The House Price Index (HPI) is a time series data on average annual base 100 on 2015. Housing prices, and property prices in general, may evolve differently from the prices of other consumer goods and all economic actors need to be aware of these differences. The calculation of house price indices responds to a national and European imperative. They aim at enriching statistical information on the housing market in France. Data have been published since early 2013 with a history dating back to 2000. The HPI is composed of two indices: the old housing price index and the new housing price index. This is a broad measure of the movement of single-family property prices in France metropolitan area. This sample is chosen because aside from serving as an indicator of house price trends, it also functions as an analytical tool for estimating changes in the rates of mortgage defaults, prepayments, and housing affordability. The HPI is one of many economic indicators that investors use to keep a pulse on broader economic trends and potential shifts in the stock market.

3.2. Consumer Price Index (CPI)

The original consumer price index (CPI) obtained is a time series data on average annual base 100 on 2015 and not seasonally adjusted. The CPI measures the overall change in consumer prices based on a representative basket of goods and services over time. Housing rents are used to estimate the change in shelter costs including owner-occupied housing that account for nearly a third of the CPI.

3.3. Unemployment Rate

3.4. Interest Rates

3.5. Sampling Designs

In time series analysis, there are two main sampling schemes to be considered, deterministic and random. In deterministic sampling, the sampling points are chosen according to a deterministic rule, such as periodic sampling. In random sampling, the sampling points are chosen according to a randomized rule. This project follows the random sampling scheme. There are several reasons for considering random sampling schemes: they are easier to analyze, they provide bounds to the performance of the best deterministic designs, and in certain cases they are imposed by the very nature of the information gathering process, such as when an observation can be taken only at a time a certain event occurs, or, when there are imperfections in deterministic sampling schemes. In short, each of the four samples used in this project are with chosen independently from each other for their sampling points. The sampling design for this project is summarized by the following table.

Table 1. Sampling designs of the project

No.	Time Series Variable	Sampling points	Original Model	Model Used in Project	Sample size
1.	HPI	1996 – 2022	Quarterly	Quarterly	107
2.	CPI	1996 – 2022	Monthly	Quarterly	107
3.	Unemployment rate	1996 - 2022	Quarterly	Quarterly	107
4.	Interest rate	1996 – 2022	Daily	Quarterly	107

Moreover, in this project, the data comprises of both time-series and some cross-sectional elements as the time series data are taken from multiple subjects, i.e., HPI, CPI, unemployment rate and interest rate. Therefore, the dataset used in this project are considered as panel of data or longitudinal data. The use of panel data allows the project to embody information across both time and space. More importantly, the panel dataset used in this project keeps the same entities and measures some quantities about them over time. According to the design, the panels are balanced which means that the project has the same number of time series observations for each cross-sectional unit. Lastly, the project has a time-fixed effects model rather than an entity-fixed effects model.

5. Research Framework

In this project, we focus on a forecasting technique called time series analysis. A time series data set is a collection of historical observations recorded over time. So, time series data are always presented in chronological order. For a time series, it is important to know the time period that corresponds to each observation.

6. Statistical Framework

In this project, a univariate time series approach is used. The term "univariate time series" refers to a time series that consists of single (scalar) observations recorded sequentially over equal time increments. Although a univariate time series data set is usually given as a single column of numbers, time is in fact an implicit variable in the time series. If the data are equi-paced, the time variable, or index, does not need to be explicitly given. The time variable may sometimes be explicitly used for plotting the series. However, it is not used in the time series model itself.

To make forecasts, time series analysis examines historical data for patterns that can be modeled, or explained, with mathematical equations. Then, forecasts are developed using these equations to project the historical patterns into the future. An important assumption of forecasts using time series data is that the future will replicate the past historical patterns. This assumption is not always true. Therefore, a constant monitoring of the environment for events that may disrupt the pattern is essential. Various statistical techniques for time series are used in this project: stationarity test, autoregressive process, random walk process, auto correlation, partial autocorrelation, autoregressive and dynamic autoregressive models.

7. Python Design

How is your Python program designed? In our python Code we start at first importing several packages that are needed for the rest of the code to work. We import some well-known packages such as numpy and pandas but also some functions from subpackages of the package statsmodels that are especially used in time series analysis (AR and VAR functions). After we finished our codings and therefor knew that it worked we surpressed the warnings in order make the console output more readable. In the next section we processed our data by reading it into pandas DataFrames. There we transformed the interest data from monthly into quarterly data and dropped the last row of it in order to have the same sample size.

