Abdulrahman AlSaran

Advanced Data Analysis Using R:

Analysis of PCE on Gasoline and Energy Goods

Research Question

- How has the Price Indexes and Personal Consumption Expenditures for Gasoline and other energy goods evolved before and during the COVID-19 pandemic and after in the US,and what does this reveal about consumer financial behavior and economic resilience in the US?

```
#Calling the full datasets
PCE_US_Major_Products <- read_csv("PCE_by_Major_Type_of_Product.csv")
```

```
## Rows: 32 Columns: 95
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (59): type, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, ...
## num (36): 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
PI_US_Major_Products <- read_csv("PI_for_PCE_by_Major_Type_of_Product.csv")
```

```
## Rows: 31 Columns: 95
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (59): type, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, ...
## dbl (36): 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Changing names of Price index data to start With "PI_"
PI_US_Major_Products <- PI_US_Major_Products %>%
  mutate(type = paste0("PI_", type))

#Joining
PI_PCE_US_Major_Products <- PCE_US_Major_Products %>%
full_join(PI_US_Major_Products)
```

```
## Joining with `by = join_by(type, `1929`, `1930`, `1931`, `1932`, `1933`,
## `1934`, `1935`, `1936`, `1937`, `1938`, `1939`, `1940`, `1941`, `1942`, `1943`,
## `1944`, `1945`, `1946`, `1947`, `1948`, `1949`, `1950`, `1951`, `1952`, `1953`,
## `1954`, `1955`, `1956`, `1957`, `1958`, `1959`, `1960`, `1961`, `1962`, `1963`,
## `1964`, `1965`, `1966`, `1967`, `1968`, `1969`, `1970`, `1971`, `1972`, `1973`,
## `1974`, `1975`, `1976`, `1977`, `1978`, `1979`, `1980`, `1981`, `1982`, `1983`,
## `1984`, `1985`, `1986`, `1987`, `1988`, `1989`, `1990`, `1991`, `1992`, `1993`,
## `1994`, `1995`, `1996`, `1997`, `1998`, `1999`, `2000`, `2001`, `2002`, `2003`,
## `2004`, `2005`, `2006`, `2007`, `2008`, `2009`, `2010`, `2011`, `2012`, `2013`,
## `2014`, `2015`, `2016`, `2017`, `2018`, `2019`, `2020`, `2021`, `2022`)`
```

```
colnames(PI_PCE_US_Major_Products)
```

- The First columns called "type" represents the major products types that that have the measures of PCE over the years and the products that start with "PI_" means the price index.
- The second columns till the last represents years from 1929-2022 that PCE is measured over.

```
# install.packages("janitor")
library(janitor)
PI_PCE_US_Major_Products <-
  clean_names(PI_PCE_US_Major_Products)
```

```
#converting all years columns into CHARCTER type for avoiding pivoting errors
PI_PCE_US_Major_Products <- PI_PCE_US_Major_Products %>%
  mutate(across(-1, as.character))
# Pivoting the data
Pivoted_PI_PCE_US_Major_Products <- PI_PCE_US_Major_Products %>%
pivot_longer(cols = -1,
names_to = "year", values_to = "pce")

# filter Only "Gasoline and other energy goods"
PI_PCE_US_Gasoline_Energy <- Pivoted_PI_PCE_US_Major_Products %>%
  filter(`type` %in% c("Gasoline and other energy goods",
                       "PI_Gasoline and other energy goods"))
#Removing the x at the beginning created by clean_names() to help turn values to numeric
#later for graphing
PI_PCE_US_Gasoline_Energy$year <- sub("^x", "", PI_PCE_US_Gasoline_Energy$year)

#making the year & PCE  values into numeric
PI_PCE_US_Gasoline_Energy$year <- as.numeric(as.character(PI_PCE_US_Gasoline_Energy$year))
PI_PCE_US_Gasoline_Energy$pce <- as.numeric(as.character(PI_PCE_US_Gasoline_Energy$pce))

#Using glimpse()
glimpse(PI_PCE_US_Gasoline_Energy)
```

```
## Rows: 188
## Columns: 3
## $ type <chr> "Gasoline and other energy goods", "Gasoline and other energy goo~
## $ year <dbl> 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939,~
## $ pce  <dbl> 3.4, 3.3, 2.8, 2.6, 2.6, 2.9, 3.0, 3.3, 3.5, 3.4, 3.6, 3.8, 4.3, ~
```

