Rakesh Jayswal
G46567594

ECON -6376

Impact of Macroeconomic Indicators on Housing Prices

- **Time Series**: GDP, Consumer Price Index, Mortgage Interest Rate and Housing price Index.
- **Idea**: Analyzing how changes in GDP, interest rates, and inflation affect housing prices over time.

Acknowledgment

I would like to express my sincere gratitude to Professor Dr. Garry Cornwall for his invaluable guidance and support in the development of this paper. His encouragement and insightful feedback have been instrumental in helping me refine my research and deepen my understanding of macroeconomic factors affecting housing prices. I am truly grateful for his mentorship throughout this project.

1. Write a few paragraphs explaining your question and why you find it interesting. What do you expect to learn from it?

The question of how macroeconomic indicators such as GDP, Consumer Price Index (CPI), mortgage rates, and the Housing Price Index (HPI) impact housing prices is fascinating because it reveals the complex interplay between the economy and the real estate market. Housing is not only a basic need but also a key asset that shapes personal wealth, financial stability, and overall economic growth. Understanding what drives housing prices can provide essential insights for policymakers, investors, and individuals alike. By studying these relationships, we can anticipate how shifts in economic conditions might influence housing affordability, demand, and market stability, especially under varying economic policies or external shocks.

I find this question particularly intriguing because it examines factors that directly affect people's lives and their economic well-being. For example, GDP growth often corresponds with rising incomes, potentially boosting demand for housing, while CPI reflects inflation levels, influencing both purchasing power and construction costs. Mortgage rates, being the cost of borrowing, directly impact people's ability to afford homes, thereby affecting housing demand. The Housing Price Index provides a measure of changes in housing values, serving as a crucial indicator of real estate market trends. By analyzing these variables together, I hope to gain a clearer understanding of how sensitive housing prices are to macroeconomic fluctuations.

Through this study, I expect to learn about the extent to which each of these indicators influences housing prices and to what degree they interact with each other. This knowledge can provide valuable insights for predicting future trends in housing affordability and economic stability, informing both investment decisions and public policy. Ultimately, I aim to uncover patterns and relationships that can serve as a basis for more informed decision-making in real estate and economic planning

i) Impact of Macroeconomic Indicators on Housing Prices

In my review of the paper "Impact of Macroeconomic Indicators on Housing Prices" by Satish Mohan and colleagues, I explored the effects of key macroeconomic variables—such as crude oil prices, the 30-year mortgage interest rate (IR), Consumer Price Index (CPI), Dow Jones Industrial Average (DJIA), and the unemployment rate (UR)—on housing prices over time. By using a Vector Autoregression (VAR) model and time-series data from the Town of Amherst, New York, they discovered that mortgage interest rates and unemployment rates have the most significant impact on housing price fluctuations. Additionally, I found that recent housing prices play a critical role in shaping future price expectations, highlighting the self-reinforcing nature of housing market trends. This study offered valuable insights into the complex relationship between economic indicators and housing prices, deepening my understanding of how these factors interact within the real estate market.

ii) **Forecasting House Prices in OECD Economies**

In reviewing the paper "Forecasting House Prices in OECD Economies" by Kishor and Marfatia (2017) for my topic on the Impact of Macroeconomic Indicators on Housing Prices, I found substantial insights into how domestic and global economic indicators influence housing markets. The authors provide a comprehensive analysis of how interest rates, income levels, industrial production, and stock market performance affect housing price forecasts across 16 OECD countries. Their findings indicate that interest rates, reflecting monetary policy, have the strongest predictive power, followed by income and stock market dynamics, suggesting that housing markets are highly sensitive to financial conditions. Additionally, the study highlights the global interdependence of housing markets, where US and aggregate OECD house prices significantly impact national housing price forecasts, demonstrating spillover effects in the global economy. This paper not only deepens the understanding of individual economic indicators but also emphasizes the importance of a global perspective in housing market forecasts, making it a valuable reference for exploring how macroeconomic variables shape housing prices in interconnected economies.

3. Identify data sources and obtain each of the time series in your question.
    - Remember, you need to have four time series.
    - Your series should be at least 100 observations in length and of the same frequency.
    - Avoid stocks and indices, as they are typically random walks.

4. Write a few paragraphs about the data you have collected.
    - Who collects it and how is it collected? Survey, administrative data, etc.
    - What are the limitations of the data? Are there imputed data, missing data, measurement issues, etc.? You may need to search the technical notes for the data to find this type of information. It can be important!
    - Is it seasonally adjusted? Is it transformed in any way (e.g., year-over-year growth, log transformation, etc.)?
    - What are the units of measurement? Are they consistent across series?

I. Who collects the data and how is it collected?

The data from FRED (Federal Reserve Economic Data) is collected and maintained by various government agencies and institutions, then made available via the FRED platform managed by the Federal Reserve Bank of St. Louis. The GDP data is provided by the U.S. Bureau of Economic Analysis (BEA) as part of their national accounts data. Consumer Price Index (CPI) data is collected by the U.S. Bureau of Labor Statistics (BLS) through surveys on consumer expenditures. Mortgage Interest Rates data comes from the Federal Home Loan Mortgage Corporation (Freddie Mac), based on weekly national surveys of mortgage lenders. Housing Price Index (HPI) data is gathered by S&P Dow Jones Indices through its Case-Shiller Home Price Index, derived from property transactions and mortgage data.

II. What are the limitations of the data?

Each dataset has its own limitations. For GDP, a limitation is that it is only available quarterly, which requires interpolation to generate monthly values, potentially leading to inaccuracies in the short term. For the CPI, the accuracy depends on the survey methods and the representativeness of the sample. Additionally, the CPI may not fully capture price changes for certain goods or services in real-time, and imputed data is sometimes used when survey responses are incomplete. Mortgage Interest Rate data may suffer from sampling bias since it's based on surveys of lenders who may not fully represent the national mortgage market. Lastly, HPI data can be influenced by regional market variations, and there may be missing data in some areas due to incomplete transaction records. Additionally, all these datasets are subject to revisions over time as more accurate data becomes available.

