Time Series Modeling and Prediction of All-Transaction House Price Index (HPI) in Connecticut

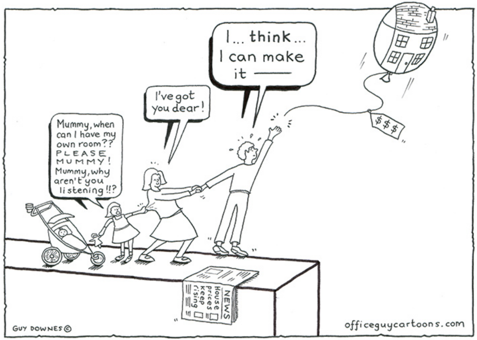
Jingqian Xu

November 23, 2022

Outlines

- Background Information
- Description of Data
- Description of Objective
- Data Analysis
  - Exploratory Data Analysis (EDA)
  - Model Construction
  - Estimation of Model
  - Diagnostic Checking
  - Prediction
- Limitations of Analysis
- Summary
- Reference

Background Information


I... think... I can make it ——
I've got you dear!
Mummy, when can I have my own room?? PLEASE MUMMY! Mummy, why aren't you listening!!?
$$$
NEWS House prices keep rising
GUY DOWNES ©
officeguycartoons.com

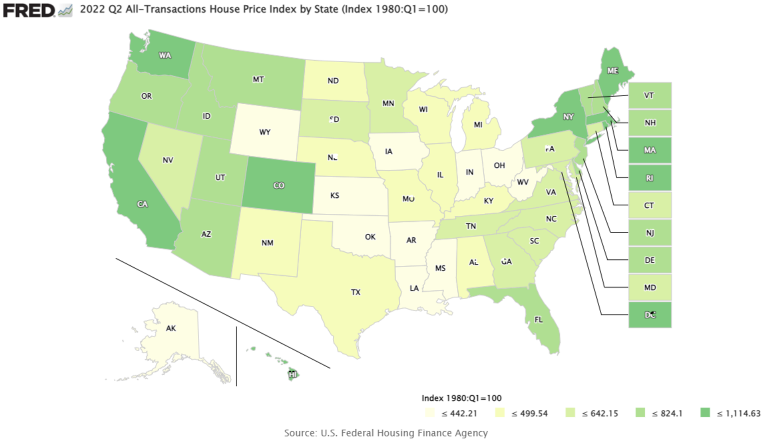
The rise in home prices year-over-year for November 2021 is **17.5%** as reported by the FHFA on Jan.25, 2022.

In September 2022, home prices in Connecticut were up **10.0%** compared to last year, selling for a median price of $365,400.

Background Information

- Housing Price Index (HPI) from The U.S. Federal Housing Finance Agency (FHFA) is the reliable house price indexes that measure changes in single-family home values based on data from all 50 states and over 400 American cities that extend back to the mid-1970s.
- HPI incorporates tens of millions of home sales, offers insights about house price fluctuations and serves as a timely, accurate indicator of house price trend at various geographic levels.
- Real house price at the State level are driven by fundamentals such as real per capital disposable income, net borrowing cost and State level population growth [**1**].

Background Information


FRED 2022 Q2 All-Transactions House Price Index by State (Index 1980:Q1=100)

Source: U.S. Federal Housing Finance Agency

Figure: 2022 Q2 All-Transactions House Price Index by State[2]

Although HPI in Connecticut increases by 7 times since 1980, Connecticut is one of states having less HPI increment in the New England.

Description of Data

- The HPI data resource is U.S. Federal Housing Finance Agency (FHFA®)[2].
- The explanatory economic indicators (per captia personal income in Connecticut[3], population growth rate in Connecticut and 10-Year Real Interest Rate in the U.S.[4]) we may use comes from the Federal Reserve Economic Data(FRED®).
- The file CTSTHPI.csv contains quarterly time series data of HPI and the date they are collected. HPI is not seasonally adjusted, and the first record of HPI is 100 in the first quarter of 1980.
- Total 190 HPI observations are collected from January 1st 1975 to April 1st 2022 without missing value.

