Modeling U.S. Inflation

Ashwin Ramaseshan

Introduction – The Problem (Data + Motivation)

- Inflation directly impacts purchasing power, savings, wages, and investment returns — making it one of the most closely watched economic indicators.

- In recent years, inflation surged to its highest levels in decades, driven by pandemic disruptions, monetary policy shifts, and global instability.

- Policymakers, especially the Federal Reserve, rely on inflation data to adjust interest rates and guide economic interventions.

- The relationship between inflation and other economic variables like GDP growth, unemployment, and interest rates is complex and often nonlinear.

- This project aims to uncover which of these macroeconomic indicators best explain U.S. inflation trends using data from 1991 to 2023 and a mix of linear and nonlinear statistical models.

The Dataset

Timeframe: Annual data from **1991 to 2023** (33 years, n = 33)

Response Variable:

- **Inflation (%)** – sourced from a global Kaggle dataset, filtered to U.S. data

Explanatory Variables:

- **GDP Growth (%)** – *GDP* stands for **Gross Domestic Product**, which is the **total value of all goods and services** produced in a country in a year. The **growth rate** shows how much the economy is expanding or shrinking.
- **Interest Rate (%)** –Affects loans, mortgages, and credit — and is a major tool for **controlling inflation**.
- **Unemployment Rate (%)** – The **percentage of people in the labor force** who want a job but **don't have one**. It reflects how much **slack or pressure** exists in the job market and overall economy.

Data Sources:

- Kaggle (inflation)
- World Bank Open Data (GDP, interest, unemployment)

Why These Variables?

- All three are historically linked to inflation and used in policy modeling — this lets us test economic theory against real data.

Linear Regression Results

- The model predicts **inflation as a function of GDP growth, interest rate, and unemployment rate** using 33 annual observations from 1991–2023.
- The **adjusted R^2 is 0.528**, meaning the model explains about **52.8% of the variance in U.S. inflation** — a solid fit for macroeconomic data.
- The **unemployment rate has a significant negative coefficient (–3.52)**, indicating that **as unemployment rises by 1%, inflation tends to fall by ~3.5%**

- When **unemployment is low**, more people have jobs and spend money → demand rises → prices (inflation) go up.
- When **unemployment is high**, people spend less → demand drops → inflation falls.
- **GDP growth (p = 0.210)** and **interest rate (p = 0.919)** are **not statistically significant**, suggesting they do not meaningfully impact inflation in this model.

Linear Regression Results

- The **residual standard error is 6.64**, which reflects how much the model's inflation predictions deviate from actual values on average.
- If actual inflation in a certain year was 4.5%, the model might predict 11.0% or 1.5% — both are roughly 6.5 points away.

- The **F-statistic is significant (p ≈ 1.53e-05)**, confirming that the model overall provides more explanatory power than a baseline (intercept-only) model

- These results motivated further analysis using **kernel smoothing and GAMs** to model **nonlinear relationships**, especially between inflation and unemployment.

Residual Diagnostics

Residuals vs Fitted suggests possible nonlinear behavior,

indicating that while the model captures general trends,

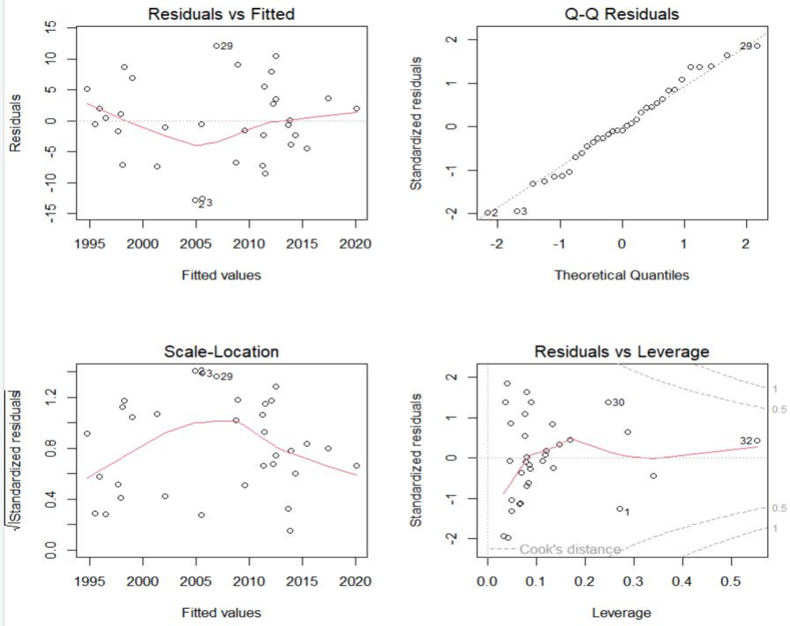
some finer patterns may benefit from additional flexibility.