III. Is the data seasonally adjusted?

Yes, most of the data collected from FRED is seasonally adjusted to remove effects of predictable seasonal patterns. GDP is seasonally adjusted by the BEA to account for regular fluctuations in economic activity, such as holiday shopping or weather-related slowdowns. CPI data is also typically adjusted for seasonality, smoothing out regular seasonal price changes like food and clothing costs. Mortgage Interest Rates are not always seasonally adjusted since rates can fluctuate due to other macroeconomic factors, but seasonal trends (such as higher activity in spring and summer real estate markets) can influence rates. The Housing Price Index (HPI) is seasonally adjusted to account for regular, predictable variations in home prices throughout the year.

IV. What are the units of measurement?

The units of measurement vary across the datasets but are consistent within each series. GDP is measured in billions of chained (inflation-adjusted) U.S. dollars, providing a constant measure of economic output. CPI is presented as an index (with a base year, such as 1982-1984 = 100), reflecting the relative change in consumer prices over time. Mortgage Interest Rates are measured as a percentage, representing the average annual interest rate for 30-year fixed-rate mortgages. HPI is also an index, where a base value (e.g., 100) represents the price level at a chosen starting period. The units are consistent across observations within each series, which helps in making comparative analyses over time.

5. Define strict and weak stationarity. Determine if each time series is stationary. Use all available evidence to support your conclusion.

A time series is considered strictly stationary if the joint probability distribution remains the same for any two time points. This implies that statistical properties, such as the mean, variance, and autocorrelation structure, are constant over time.

Mathematically:

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h}) \quad \forall h, \forall t_1, t_2, \ldots, t_k.$$

On the other hand, weak stationarity, also known as covariance stationarity, is a less stringent form that is more commonly used in time series analysis. A time series is weakly stationary if it meets three conditions: it has a constant mean over time, a constant variance over time, and the autocovariance (or autocorrelation) between two points depends only on the lag between them, not on the specific time at which the observations are made. Weak stationarity is easier to test and more frequently applied in practical scenario.

Mathematically,

- Constant Mean: The expected value E[Xt] must be constant for all t:

$$\mathbb{E}[X_t] = \mu, \quad \forall t$$

- Constant Variance: The variance Var(Xt) must be constant over time:

$$\text{Var}(X_t) = \sigma^2, \quad \forall t$$

- Autocovariance Depends Only on Lag: The covariance between Xt and X t+h depends only on the lag hh, not on the time t:

$$\text{Cov}(X_t, X_{t+h}) = \gamma(h), \quad \forall t, h$$

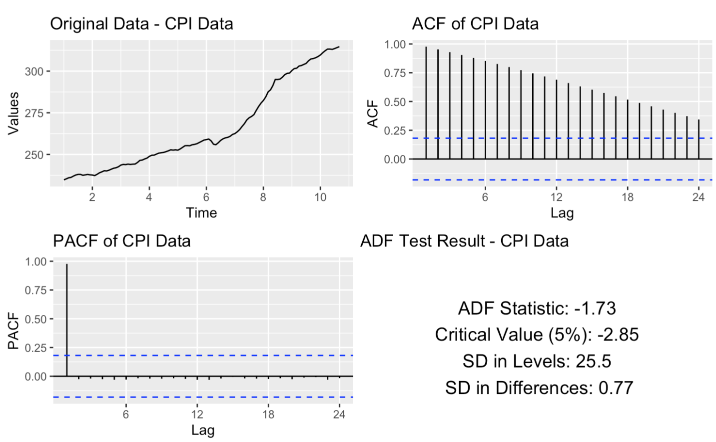


Fig i. Consumer Price Index

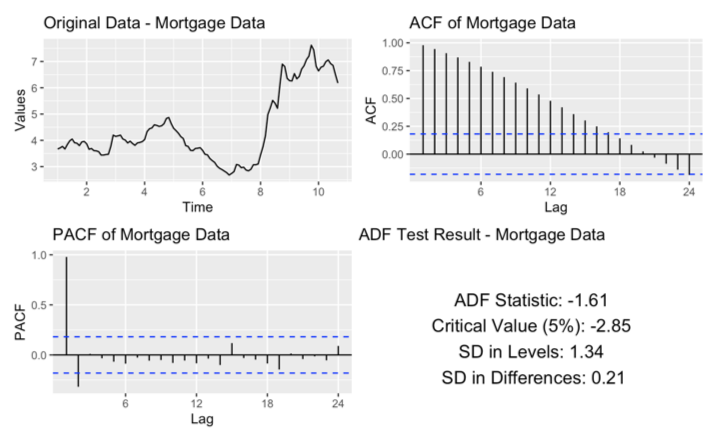


Fig ii. Mortgage

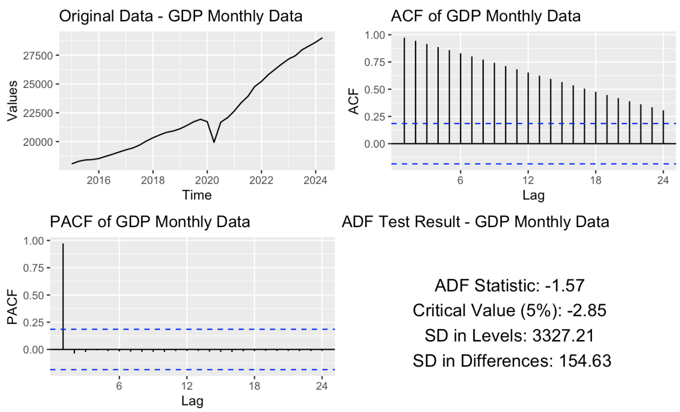


Fig iii. Gross Domestic Product

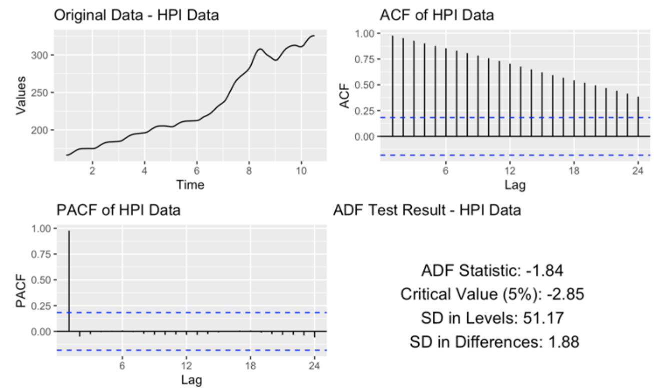


Fig iv. Housing Price Index

In my analysis of the four datasets (CPI, mortgage, GDP, and HPI), I consistently observed evidence of non-stationarity across all the time series. Beginning with the CPI data, the original time series plot shows a strong upward trend, indicating a clear violation of stationarity assumptions. The ACF plot exhibits a slow decay over time, suggesting that past values have a long-lasting influence on current values, typical of trending data. The PACF plot also shows significant spikes at early lags, implying short-term dependencies. The ADF test result, with a statistic of -1.73, is higher than the critical value of -2.85, indicating that I cannot reject the null hypothesis of a unit root. This suggests the data is non-stationary, further supported by the large standard deviation in levels (25.5) compared to differences (0.77), highlighting the potential for differencing to stabilize the data.

