Discover the Influence of Econmic Variables on CPI

Ruiming Min[1], Zicheng Wang[2], Qi Yang[3], & Shumeng Zhang[4]

[2] UIUC

Author Note

Abstract

This study employs time series analysis to investigate the Consumer Price Index (CPI), a crucial economic indicator reflecting changes in the average price level of goods and services. The main result reveals that incorporating economic variables, including the unemployment rate, real consumption, real government spending, and real investment, significantly enhances the prediction of CPI in the time series analysis model. The main result further confirms that positive coefficients for real consumption, real government spending, and real investment while negative for the unemployment rate indicate that these factors contribute to CPI growth aligning with economic theory. The CPI model's statistical significance, accuracy in fitting historical data, and precision in forecasting CPI values highlight the influential role of economic variables in shaping the CPI. This study provides valuable insights into the relationship between economic variables and CPI that may contribute to future research in an economic related field.

Discover the Influence of Econmic Variables on CPI

Introduction

Economics is a highly applicable field for utilization of time series analysis. The volume of data is abundant and inherently index with time. By using the time series statistical models, we can leverage historical data and catch trends and patterns for the data and make reasonable predictions.

Consumer Price Index (CPI)

In this project, we are analyzing the Consumer Price Index (CPI), a widely-used economic indicator that measures changes in the average price level of a basket of goods in an economic entity. It is commonly used to track inflation and assess price changes. A positive CPI value indicates inflation, while a negative value indicates deflation in the market. The goods' prices selected to calculate this index typically represent the overall prices of goods and services in society. CPI is calculated relative to the price in the previous time period, which is assigned a value of 100. The percentage difference between the current price and the previous price is reflected in the index using a weighted average formula. Generally, a CPI increase of 3% is considered inflation, while values over 5% may indicate significant inflation in the economic entity. There are also some variants of CPI. Such as core CPI which excludes volatile food and energy prices, and regional CPIs, which focus on CPI values in specific geographical areas. Given the critical role CPI plays in the economic landscape, we will introduce our approach to explore temporal and seasonal data patterns of CPI.

Related Research

One of the seminal paper in this field is by James H. Stock[**STOCK1999293**]. His research primarily centers around Phillips curve, which describes the connection of inflation and unemployment rate. In our own research, employment stands out as a crucial metrics

that we analyze. In recent years, there are some researches [**NGUYEN2023e20730**] [**BARKAN20231145**] employing machine learning and deep neural network for the prediction. These models exists strong performance, particularly with large dataset and when forecasting extended timeframes. However, it is important to note that due to the nature of these complex machine learning models, they often lack interpretability or the ability to provide clear explanations for their predictions.

Data Sources and Analysis Approach

Our primary data sources includes critical economic indicators, including the unemployment rate, government expenditure, investment, and consumption. These indicators play a pivotal role in measuring the level of economic activity within the entity. Furthermore, in our analytical approach, we use lagged and transformed Consumer Purchasing Index values, in conjunction with time-related variables, to enrich the comprehensiveness and precision of our assessments. It's worth noting that all these data elements are aligned with our response variable and are collected on a monthly basis.

Methods

To address the problem at hand, we used the following three steps to analyze the data: 1. detrending, 2. lag analysis, 3. adding the economic variables.

Detrending

Observing the original data, the CPI has a significantly exponential trend. So we first use a exponential model to fit the data. Moreover, according to the nature of the CPI, which is highly related to inflation rate, and the avarage of inflation rate, we choose 1.05 as the base of the exponential model.

