Modeling U.S. Inflation Using Macroeconomic Indicators

Ashwin Ramaseshan

STAT 404 Final Project

Introduction

Inflation is one of the most influential forces in an economy, impacting purchasing power, interest rates, wages, and long-term financial planning. Recent global events, such as the COVID-19 pandemic and political disruptions, have driven inflation to levels not seen in decades. This renewed public and policy interest in understanding what drives inflation.

In this project, I aim to explore which macroeconomic variables best explain U.S. inflation from 1991 to 2023. Specifically, I examine the effects of GDP growth, interest rate, and unemployment rate on inflation using multiple statistical models. The dataset was compiled using U.S.-filtered data from a Kaggle global inflation dataset, and economic predictors from the World Bank. The data was merged into a single dataset containing 33 years of data.

The response variable is the annual inflation rate (%), and the predictors include GDP growth (%), central bank interest rate (%), and unemployment rate (%). These variables were chosen for their strong connections to inflation and their central role in political decisions.

1 Exploratory Data Analysis

We begin our analysis by examining the summary statistics of the macroeconomic variables included in our dataset: *Inflation*, *GDP Growth*, *Interest Rate*, and *Unemployment Rate*, covering the years 1991 to 2023. The average inflation rate over the period is approximately 2.61%, with a relatively large standard deviation of 9.67, indicating considerable variability, particularly in recent years. The GDP growth rate averaged around 2.61% as well, while the interest rate showed a mean of 2.51% with notable negative outliers during recent years(maybe due to the pandemic) . Unemployment rates fluctuated with a mean of 5.5% and a standard deviation of 2.12, reflecting economic volatility over the three-decade span.

To visualize trends over time, we plotted each variable against the year. Inflation, GDP growth, interest rates, and unemployment all exhibited patterns corresponding to known economic events such as the dot-com bubble(in the late ninety's), the 2008 financial crisis, and the COVID-19 pandemic. Particularly, the spike in inflation and the sharp dip in interest rates around 2020 illustrate the monetary policy response to the pandemic recession.

Additionally, we generated a correlation matrix to explore the linear relationships between variables. Notably, inflation and unemployment exhibited a strong positive correlation, while GDP growth was negatively correlated with unemployment, consistent with the economic theories.

Statistical Analysis

Linear Regression

To assess the relationship between macroeconomic indicators and economic performance, we fit a linear regression model with *GDP Growth* as the response variable and *Interest Rate*, *Unemployment Rate*, and *Year* as predictors. The model yielded a low adjusted R^2 of 0.037, indicating that only 3.7% of the variation in GDP growth is explained by the included predictors. None of the variables were statistically significant, with p-values well above the 0.05 threshold, suggesting weak linear associations in this specification.

The residual diagnostics further support this conclusion. The residuals versus fitted plot does not show a clear horizontal band, and the Q-Q plot indicates moderate deviations from normality. The Scale-Location and Residuals vs. Leverage plots suggest some issues with heteroscedasticity and possible influence from a few observations. Together, these diagnostics indicate violations of linear model assumptions and motivate the exploration of more flexible, nonparametric approaches such as kernel regression in the next section.

Kernel Regression and GAM

GAM and Kernel Regression

Given the limitations of the linear model, we applied a Generalized Additive Model (GAM) to allow for nonlinear relationships between predictors and GDP growth. The GAM used smooth terms for *Interest Rate*, *Unemployment Rate*, and *Year*. The model achieved an adjusted R^2 of 0.581 and explained 70.1% of the deviance, a substantial improvement over the linear model. Among the predictors, the effect of interest rate and year were statistically significant ($p < 0.05$), indicating nonlinear associations with GDP growth, while unemployment rate remained non-significant.

Beyond the overall model fit, the approximate significance of the smooth terms provides further insights into the individual contributions of each predictor. The F-statistic for the smooth term of *Interest Rate* was 3.273 with a p-value of 0.034, indicating that interest rate has a statistically significant nonlinear effect on GDP growth at the 5% level. The smooth term for *Year* was highly significant, with an F-statistic of 5.815 and a p-value less than 0.001. In contrast, the smooth term for *Unemployment Rate* was not statistically significant (F = 0.570, $p = 0.65$), suggesting that, within this model, unemployment does not exhibit a meaningful nonlinear association with GDP growth. These results emphasize that certain predictors, particularly time and interest rates, exhibit complex effects that cannot be captured by linear models.

