

# STAT429 Project Proposal

2025-10-19

## UG section

### Names:

Samuel Wu (samuel54)

Molin Yang (moliny2)

Rongsheng Zhang(rz36)

### Questions:

We would like to build a model to estimate the CPI of used cars and trucks as the response variable using three different predictors including CPI for new cars, motor fuels, and also federal funds effective rates. We will also include the influence of COVID-19 pandemic by adding a breakpoint at 2020.1 (Jan 2020).

### Data Description:

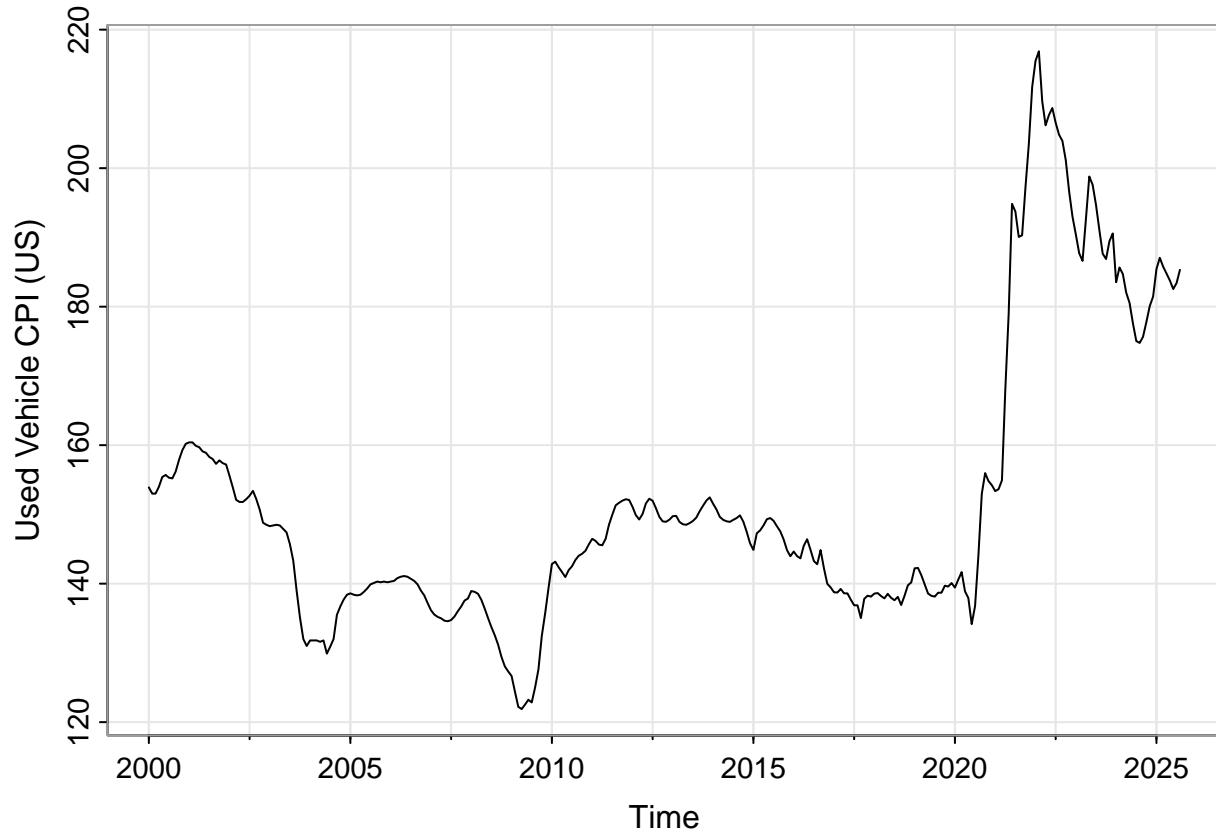
We obtained data from <https://fred.stlouisfed.org/>. We use monthly, seasonally-adjusted time-series data from the United States covering the 21st century up to August 2025 to build a regression model estimating the CPI for used cars and trucks (series ID CUSR0000SETA02). This series is published by the U.S. Bureau of Labor Statistics via the FRED database, indexed to 1982-1984 = 100. It functions as the dependent (response) variable in the analysis, capturing the price movements for used cars and trucks in U.S. urban consumer markets. The three predictor (independent) variables are: the CPI for new vehicles (series ID CUSR0000SETA01) which reflects the new-vehicle price environment; the CPI for motor fuel (series ID CUSR0000SETB) capturing the cost of operating vehicles and thus affecting demand for used cars; and the federal funds effective rate (series ID FEDFUNDS) which represents the broader financing/credit cost environment that influences vehicle-purchase decisions. All four series share a monthly frequency, allowing consistent alignment for regression over the specified period.