CPI vs. Time

CPI and trend vs. Time
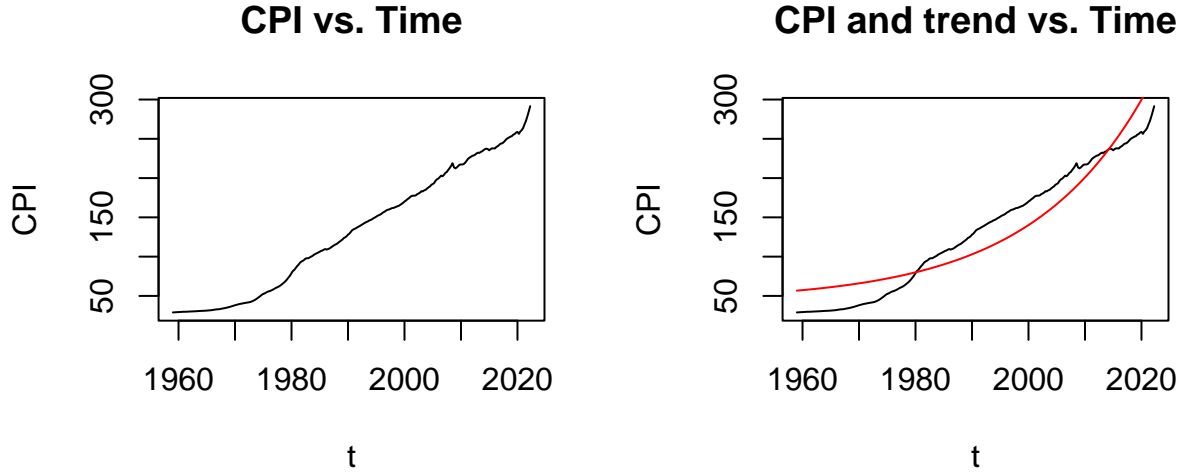
Adding the economic variables

After the detrending and adding the first order lag of CPI, the time series of CPI, $\{CPI(t)\}$, has became $CPI(t) = T(t) + CPI(t-1) + Y_t$.

Y_t vs. Time

Y_t/CPI vs. Time
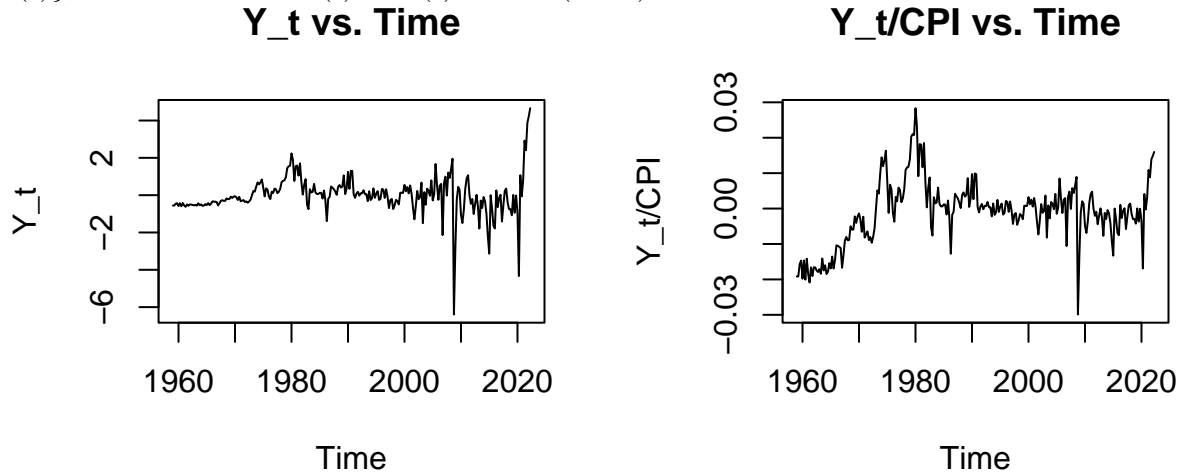
The plot above shows that Y_t still dose not behave like a white noise. Additionally, the significant depressions at 2008 and 2020 and the significant peak at 1979 reflexed the influence of the economic crisis[**gross2019iran**] [**williams2010uncontrolled**] [**forbes_2019_strange_new_world**]. So we add some economic variables to the model to regress the final model. Also, we add the sine functions of time to the model to fit the seasonal effect of the business cycle.

According to the nature of CPI [**blanchard2004macroeconomics**], $CPI(t) = \frac{\sum_{i \in \mathcal{P}} P_{i,t} Q_i}{\sum_{i \in \mathcal{P}} P_{i,0} Q_i}$, where $\mathcal{P}$ is the set of all goods and services, $P_{i,t}$ is the price of good

or service i at time t, and Q_i is the quantity of good or service i at the base period. Since

the Wage-Setting Relation, $W = \mathcal{A}P^e F(u, z)$, where W is the nominal wage, $\mathcal{A}$ is the price

mark-up, $F(u, z)$ is the function of unemployment rate and other variables, u is the

unemployment rate, and z is the output(GDP), and the Price-Setting Relation,

$P = (1 + m)\frac{W}{\mathcal{A}}$, where m is the desired mark-up, it is concluded as

$$CPI(t) = \frac{\sum_{i \in \mathcal{P}}(1 + m_i)\frac{W_i}{\mathcal{A}}Q_i}{\sum_{i \in \mathcal{P}} P_{i,0}Q_i} = \frac{\sum_{i \in \mathcal{P}}(1 + m_i)P^e_{i,t}F(u, z)Q_i}{\sum_{i \in \mathcal{P}} P_{i,0}Q_i}$$

.