Then for every dataset we created a Year-Quarter column named "Period". Further, we merged all timeseries into one DataFrame called Total of whom we also changed the column names. Then our python code generates a copy of the this Dataframes and then convertes the time series in more stationary time series by calculating their returns or their first-order differennces. Finally, we dropped the first row of the newly created DataFrame because of the Nas created. After this processing we visualized both the untransformed as well as the transformed time-series. For the Consumer Price Index (CPI) we found out that taking the returns yields a in our case quaterly Inflation that is a very stationary time series.

The visualization is then followed by the Unit Root Test that we used to check statistically the stationarity of the respective time series. We concluded that our transformation was necessary and that after the transformation all time series were stationary enough (We rejected the null hypothesis of nonstationarity at least at the 5% level) such that we could include them in further statistical models. In the next section of the code we plotted the Autocorrelation and Partial Autocorrelation across lags of the respective time series. We plotted both for each time series and specified the maximum number of lags for each plot as twenty. Then we dedicated our Code to the topic of estimation and prediction and implemented a Vector Autoregression (VAR) model. Therefore, we have split our Data in a training and test dataset. For the test dataset we used the last 40 quarters (10 years) and for the training dataset the remaining 66 quarters. Then we have run a for loop to ouput the Grid of Orders. We used that to find the order where the model best predicts the data and does not overfit. We then used the model to predict the values of the Housing Price Index in the test dataset and transformed them to the right scale again. Afterwards we plotted the prediction and the actual values to be able to see the models fit for each time series we did the same. In the last section of the code we standardized all untransformed time series and visualized them together in one plot

8. Results and Discussion

Processing the data for the first time reveals that they are non-stationary. This observation was done by plotting the line graphs for each of the time series used in this project. The results are shown by the following figure.

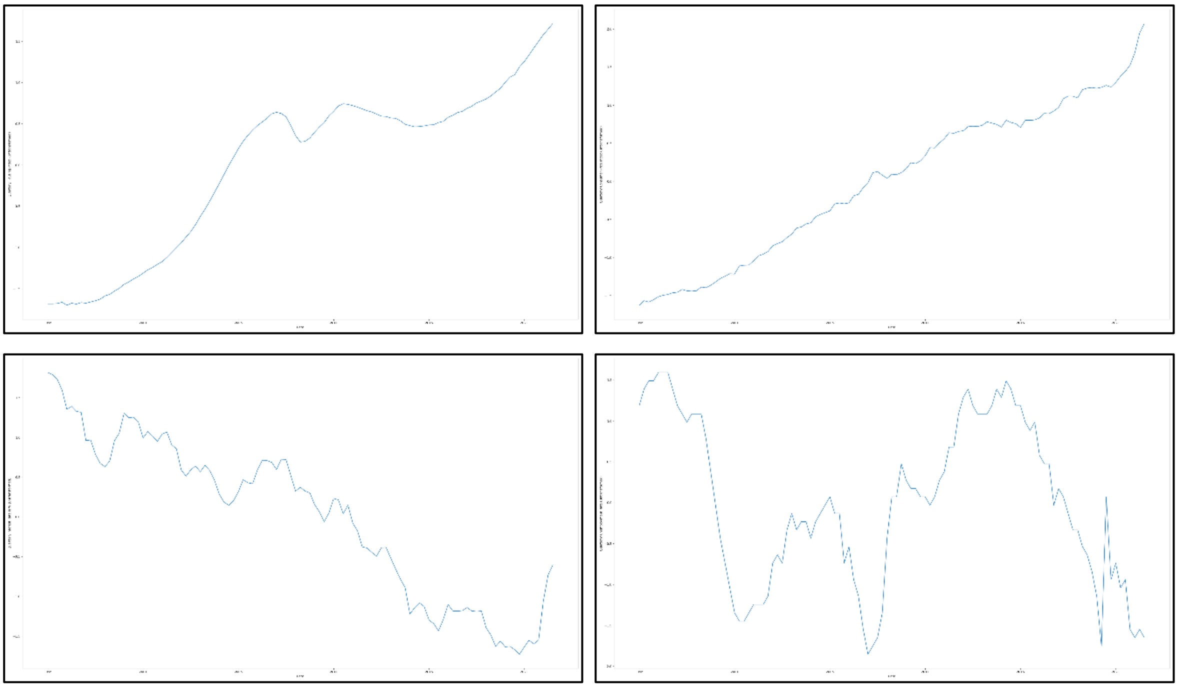

Fig. 1. Untransformed data graphs of HPI, CPI, unemployment rate, and interest rate.

