FORECASTING TURKEY'S INFLATION

A PREPRINT

Arzu Isik Topbas
arzuu.isik@gmail.com
Dynamic Modeling and Forecasting in Big Data

December 20, 2023

ABSTRACT

This project presents a thorough analysis of inflation in Turkey by leveraging a diverse dataset obtained from reputable sources, including the OECD, World Bank, Turkish Statistical Institute, and government portals. The dataset encompasses crucial economic indicators such as the Consumer Price Index (CPI), gold and oil prices, unemployment rates, exchange rates, and interest rates, covering the monthly inflation rate data from January 2005 to September 2023. Employing various time series models and regression techniques, both univariate and multivariate approaches are considered for forecasting inflation. The findings underscore the challenges posed by non-stationarity in the inflation data and provide a comparative assessment of the performance of different models in predicting future inflation trends. The primary objective is to examine and compare the effectiveness of these methods in forecasting the inflation rate over a four-month horizon, offering valuable insights into potential trends and fluctuations in the Turkish economy.

Keywords Inflation · Turkey · More

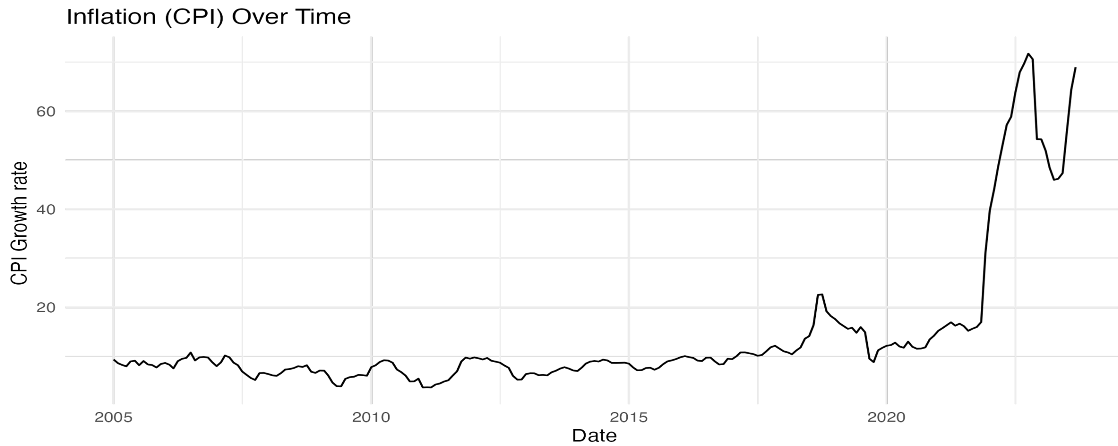
1 Introduction

Inflation is a critical economic indicator with profound implications for monetary policy, fiscal planning, and overall economic stability. This project focuses on Turkey, where inflation has exhibited a notable upward trend in recent years, reaching unprecedented levels. The dataset encompasses diverse variables, allowing for a nuanced analysis of the factors influencing inflation. Through the rigorous application of forecasting models such as ARIMA, ARMAX, linear regression, structural time series models, and dynamic factor models, we meticulously evaluated their performance. The final comparison results highlight the effectiveness of forecasting models, with the "Structure Change - Segmented Linear" and "Structure Change - Linear" models emerging as particularly robust performers. By evaluating the performance of these models, we aim to provide insights into the dynamics of inflation in Turkey and contribute valuable information for policymakers, economists, and businesses operating in the region.

2 Methodology and Data

2.1 Data

A comprehensive dataset has been compiled, drawing from various reputable sources to analyze key economic indicators. The inflation data, retrieved from the OECD, provides insights into the Consumer Price Index (CPI), serving as a crucial measure of general price level changes over time. The gold and oil datasets, sourced from the World Bank, shed light on the trends and fluctuations in the prices of these essential commodities, offering valuable information for understanding economic stability and resource dependencies. The unemployment statistics, obtained from the Turkish Statistical Institute, contribute essential labor market insights, while the exchange rate data from the Turkish government's official portal and the interest rate data from the St. Louis Fed's FRED platform provide a comprehensive view of monetary policy and its impact on the economy. Appendix A.1

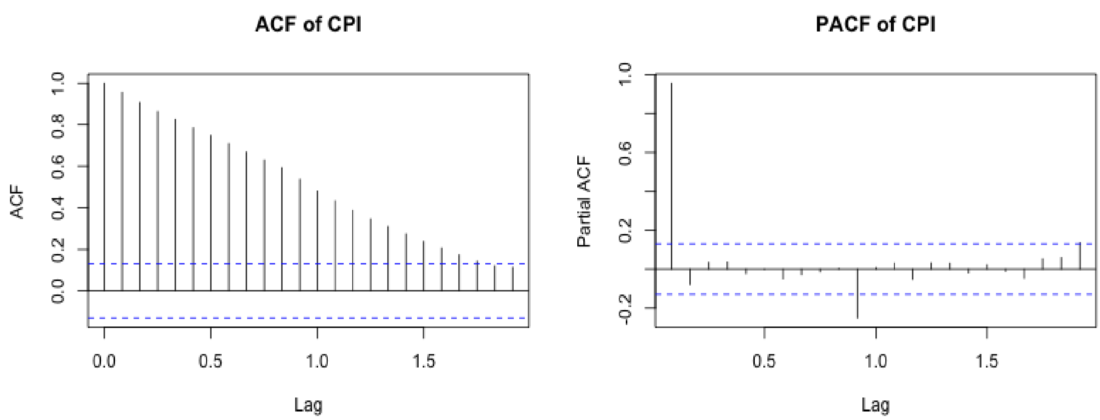


Turkey's inflation rate has been steadily increasing since 2015, with a sharp rise in 2022 and 2023. As of November 2023, Turkey's annual inflation rate is 61.98%, the highest it has been in decades.