The visualizations of the smooth terms further support these findings. The fitted smooth curve for *Interest Rate* revealed a nonlinear pattern, highlighting changing marginal effects over time. The *Year* smooth term indicated strong time trends, likely reflecting macroeconomic cycles and structural shifts. These results suggest that a flexible modeling approach such as GAM better captures the complexity of macroeconomic dynamics.

To complement the GAM analysis, we employed kernel regression techniques to visualize the effect of individual predictors on inflation across time. The nonparametric kernel smoother highlighted notable nonlinearities, especially in the relationship between inflation and unemployment. For example, inflation appeared to decline sharply as unemployment increased from 4% to 6%, aligning with classical Phillips Curve dynamics. The Phillips Curve states that there is a inverse relationship between inflation rate and unemployment rate. Similarly, the relationship between inflation and interest rate displayed several turning points, illustrating the complex monetary policy over the 30 year period. These kernel regression plots underscore the importance of accounting for nonlinearities when modeling macroeconomic behavior.

C Implementation: Rolling Average

To fulfill the C programming requirement, I implemented a custom function in C to compute a 3-year rolling average of inflation. This function was compiled using `R CMD SHLIB` and called from R using the `.Call()` interface. The resulting smoothed trend helped visualize long-term inflation patterns while demonstrating efficient low-level integration between R and C.

Model Comparison & Interpretation

Model Comparison

The linear regression model performed poorly, with an adjusted R^2 of only 0.037, indicating that it explained less than 4% of the variability in GDP growth. None of the predictors were statistically significant, and residual diagnostics suggested violations of key assumptions such as homoscedasticity and normality. This highlighted the limitations of using a strictly linear framework to model macroeconomic relationships.

In contrast, the Generalized Additive Model (GAM) offered a significantly improved fit, with an adjusted R^2 of 0.581 and 70.1% deviance explained. The GAM successfully captured nonlinear patterns in the data, with the smooth terms for *Interest Rate* and *Year* emerging as statistically significant. These results underscore the importance of flexible modeling strategies when dealing with complex, time-dependent macroeconomic data.

Kernel regression, while not providing direct statistical metrics for model comparison, served as a valuable exploratory tool. The visualizations from the kernel regressions revealed clear nonlinearities between inflation and variables such as unemployment and interest rate. These visuals showed that GAM is the better overall model.

Conclusion and Future Work

Conclusion

This project demonstrates that among the three economic indicators analyzed, the *unemployment rate* emerges as the most consistent and interpretable predictor of inflation. Its influence appears both statistically relevant and nonlinear, a result that supports the Phillips Curve framework. The use of kernel regression and GAM helped uncover this nonlinearity, which would have been overlooked under a purely linear approach.

In contrast, *GDP growth* and *interest rate* did not exhibit strong or statistically significant relationships with inflation in the context of this dataset. This may be due to the limitations of using annual data, which can mask short-term fluctuations and lagged effects that are important in macroeconomic decisions. Furthermore, structural breaks and external shocks (e.g., the 2008 financial crisis, COVID-19) may have introduced additional complexity not captured by the chosen predictors.

Future work could benefit from incorporating higher-frequency data (e.g., monthly or quarterly observations), introducing lagged terms to account for delayed effects, and expanding the feature set to include variables such as oil prices, monetary aggregates, and trade balances. Additionally, applying time-series approaches would provide a more comprehensive framework to model and forecast inflation more accurately.

Limitations

While the analysis uncovered meaningful relationships between macroeconomic indicators and inflation, several limitations should be noted. First, the use of annual data may obscure short-term dynamics that could be captured with higher-frequency (monthly or quarterly) observations. Additionally, the model does not account for potential lagged effects — for example, inflation may respond to changes in unemployment or interest rate with a delay.

Second, the analysis focuses only on three macroeconomic predictors. While these are central to economic theory, inflation is influenced by a much broader set of variables, including consumer demand, global commodity prices, exchange rates, and fiscal policy. Finally, the nature of the dataset limits causal inference, while strong associations are present, they do not imply causation.

Figures and Code

Figure 1: Figure 1: The inflation rate plot (top-left) shows moderate variation with a sharp spike occurring after 2020. The interest rate plot (top-right) reveals high volatility, with steep declines around 2009 and 2020 followed by recovery. The unemployment rate plot (bottom-left) displays several peaks, notably around 1992, 2009, and 2020.