Time series analysis would require for the data to be strong in stationary. Theoretically, strict stationarity means joint probability distribution of the process remains the same throughout time. Thus, the joint probability density function (pdf) of any set of T observations $Y_{t1}, Y_{t2}, \ldots, Y_{tT}$ must be the same as the pdf of Y_{t1+k}, Y_{t2+k} , …, Y_{tT+k} for any t, k (process shifted ahead k periods). The joint pdf is time-invariant. On the other hand, a time-series is said to be weakly stationary if the mean ($E(Y_t)$), variance ($\sigma^2(Y_t)$), and covariance ($cov(Y_t, Y_{t-k})$) are independent of t. The notion of weak stationarity is more commonly used in practice.

The above plotted graphs for the different variables namely, HPI, CPI, unemployment rate and interest rate are non-stationary i.e., their statistical properties like mean, variance, covariance varies with time, or these stats properties are the function of time. In other words, stationarity in time series also means series without any non-stationary properties such as a trend. Stationary series is easier for statistical models to predict effectively and precisely. So, for the same reason the dataset was then transformed so that they exhibit the stationary behavior. Firstly, to confirm the non-stationarities of the data set, unit root tests are performed under each time series used. One of the most popular tests is the Dickey–Fuller test. The Dickey–Fuller test takes the following form:

$$\Delta yt = (\rho - 1)y_{t-1} + \epsilon t = \delta y_{t-1} + \epsilon_t$$

where: $\Delta y_t = y_t - y_{t-1}$ and $\delta \equiv \rho - 1$.

This model can be estimated by linear regression and testing for a unit root (non-stationarity) is equivalent to testing $\delta = 0$. If $\delta = 0$, then this implies that $\rho = 1$ and thus $y_t = y_{t-1} + \epsilon_t$, which is a non-stationary random walk process. In other words, if we cannot reject the null hypothesis that δ is not different from 0, then the time series is non-stationary. In this project, the analysis time series are found to be non-stationary, so then we took the first difference $\Delta y_t = y_t - y_{t-1}$ to transform it to a stationary time series before doing further statistical analysis. Taking the first difference helps remove the properties that make a time-series non-stationary (e.g. trend). The following is the dataset after transformation. After transforming the variables, the trends are removed from the above plots which will help us in predicting the data effectively.