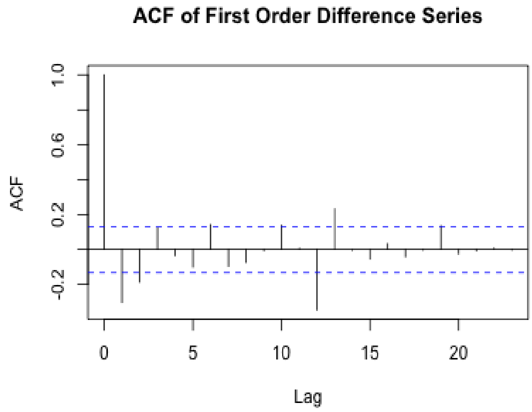
2.2 Stationarity Test

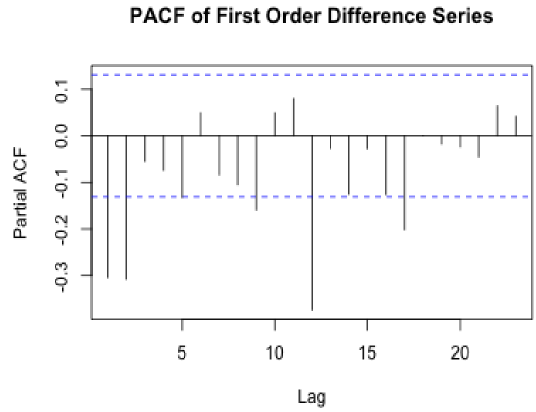
The Augmented Dickey-Fuller (ADF) and Elliot, Rothenberg and Stock (ERS) unit root tests were employed to assess the stationarity of Turkey's Consumer Price Index (CPI). Both tests revealed that the series possesses a unit root, indicating non-stationarity. This signifies that the CPI's mean and variance are not constant over time, showcasing an upward trend consistent with the recent high inflation rates experienced in Turkey.

Further analysis of the first difference of the CPI confirms this conclusion. The ACF plot exhibits a prominent spike at lag 1, signifying a strong positive autocorrelation in the differenced series, which deviates from white noise. The PACF plot also reveals a significant spike at lag 1, confirming the presence of an AR(1) component, consistent with the unit root findings.



Therefore, these results demonstrate that Turkey's CPI is non-stationary. This presents challenges for statistical analysis and forecasting, necessitating differencing the series to achieve stationarity before performing further statistical procedures.





3 Result

The analysis of inflation in Turkey yielded noteworthy findings, in this report we assess the performance of various time series forecasting models using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The models are ranked based on their predictive accuracy for the target variable.

Table 1: Model Performance Metrics

Model	**MAE**	**RMSE**
Structure Change - Segmented Linear	0.06	0.09
Structure Change - Linear	0.06	0.10
VAR Growth A.15	1.01	1.43
VAR Level A.14	1.79	1.84
Autoregressive	2.44	2.67
StructTS Trend	3.09	4.66
ARMAX A.9	4.51	5.52
Linear Regression (Year) A.8	5.07	6.18
Linear Regression A.6	5.18	6.19
Linear Regression (Month) A.7	5.35	6.29
ARIMA $c(1, 0, 0)$ A.10	5.51	6.30

The comparative analysis of various models for forecasting inflation in Turkey reveals distinctive performance metrics based on Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Notably, the models "Structure Change - Segmented Linear" and "Structure Change - Linear" exhibit the most accurate predictions with the lowest MAE of 0.06 and 0.06, and RMSE of 0.09 and 0.10, respectively. These structural change models showcase a remarkable ability to capture inflation dynamics.

Moving to VAR models, "VAR Growth" outperforms "VAR Level" with a lower MAE (1.01 compared to 1.79) and RMSE (1.43 compared to 1.84). The Autoregressive (AR) model follows, displaying higher errors than the VAR models but still providing reasonable forecasting accuracy, with a MAE of 2.44 and RMSE of 2.67.

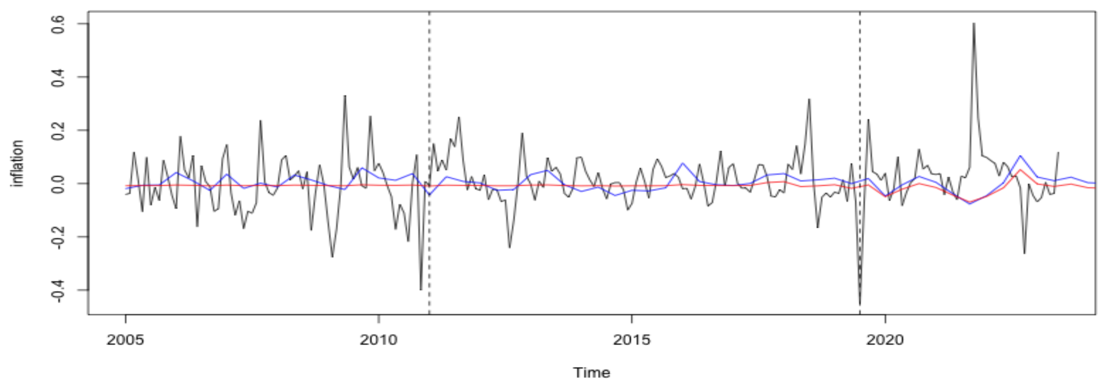
Incorporating structural time series components, the "StructTS Trend" model demonstrates accuracy but with higher errors (MAE: 3.09, RMSE: 4.66). The ARMAX model, integrating autoregressive and exogenous variables, shows moderate accuracy, recording a MAE of 4.51 and RMSE of 5.52.

Linear regression models, including "Linear Regression (Year)," "Linear Regression," and "Linear Regression (Month)," present progressively higher MAE and RMSE values, indicating a trend of decreasing predictive performance. "Linear Regression (Year)" performs slightly better than the others within this category.