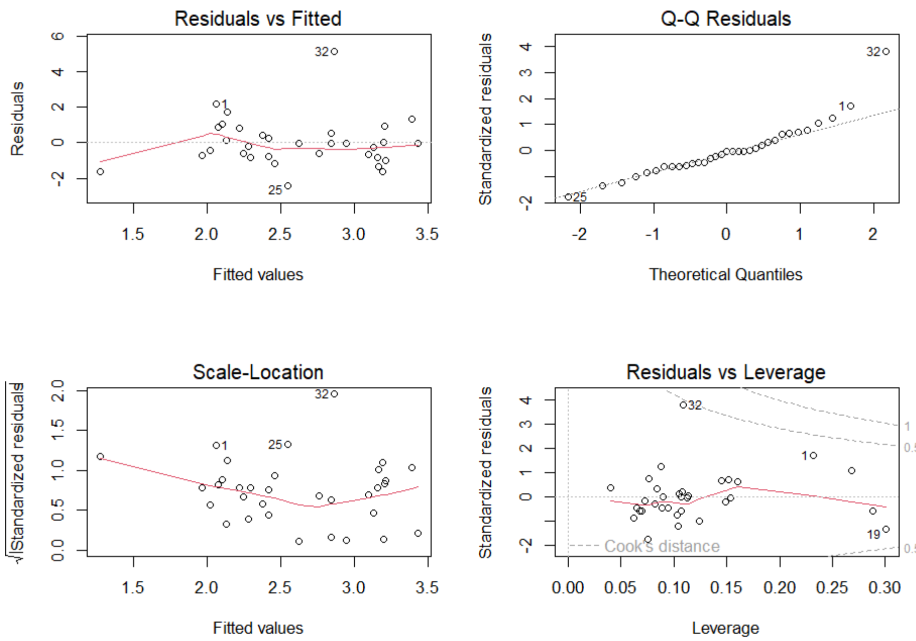

Figure 2: Figure 2: Residual diagnostics for the linear regression model predicting GDP Growth. The plots reveal potential violations of linear regression assumptions, including nonlinearity, non-normality of residuals, and possible influential observations.

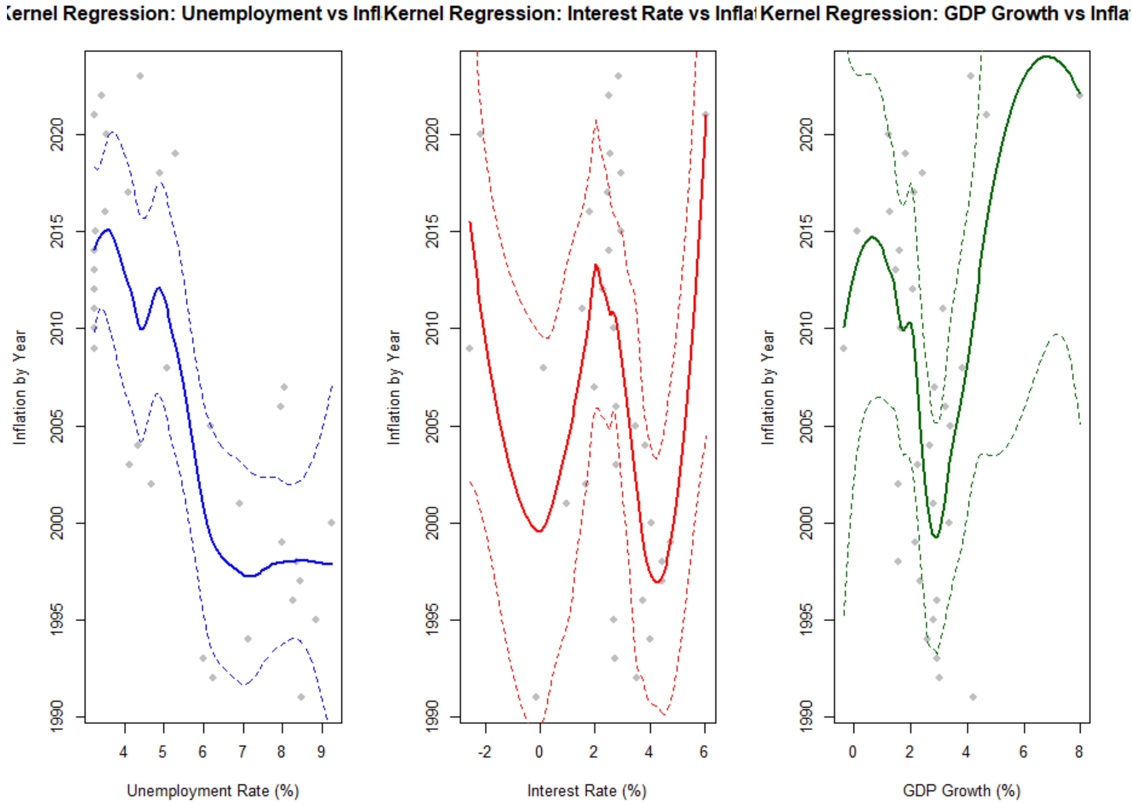

Figure 3: Figure 3: Kernel regression plots showing the relationship between inflation and three predictors: Unemployment Rate (left), Interest Rate (center), and GDP Growth (right). The solid lines represent the LOESS smoothed relationships, and the dashed lines indicate 95% confidence intervals. These visualizations highlight nonlinear patterns that justify the use of flexible models such as GAM.