For the mortgage data, I observed similar patterns. The original time series plot shows fluctuations with a trend, and the ACF exhibits a gradual decay, pointing to long-term dependencies. The PACF plot indicates significant autocorrelation at the first few lags. The ADF test statistic of -1.61, once again greater than the critical value, fails to reject the null hypothesis of a unit root, confirming non-stationarity. The standard deviation in levels (1.34) is much larger than in differences (0.21), suggesting that differencing would help achieve stationarity.

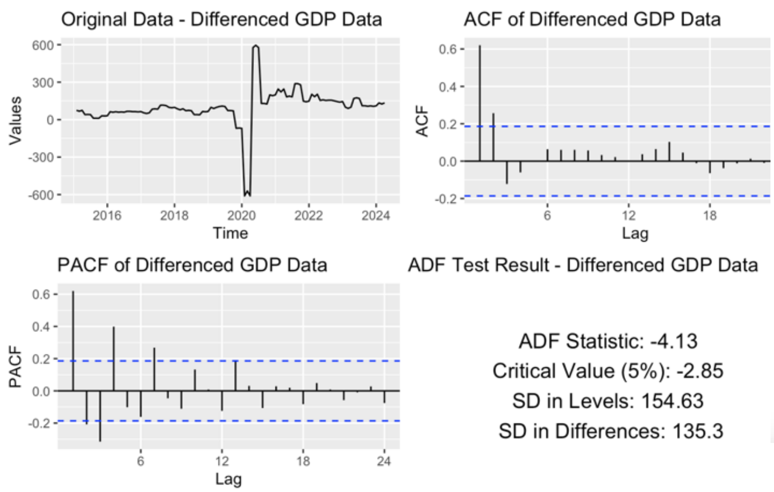
In the GDP data, the original time series shows a strong upward trend with a noticeable dip around 2020, likely reflecting the economic impact of the COVID-19 pandemic. The ACF plot exhibits the same slow decay, signaling non-stationarity. The PACF shows sharp spikes at the initial lags, reinforcing the presence of short-term dependencies. The ADF test statistic of -1.57 is higher than the critical value, meaning I cannot reject the null hypothesis of non-stationarity. The large standard deviation in levels (3327.21) compared to differences (154.63) further emphasizes the need for differencing to stabilize the series.

Finally, the HPI data follows the same pattern. The original time series reveals an upward trend, and the ACF shows a slow decay over the lags, suggesting persistent autocorrelations. The PACF plot highlights significant autocorrelation at the first lags. The ADF test statistic of -1.84, being greater than the critical value, again confirms the presence of a unit root. The comparison between the standard deviation in levels (51.17) and in differences (1.88) suggests that differencing will help stabilize the data.

In summary, across all four datasets, the combination of an upward trend in the time series, a slowly decaying ACF, significant PACF spikes at initial lags, and ADF test results consistently indicate non-stationarity. So I

Differentiated the data in each case to achieve stationarity, which is essential for accurate time series modeling and forecasting.

Differencing Time Series data to make stationary



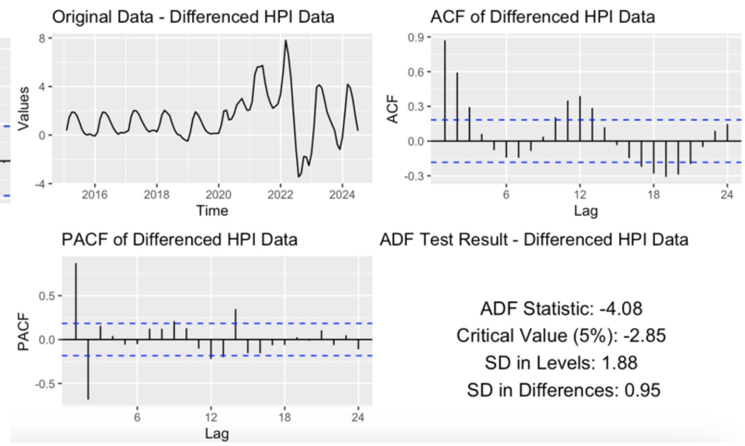
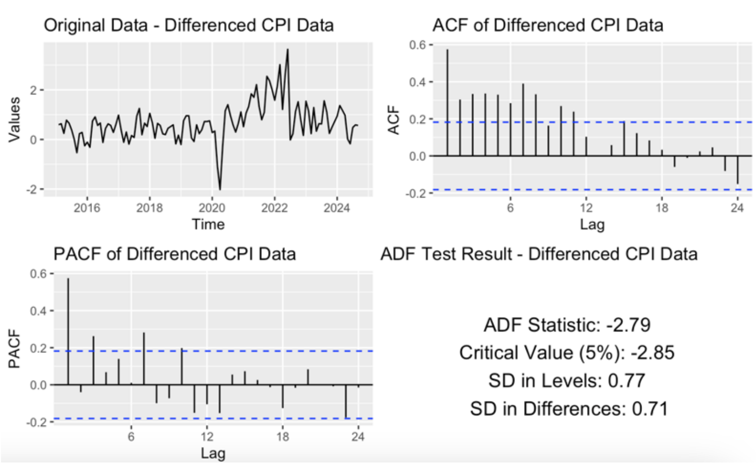