Among the models considered, the ARIMA model with a structure of $c(1, 0, 0)$ concludes the analysis with a MAE of 5.35 and RMSE of 6.29, indicating satisfactory but comparatively less accurate forecasting performance.

3.1 Breakpoint Analysis

In 2011, Turkey grappled with the Syrian Civil War's impact as it hosted a significant number of Syrian refugees fleeing violence. Concurrently, the country faced two devastating earthquakes in the eastern region, particularly in Van province, highlighting its vulnerability to natural disasters. Additionally, In 2019, the world faced an unprecedented challenge with the outbreak of the COVID-19 pandemic caused by the novel coronavirus, SARS-CoV-2.



In summary, the choice of a forecasting model for inflation in Turkey should be guided by a careful consideration of the specific characteristics of the time series data and the desired balance between accuracy and model complexity. The analysis suggests that models incorporating structural changes and VAR models tend to offer better predictive performance, and further exploration and refinement, especially focusing on these models, may lead to enhanced forecasting outcomes. Additionally, incorporating exogenous variables and conducting a deeper analysis of model suitability are recommended for a comprehensive forecasting solution.

4 Future Work

In order to improve the depth and efficacy of the analysis, this project's future work may go various different directions for investigation and improvement. Here are some potential directions for future research:

Exogenous Variable Integration: One can investigate how more exogenous variables affect inflation forecasts. Inflation patterns may be influenced by variables like governmental fiscal policies, international economic data, or geopolitical events. By include these factors in the models, their prediction power may be increased, and a more thorough grasp of the factors influencing inflation in Turkey may result.

Machine Learning Approaches: We can investigate the use of cutting-edge machine learning methods to anticipate inflation. Prediction accuracy may be increased by using algorithms like neural networks, gradient boosting, or deep learning architectures, which are capable of identifying complex patterns and nonlinear correlations in the data. Examine how well these machine learning models perform in comparison to the conventional time series and regression models that were employed for this project.

Long-Term Forecasting: The forecasting horizon can be extended beyond the present four months. In evaluating the models' performance for longer-term forecasts, take into account the possible difficulties involved in projecting further into the future. Examine how the accuracy of your forecasts for the short and long terms may be traded off, and look at ways to reduce the uncertainty in your longer-term projections.

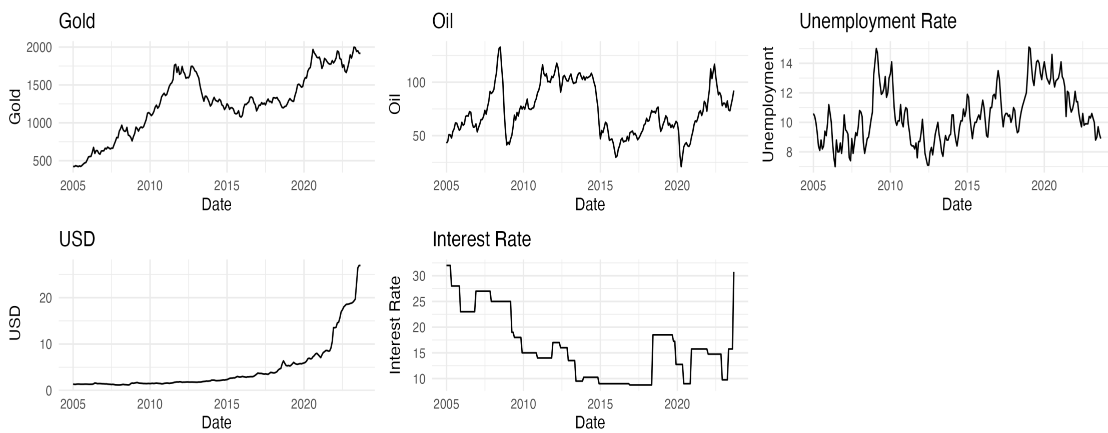
A Appendix

A.1 Data Source

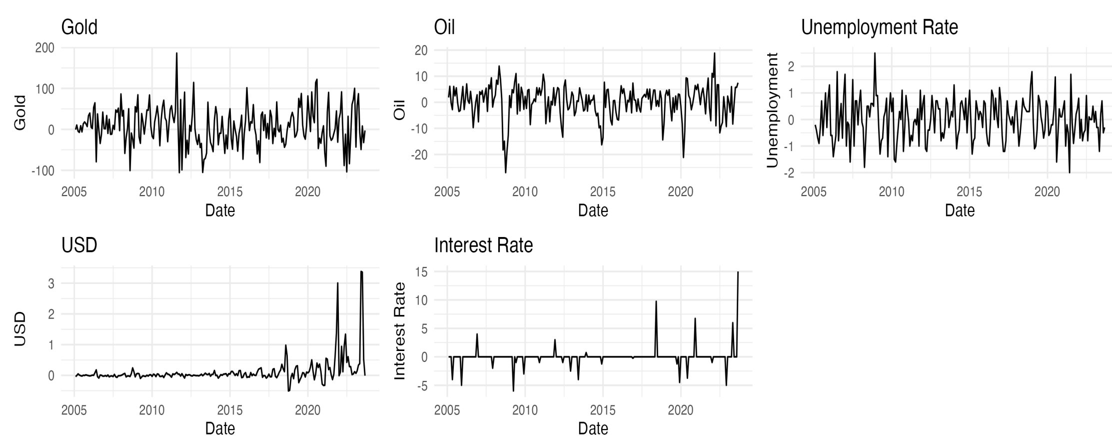
Variables	**Source**
Inflation	OECD
Gold Data	World Bank
Oil	World Bank
Unemployment	TUIK
Exchange Rate	Central Bank of the Republic of Türkiye
Interest Rate	FRED