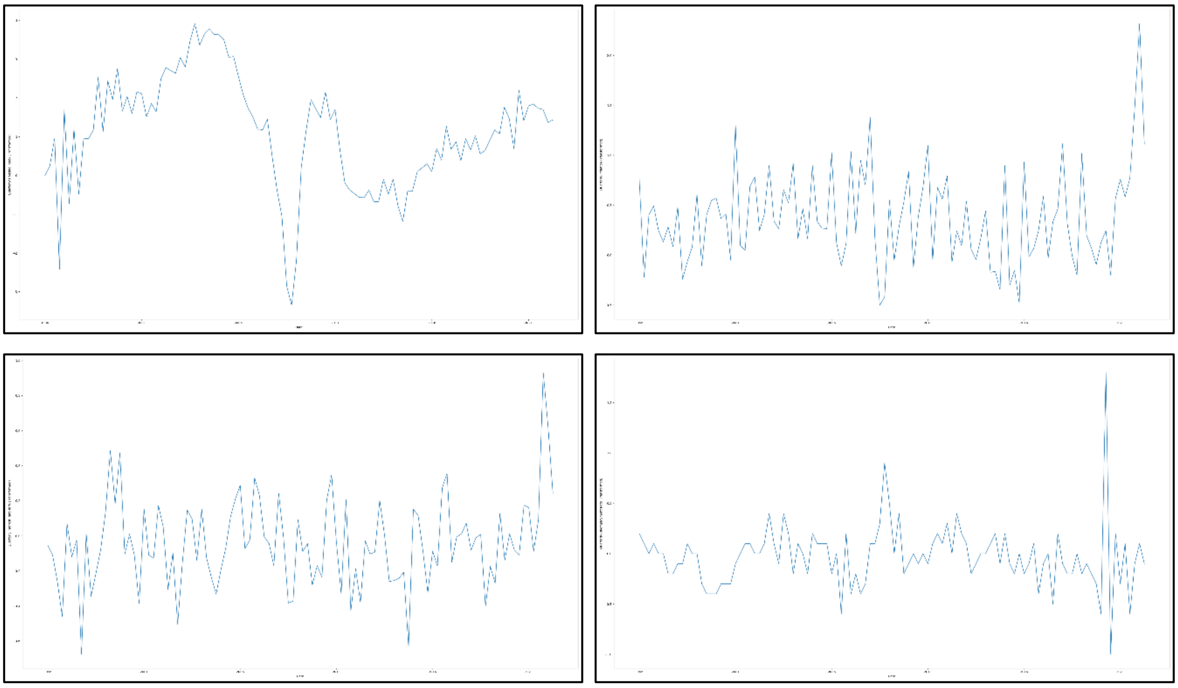

Fig. 2. Transformed data graphs of HPI, CPI, unemployment rate, and interest rate.

Autocorrelation is the correlation between two observations at different points in a time series. For example, values that are separated by an interval might have a strong positive or negative correlation. When these correlations are present, they indicate that past values influence the current value. Analysts use the autocorrelation and partial autocorrelation functions to understand the properties of time series data, fit the appropriate models, and make forecasts. Autocorrelation coefficient for lag k:

$$\rho(k) = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-k})}} = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} = \frac{\gamma(k)}{\gamma(0)}$$

By definition, $-1 < \rho(k) < 1$ and $\rho(0) = 1$ Autocorrelation function (ACF): sequence of autocorrelations $\rho(k)$ for all k = 0, 1, 2, ... Autocorrelogram is the graphical representation of the empirical ACF.

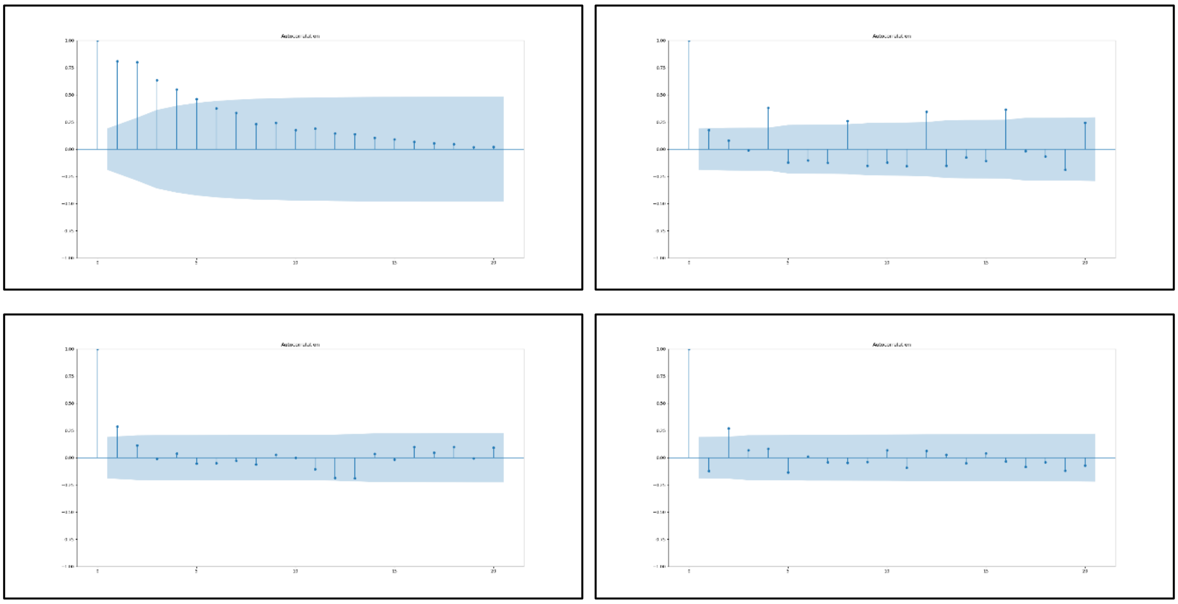

Fig. 3 ACF for HPI, CPI, unemployment rate, and interest rate.