Fig vi. Housing Price Index

Fig v Gross Domestic Product



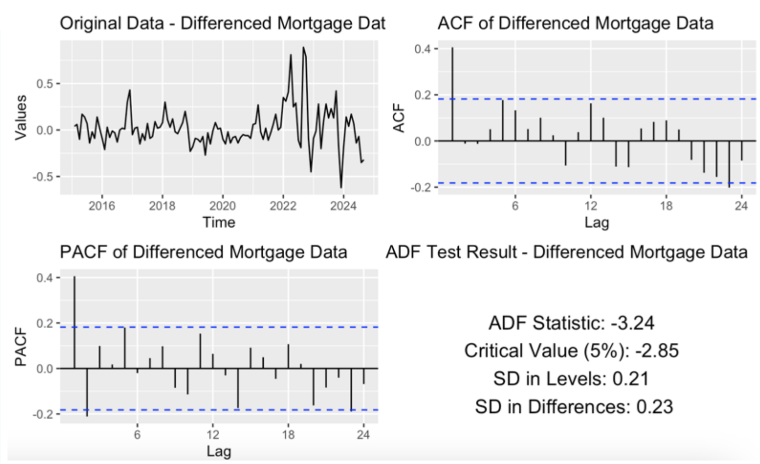


Fig vii. Consumer Price Index data

Fig viii, Mortgage Data

In analyzing the differenced GDP, HPI, CPI, and mortgage data, the findings consistently show important indicators of stationarity and autocorrelation, which support the conclusion that each dataset is suitable for further time series modeling.

For the differenced GDP data, the ADF test statistic of -4.13 is well below the critical value of -2.85, rejecting the null hypothesis of a unit root and confirming the data's stationarity. The ACF plot reveals significant autocorrelations at early lags but diminishing values after a few lags, which is further supported by the PACF plot showing partial correlations that taper off as the number of lags increases.

Similarly, the differenced HPI data shows a stationary series with an ADF statistic of -4.08, also below the critical value of -2.85. The ACF and PACF plots indicate the presence of autocorrelations at initial lags, with the PACF suggesting the potential for an autoregressive component, which could be incorporated in future modeling.

For the differenced CPI data, the ADF test statistic of -2.79 is close to but slightly above the critical value, meaning that at the 5% significance level, the null hypothesis of non-stationarity cannot be rejected. Despite this, the difference in standard deviations between the original and differenced series, along with the ACF and PACF plots showing short-term autocorrelations, suggests that the series is near stationarity and could potentially benefit from additional differencing or transformation to achieve full stationarity.

Lastly, the differenced mortgage data yields an ADF test statistic of -3.24, which is well below the critical value of -2.85, clearly indicating stationarity. The ACF and PACF plots exhibit significant autocorrelations at the early lags, further supporting the use of an autoregressive model to capture the short-term dependencies in the series.

Overall, I can conclude that the ADF test results, along with the patterns in the ACF and PACF plots, consistently show that differencing has achieved stationarity across all datasets.

6. Choose one of your time series. Split the data into a train (1:(T-6)) and test ((T-5):T) set it. Model the training set using the Box-Jenkins methodology (ARIMA).
   - Define the model selection criteria you will use to determine the best model and describe its relative strengths/weaknesses.

To determine the best model for my data, I will use the Akaike Information Criterion (AIC) as the model selection criterion. AIC is widely used for selecting models in time series analysis as it balances the goodness of fit with model complexity. It penalizes models that have too many parameters, helping to avoid overfitting, but without being too strict. The lower the AIC value, the better the model fits the data while maintaining simplicity. One of the key strengths of AIC is its ease of computation and its ability to guide the selection of models that perform well in terms of fit.

Mathematically,

$$\mathrm{AIC} = -2\ln(L) + 2k$$

Where, L is the maximum likelihood of the model, representing how well the model fits the data and k is the number of parameters in the model.

However, a potential weakness is that AIC may still favor more complex models compared to other criteria like the Bayesian Information Criterion (BIC), which could sometimes result in overfitting. In small samples, it may show a bias toward more complex models, potentially leading to overfitting, another limitation is that AIC provides only a relative measure of model quality, allowing for comparison between models on the same dataset but not offering an absolute assessment of fit. Finally, the AIC assumes that one of the models being compared is the true model, which may not be valid in all cases, especially when none of the models fully capture the underlying data structure. Despite these weaknesses, AIC remains a widely used and robust criterion for model selection in many applications

I choose to go with GDP data.

- Evaluate the ACF and PACF of the stationary data to determine what model to start your analysis with, be specific about why you choose your starting point.

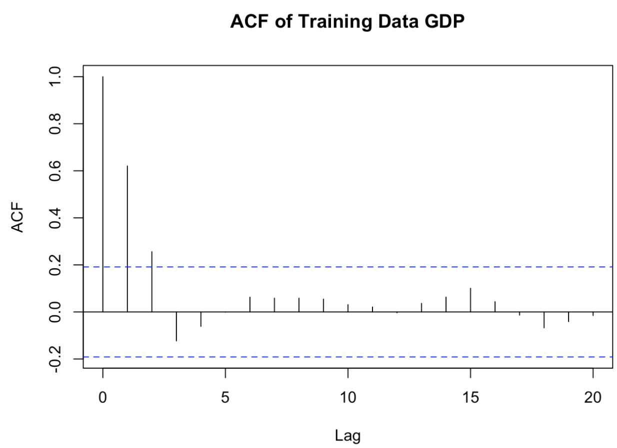


Fig ix. ACF of Training Data GDP

Fig x. PACF of Training Data GDP

Based on the analysis of the ACF and PACF plots, ACF decays faster than PACF so this is MA model, and the lags are significant here so, I concluded that an ARIMA model moving average (MA) components is appropriate for this time series.

Similarly, the ACF plot showed a strong spike at lag 1 followed by a quick decay, indicating that the data exhibits more of moving average behavior. This led me to consider including MA terms in the model, particularly up to MA (2) MA (3) and MA(4), as the ACF values drop off rapidly after lag 1.

- Estimate the starting model and move from general to specific using the ARIMA framework. Provide a table of at least three different models including your "best" model as indicated by your information criteria. Be detailed in explaining your decision process.