Table 2: Data Sources

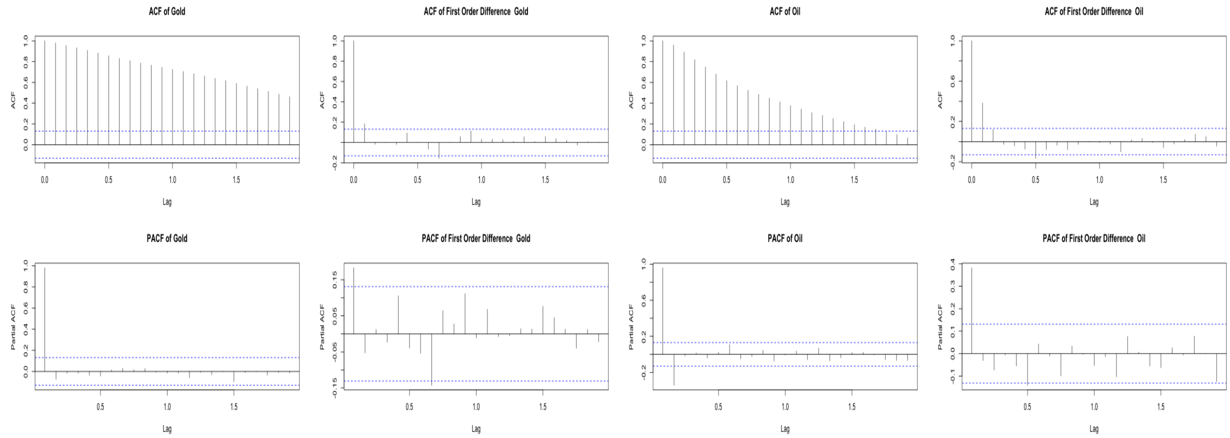
A.2 Independent Variables Trends

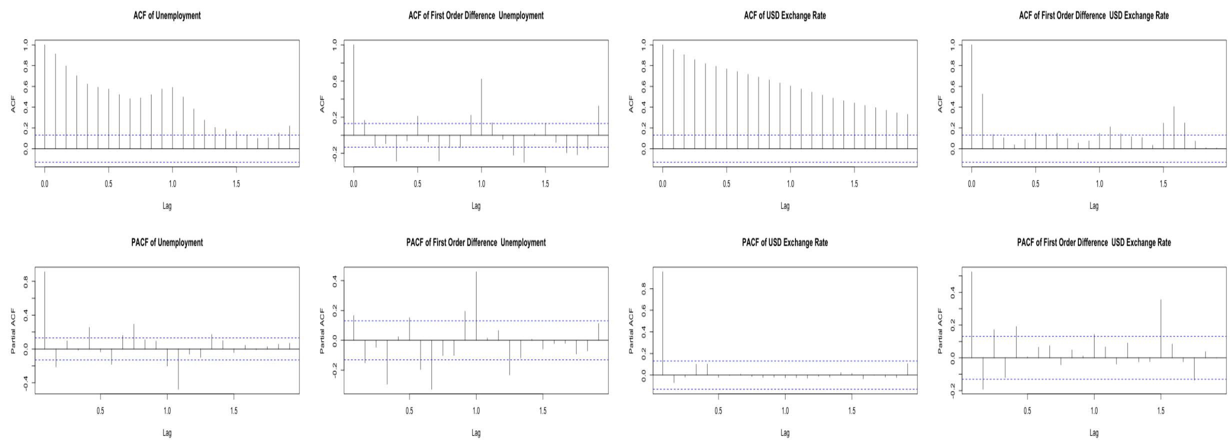


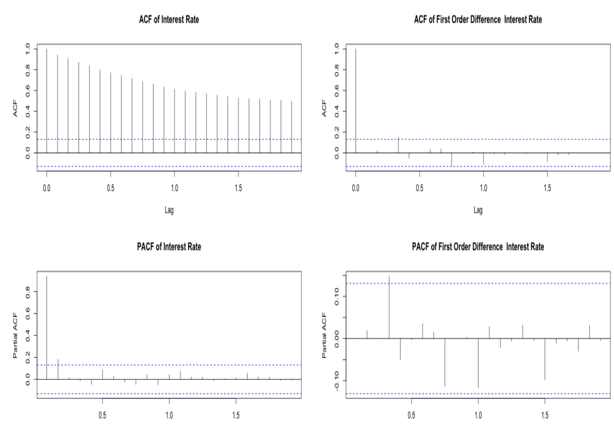
A.3 Independent Variables - first difference



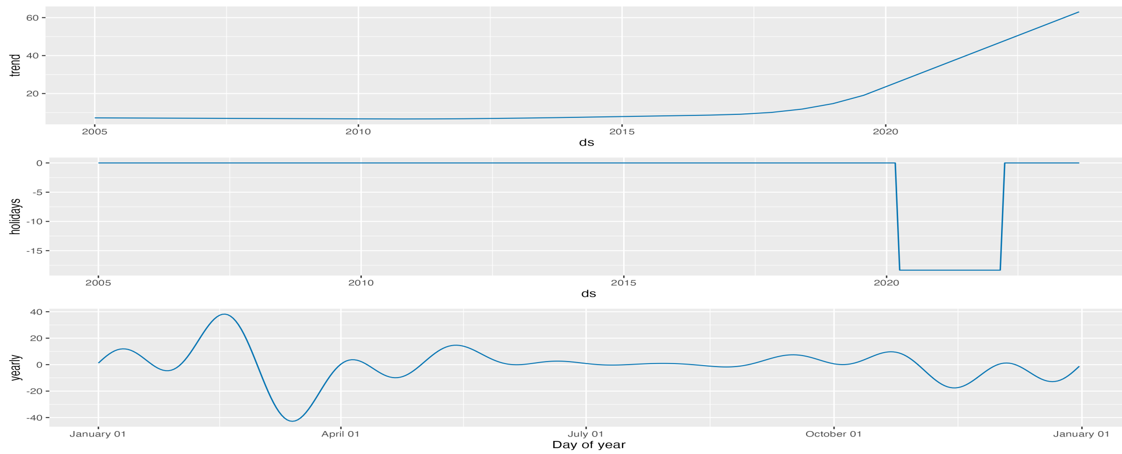
A.4 Independent Variables - Check Stationarity