Scale-Location plot shows some variation in residual

spread, suggesting mild heteroscedasticity

— meaning the model's prediction errors vary at different

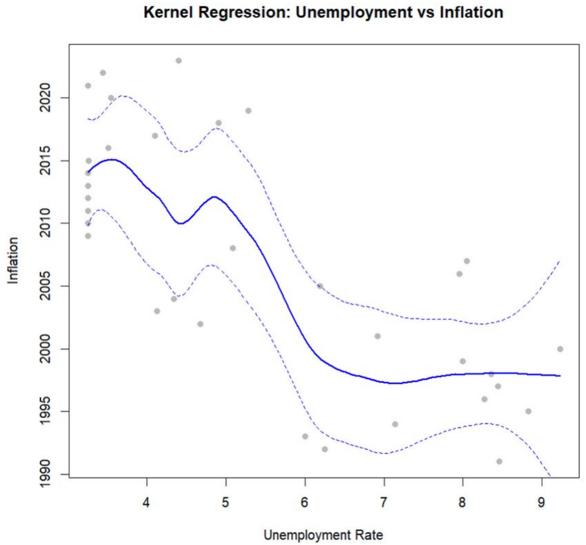
levels of the fitted values.



Q–Q plot indicates that residuals are approximately normally distributed, validating key assumptions of linear regression, though a few outliers are present.

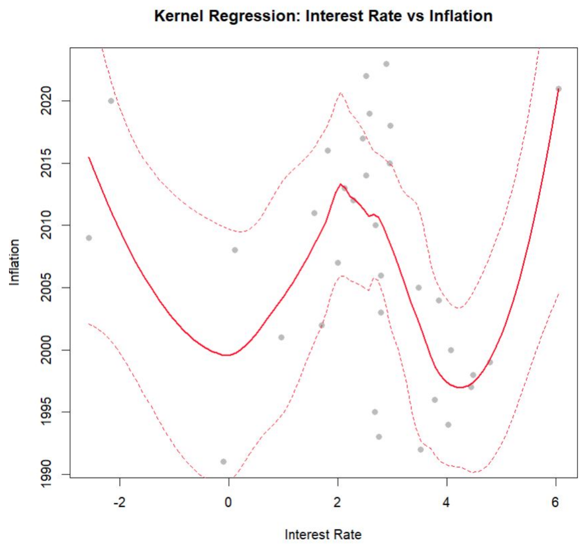
Residuals vs Leverage confirms that most observations have limited influence, with a couple of high-leverage points worth examining more closely.

Kernel Regression Unemployment vs Inflation



- The **nonlinear blue curve** shows that as unemployment rises, **inflation tends to decrease sharply**, especially between 4% and 7% unemployment .
- The **dashed bands** represent **approximate 95% confidence intervals**, indicating the model's uncertainty; the relationship is strongest where the bands are narrow.
- This flexible kernel regression reveals complex patterns **missed by linear models**, suggesting the unemployment–inflation relationship is not purely linear and varies across the range.