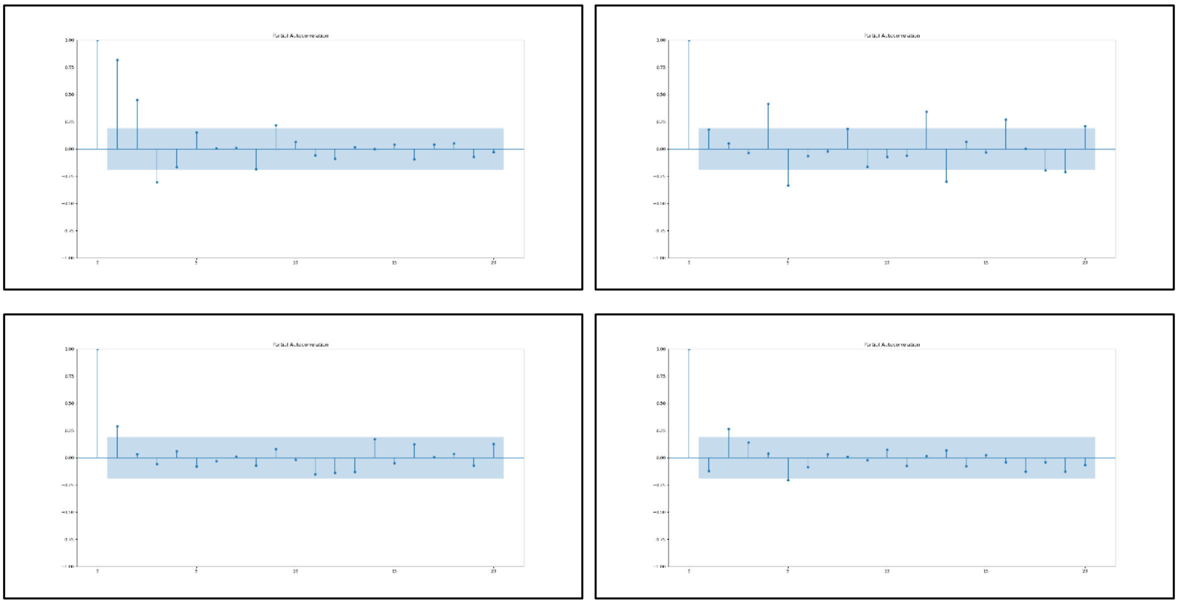

Fig. 4 PACF for HPI, CPI, unemployment rate, and interest rate.

It can be shown that the autocorrelation function (ACF) of $AR(\rho)$ are not 0 for finite k and will only decay to 0 as k $\rightarrow \infty$. It can be easily seen that the partial autocorrelation function $(PACF)$ of $AR(\rho)$ are not 0 for k = 0, 1, 2, ..., p and then 0 for all $k > p$. Autocorrelation function (ACF) and Partial Autocorrelation Function (PACF, also called Partial ACF) are important functions in analyzing a time series. They generally produce plots that are very important in finding the values p, q and r for Autoregressive (AR) and Moving Average (MA) models. An ACF measures and plots the average correlation between data points in time series and previous values of the series measured for different lag lengths. A PACF is similar to an ACF except that each partial correlation controls for any correlation between observations of a shorter lag length.

Lastly, we use our time series to do forecasting. The forecasting occurs by making scientific predictions based on historical time stamped data. In this project, it involves building models through historical analysis and using them to make observations and drive future strategic decision-making. The results are as follows.