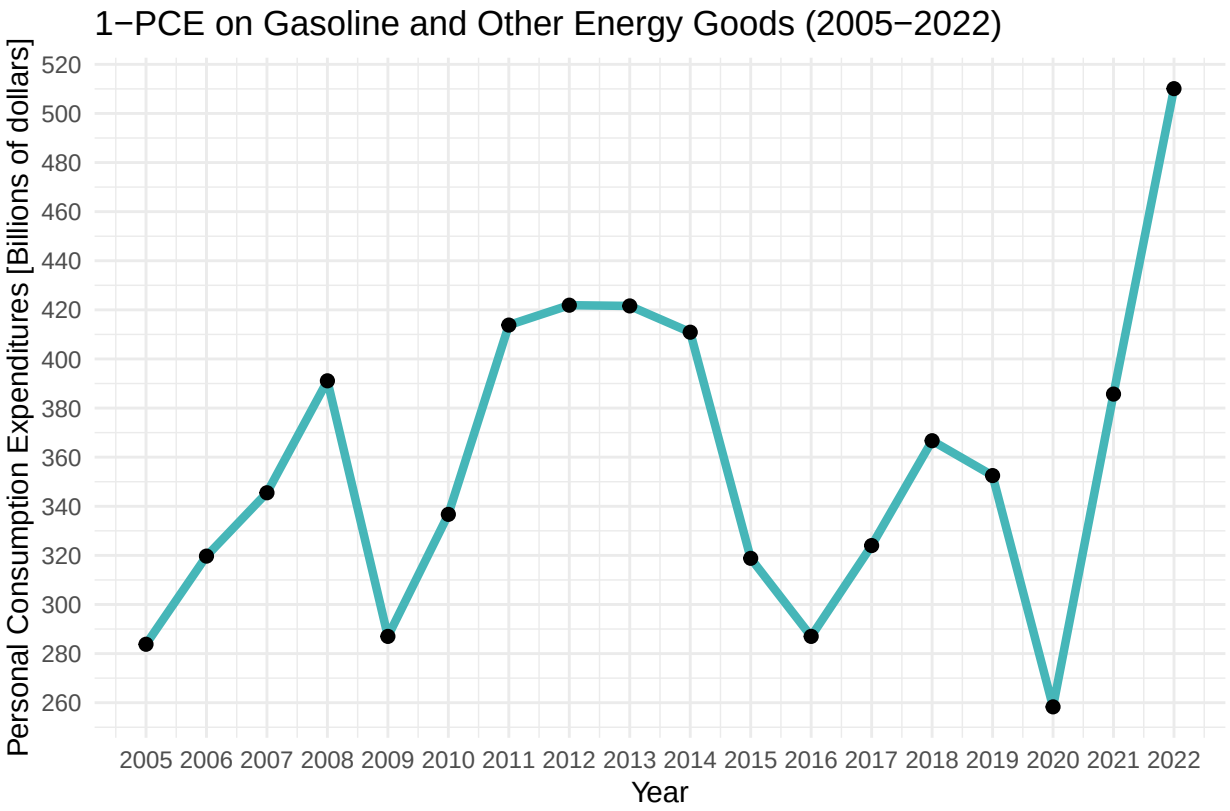
- After using glimpse() on the transformed dataset :