### Visual Representation:

```
library(astsa)

used.cars.CPI = read.csv("used cars CPI.csv")
colnames(used.cars.CPI)[1] = "Time"
colnames(used.cars.CPI)[2] = "Used Cars CPI"

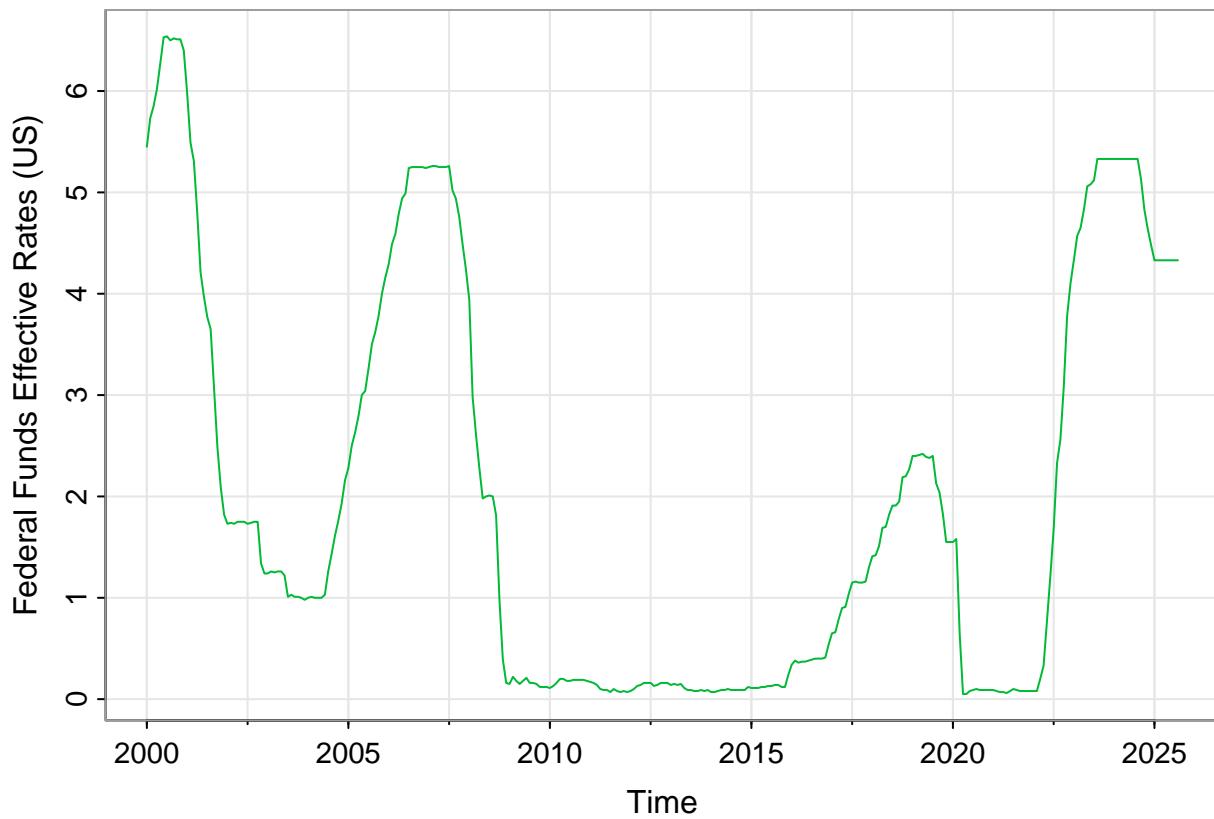
cpi.used.vehicle = ts(used.cars.CPI[2], start = c(2000, 1), frequency = 12)
tsplot(cpi.used.vehicle, col = 1, xlab = "Time", ylab = "Used Vehicle CPI (US)")
```



```
library(astsa)

effective.rates = read.csv("federal funds effective rates.csv")
colnames(effective.rates)[1] = "Time"
colnames(effective.rates)[2] = "Federal Funds Effective Rates"

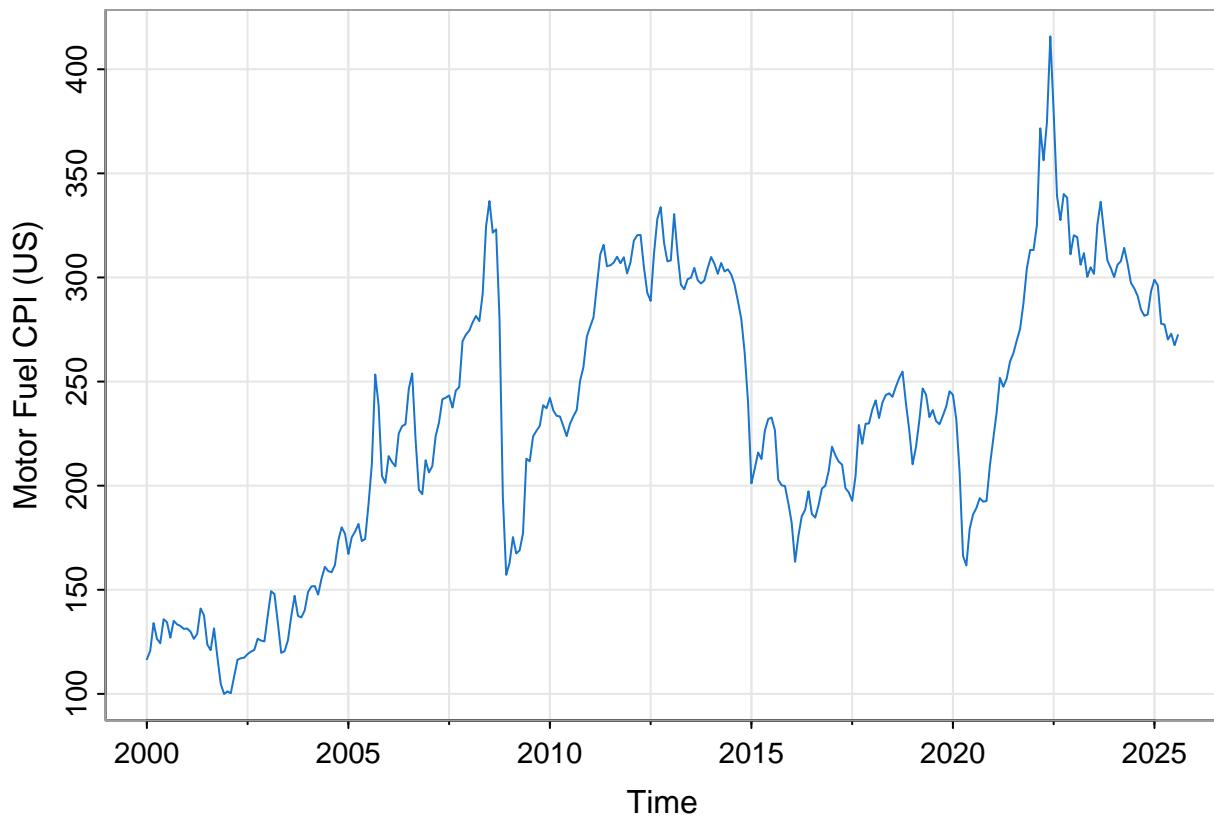
ff.effective.rates = ts(effective.rates[2], start = c(2000, 1), frequency = 12)