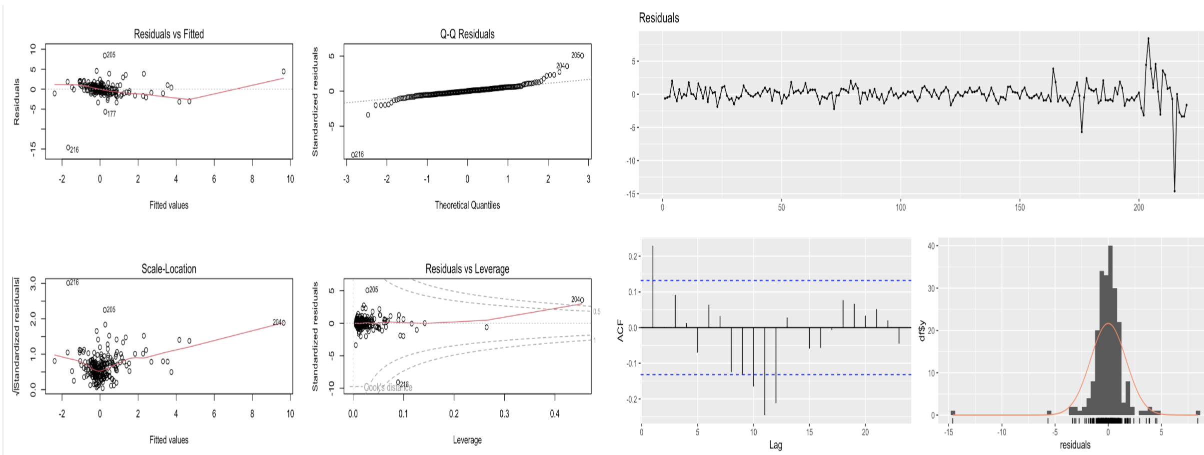
A.5 Decomposition of Prophet Forecast: Trend, Seasonality, and Holidays



A.6 Linear Regression Model

```
lm(inflation ~ gold_data + oil + unemployment + usd +
int_rate, data=train_data)
```

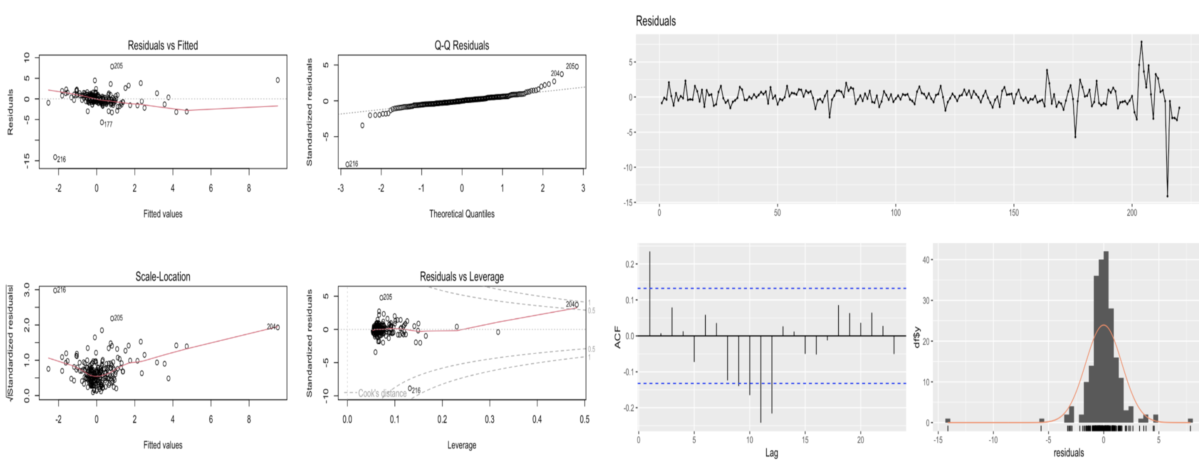
```
Call:
lm(formula = inflation ~ gold_data + oil + unemployment + usd +
    int_rate, data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-14.6249  -0.6207   0.0196   0.6222   8.4220

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.062111   0.120086  -0.517  0.60553
gold_data    -0.004993   0.002491  -2.004  0.04628 *
oil           0.051394   0.018547   2.771  0.00608 **
unemployment  0.121464   0.151214   0.803  0.42271
usd           3.282196   0.386312   8.496 3.33e-15 ***
int_rate      0.197227   0.086739   2.274  0.02397 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.687 on 214 degrees of freedom
Multiple R-squared:  0.2977,	Adjusted R-squared:  0.2813
F-statistic: 18.14 on 5 and 214 DF,  p-value: 5.358e-15
```



A.7 Linear Regression Model - factor month

```
lm(inflation ~ gold_data + oil + unemployment + usd + int_rate + month, data = train_data)
```