To simplify the formular, we assume $m_i = m \; \forall i$ and $P^e_{i,t} = P_{i,t-1}$, we get

$CPI(t) = CPI(t-1)(1 + m)F(u, z)$. By the assumption of m is small and

$F(u, z) = 1 - \alpha u + \beta z$, we get $CPI(t) = CPI(t-1)(1 + m)(1 - \alpha u + \beta z)$

$\cong CPI(t-1)(1 + m - \alpha u - \beta z)$. We could conclude

$CPI(t) = CPI(t-1)(1 + m - \alpha u + \beta z)$. For easy-computation, we assume CPI(t-1) is a

constent we get $CPI(t) = CPI_0(1 + m - \alpha u + \beta z)$. Therefore we add the unemployment

rate, the real consumption, real grovement spending, and the real investment to the model.

Moreover, we add the sine functions of time to the model to fit the seasonal effect of

the business cycle and the square of time to fit the trend of the CPI.

To summarize the first model, we get the model as follows:

$$
\begin{aligned}
CPI(t) =\, & a_0 + a_1 \cdot 1.05^t + a_2(t - 1959)^2 && \text{(trend)} \\
& + a_3 \sin\left(\frac{2\pi(t - 1959)}{40}\right) + a_4 \sin\left(\frac{2\pi(t - 1980)}{24}\right) + a_5 \sin\left(\frac{2\pi(t - 1980)}{48}\right) && \text{(business cycle)} \\
& + b_1 C_t + b_2 u_t + b_3 G_t + b_4 I_t + W_t && \text{(economic variables)}
\end{aligned}
$$

Residuals vs. Time

Residuals/CPI vs. Time
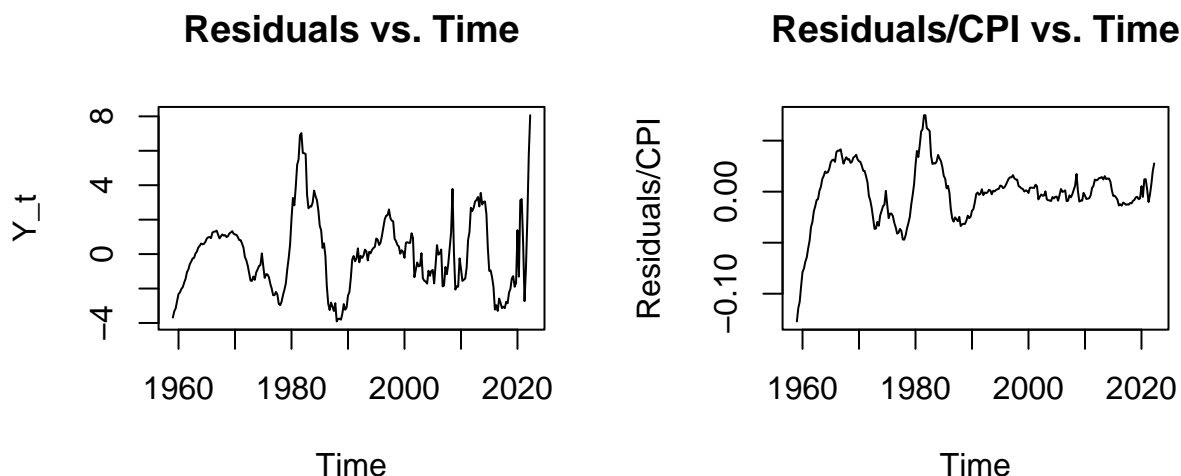
The plot above shows that residuals still dose not behave like a white noise. Therefore, we will use ACf to figure out weather or not the data has lags and improve our model.

Lag Analysis

After conclude the first model, we use PACF to figure out weather or not the data has lags. From the plot below, the data has one lags. Therefore, the CPI may be a AR(1) process and we improve our model by adding the first oreder lag of CPI to the model.