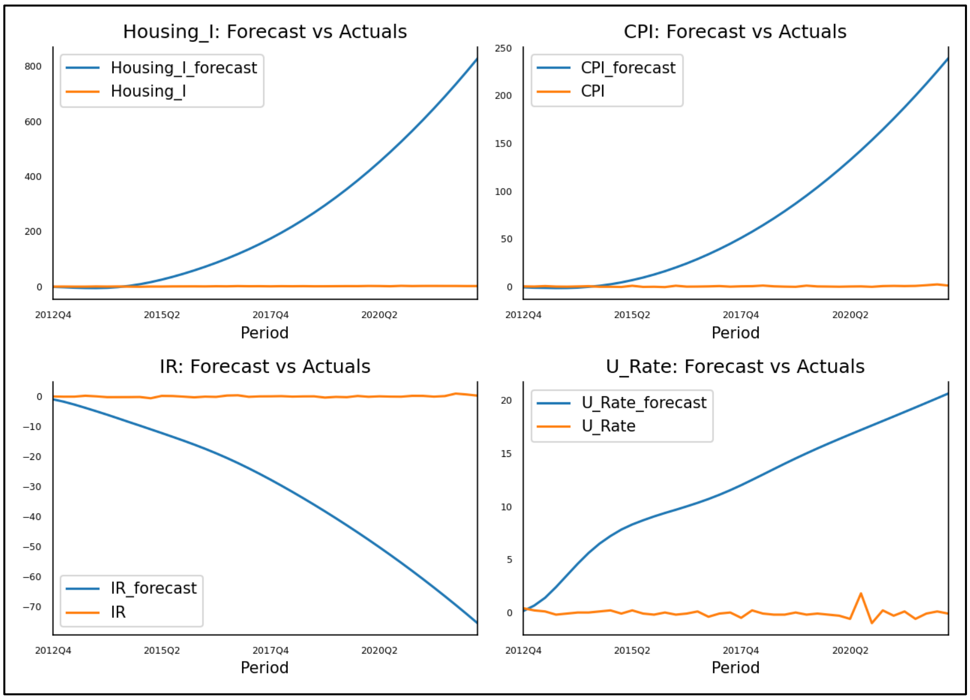

Fig. 5 Forecasting HPI, CPI, unemployment rate, and interest rate.

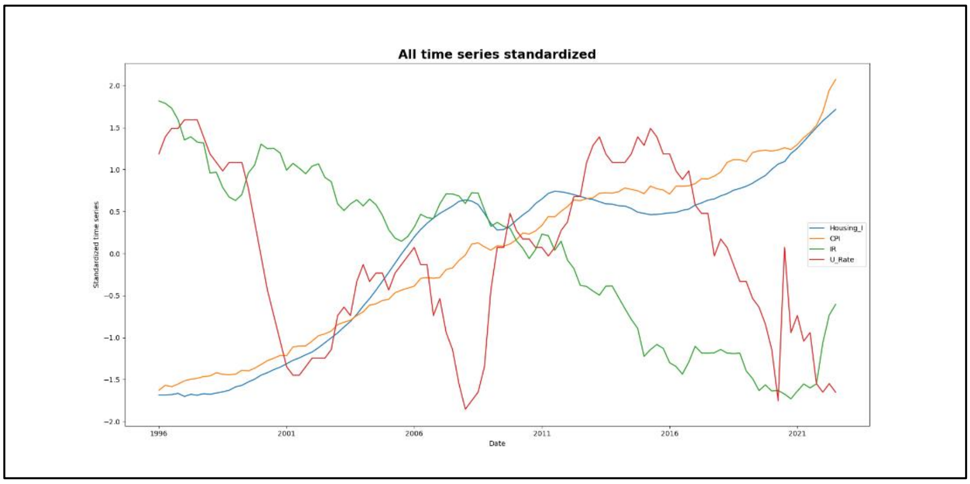

Fig. 6. All time series standardized.

Standardization is an operation that used to normalize the four time series. Let there be a time series (a list of data points). Let V and W be the mean and standard deviation of the values in the time series, respectively. The standardization of a time series is obtained by replacing each data point Z by (Z-V)/ (W).

9. Conclusion

During the crisis period when economy moves into a recession phase then central banks adopts accommodative stance and by reducing the interest rate and injecting liquidity in the economy it tries to promote the growth. As a result of this housing prices goes up again due to cheaply available liquidity and reduced interest rate.

10. List of References

BANQUE. "Project." Webstat.banque-France.fr, 2022, webstat.banque-france.fr/en/#/quickview/SERIES_KEY/ECOFI.INR.FR.FIGB_PA._Z.D/node_EQU_5384916_AND_DATASET_EQU_ECOFI_AND_periodSortOrder_EQU_DESC. Accessed 21 Dec. 2022.

INSEE. "Titre | Insee." Www.insee.fr, 2022, www.insee.fr/en/statistiques/series/105071770?. Accessed 21 Dec. 2022.

---. "Titre | Insee." Www.insee.fr, 2022, www.insee.fr/en/statistiques/serie/001688526. Accessed 21 Dec. 2022.