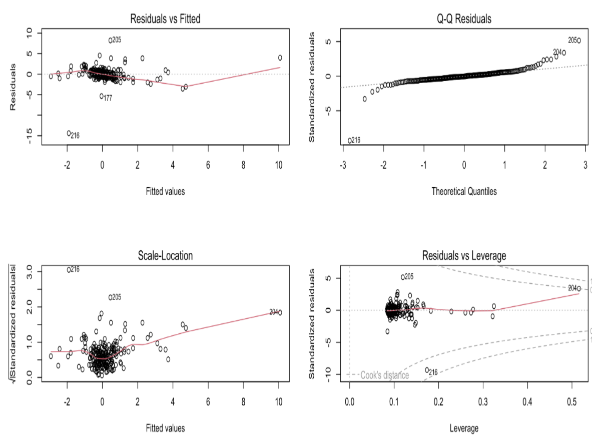
```
Call:
lm(formula = inflation ~ gold_data + oil + unemployment + usd +
    int_rate + month, data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-14.1219  -0.6231   0.0044   0.7431   7.8589

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.156898   0.396445   0.396   0.6927
gold_data    -0.005891   0.002604  -2.263   0.0247 *
oil           0.049442   0.019217   2.573   0.0108 *
unemployment  0.009590   0.215165   0.045   0.9645
usd           3.392963   0.395949   8.569 2.63e-15 ***
int_rate      0.184506   0.088758   2.079   0.0389 *
month02      -0.589777   0.577196  -1.022   0.3081
month03      -0.266995   0.580890  -0.460   0.6463
month04      -0.424596   0.572769  -0.741   0.4594
month05      -0.378568   0.566343  -0.668   0.5046
month06       0.045527   0.590164   0.077   0.9386
month07      -0.062155   0.569339  -0.109   0.9132
month08      -0.236181   0.568976  -0.415   0.6785
month09      -0.222332   0.568412  -0.391   0.6961
month10      -0.303024   0.567880  -0.534   0.5942
month11      -0.687733   0.600613  -1.145   0.2535
month12       0.466425   0.590722   0.790   0.4307
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.709 on 203 degrees of freedom
Multiple R-squared:  0.3167,    Adjusted R-squared:  0.2629
F-statistic: 5.881 on 16 and 203 DF,  p-value: 1.753e-10
```

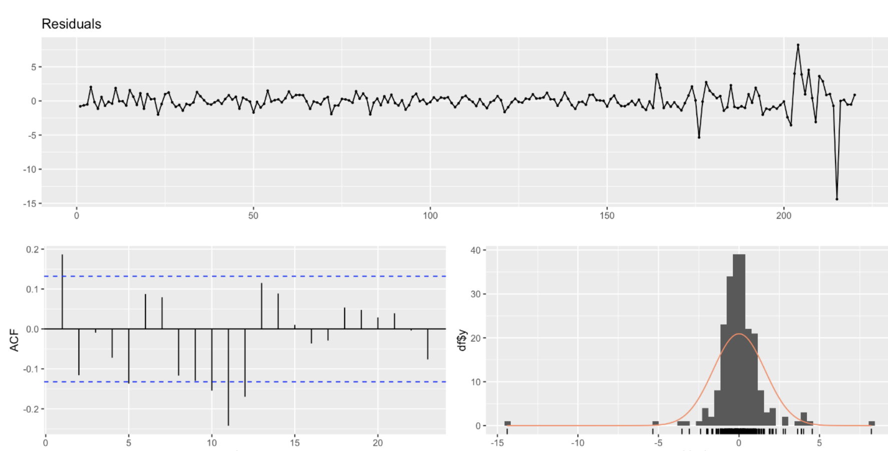


A.8 Linear Regression Model - factor year

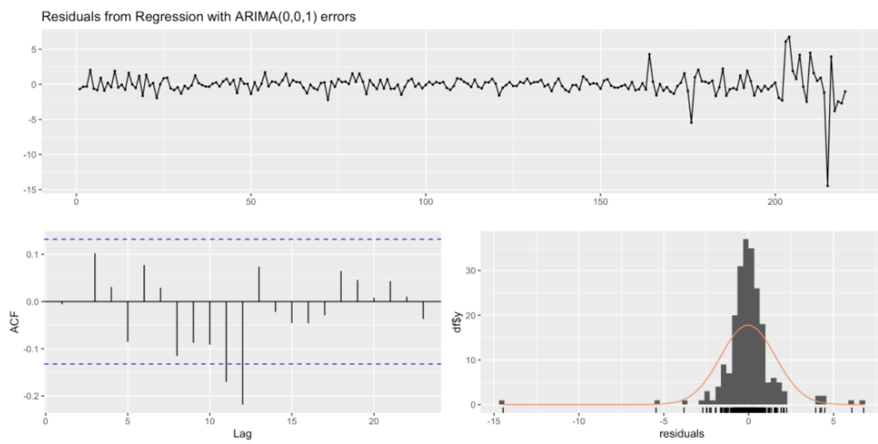
```
lm(inflation ~ gold_data + oil + unemployment + usd + int_rate + year, data = train_data)
```

```
Call:
lm(formula = inflation ~ gold_data + oil + unemployment + usd +
    int_rate + year, data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-14.3960  -0.6111  -0.0494   0.5531   8.2178

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.096025   0.496976   0.193  0.84699
gold_data    -0.004495   0.002655  -1.693  0.09209 .
oil           0.049244   0.019666   2.504  0.01310 *
unemployment  0.088498   0.155938   0.568  0.57101
usd           3.334500   0.446254   7.472 2.55e-12 ***
int_rate      0.246677   0.092553   2.665  0.00833 **
year2006     -0.185515   0.702915  -0.264  0.79212
year2007     -0.177520   0.696874  -0.255  0.79919
year2008      0.002956   0.706192   0.004  0.99666
year2009      0.167748   0.694828   0.241  0.80948
year2010     -0.400708   0.698414  -0.574  0.56680
year2011      0.336725   0.702252   0.479  0.63212
year2012     -0.274133   0.696134  -0.394  0.69416
year2013     -0.257682   0.705556  -0.365  0.71534
year2014      0.241708   0.706814   0.342  0.73274
year2015     -0.158479   0.701816  -0.226  0.82158
year2016     -0.402949   0.697743  -0.578  0.56426
year2017      0.046303   0.698599   0.066  0.94722
year2018     -0.198863   0.713983  -0.279  0.78090
year2019     -0.496029   0.696354  -0.712  0.47711
year2020     -0.160730   0.705713  -0.228  0.82007
year2021      0.133408   0.732993   0.182  0.85577
year2022     -0.181784   0.721815  -0.252  0.80143
year2023     -3.032947   1.014171  -2.991  0.00314 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.699 on 196 degrees of freedom
Multiple R-squared:  0.3476,    Adjusted R-squared:  0.271
F-statistic:  4.54 on 23 and 196 DF,  p-value: 1.546e-09
```