a. No.Rows : 18
b. data Types : Character , Numeric
c. Columns : type, year,pce

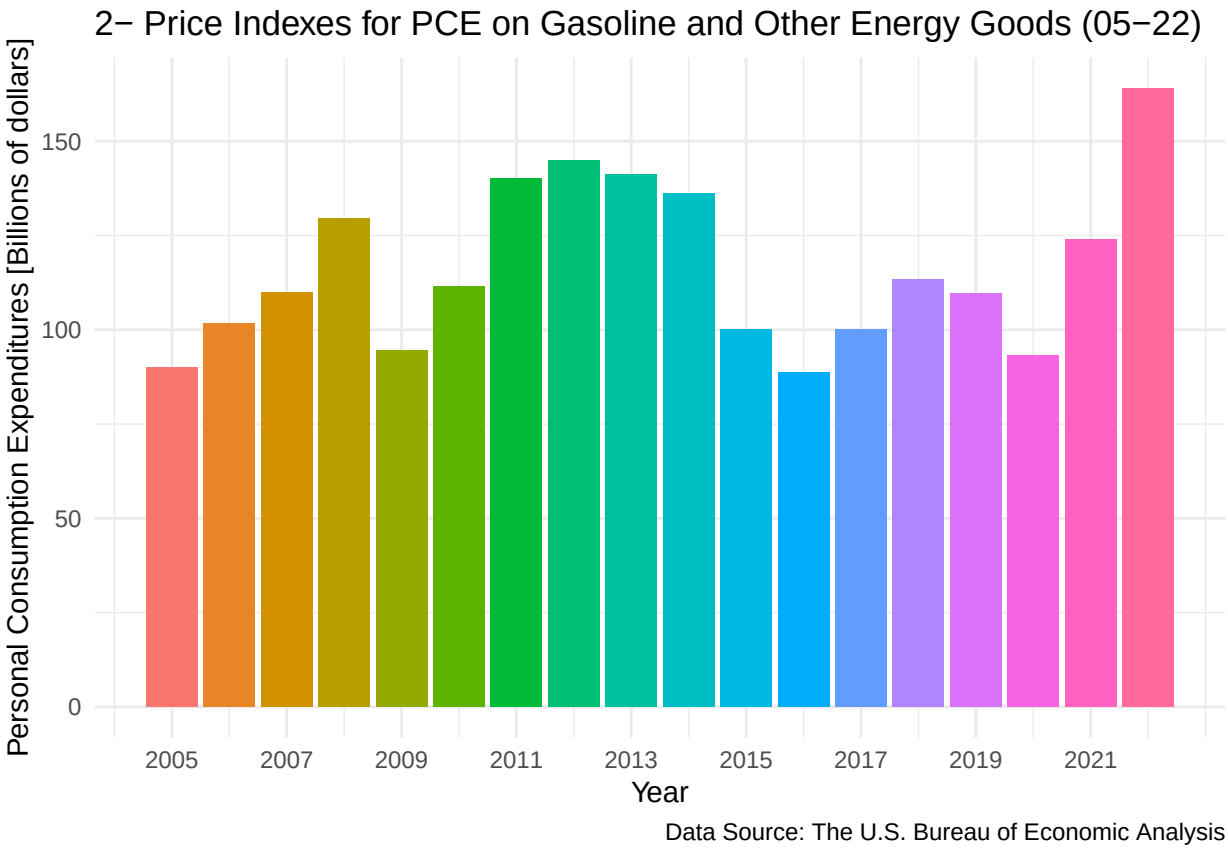
```
#Using a function that shows a five number summary of my variables of interest.
summary(PI_PCE_US_Gasoline_Energy)
```

```
##      type                year           pce
##  Length:188         Min.   :1929   Min.   :  2.600
##  Class :character   1st Qu.:1952   1st Qu.:  9.139
##  Mode  :character   Median :1976   Median : 30.570
##                     Mean   :1976   Mean   : 76.737
##                     3rd Qu.:1999   3rd Qu.:108.698
##                     Max.   :2022   Max.   :510.100
```

```
#1-Graph for PCE on Gasoline and Other Energy Goods (2005-2022)
PI_PCE_US_Gasoline_Energy %>%
  filter(`type` %in% "Gasoline and other energy goods"& year >=2005 & year <=2022) %>%
ggplot( aes(x = year, y = pce)) +
  geom_line(aes(group=1), color="#46B6B8", size = 1.5) +
  geom_point(color="black", size=2)+
  labs(title = "1-PCE on Gasoline and Other Energy Goods (2005-2022)",
       x = "Year",
       y = "Personal Consumption Expenditures [Billions of dollars]",
       caption = "Data Source: The U.S. Bureau of Economic Analysis") +
  scale_x_continuous(breaks = seq(min(PI_PCE_US_Gasoline_Energy$year),
                                  max(PI_PCE_US_Gasoline_Energy$year), by = 1)) +
   scale_y_continuous(breaks = pretty(PI_PCE_US_Gasoline_Energy$pce, n = 20))+
  theme_minimal()
```




```
#2- Price Indexes for PCE on Gasoline and Other Energy Goods (2005-2022)
PI_PCE_US_Gasoline_Energy %>%
  filter(`type` %in% "PI_Gasoline and other energy goods" & year >=2005 & year <=2022) %>%
ggplot(aes(x = year, y = pce, fill = as.factor(year))) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  labs(title = "2- Price Indexes for PCE on Gasoline and Other Energy Goods (05-22)",
       x = "Year",
       y = "Personal Consumption Expenditures [Billions of dollars]",
       caption = "Data Source: The U.S. Bureau of Economic Analysis") +
  scale_x_continuous(breaks = seq(min(PI_PCE_US_Gasoline_Energy$year),
                                  max(PI_PCE_US_Gasoline_Energy$year), by = 2))  +
  theme_minimal()
```



- Advanced Data Analysis Using R: Analysis of PCE on Gasoline and Energy Goods

For the rest of the report i will choose the PCE on Gasoline and Other Energy Goods variable to answer them & i choose the period from 2005 - 2022 for analsyis.