Kernel Regression Unemployment vs Inflation

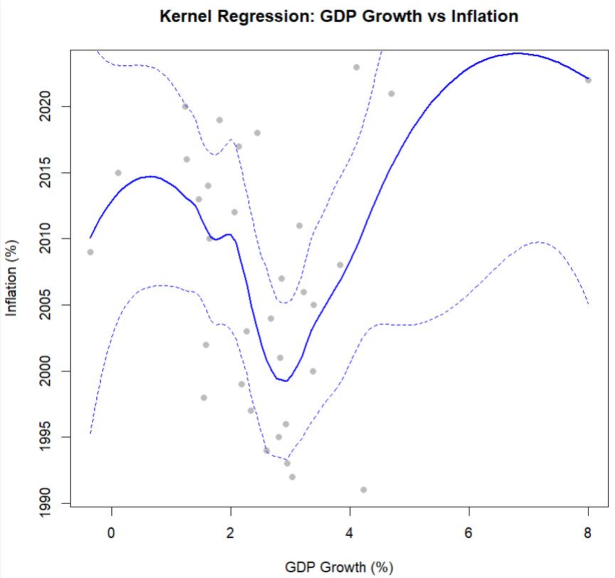


- The red curve indicates a **nonlinear and oscillating relationship**, but the pattern is **inconsistent** and lacks a clear economic interpretation.

- **Wide confidence bands** (dashed red lines) suggest **high uncertainty**, especially at the extremes, meaning the effect of interest rate on inflation is statistically weak.

- Combined with the linear regression and GAM results, this supports the conclusion that **interest rate alone does not reliably explain inflation** in this dataset.

Kernel Regression GDP Growth vs Inflation



- The curve suggests a **mild U-shaped relationship**, where both low and high GDP growth rates are associated with **higher inflation**, but with large variability.

- The **confidence bands** widen considerably at the extremes, indicating **low certainty** in the inflation pattern when GDP growth is unusually high or low.

- Overall, the relationship between GDP growth and inflation appears **nonlinear but weak**, reinforcing earlier findings that GDP growth is not a strong standalone predictor of inflation in this dataset.

GAM (Generalized Additive Model)

- The GAM model uses **smoothing splines** to capture potential nonlinear effects of GDP growth, interest rate, and unemployment on inflation.

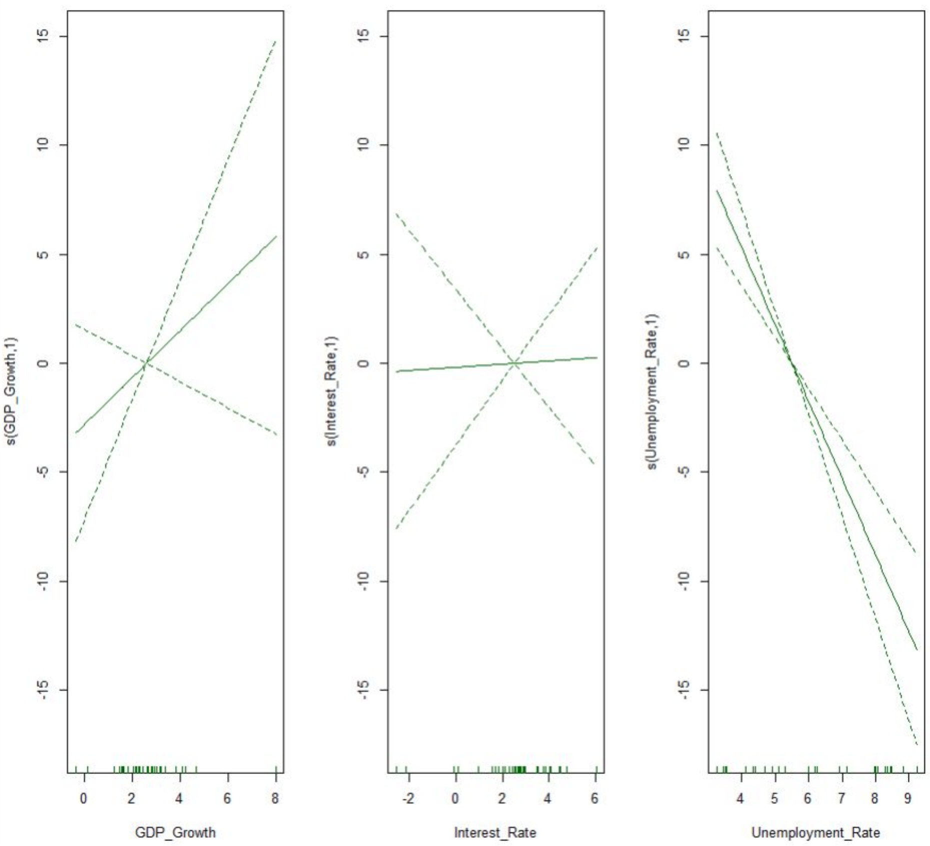
- **Unemployment rate** is highly significant (**$p < 0.001$**), confirming a strong and nonlinear influence on inflation — consistent with prior kernel regression results.