Residuals of first model vs. Time

PACF of Residuals of first mode
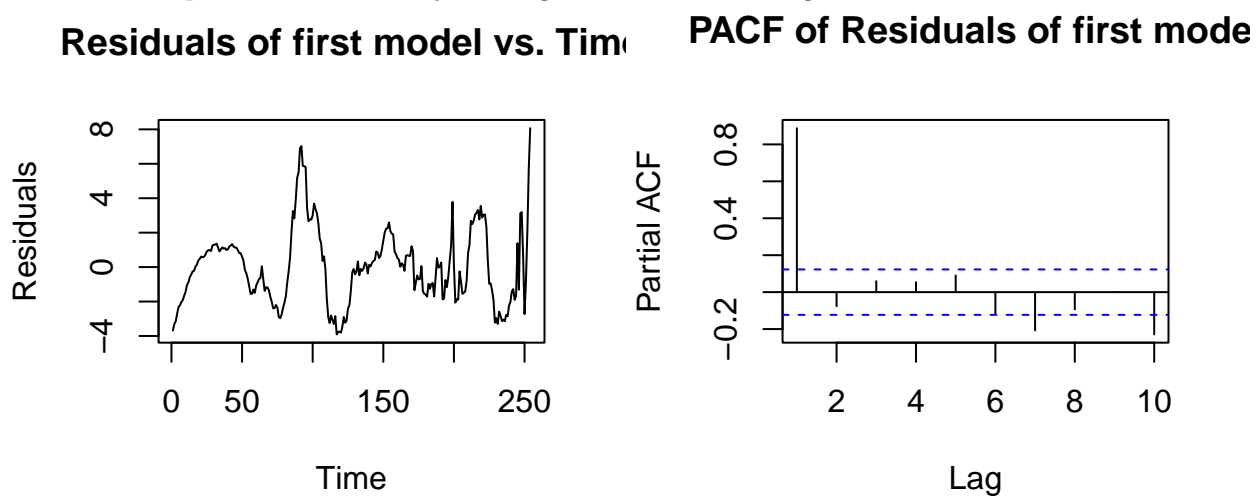
Moreover, we delete the square of time from the model since it is not significant.

To summarize the second model, we get the final model as follows:

$$CPI(t) = a_0 + a_1 \cdot 1.05^t + a_2 CPI(t-1) \qquad \text{(trend and lags)}$$

$$+ a_3 \sin\left(\frac{2\pi(t-1959)}{40}\right) + a_4 \sin\left(\frac{2\pi(t-1980)}{24}\right) + a_5 \sin\left(\frac{2\pi(t-1980)}{48}\right) \quad \text{(business cycle)}$$

$$+ b_1 C_t + b_2 u_t + b_3 G_t + b_4 I_t + W_t \qquad \text{(economic variables)}$$

where W_t is a white noise.

Results

Fitting Results

The fitted result of the model is shown below.

| Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.159	4.662	4.632	5.90	$\times\ 10^{-6}$
I(1.05^t)	-8.779×10^{-42}	1.245	-7.050	1.83	$\times\ 10^{-11}$
sin_1	-3.881	6.540	-5.934	1.01	$\times\ 10^{-8}$
sin_2	-1.028	1.593	-6.455	5.80	$\times\ 10^{-10}$
sin_3	-3.760	9.484	-3.965	9.66	$\times\ 10^{-5}$
CPI_lag1	9.735×10^{-1}	1.351	72.033	<2	$\times\ 10^{-16}$
unemp_rate	1.847×10^{-1}	5.406	3.416	0.000745	
grove_exp	6.250×10^{-4}	2.178	2.870	0.004470	
invest	7.797×10^{-4}	3.737	2.087	0.037966	
consump_real	3.423×10^{-3}	3.337	10.259	<2	$\times\ 10^{-16}$

Residual standard error: 0.8011 on 244 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 2.688e+05 on 9 and 244 DF, p-value: < 2.2e-16

After we add the selected economic variables including the unemployment rate, real consumption, real government spending, and real investment to the model. We generate an

overall table above to examine the fit of the variables. From the table above, we can find out all the added economic variables have p-values 0.000745 for the unemployment rate,$< 2 * 10^{-16}$ for real consumption, 0.004470 for real government pending,0.037966 for real investment, therefore, we can conclude that all the added economic variables in the table are statistically significant. For the entire model, we can find out the R squared is 0.9999 which is large enough to conclude that the added economic variables fit the data well, and the p-value for the F test is less than $\alpha(0.05)$, therefore, we can conclude that the model is statistically significant. Based on the result above, we can conclude that CPI will be influenced by the unemployment rate, real consumption, real government spending, and real investment.