From the ACF and PACF, I decided to test ARIMA models with different combinations of AR and MA terms, from general to specific as ARIMA (0,1,4), ARIMA (0,1,3), and ARIMA (0,1,2). These models incorporate no autoregressive components but do have moving average components, allowing me to capture the complex patterns in the time series based on both past values and past forecast errors.

```
Series: train_data
ARIMA(0,1,2)

Coefficients:
          ma1      ma2
      -0.4898  -0.4104
s.e.   0.0779   0.0714

sigma^2 = 18169:  log likelihood = -657.32
AIC=1320.64   AICc=1320.88   BIC=1328.57

Training set error measures:
                   ME     RMSE      MAE     MPE     MAPE     MASE      ACF1
Training set 8.20394 132.8539 55.39587 5.02372 36.16771 1.506378 0.1898946
```

```
Series: train_data
ARIMA(0,1,3)

Coefficients:
          ma1     ma2      ma3
      -0.0596  0.0359  -0.9000
s.e.   0.0530  0.0562   0.0494

sigma^2 = 9421:  log likelihood = -624.39
AIC=1256.79   AICc=1257.19   BIC=1267.36

Training set error measures:
                   ME     RMSE      MAE      MPE    MAPE      MASE       ACF1
Training set 8.453637 95.19399 34.10141 6.544396 22.9861 0.9273183 0.01176946
```

```
Series: train_data
ARIMA(0,1,4)

Coefficients:
          ma1     ma2      ma3      ma4
      -0.0392  0.0369  -0.8991  -0.0248
s.e.   0.0994  0.0556   0.0500   0.1022

sigma^2 = 9508:  log likelihood = -624.36
AIC=1258.73   AICc=1259.34   BIC=1271.95

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE         ACF1
Training set 8.465925 95.15829 33.92769 6.554687 22.91568 0.9225944 -0.008934958
```

Outputs: Arima Models

Table 1: Comparison of Arima Models

Parameter	ARIMA.0.1.2.	ARIMA.0.1.3.	ARIMA.0.1.4.
AIC	1320.6400	1256.7900	1258.7300
BIC	1328.5700	1267.3600	1271.9500
RMSE	132.8539	95.1940	95.1583
MAE	55.3959	34.1014	33.9277
MAPE	36.1677	22.9861	22.9157
θ_1 (MA term)	-0.4898	-0.0596	-0.0392
θ_2 (MA term)	-0.4104	0.0359	0.0369
θ_3 (MA term)	NA	-0.9000	-0.8991
θ_4 (MA term)	NA	NA	-0.0248

To determine the best ARIMA model, I focused primarily on the AIC values. Starting with ARIMA (0,1,2), I found that its AIC was 1320.64, which suggested that while it provided a reasonable fit, there was room for improvement. I then fitted ARIMA (0,1,3), which showed a significant reduction in AIC to 1256.79, indicating a much better fit. Finally, I tested ARIMA (0,1,4), but its AIC increased slightly to 1258.73, suggesting that the additional complexity of the model did not lead to a significant improvement in fit. Based on these AIC comparisons, I selected ARIMA (0,1,3) as the best model, as it had the lowest AIC, balancing fit and complexity effectively

7. Define serial correlation and how to test for it. Include the appropriate maths in your explanation. Test your models for serial correlation and discuss the results and how they impact the choice of "best" model.

Serial correlation refers to the correlation between the residuals (errors) of a time series model across different time periods. It occurs when residuals are not independent but instead display patterns over time, indicating that the model has not fully captured the underlying time-dependent structure of the data. Detecting serial correlation is important because it suggests that the model's residuals contain information that the model has not accounted for, leading to inefficiencies in prediction and analysis.

One common method to test for serial correlation is the Box-Ljung test. This statistical test evaluates whether the residuals of a time series model are independently distributed by examining autocorrelation over a specified number of lags.

Mathematically,

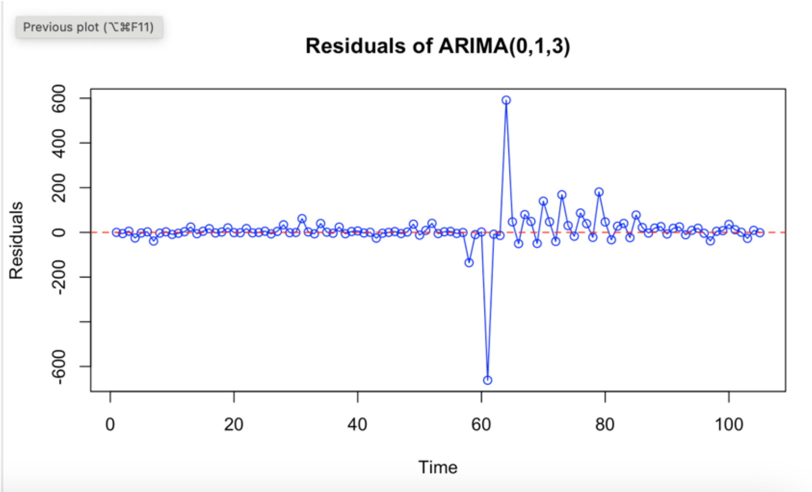
$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