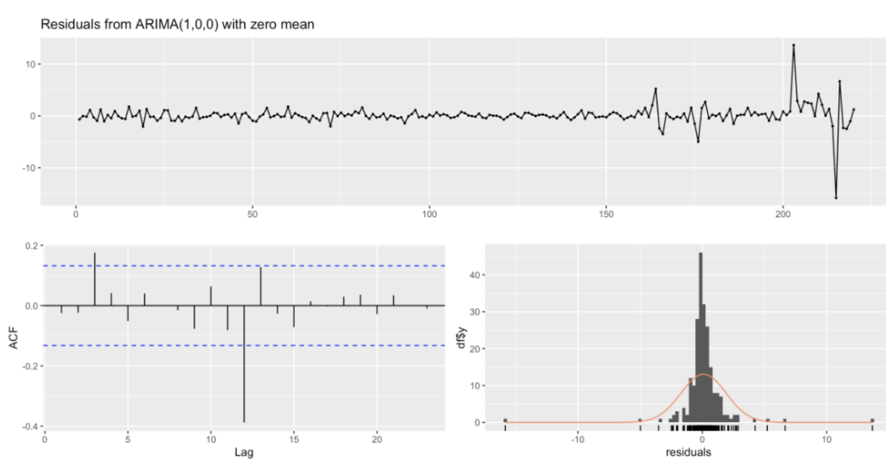
A.9 ARIMAX



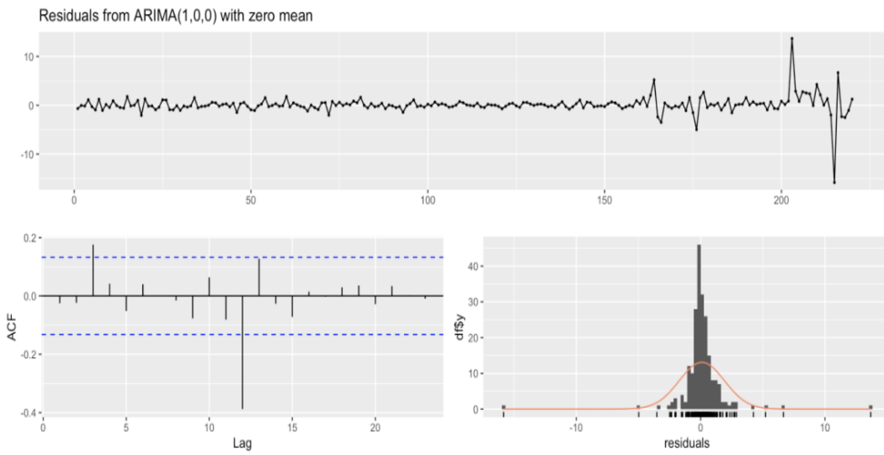
A.10 ARIMA c(1,0,0)

```
Series: diff_train_data
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1    mean
      0.4065  0.1649
s.e.  0.0614  0.2053

sigma^2 = 3.316:  log likelihood = -443.11
AIC=892.21   AICc=892.32   BIC=902.39

Training set error measures:
                     ME     RMSE       MAE       MPE   MAPE      MASE
Training set 0.00206075 1.812604 0.8462958 -11.29707 278.02 0.8083294
                   ACF1
Training set -0.0206396
MAE for ARIMA c(1, 0, 0): 5.513447
RMSE for ARIMA c(1, 0, 0): 6.306589
```