Description of Objective

1. Conduct Exploratory Data Analysis (EDA) of raw HPI data and find the proper type of time series models to fit the pattern of HPI well and explain the variability of HPI well.
    - The first type of model are those have X_t but no other explanatory variables, such as models from ARIMA family.
    - The second type of model is regression model with exogenous explainable variables, such as ARIMAX model.
2. Conduct model selection, estimate the model parameters, check the model adequacy and make prediction.
3. Future HPI prediction, interpretation of prediction result and draw some insightful conclusion.

Exploratory Data Analysis (EDA)

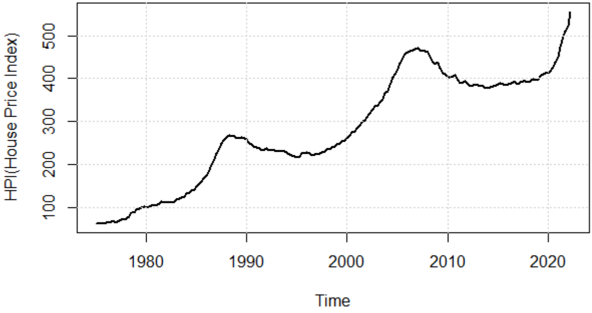


Figure: All-Transactions House Price Index for Connecticut from 1975 to 2022

We find the seasonality of HPI, and it repeats its pattern around every 20 years. HPI increases in the first decade and drop in another half. Also, the HPI in Connecticut keeps increasing from early 1970s to 2020.

Exploratory Data Analysis (EDA)

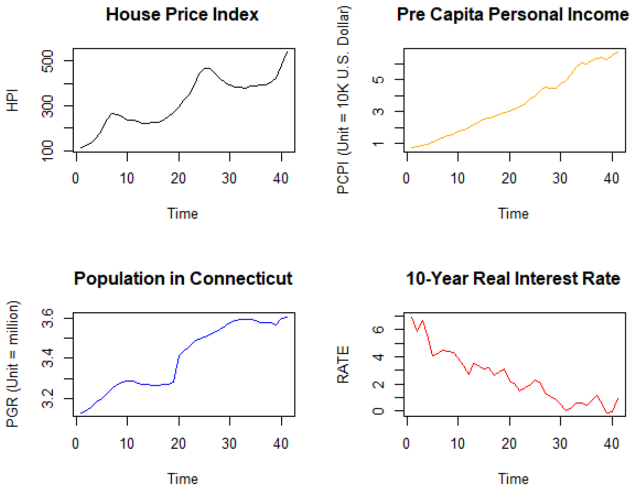


Figure: Time Series for both Response and Exogenous Variables

Exploratory Data Analysis (EDA)

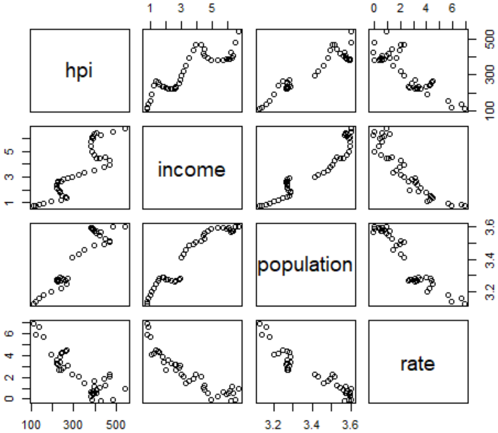


Figure: Covariance Matrix for Response and Exogenous Variables

Data Pre-processing

- Checking the missing value of raw data and impute missing values.
- The time series of economic and demographic data having various time interval. The per capital disposable income and population in the Connecticut is yearly collected from 1975.1.1 to 2021.1.1 and from 1900.12.1 to 2021.12.1 separately. We will unify the time interval for modeling purpose.
- 10-year interest rate is monthly collected from 1982.1.1 to 2022.10.1. We will use rolling average method to transform the monthly data into annual data. Similarly, we will transform quarterly HPI data into annual data.