Organization for Economic Co-operation and Development. "Consumer Price Index of All Items in France." FRED, Federal Reserve Bank of St. Louis, 1 Jan. 1955, fred.stlouisfed.org/series/FRACPIALLMINMEI#0.

11. Appendix

```
Python 3.9.12 (main, Apr 4 2022, 05:22:27) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 8.2.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/horst/Lecture 1/Group Project 2.py')
Untransformed series (HI, CPI, IR, U_Rate):
DF stat: 6.910845363021661 DF P-value: 1.0
DF stat: 7.914185404364599 DF P-value: 1.0
DF stat: -2.489421156183509 DF P-value: 0.012364156047629297
DF stat: -1.0235570245632326 DF P-value: 0.2787046567575674
Transformed series (HI, CPI, IR, U_Rate):
DF stat: -2.494694107203914 DF P-value: 0.01218347545001427
DF stat: -5.820332926414889 DF P-value: 2.2608077977840842e-08
DF stat: -7.346901667878562 DF P-value: 1.0031808717869325e-11
DF stat: -11.452104306196478 DF P-value: 8.542455234018851e-21
Order = 1
AIC: -7.749776566058559
BIC: -7.080734329167593

Order = 2
```

```
AIC: -7.795252750889
BIC: -6.580881016499186

Order = 3
AIC: -7.978288312872632
BIC: -6.209351713311366

Order = 4
AIC: -8.787967983454447
BIC: -6.454981883727572

Order = 5
AIC: -8.887669644725467
BIC: -5.9808925202900864

Order = 6
AIC: -8.677644145680166
BIC: -5.187069875309999

Order = 7
AIC: -8.727258397178925
BIC: -4.642608507465985

Order = 8
AIC: -9.269161609398427
BIC: -4.579877516430717

Order = 9
AIC: -9.729532017899505
BIC: -4.424767322469444

Order = 10
AIC: -10.204770004746294
BIC: -4.273382910450499

Summary of Regression Results
==================================
Model: VAR
Method: OLS
Date: Wed, 21, Dec, 2022
Time: 19:22:13
--------------------------------------------------------------------
No. of Equations: 4.00000 BIC: -5.98089
Nobs: 61.0000 HQIC: -7.74848
Log likelihood: 8.85292 FPE: 0.000155260
AIC: -8.88767 Det(Omega_mle): 4.75471e-05
--------------------------------------------------------------------
Results for equation Housing_I
===============================================================================
coefficient std. error t-stat prob
-------------------------------------------------------------------------------
const 1.035358 0.251752 4.113 0.000
L1.Housing_I 0.607150 0.130206 4.663 0.000
L1.CPI -0.394985 0.268226 -1.473 0.141
L1.IR -0.234873 0.457749 -0.513 0.608
L1.U_Rate 0.495278 0.442016 1.120 0.263
L2.Housing_I 0.627290 0.124197 5.051 0.000
L2.CPI -0.857509 0.235125 -3.647 0.000
L2.IR -0.009267 0.413293 -0.022 0.982
```

```
L2.U_Rate 0.482397 0.414954 1.163 0.245
L3.Housing_I -0.174365 0.138287 -1.261 0.207
L3.CPI -0.776433 0.246490 -3.150 0.002
L3.IR -0.290357 0.431489 -0.673 0.501
L3.U_Rate 0.679053 0.397731 1.707 0.088
L4.Housing_I 0.001533 0.123860 0.012 0.990
L4.CPI -0.616325 0.291365 -2.115 0.034
L4.IR 0.105912 0.425350 0.249 0.803
L4.U_Rate 0.355034 0.415346 0.855 0.393
L5.Housing_I 0.027506 0.116521 0.236 0.813
L5.CPI -0.238631 0.329771 -0.724 0.469
L5.IR 1.224989 0.418364 2.928 0.003
L5.U_Rate -0.480407 0.400494 -1.200 0.230
===============================================================================

Results for equation CPI
===============================================================================
coefficient std. error t-stat prob
-------------------------------------------------------------------------------
const 0.280236 0.176442 1.588 0.112