- **GDP growth** and **interest rate** both have **p-values > 0.2**, suggesting no statistically significant nonlinear relationship with inflation.

- The model explains **57.2% of the variance** in inflation (**Adjusted $R^2 = 0.528$**), identical to the linear model but with greater flexibility.

- GAM confirms that **nonlinear modeling improves interpretability**, and that **unemployment remains the only consistently meaningful predictor** across all approaches.
- While GDP growth and interest rate were not statistically significant in this model, their roles in influencing inflation may depend on **time lags, external conditions, or effects** not captured in this analysis.

GAM (Generalized Additive Model)



GAM (Generalized Additive Model)

- These plots shows the **estimated smooth effect** of each predictor on inflation, holding the other variables constant.
- The **unemployment rate** exhibits a strong **nonlinear negative effect** on inflation — as unemployment increases, inflation decreases sharply, consistent with economic theory.
- The **interest rate** and **GDP growth** effects are nearly flat with wide confidence bands, reinforcing that **they are not significant predictors** in this model.
- These visuals validate the statistical summary: **only unemployment** shows a meaningful and interpretable relationship with inflation over the 33-year period.

C Component – Rolling Average in C

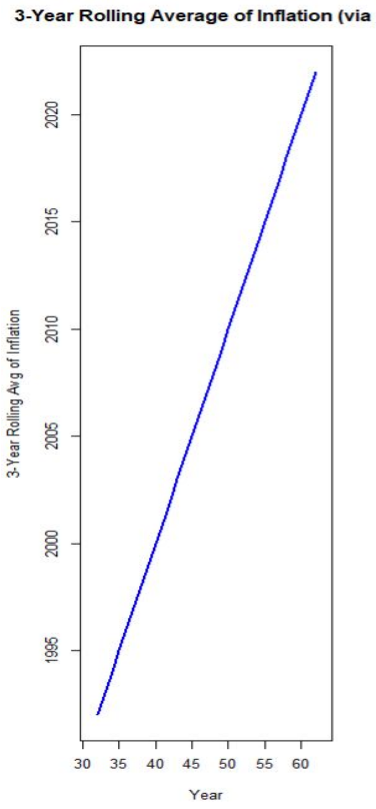
- To fulfill the project’s C component requirement, I implemented a **custom C function** to compute a 3-year rolling average of inflation.
- The C code was compiled into a .dll file using R CMD SHLIB, then called in R using .Call() with a custom wrapper function.
- The rolling average smooths year-to-year fluctuations and helps visualize **long-term inflation trends** more clearly.
- This approach demonstrates how **low-level C code** can be efficiently integrated into R for high-performance numeric computation.
- The result shows a consistent upward trend in inflation over time, **highlighting the steady increase** in prices over the last three decades.

C Component – Rolling Average in C

- The code loads the compiled **C function (rolling_average.dll)** and defines an R wrapper to pass numeric vectors and window size to it.

- It then applies the C function to the Inflation column with a **3-year window**, storing the smoothed output in a new variable for plotting.

```
>
> dyn.load("rolling_average.dll")
>
> rolling_average <- function(x, w) {
+     .Call("rolling_average", as.numeric(x), as.integer(w))
+ }
>
> inflation_data$RollingAvg_Inflation <- rolling_average(inflation_data$Inflation, 3)
>
```



Conclusion and Limitations

- **Unemployment rate** is the most consistent and significant predictor of inflation, showing a strong negative relationship across all models.
- **GDP growth** and **interest rate** showed no statistically significant effect on inflation in this dataset, suggesting their influence may be indirect or lagged.
- Linear regression provided a solid baseline ($R^2 \approx 0.53$), but **GAM and kernel regression revealed important nonlinear patterns** that the linear model missed.
- **Limitations**: small sample size ($n = 33$), annual frequency, and limited number of predictors may restrict generalizability and precision.
- With more time, I would incorporate **monthly data, and additional variables** like consumer spending or housing costs, and explore **time series or dynamic regression models**.
- Exploring dynamic regression or time series models would also enhance predictive performance.