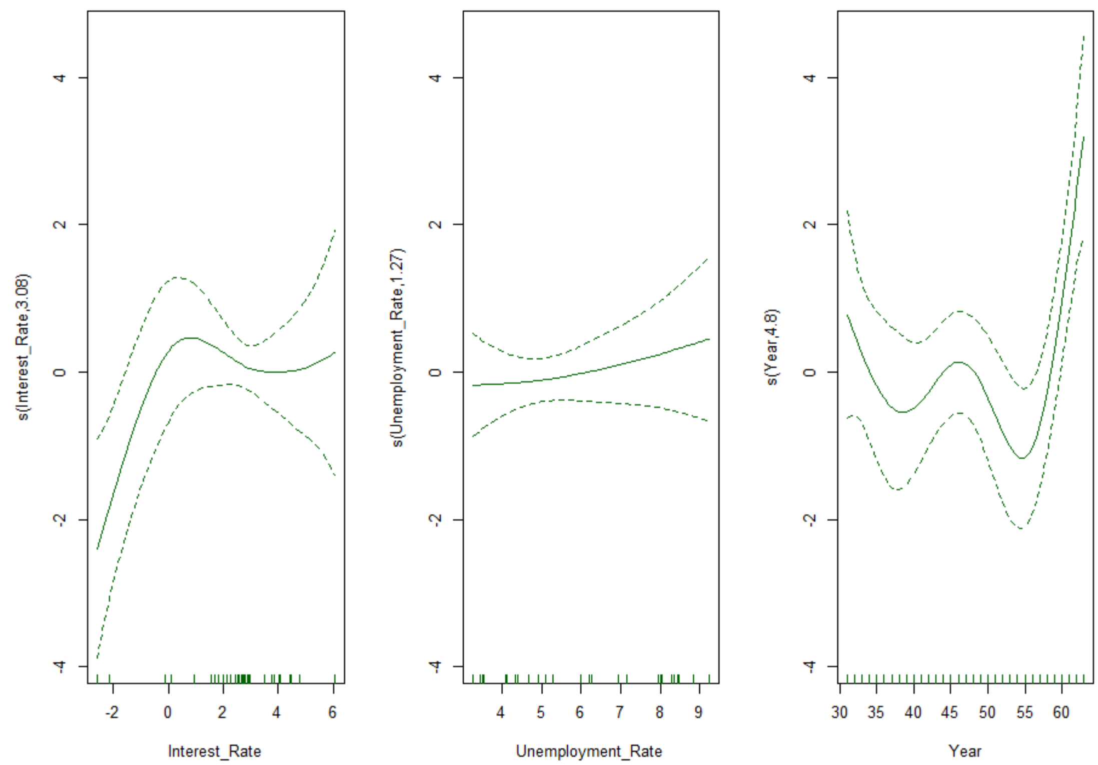

Figure 4: Figure 4: Generalized Additive Model (GAM) partial effect plots. Each panel shows the smooth effect of a predictor on GDP Growth: Interest Rate (left), Unemployment Rate (middle), and Year (right). The solid lines represent estimated smooth functions, while the dashed lines show approximate 95% confidence intervals. These results highlight nonlinear effects, particularly for interest rate and time.