Model Construction

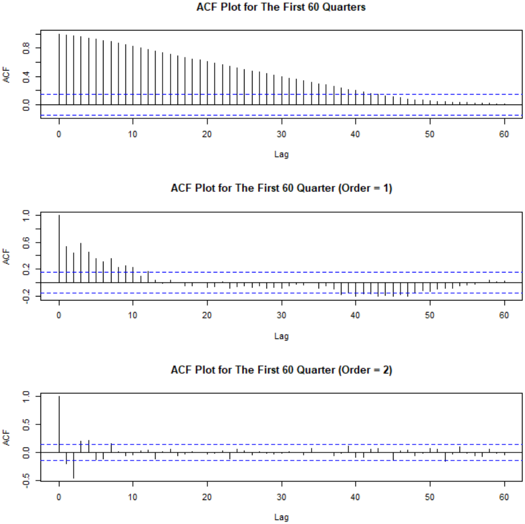


Figure: ACF Plot Under different Order

Model Construction

- Time series X_t is said to follow an integrated autoregressive-moving average model ($ARIMA$) if the d th difference $W_t = \nabla^d X_t$ is a stationary $ARMA$ process. If W_t is $ARMA(p, q)$, we say that X_t is $ARIMA(p, d, q)$.
- The differencing operation on an arbitrary time series X_t is defined as:

$$\nabla X_t = X_t - X_{t-1}$$

$$\nabla^2 X_t = \nabla(\nabla X_t) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$

- The $ARIMA(p, d, q)$ is expressed as:

$$\nabla^d X_t = \theta_0 + \sum_{i=1}^{p} \phi_i \nabla^d X_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-1}$$

$$a_t \sim i.i.d.N(0, \sigma^2), \sigma^2 < \infty$$

.

Model Construction

- The time series is not stationary since the noticeable increment trend is detected and ACF slowly decays in the following raw data ACF plot. So, we transform the time series into stationary time series by using the first order difference and second order difference.
- The second order difference is helpful and ACFs after lag 5 basically lay inside the confidence limit. So, we consider the d in our model as 2. Also, we may consider MA (5) model based on the ACF plot. Also, we draw the time series plot and PACF plot to identify p and q. Based on the plots below, we may consider AR (2) model.

Model Fitting

- Above procedure is rough. So, we consider a more data driven approach and select best model based on the best in-sample model selection criteria (e.g. BIC) So, we find $ARIMA(2, 2, 2)$ model has the lowest BIC equal to 5.5832. It's written as:

$$\nabla^2 X_t = \theta_0 + \sum_{i=1}^{2} \phi_i \nabla^2 X_{t-i} + a_t - \sum_{i=1}^{2} \theta_i a_{t-1}$$

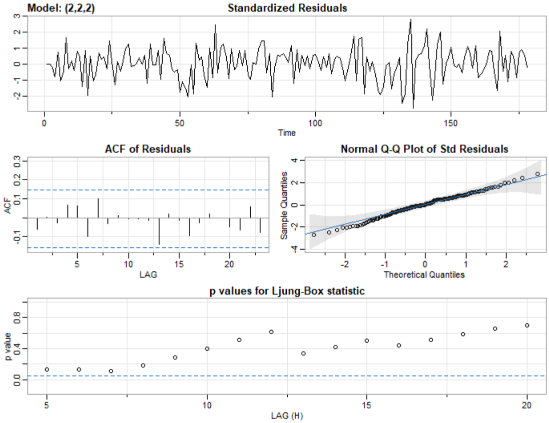
  .

- We use the conditional least squared estimation to get the fitted model:

$$\nabla^2 X_t = -0.4802\nabla^2 X_{t-1} - 0.6436\nabla^2 X_{t-2} + a_t + 1.6236 a_{t-1} - 0.6792 a_{t-2}$$

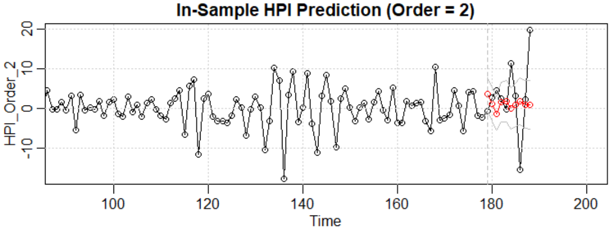
- Is the model adequate?