Fitted CPI & real CPI vs. Time
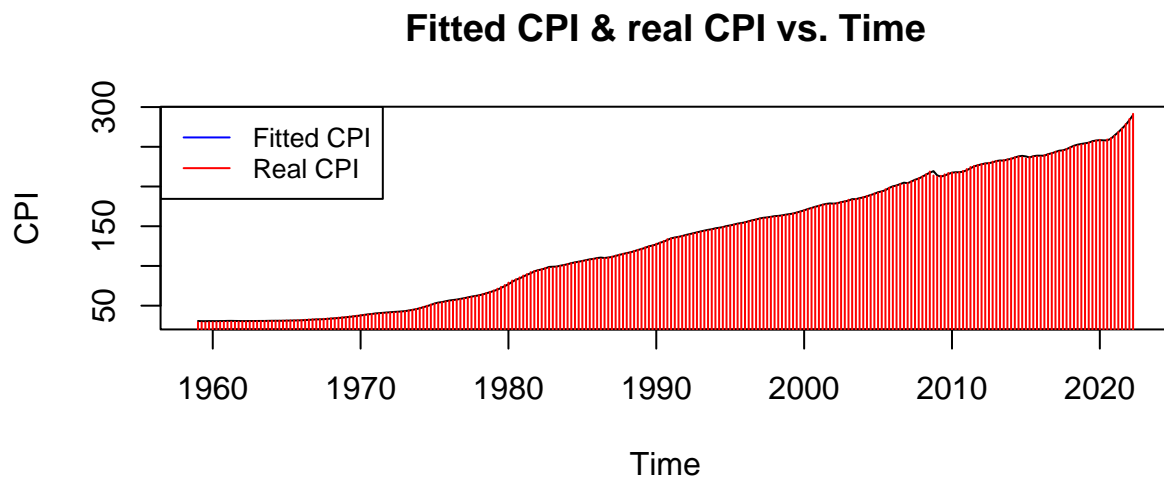
The graph above is the comparison between our model after adding the economic variables and the real CPI versus the time that we can find in the data. The model seems to completely fit with the real CPI. Therefore, the graph further confirms our research and the added economic variables are essential for fitting a CPI model to contribute significantly to the model's precision.

**Residuals vs. Time**   **PACF of Residuals of final mode**
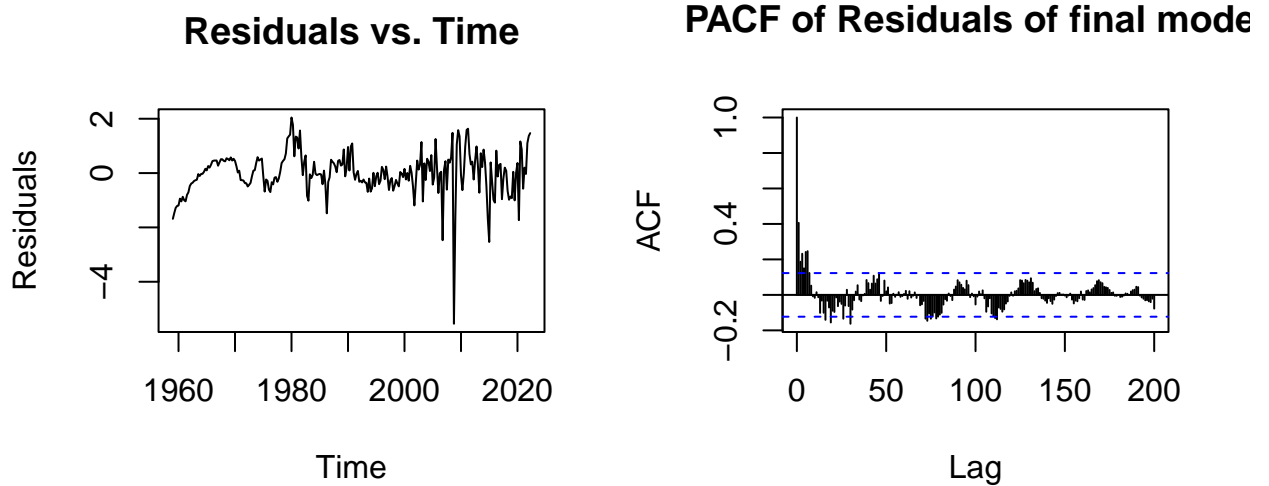
To check our model, we perform an analysis of residuals to ensure that the CPI model's predictions are unbiased. We can find out the overall residuals plot shows the difference between predicted CPI values and the real CPI against time. Although a depression in 2010 and a relatively upward trend the 1960 to 1970 is noticeable, the residuals overall resemble white noise at other points, indicating a better fit of the model.[**inflation1960-1970**] The reason for the upward trend from 1960 to 1970 may be because of the robust consumer spending after World War II, the low unemployment rate, and Vietnam War spending. All these factors are included in the economic variables added to the model, therefore, these may be the reason for the upward trend. Also, for the noticeable depression before 2010, the reason may be the financial crisis in 2008[**financial_crisis**]. These are the cases that are worth the discussion in the future. Additionally, the partial autocorrelation function (PACF) plot demonstrates that the correlations of residuals largely fall within the confidence interval, affirming the CPI model is significant.