- As you can see in Graph 1 the PCE started of with a rise until a sudden dip from 390 billion to 290 billion from 2008-2009, then A subsequent rise of 29.27% is observed from 2009 to 2011, followed by a relatively steady phase until 2014, then Another notable dip of about 29.27% occurs between 2014 and 2016.
- But,the biggest dip in number happened between 2019-2020 and its when the pandemic hits.
- Then from 2020 -2022 after the pandemic the PCE has increased by 52.9% which is really a huge increase.

- Hypothesis:

The 2008-2009 dip is likely attributed to the financial crisis during that period, impacting consumer spending and economic activities adversely.

The 2014-2016 dip may be associated with a global increase in oil supply, primarily due to booming U.S. shale oil production, coupled with declining demand, particularly from mid-2015 to early 2016.

The exponential increase in PCE from 2020 to 2022 could be driven by economic recovery post the pandemic-induced recession, reviving energy demand, alongside a surge in crude oil prices during this period.

- Future Exploration:

Exploring the correlation between economic recessions, global oil supply, and PCE on Gasoline and Other Energy Goods could provide deeper insights.

Investigating the impact of government energy policies and consumer behavior on PCE during various economic conditions could be an avenue for further research.

Statistical analysis to validate the observed trends and to understand the magnitude and significance of the impact of economic conditions on PCE would be beneficial.

Why i choose this period?

I choose the period 2005 to 2022 to understand things like the 2008 crises, and the 2014-2016 booming U.S. shale oil production to find hints at consumer spending's susceptibility to global economic conditions. Which Also helps me to analyse the Pandemic Onset at 2019-2020 when A sharp decline in PCE aligns

with reduced mobility and economic contraction, and Post-Pandemic Recovery 2020-2022 The significant PCE increase possibly reflects economic recovery and anticipatory buying due to potential price hikes.

Regarding any potential answers or raise more questions, i reached some Hypothetical Conclusions:
1- Trends suggest a correlation between economic conditions and PCE on energy goods.
2- The post-pandemic PCE rise could indicate economic resilience and altered consumer behavior.

As for New questions these came to mind regarding the findings:
1- What policies influenced these PCE trends?
2- How did consumer behavior towards energy consumption evolve during and post-pandemic?

In conclusion ,The preliminary analysis lays groundwork for deeper investigation into economic resilience, consumer behavior, and policy implications concerning energy goods' consumption during varying economic scenarios.

#After diving into data exploration, it's time to operationalize the insights.

- I selected the Personal Consumption Expenditures (PCE) on Gasoline And Other Energy Goods from 1929-2022 rather than starting from 2005 to get more accurate results as the dependent variable (Y), and wages and salaries across the same time frame as the independent variable (X).

- Rationale:
  Analyzing the impact of wage fluctuations on energy spending across a period with significant economic events like the 2008 financial crisis and the COVID-19 pandemic can offer insights into consumer financial behavior For instance, are consumers cutting back on energy spending when earnings are lower? Or are they continuing to spend.

understanding this relationship may help decision makers devise measures to stabilize income or energy prices during economic downturns to support consumer spending and overall economic stability.

This analysis aligns with the original research question, shedding light on economic resilience in the US.

Regression formula:

$$PCE|Gasoline|EnergyGoods = \beta_0 + \beta_1 WagesAndSalaries + \epsilon$$

1-Cleaning and Transforming the data for the regression

```
#Preparing variables in the regression
Regression_data <- PI_PCE_US_Major_Products %>%
    filter(`type` %in% c("Gasoline and other energy goods",
                         "PI_Gasoline and other energy goods",
                         "Wages and salaries"))

# Change the format to better suit a regression using library(reshape2)
melted_data <- melt(Regression_data, id.vars = "type")
Regression_data <- dcast(melted_data, variable ~ type, value.var = "value")
Regression_data$variable <- sub("^x", "", Regression_data$variable)
Regression_data$variable <- as.numeric(as.character(Regression_data$variable))
Regression_data <- Regression_data %>% rename(year = variable)
Regression_data$`PI_Gasoline and other energy goods`<-as.numeric(
as.character( Regression_data$`PI_Gasoline and other energy goods`))


#Clean column names
Regression_data <- clean_names(Regression_data)

#converting millions to billions
Regression_data$wages_and_salaries <- as.numeric(as.character(Regression_data$wages_and_salaries))
Regression_data$wages_and_salaries <- Regression_data$wages_and_salaries/1000
```