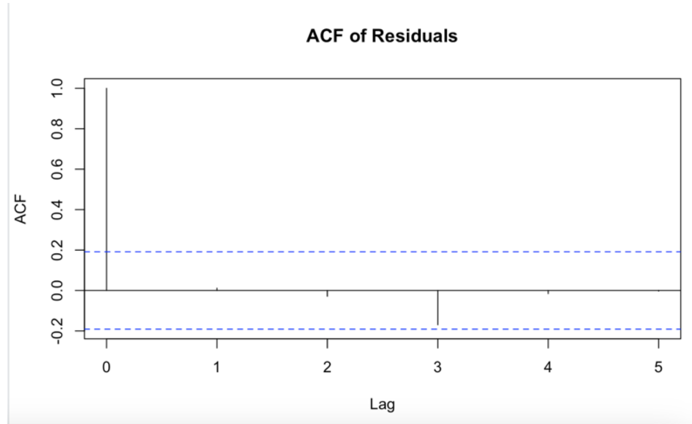
Were,

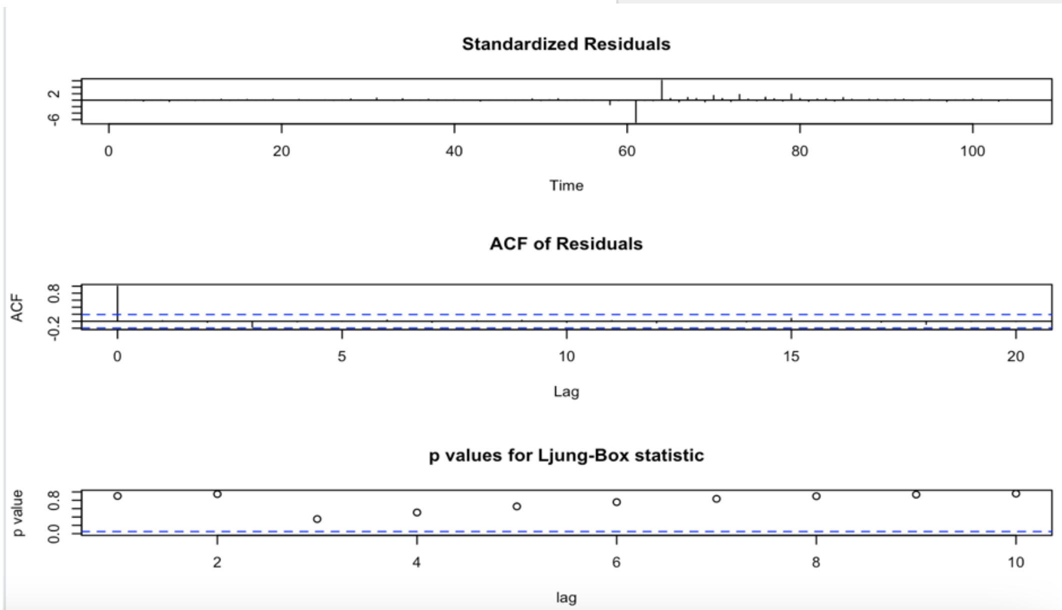
- n is the number of observations,
- h is the number of lags being tested,
- ρ^k is the sample autocorrelation at lag k.

The null hypothesis (H_0) assumes that the residuals are independently distributed (no serial correlation), while the alternative hypothesis (H_1) suggests the presence of significant serial correlation. If the p-value of the Box-Ljung test is greater than 0.05, we fail to reject the null hypothesis, indicating no significant serial correlation in the residuals.

This can be visualized using the tsdiag method, a valuable diagnostic tool in time series analysis for evaluating the adequacy of models like ARIMA, particularly in assessing the presence of serial correlation in residuals. It generates several diagnostic plots that allow for visual inspection of the residuals, offering a complementary approach to formal statistical tests. Among the key outputs of tsdiag are the standardized residuals plot, the autocorrelation function (ACF) plot of residuals, and the p-values from the Box-Ljung test. The standardized residuals plot displays how residuals behave over time, with random distribution around zero indicating that the model has effectively captured the time series structure. Patterns or trends in this plot, however, would suggest serial correlation, indicating a need for model adjustment. The ACF plot shows the degree of correlation in the residuals at different time lags, and for a well-fitted model, the ACF should exhibit no significant spikes outside of the confidence bounds. Significant autocorrelations in this plot would suggest that the model has not adequately captured time-dependent patterns, revealing the presence of serial correlation. The plot of p-values from the Box-Ljung test helps determine whether residuals exhibit significant autocorrelation across different lags, with high p-values indicating no significant autocorrelation. This visual diagnostic approach allows for a more nuanced assessment of a model’s fit, beyond the results of formal tests. By examining these diagnostics, analysts can better understand whether a model's residuals truly behave like white noise, supporting the selection of the most suitable model for the data. Thus, tsdiag serves as an important tool in verifying model assumptions and guiding the selection of the best model for accurate time series analysis.







Ljung-Box Test Result for ARIMA(0,1,3)

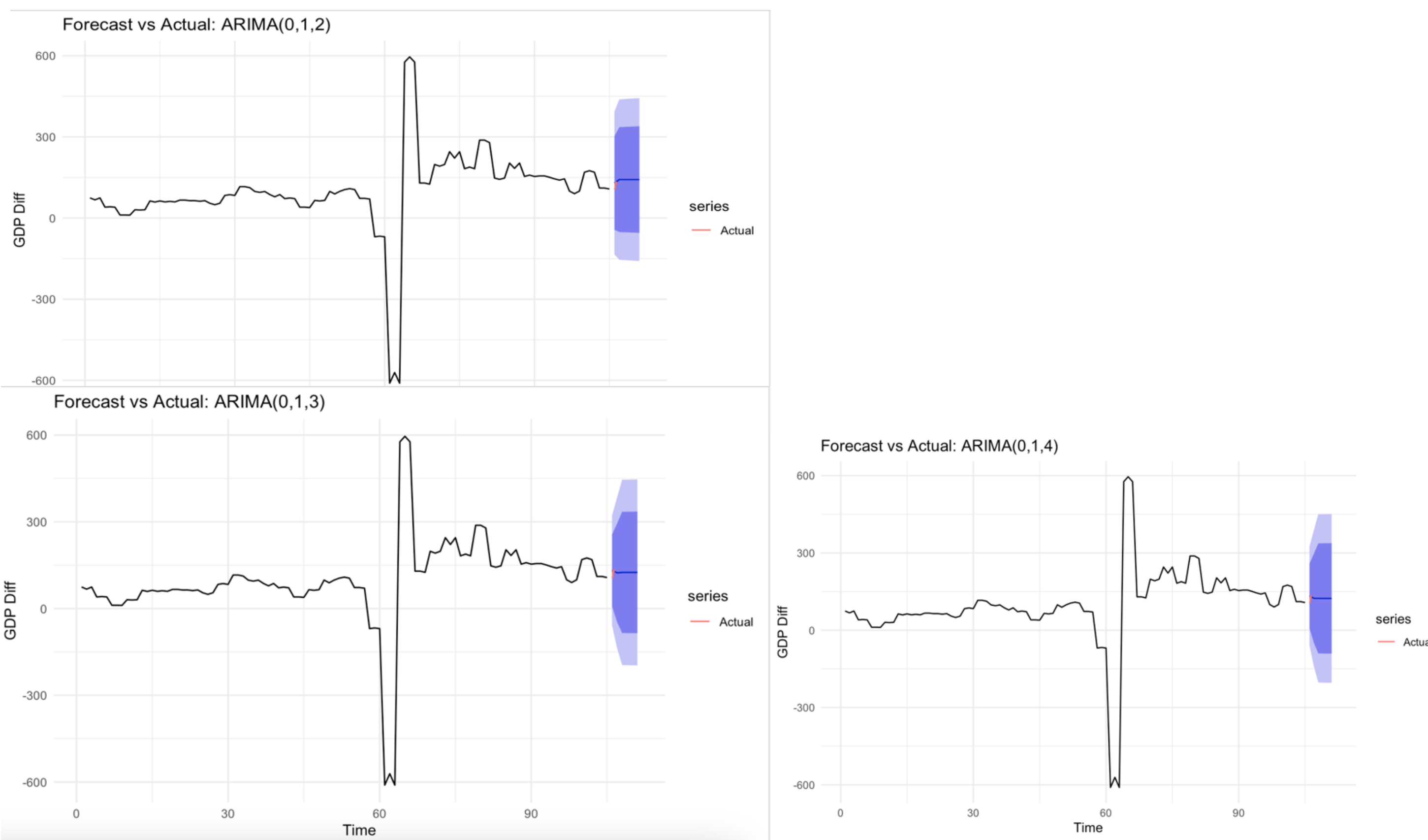
	Model	Statistic	df	p-value
X-squared	ARIMA(0,1,3)	3.3081	5	0.6526