L1.Housing_I 0.183665 0.091256 2.013 0.044
L1.CPI 0.053855 0.187988 0.286 0.775
L1.IR -0.100606 0.320816 -0.314 0.754
L1.U_Rate 0.039859 0.309789 0.129 0.898
L2.Housing_I -0.046224 0.087044 -0.531 0.595
L2.CPI -0.125817 0.164789 -0.764 0.445
L2.IR 0.127153 0.289659 0.439 0.661
L2.U_Rate -0.157556 0.290823 -0.542 0.588
L3.Housing_I -0.098395 0.096919 -1.015 0.310
L3.CPI -0.098011 0.172754 -0.567 0.570
L3.IR 0.330926 0.302412 1.094 0.274
L3.U_Rate 0.273942 0.278752 0.983 0.326
L4.Housing_I 0.038128 0.086808 0.439 0.661
L4.CPI 0.498482 0.204205 2.441 0.015
L4.IR 0.115277 0.298109 0.387 0.699
L4.U_Rate -0.180915 0.291098 -0.621 0.534
L5.Housing_I -0.013093 0.081664 -0.160 0.873
L5.CPI -0.218560 0.231122 -0.946 0.344
L5.IR 0.047553 0.293213 0.162 0.871
L5.U_Rate -0.038032 0.280689 -0.135 0.892
===============================================================================

Results for equation IR
===============================================================================
coefficient std. error t-stat prob
-------------------------------------------------------------------------------
const 0.203866 0.097993 2.080 0.037
L1.Housing_I -0.073073 0.050682 -1.442 0.149
L1.CPI -0.160779 0.104405 -1.540 0.124
L1.IR 0.288848 0.178176 1.621 0.105
L1.U_Rate -0.013184 0.172052 -0.077 0.939
L2.Housing_I 0.021813 0.048343 0.451 0.652
L2.CPI -0.139725 0.091521 -1.527 0.127
L2.IR -0.133334 0.160872 -0.829 0.407
L2.U_Rate 0.109565 0.161518 0.678 0.498
L3.Housing_I 0.021082 0.053827 0.392 0.695
L3.CPI -0.249740 0.095945 -2.603 0.009
L3.IR 0.147016 0.167955 0.875 0.381
L3.U_Rate 0.008981 0.154814 0.058 0.954
```

```
L4.Housing_I 0.166887 0.048212 3.462 0.001
L4.CPI -0.141173 0.113412 -1.245 0.213
L4.IR -0.267358 0.165565 -1.615 0.106
L4.U_Rate -0.055830 0.161671 -0.345 0.730
L5.Housing_I -0.069434 0.045355 -1.531 0.126
L5.CPI -0.219748 0.128362 -1.712 0.087
L5.IR 0.140866 0.162846 0.865 0.387
L5.U_Rate 0.030245 0.155890 0.194 0.846
===============================================================================

Results for equation U_Rate
===============================================================================
coefficient std. error t-stat prob
-------------------------------------------------------------------------------
const -0.273492 0.088445 -3.092 0.002
L1.Housing_I -0.007804 0.045744 -0.171 0.865
L1.CPI 0.001383 0.094232 0.015 0.988
L1.IR -0.239558 0.160815 -1.490 0.136
L1.U_Rate -0.080860 0.155288 -0.521 0.603
L2.Housing_I -0.003842 0.043633 -0.088 0.930
L2.CPI 0.070869 0.082604 0.858 0.391
L2.IR 0.020262 0.145197 0.140 0.889
L2.U_Rate 0.274035 0.145781 1.880 0.060
L3.Housing_I 0.023415 0.048583 0.482 0.630
L3.CPI 0.329051 0.086596 3.800 0.000
L3.IR -0.274821 0.151590 -1.813 0.070
L3.U_Rate 0.327618 0.139730 2.345 0.019
L4.Housing_I -0.059018 0.043514 -1.356 0.175
L4.CPI 0.240975 0.102362 2.354 0.019
L4.IR 0.027086 0.149433 0.181 0.856
L4.U_Rate -0.134095 0.145919 -0.919 0.358
L5.Housing_I 0.002359 0.040936 0.058 0.954
L5.CPI 0.151291 0.115855 1.306 0.192
L5.IR -0.075116 0.146979 -0.511 0.609
L5.U_Rate -0.151092 0.140701 -1.074 0.283
===============================================================================

Correlation matrix of residuals
Housing_I CPI IR U_Rate
Housing_I 1.000000 0.059418 -0.052663 -0.180325
CPI 0.059418 1.000000 0.543692 -0.043041
IR -0.052663 0.543692 1.000000 0.050811
U_Rate -0.180325 -0.043041 0.050811 1.000000



5

In [2]:
```