tsplot(ff.effective.rates, col = 3, xlab = "Time", ylab = "Federal Funds Effective Rates (US)")
```



```
library(astsa)

motor.fuel.CPI = read.csv("motor fuel CPI.csv")
colnames(motor.fuel.CPI)[1] = "Time"
colnames(motor.fuel.CPI)[2] = "Motor Fuel CPI"

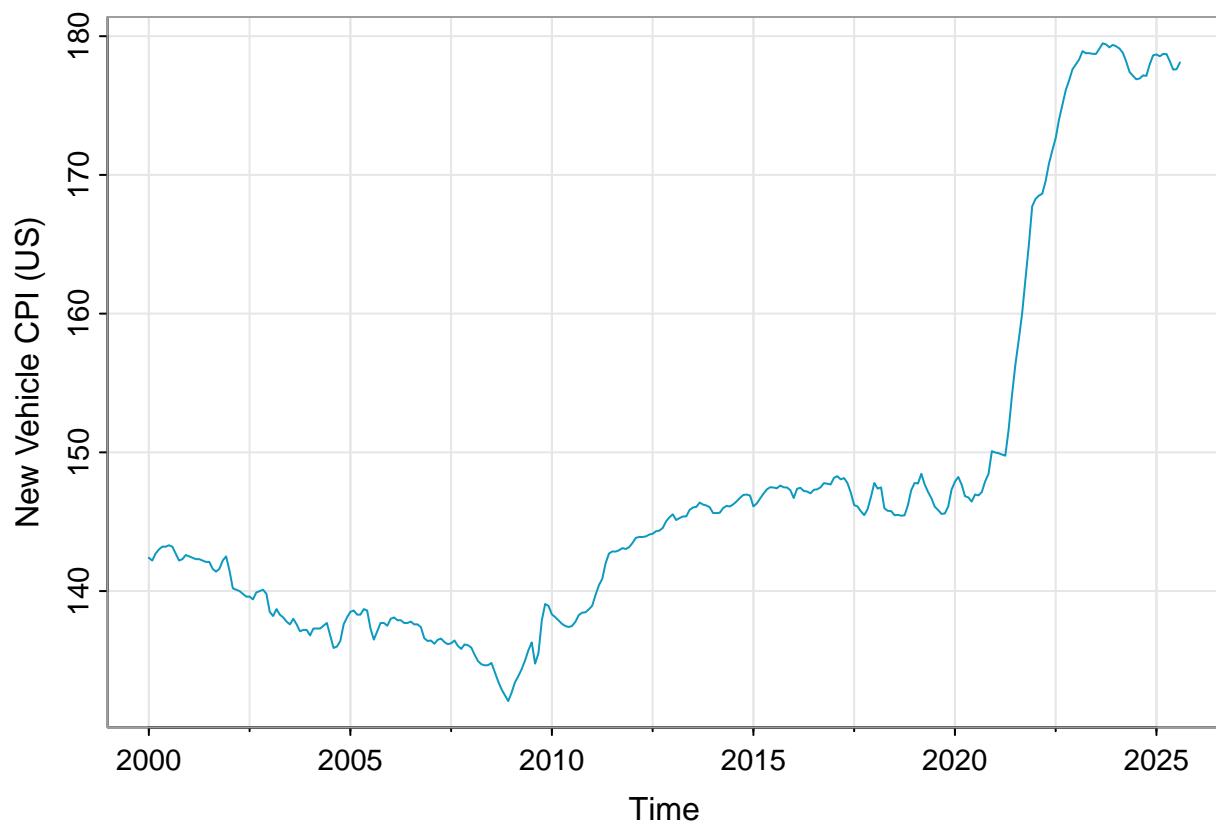
cpi.fuel = ts(motor.fuel.CPI[2], start = c(2000, 1), frequency = 12)
tsplot(cpi.fuel, col = 4, xlab = "Time", ylab = "Motor Fuel CPI (US)")
```



```
library(astsa)

new.vehicle.CPI = read.csv("new vehicle CPI.csv")
colnames(new.vehicle.CPI)[1] = "Time"
colnames(new.vehicle.CPI)[2] = "New Vehicle CPI"

cpi.new.vehicle = ts(new.vehicle.CPI[2], start = c(2000, 1), frequency = 12)
tsplot(cpi.new.vehicle, col = 5, xlab = "Time", ylab = "New Vehicle CPI (US)")
```



## Analysis Plan:

### Analysis A tasks

1. We are going to download the datasets from the internet and run the preliminary analysis.
2. And then we are going to select two models (one ordinary linear regression model and the other one is going to be a linear regression model with dummy variables).
3. Then by using the summary function we are going to analyze the result of our model
4. Then by comparing multiple metrics such as lack of fit,  $R^2$ , etc we are going to select the best model in the end.
5. After that we are going to use the predict function to predict the future used car prices for 5 future values.