In evaluating the serial correlation in the residuals of my ARIMA models, I utilized the tsdiag method, the autocorrelation function (ACF) of the residuals, and the results from the Box-Ljung test with a lag of 5. The tsdiag output included a plot of the standardized residuals, the ACF of the residuals, and the p-values for the Ljung-Box statistic. The standardized residuals plot showed that the residuals fluctuate around zero without any discernible pattern, indicating that the model captures the underlying data structure well.

The ACF plot of the residuals revealed that the autocorrelations for lags 1 to 5 fell within the confidence bounds, confirming that there is no significant serial correlation in the residuals. This aligns with the Box-Ljung test results, which indicated a test statistic of 3.30, degrees of freedom of 5, and a high p-value of 0.6526. Such a high p-value suggests that I fail to reject the null hypothesis, further supporting the conclusion that the residuals behave like white noise.

Together, these diagnostics from tsdiag, the residual ACF plot, and the Box-Ljung test results provide strong evidence that the ARIMA (0,1,3) model adequately captures the dynamics of the differentiated GDP data. The absence of serial correlation in the residuals, along with the model's simplicity and low AIC value of 1256.79 compared to ARIMA (0,1,4) and ARIMA (0,1,2), reinforces my decision to select the ARIMA(0,1,3) model as the best fit for my analysis. This comprehensive evaluation ensures that the chosen model is robust and effective for forecasting and further analysis.

8. Using each of the three models you provided in the table, forecast the next six observations of the time series and evaluate the forecast.
    - You will need to choose a loss function for evaluating the forecast and define it.
    - Which model is the better forecast based on your loss function?



Plots: Forecast Vs Actual for Arima Models

Table III Forecasted Values and MAE for ARIMA Models

Model	Forecast 1	Forecast 2	Forecast 3	Forecast 4	Forecast 5	Forecast 6	Mean Absolute Error (MAE)
ARIMA(0,1,2)	129.5366	142.3009	142.3009	142.3009	142.3009	142.3009	20.22
ARIMA(0,1,3) (Best Model)	131.8674	123.7059	125.0925	125.0925	125.0925	125.0925	11.82
ARIMA(0,1,4)	133.0133	123.8820	123.4532	123.4488	123.4488	123.4488	12.59

In this analysis, I used three ARIMA models—ARIMA (0,1,2), ARIMA (0,1,3), and ARIMA (0,1,4) to forecast the next six observations of the differenced GDP data. To evaluates the forecast accuracy, I chose Mean Absolute Error (MAE) as the loss function. MAE measures the average magnitude of forecast errors without considering their direction, making it an intuitive metric for assessing forecast accuracy. The MAE is calculated by taking the absolute differences between forecasted values and actual values, then averaging these differences across all observations.

Mathematically,

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

Where yt represents the actual observed values, y^t represents the forecasted values, and n is the total number of observations. A lower MAE indicates better forecast accuracy, as it suggests that the forecasted values are closer, on average, to the actual values. MAE is useful because it provides a clear view of average error magnitude without being overly influenced by extreme values, making it a reliable choice for evaluating the forecast.

After generating forecasts for each model, I calculated the MAE for each set of forecasted values. The results showed that the ARIMA (0,1,3) model had the lowest MAE at 11.82, which indicates that this model produced the smallest average forecast error among the three. This made ARIMA (0,1,3) the best model based on my chosen evaluation criterion, as it provided the most accurate predictions for the next six observations. In summary, by focusing on minimizing MAE, I was able to identify ARIMA (0,1,3) as the most reliable model for forecasting this time series

9. Define the concept of forecast combinations and provide at least two specific examples with a detailed explanation.

Rakesh Jayswal
G46567594

Forecast combinations are a method used to improve prediction accuracy by combining the outputs of several forecasting models. Instead of relying on just one forecast, this approach brings together multiple forecasts to take advantage of each model's strengths and minimize their individual weaknesses. The idea is that different models may capture unique patterns in the data, so combining them can result in a more accurate overall forecast. By pooling together different forecasts, this method reduces the impact of errors from any single model, leading to more robust and reliable predictions.

Mathematically,

$$\hat{y}_{c,t} = w_1 \hat{y}_{1,t} + w_2 \hat{y}_{2,t}$$

Where w1 and w2 are weights that sum up to 1: (w1+w2=1).

if we have more than two models (say, n models), the combined forecast becomes,

$$\hat{y}_{c,t} = \sum_{i=1}^{n} w_i \hat{y}_{i,t}$$

where $\sum_{i=1}^{n} w_i = 1.$

Example 1: Simple Average Combination

The simple average combination is one of the most straightforward methods, where each forecast is given an equal weight. For example, if we have three models y^1,t,y^2,t and y^3,, the combined forecast is calculated as

$$\hat{y}_{c,t} = \frac{1}{3} \left(\hat{y}_{1,t} + \hat{y}_{2,t} + \hat{y}_{3,t} \right)$$

In this case, each forecast contributes equally, with weights w1=w2=w3=1. This approach is useful when we do not have a strong reason to favor one model over another and want to balance them equally.