Forecasting Results

The forecasting result of the model is shown below. The analysis of the percentage prediction errors over time reveals a promising performance of the model in predicting the CPI. The plot illustrates that the errors are consistently small across the entire time range,

with the majority of the percentage errors approximately around zero. This indicates a remarkable alignment between the predicted CPI values and the actual CPI values during the test period. Furthermore, the mean squared error is $7.185918 * 10^{-5}$ which confirms the model's precision. The small errors and low mean squared error suggest that the model has effectively captured the CPI data, showing its reliability in forecasting future CPI values. , the prediction is also a confirmation of the importance of the added economic variables the unemployment rate, real consumption, real government spending, and real investment in the model.

Precnet of Predict Error vs. Time
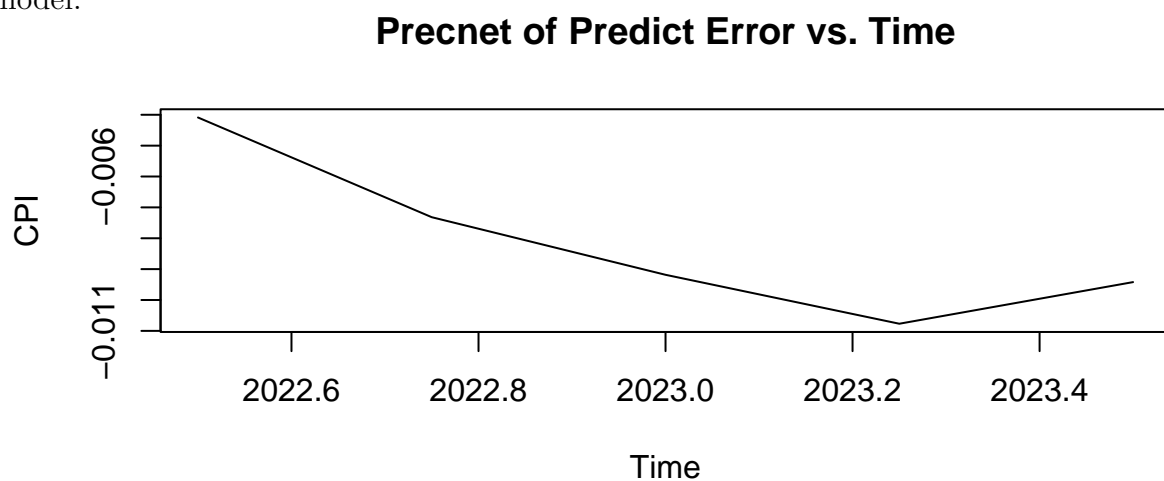
And the mean square error of the model is 7.185918e-05

rest. In conclusion, the model's statistical significance, accuracy in fitting historical data, and precision in forecasting CPI values highlight the influential role of economic variables in shaping the CPI. The positive coefficient estimates for real consumption, real government spending, and real investment indicate that an increase in these variables is associated with a rise in the CPI. This aligns with standard economic theory, where increased consumer spending, government expenditure, and investment often contribute to overall economic growth and can lead to higher inflation.Conversely, the negative coefficient estimate for the unemployment rate suggests an inverse relationship with the CPI. Higher unemployment rates tend to be associated with economic downturns, leading to decreased consumer spending and overall demand, which can put downward pressure on prices and inflation. As the analysis encompasses both statistical and practical significance,

these findings contribute valuable insights for understanding the relationship between economic variables and the CPI, providing insights for investigation further.

Discussion

A synopsis of your findings and any limitations your study may suffer from.

References

Appendix (Optional)

Any R codes or less important R outputs that you wanted to keep- can go in here.