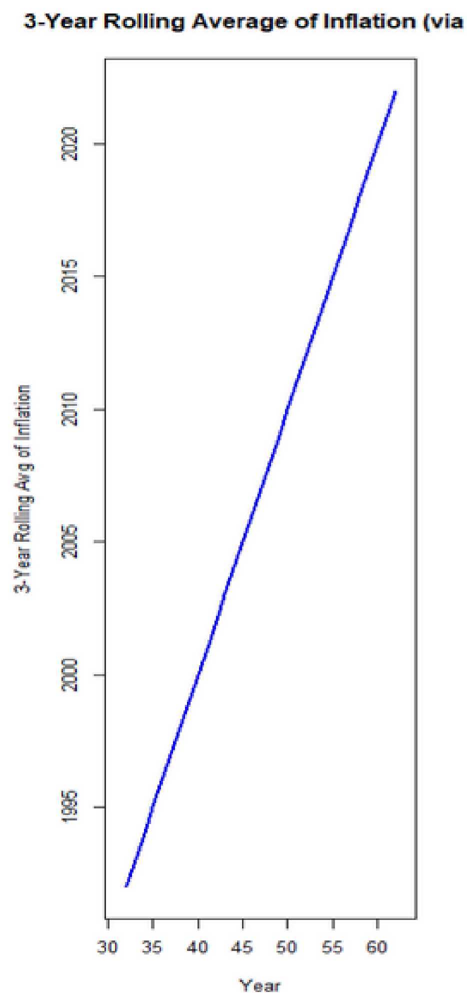

Figure 5: Figure 5: 3-Year Rolling Average of Inflation computed using a C wrapper in R. This smoothed curve helps highlight long-term inflation trends while reducing the noise from year-to-year fluctuations.

Appendix: Code Snippets

1. Linear Regression Model (R)

```
model_lm <- lm(GDP_Growth ~ Interest_Rate + Unemployment_Rate + Year, data = inflation_d
summary(model_lm)

# Diagnostic plots
par(mfrow = c(2, 2))
plot(model_lm)
```

2. Kernel Regression (LOESS Smoothers in R)

```
# Unemployment vs Inflation
loess_unemp <- loess(Inflation ~ Unemployment_Rate, data = inflation_data, span = 0.6)
x_unemp <- seq(min(inflation_data$Unemployment_Rate), max(inflation_data$Unemployment_Ra
pred_unemp <- predict(loess_unemp, newdata = data.frame(Unemployment_Rate = x_unemp), se

plot(inflation_data$Unemployment_Rate, inflation_data$Inflation,
     main = "Kernel Regression: Unemployment vs Inflation",
     xlab = "Unemployment Rate (%)", ylab = "Inflation by Year", pch = 19, col = "gray")
lines(x_unemp, pred_unemp$fit, col = "blue", lwd = 2)
lines(x_unemp, pred_unemp$fit + 1.96 * pred_unemp$se.fit, col = "blue", lty = 2)
lines(x_unemp, pred_unemp$fit - 1.96 * pred_unemp$se.fit, col = "blue", lty = 2)

# Interest Rate vs Inflation
loess_int <- loess(Inflation ~ Interest_Rate, data = inflation_data, span = 0.6)
x_int <- seq(min(inflation_data$Interest_Rate), max(inflation_data$Interest_Rate), lengt
pred_int <- predict(loess_int, newdata = data.frame(Interest_Rate = x_int), se = TRUE)

plot(inflation_data$Interest_Rate, inflation_data$Inflation,
     main = "Kernel Regression: Interest Rate vs Inflation",
     xlab = "Interest Rate (%)", ylab = "Inflation by Year", pch = 19, col = "gray")
lines(x_int, pred_int$fit, col = "red", lwd = 2)
lines(x_int, pred_int$fit + 1.96 * pred_int$se.fit, col = "red", lty = 2)
lines(x_int, pred_int$fit - 1.96 * pred_int$se.fit, col = "red", lty = 2)

# GDP Growth vs Inflation
loess_gdp <- loess(Inflation ~ GDP_Growth, data = inflation_data, span = 0.6)
x_gdp <- seq(min(inflation_data$GDP_Growth), max(inflation_data$GDP_Growth), length.out
pred_gdp <- predict(loess_gdp, newdata = data.frame(GDP_Growth = x_gdp), se = TRUE)

plot(inflation_data$GDP_Growth, inflation_data$Inflation,
     main = "Kernel Regression: GDP Growth vs Inflation",
     xlab = "GDP Growth (%)", ylab = "Inflation by Year", pch = 19, col = "gray")
lines(x_gdp, pred_gdp$fit, col = "darkgreen", lwd = 2)
lines(x_gdp, pred_gdp$fit + 1.96 * pred_gdp$se.fit, col = "darkgreen", lty = 2)
lines(x_gdp, pred_gdp$fit - 1.96 * pred_gdp$se.fit, col = "darkgreen", lty = 2)
```

3. Generalized Additive Model (GAM)

```
library(mgcv)

model_gam <- gam(GDP_Growth ~ s(Interest_Rate) + s(Unemployment_Rate) + s(Year),
```

```
                 data = inflation_data)
summary(model_gam)

par(mfrow = c(1, 3))
plot(model_gam, se = TRUE, col = "darkgreen")
```