2- The Regression

```
PCE_Regression <- lm( gasoline_and_other_energy_goods ~ wages_and_salaries
                      ,data = Regression_data)

summary(PCE_Regression)
```

```
##
## Call:
## lm(formula = gasoline_and_other_energy_goods ~ wages_and_salaries,
##     data = Regression_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -154.322   -4.806   -1.516    0.408  123.975
##
## Coefficients:
##                    Estimate Std. Error t value            Pr(>|t|)
## (Intercept)           3.157      5.442    0.58               0.563
## wages_and_salaries   43.263      1.346   32.14 <0.0000000000000002 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38.65 on 83 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.9256, Adjusted R-squared:  0.9247
## F-statistic:  1033 on 1 and 83 DF,  p-value: < 0.00000000000000022
```

##Interpretation :

- Coefficients:

a. Intercept: When wages and salaries are zero, the expected Personal Consumption Expenditures (PCE) on Gasoline And Other Energy Goods is $3.157 billion.

b. Slope: For every 1 billion increase in wages and salaries, the PCE on Gasoline And Other Energy Goods increases by approximately $43.263 billion .

- Significance: The p-value for wages and salaries is extremely low (0.0000000000000002), indicating a statistically significant relationship at the 5% level (and even at the 1% level). This suggests that there's a statistically significant positive relationship between wages and salaries and PCE on Gasoline And Other Energy Goods.

Model Fit:

Multiple R^2 =0.9256: About 92.26% of the variability in PCE on Gasoline And Other Energy Goods is explained by wages and salaries. This is a substantial amount of explanatory power, indicating a high model fit.

Adjusted R^2 =0.9247 : after adjusting for the number of predictors, about 92.47% of the variability is explained, which still indicates a high model fit.

F-statistic (1033) with a p-value (0.00000000000000022) This also suggests the model is statistically significant at the 5% level (and even at the 1% level).

The model suggests a positive and statistically significant relationship between wages and salaries and PCE on Gasoline And Other Energy Goods with a moderate explanatory power.

- Multi-variate regression using an additional predictor :

i will add Price index for PCE on Gasoline And Other Energy Goods as a second x

$$PCE|Gasoline|EnergyGoods = \beta_0 + \beta_1 WagesAndSalaries + \beta_2 PiForPCE|Gas|EnergyGoods + \epsilon$$

```
PCE_Regression_multi <- lm( gasoline_and_other_energy_goods ~ wages_and_salaries +
                      pi_gasoline_and_other_energy_goods,data = Regression_data)

summary(PCE_Regression_multi)
```

```
##
## Call:
## lm(formula = gasoline_and_other_energy_goods ~ wages_and_salaries +
##     pi_gasoline_and_other_energy_goods, data = Regression_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.975  -1.903   0.107   1.905  12.699
##
## Coefficients:
##                                    Estimate Std. Error t value
## (Intercept)                       -13.19508    0.97483 -13.536
## wages_and_salaries                  5.10256    0.75720   6.739
## pi_gasoline_and_other_energy_goods   2.84860    0.05389  52.864
##                                             Pr(>|t|)
## (Intercept)                       < 0.0000000000000002 ***
## wages_and_salaries                       0.00000000205 ***
## pi_gasoline_and_other_energy_goods < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.564 on 82 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9978
## F-statistic: 1.93e+04 on 2 and 82 DF,  p-value: < 0.00000000000000022
```

- Coefficients:

a. Intercept: When both wages and salaries and the price index are zero, the expected PCE on Gasoline And Other Energy Goods is $2.36 billion.

b. Wages and Salaries : For every 1 billion increase in wages and salaries, the PCE on Gasoline And Other Energy Goods increases by approximately $5.10256 billion, assuming other factors remain constant.

c. Price Index : For every unit increase in the price index, the PCE on Gasoline And Other Energy Goods increases by approximately $2.84860 billion, assuming other factors remain constant.

- Significance:
a.The p-value for wages and salaries is 0.00000000205, which is statistically significant at the 5% level,

indicating that the relationship between wages and salaries and PCE on Gasoline And Other Energy Goods is statistically highly significant at the 5% level.
b.The p-value for the price index is practically 0, indicating also a highly statistically significant relationship with the PCE on Gasoline And Other Energy Goods.

- Model Fit:
  Multiple R^2 =0.9979: The model explains about 99.79% of the variability in PCE on Gasoline And Other Energy Goods.
  Adjusted R^2 =0.9978 : Adjusting for the number of predictors, about 99.78% of the variability is explained, still showing a strong model fit.
  F-statistic (19,300) with a p-value (0.00000000000000022) : This suggests the model is highly statistically significant.

Adding the Price Index for PCE on Gasoline And Other Energy Goods as a second explanatory variable greatly improved the model fit and explanatory power. This model better aligns with my research question and provides more robust insights into the factors influencing PCE on Gasoline And Other Energy Goods.