Example 2. Granger-Ramanathan combination

The Granger-Ramanathan combination is an advanced method for combining forecasts, developed by Clive Granger and Richard Ramanathan in 1984. Unlike traditional methods that use fixed weights, this approach determines optimal weights through a regression model, dynamically adjusting each model's contribution based on historical accuracy. By using this data-driven technique, the combined forecast adapts over time, often resulting in more accurate prediction

In this method, we have forecasts from multiple models, each predicting the same variable. Suppose the forecasts at time tt from models 1 through n are y^1,t, y^2,ty^2,t, ..., y^n,t, The combined forecast y^c,,t is calculated a

$$\hat{y}_{c,t} = w_1 \hat{y}_{1,t} + w_2 \hat{y}_{2,t} + \cdots + w_n \hat{y}_{n,t}$$

To determine the optimal weights, we set up a regression where the actual observed value ytyt is regressed on the forecasts from each model. This setup can be represented by

$$y_t = w_0 + w_1\hat{y}_{1,t} + w_2\hat{y}_{2,t} + \cdots + w_n\hat{y}_{n,t} + \epsilon_t$$

Here, yt is the observed value, w0is the intercept term, and ϵt represents the error term. Running this regression over historical data provides the values of w0,w1,…,wnt that minimize the forecast error, resulting in the most accurate combination of the model forecasts.

For example, if we have two models and our regression yields the weights w0=0.5, w1=0.3, and w2=0.7, our combined forecast for the next period t would be:

$$\hat{y}_{c,t} = 0.5 + 0.3\hat{y}_{1,t} + 0.7\hat{y}_{2,t}$$

The Granger-Ramanathan method effectively leverages the unique strengths of each model. Unlike a simple averaging approach, this regression-based technique allows the combined forecast to adapt based on each model's performance, providing a balanced and robust prediction that often outperforms individual models. This adaptability makes the Granger-Ramanathan combination particularly useful in complex or fluctuating datasets, where model performance may vary over time.

Why Forecast Combinations Are Effective:

Forecast combinations work because they pool the strengths of different models. Since no single model can perfectly predict the future, combining their forecasts often leads to better overall accuracy. By blending models, we can reduce the impact of outliers or biases that might be present in individual models, ultimately leading to more reliable forecasts. This method is especially useful when there's uncertainty about which model will perform best in future conditions

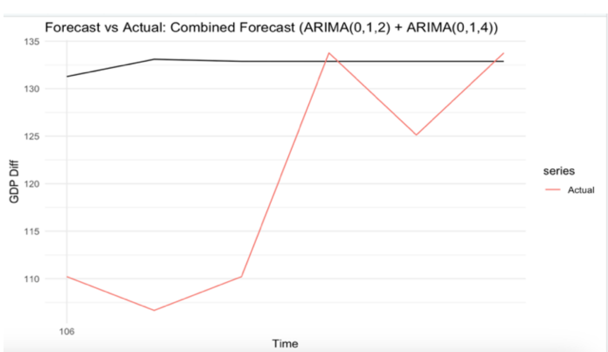


Fig: xi Combined Forecast vs Actual

Rakesh Jayswal
G46567594

10. Choosing one of the forecast combination methods create a combined forecast for the next six observations of the time series using the two models that are not the best model. Compare the combined forecast to the best model forecast and discuss the results.

Table IV Comparison of Best Model and Combined Forecast Using MAE

Model	Forecast 1	Forecast 2	Forecast 3	Forecast 4	Forecast 5	Forecast 6	Mean Absolute Error (MAE)
ARIMA(0,1,3) (Best Model)	131.8674	123.7059	125.0925	125.0925	125.0925	125.0925	11.82
Combined Forecast (ARIMA(0,1,2) + ARIMA(0,1,4))	131.2750	133.0914	132.8770	132.8749	132.8749	132.8749	13.28

In addressing the question of comparing the combined forecast to the best model forecast, I opted to create a combined forecast using the two models that were not the best: ARIMA (0,1,2) and ARIMA (0,1,4). The combined forecast was calculated as a simple average of their predictions for the next six periods. To compare the combined forecast to the best model forecast, I analyzed both sets of predictions for accuracy and behavior over the next six forecasted periods. The best model forecast, generated by ARIMA (0,1,3), produced forecasted values that were consistent and stable, providing an average trajectory closely aligned with the actual data, as indicated by its lower Mean Absolute Error (MAE). This model achieved an MAE of 11.82, demonstrating a high level of accuracy and suggesting that it captures the underlying trends in the data effectively.

On the other hand, the combined forecast, created by averaging the forecasts from ARIMA (0,1,2) and ARIMA (0,1,4), produced a smoother and moderated prediction line. The combined forecast reflected the initial values from both models but exhibited a more stable trend, starting at 131.27 and stabilizing around 132.87. This approach was intended to balance out individual biases from the two models, which each showed unique patterns in their predictions. Despite this, the combined forecast had an MAE of 13.28, which was higher than that of the best model, indicating lower accuracy.

In conclusion, while the combined forecast provided a balanced prediction by averaging the tendencies of two models, it did not outperform the ARIMA (0,1,3) model in terms of accuracy. The best model forecast was better suited to this dataset, as evidenced by its lower MAE, indicating it effectively captured the patterns in the data on its own. This suggests that, in this case, a single well-fitting model ARIMA (0,1,3)) is more reliable than a combined approach for forecasting this time series

References:

Ding, X. Macroeconomic Factors Affecting Housing Prices: Take the United States as an Example. Carey Business School, Johns Hopkins University, 100 International Drive, Baltimore, MD 21202, USA. .

Holly, S., Pesaran, M. H., & Yamagata, T. A Spatio-Temporal Model of House Prices in the USA. Faculty of Economics and CIMF, University of Cambridge, UK; University of Southern California, USA; Department of Economics and Related Studies, University of York, UK.

Adams, Z., & Füss, R. Macroeconomic Determinants of International Housing Markets. Union Investment Chair of Asset Management, European Business School (EBS), International University Schloss Reichartshausen, D-65375 Oestrich-Winkel, Germany.

Kishor, N. K., & Marfatia, H. A. Forecasting House Prices in OECD Economies