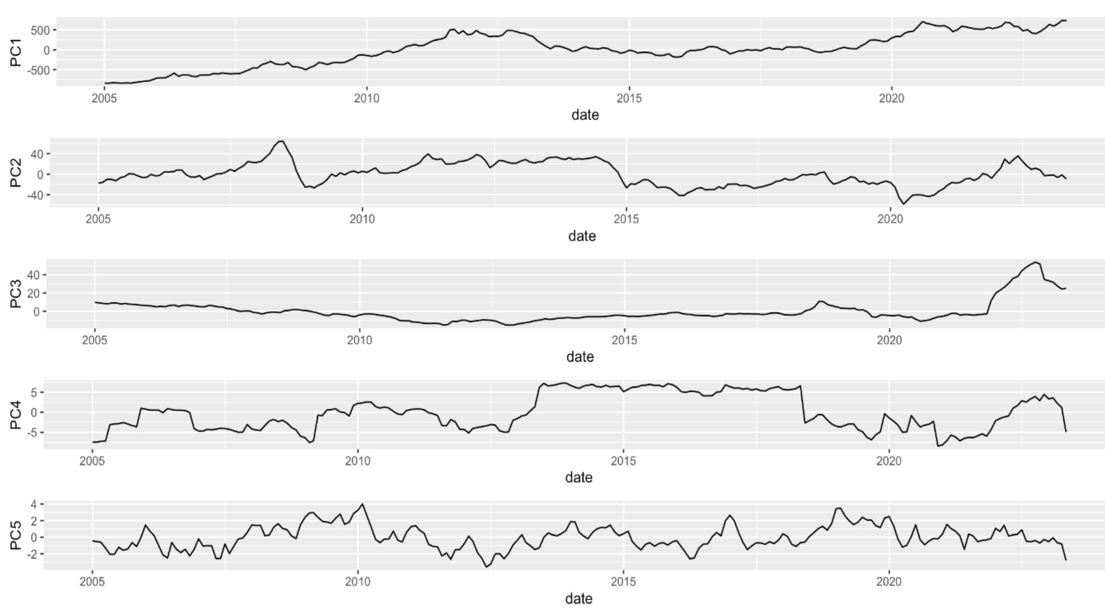
A.11 Reduced-form Arma Model

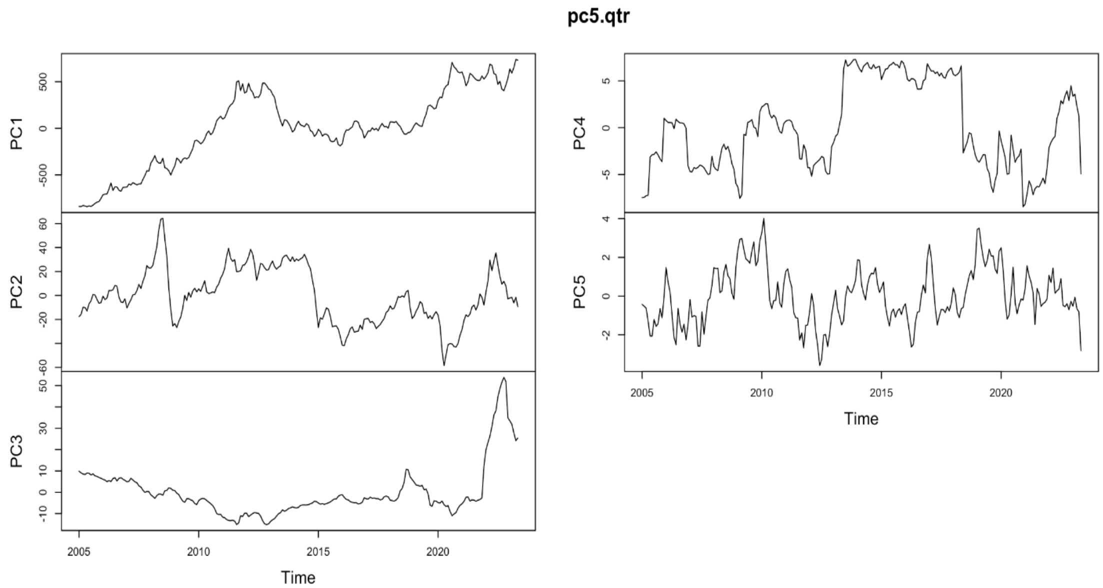


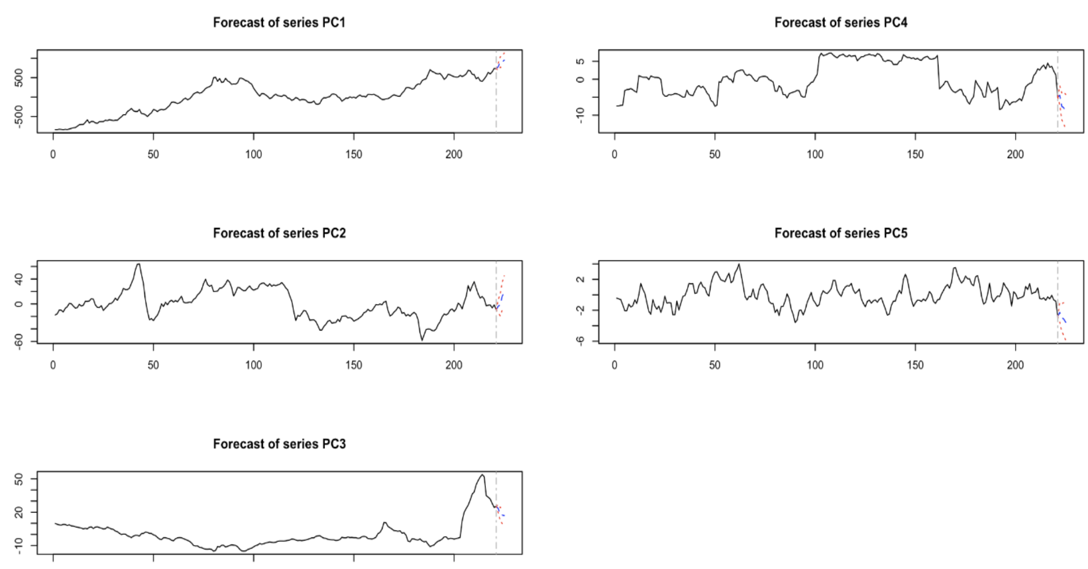
A.12 SARIMA



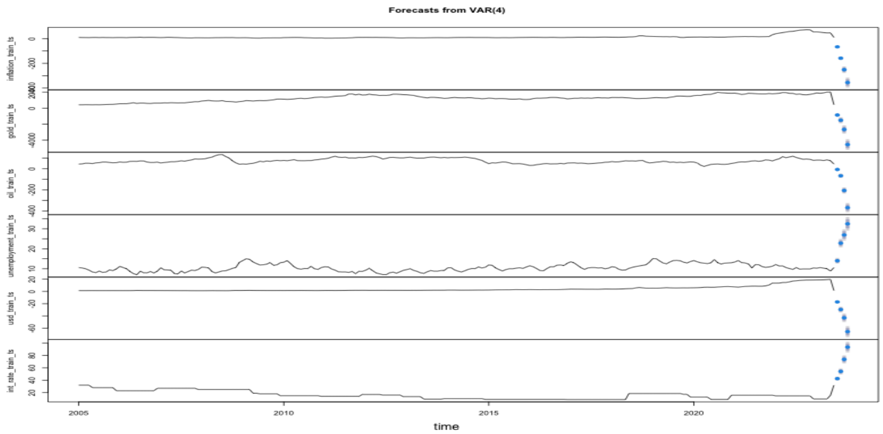
A.13 Dynamic Factor Model




pc5.qtr
PC1
PC2
PC3
PC4
PC5
2005
2010
2015
2020
Time


Forecast of series PC1
Forecast of series PC2
Forecast of series PC3
Forecast of series PC4
Forecast of series PC5
0
50
100
150
200

A.14 VAR Level



A.15 VAR Growth