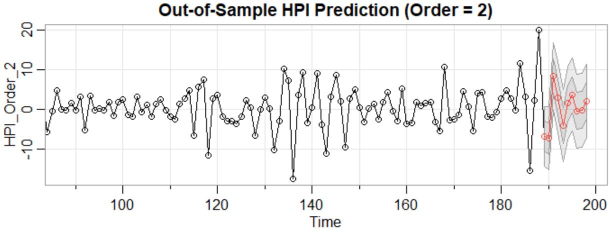
Diagnostic Checking



Since the ACFs of residual are within the confidence limit, normality of residual is valid and all p-values of Ljung-Box statistic are larger than critical value, the *ARIMA*(2, 2, 2) model is adequate.

Prediction

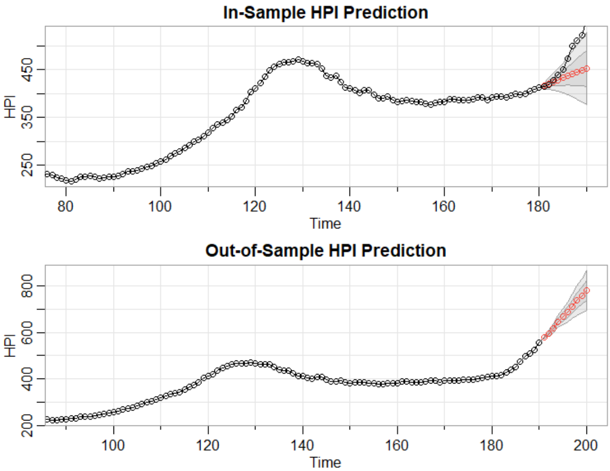




The time interval of in-sample prediction ranges from the Q4 of 2019 to the Q2 of 2022. The out-of-sample prediction is 10 steps ahead namely from Q3 2022 to Q1 2025.

Prediction

Now, we convert double difference data into actual number.



Is our prediction result reliable?

ARIMAX Model Construction

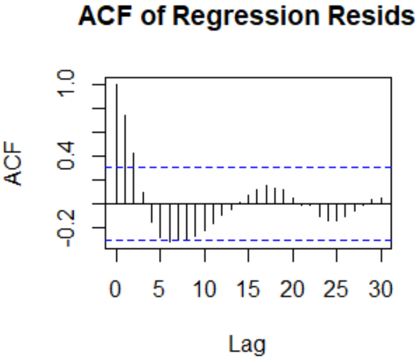
ARIMAX models refer to *ARMA* models with exogenous predictors. Suppose the data consist of an observed time series Y_t along with r deterministic, exogenous predictors $Z_{t,j}, j = 1, ..., r$. We have following *ARIMAX* model.

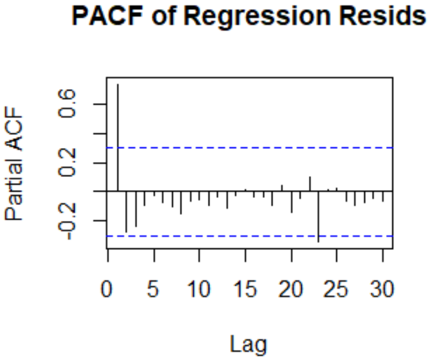
$$Y_t = \beta_0 + \theta_1 Z_{t,1} + \theta_r Z_{t,r} + X_t$$

$$\phi(B) X_t = \theta(B) a_t$$

Where $a_t(0, \sigma^2)$, only Y_t and $Z_{t,j}, j = 1, 2, 3$ are observed.
Since the con-linearity is observed from covariance matrix for response and exogenous variables. We centralized our response and residuals.

ARIMAX Model Construction


ACF of Regression Resids


PACF of Regression Resids

The sample PACF plot suggests that an AR(1) or MA(3) could be reasonable fit to the residuals X_t. Also, we choose order d = 2 as we do before. However, the selection method is rough and we get $ARIMAX(1, 2, 2)$ based on BIC.

Joint Estimation of ARIMAX model

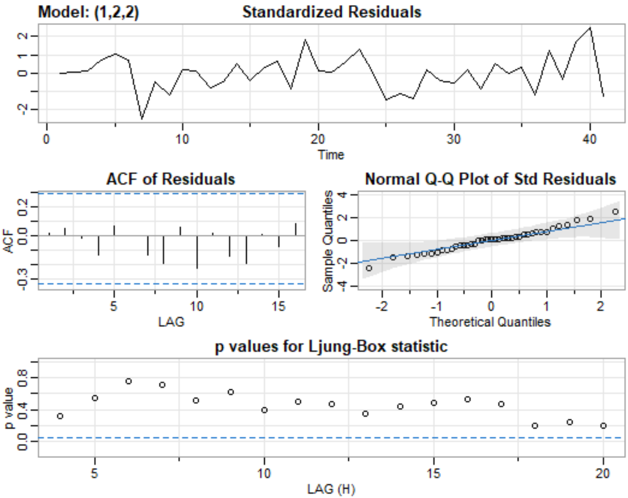
Based on the RStudio output, we can write the $ARIMAX(1,2,2)$ model as

$$y_t = 0.3399 z_{t,1} - 0.0932 z_{t,2} + 0.0330 z_{t,3} + x_t$$

$$\nabla^2 x_t = 0.6753 \nabla^2 x_{t-1} + a_t - 0.0001 a_{t-1} + 1.0000 a_{t-2}$$

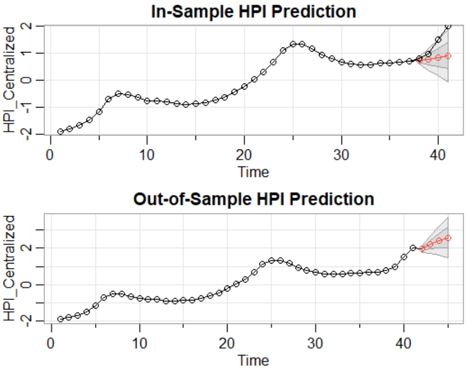
Where $a_t \sim WN(0, \sigma^2)$, lower case y_t and $z_{t,j}, j = 1,2,3$ are centralized observations.

Diagnostic Checking of ARIMAX model



For the diagnostic plot above, the $ARIMAX(1,2,2)$ model is adequate.

Prediction


In-Sample HPI Prediction

Out-of-Sample HPI Prediction

It's expected that 10-year real interest rates will continue to increase in the near future and ranges from 3.0 to 4.0. It's estimated that the population will exceed 3.7 million in 2020, and approach 3.75 million by 2025. Supposed the expected per-capita disposable income continue goes up.

Limitation of Analysis

- We find that both model can't capture the increment trend of HPI very well shown in the in-sample prediction.
- The model we choose based on BIC may not be the model having the best prediction ability.
- We can try another method to remediate the con-linearity in deterministic part of ARIMAX model like the principal component regression (PCR).
- The collected data have different length and we have to cut some part of raw data to meet the need of model construction.

Summary

1. ARIMA model shows that the HPI in Connecticut will keep moving upwards from 600 to 750 in the next 10 quarters.It indicates that the single house price in Connecticut is very likely to increase in the next 2.5 years.
2. Supposed the 10-year real interest rate increase steadily in the next couple of years, population and per-captia disposable income in Connecticut follow its expected increment pattern, the HPI will keep increasing.
3. Our findings correspond with the former research outcomes: per-captia disposable income, population is positively correlated with HPI, while rate is negatively correlated with HPI.

Reference

[1] Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata. "A spatio-temporal model of house prices in the USA." Journal of Econometrics 158, no. 1 (2010): 160-173.

[2] U.S. Federal Housing Finance Agency, All-Transactions House Price Index for Connecticut [CTSTHPI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CTSTHPI, November 9, 2022.

[3] U.S. Bureau of Economic Analysis and Federal Reserve Bank of St. Louis, Per Capita Personal Income in Connecticut [CTPCPI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CTPCPI, November 9, 2022.

[4] Federal Reserve Bank of Cleveland, 10-Year Real Interest Rate [REAINTRATREARAT10Y], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/REAINTRATREARAT10Y, November 9, 2022.