Project 1

Matt Briskey

September 3, 2023

```
knitr::opts_knit$set(root.dir = "C:/Users/16145/Google Drive/MSDA/DATA621
Advanced Analytics/Project 1")
getwd()

## [1] "C:/Users/16145/OneDrive - Franklin University/Documents"
```

Introduction

Since the OPEC oil embargo of 1973 there has been a continuing controversy in the U.S. about national and state tax policy for gasoline and other petroleum distillates. Other major disruptions in the production and/or distribution of petroleum distillates include, but are not limited to, the Iranian revolution and Iran-Iraq war in the late 1970s and early 1980s, and Persian Gulf War in 1990, the commencement of the Iraqi War in 2003, the Arab Spring of 2011, and Russia's war with Ukraine starting in 2022.

A crucial component of any such debate is a reliable model for demand, which by its very nature has a long, and what some may call dubious, history (Sims (1980), Malinvaud (1981), and Freedman et al. (1983)). As you may expect energy demand modeling is still an active field. Verwiebe et al. (2021) is a systematic literature review of 419 recent articles on modeling the demand for energy.

For this project we will be using econometric methods to analyze the U.S. demand for gasoline and some implications for tax policy, where changes in tax policy affect domestic oil producing companies such as ExxonMobil and Chevron, foreign oil producing companies such as Shell and BP, major exporters such as Saudi Arabia and Nigeria, consumers, and government revenue. Beyond the material presented in the textbook of the course, Greene (2018) bridges the gap between business economic policy studies and the field of econometrics.

The data for this project was obtained from the Federal Reserve Bank of St. Louis and the Federal Highway Administration of the U.S. Department of Transportation. The variables in the month-level R data frame **gas.RDS** are:

- ym: year-month combination in yyyy-mm-01 format, where the first month is January 2011 and the last month is December 2021.
- Y : per capita personal consumption of gasoline in gallons (at annual rates).

- P: real price of a gallon of gasoline in 2012 prices; that is 1 gallon = $3.35 at 2012 prices.(For additional information about deflating nominal values to real values see this Federal Reserve Bank of Dallas page)
- Z: per capita annual personal income in 1000's of 2012 U.S. dollars.
- y: natural logarithm of Y.
- p: natural logarithm of P.
- z: natural logarithm of Z.

There are four analytic exercises for this project. Exercise 1 has a value of 25 points, where each component has a value of 5 points. Exercise 2 has a value of 24 points, where each component has a value of 4 points. Exercise 3 has a value of 20 points, where each component has a value of 4 points. Exercise 4 has a value of 31 points, where each component has a value of 4 points except for the last one; its value is 7 points.

1. Use the R package ggplot2 for the following exercises, where the package and other contributed packages may be obtained from the Comprehensive R Archive Network (CRAN). The package "dplyr" may be used to subset data to create graphics in R.

(a) Create a time series plot of Y and P where Y is on the primary y-axis and P is on the secondary y-axis. Time is on the x-axis.

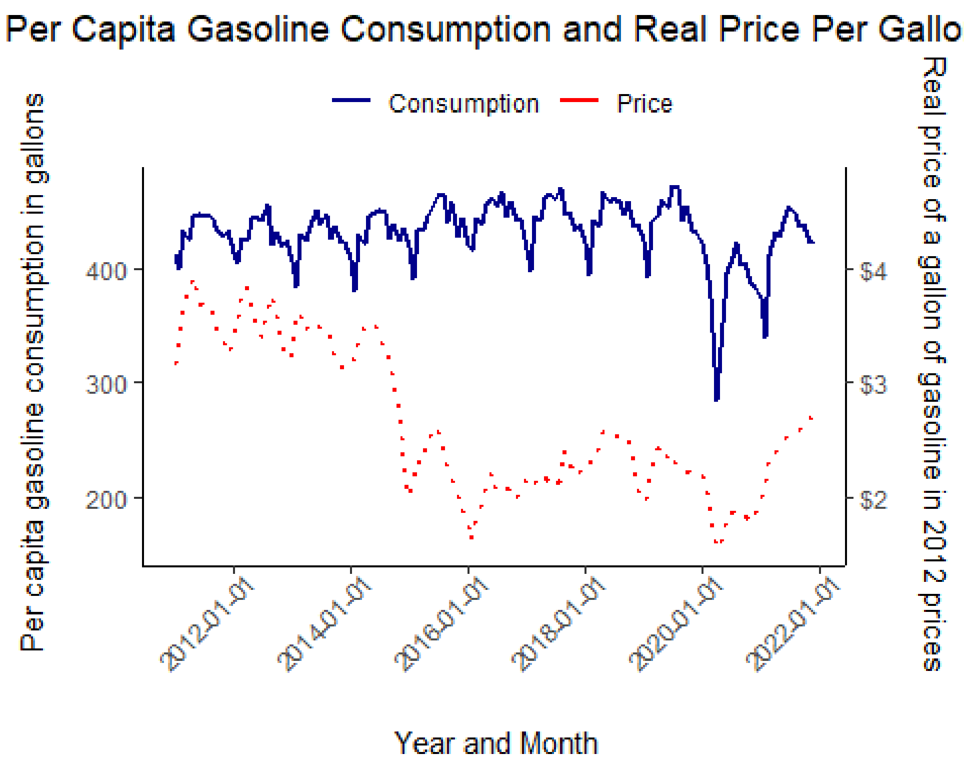
Run the code chunk below.

```
# load data
# be sure that data file is stored in the same folder as this Rmd template
file
df1 <- readRDS("Project1_gas.RDS")

# Scales for graphics
scale1 <- 100
scale2 <- 5.5

# create time series plot
ggplot() +
  geom_line(data=df1,aes(x=ym,y=Y,colour="Consumption"),
            linewidth=1) +
  geom_line(data=df1,aes(x=ym,y=P*scale1,colour="Price"),
            linetype = "dotted",
            linewidth=1) +
  scale_y_continuous(
    # Features of the first axis
    name = "Per capita gasoline consumption in gallons \n",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ ./scale1,
                        name="Real price of a gallon of gasoline in 2012
prices \n",
```

```
                              labels=scales::dollar_format()),) +
  scale_x_date(date_breaks = "2 year") +
  scale_color_manual(name = "",
                          values = c("Consumption" = "darkblue",
                                     "Price" = "red"))+
  theme_classic() +
  ggtitle("Per Capita Gasoline Consumption and Real Price Per Gallon") +
  theme(legend.position="top") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 0.9)) +
  xlab("\n Year and Month")
```



Identify and describe a few characteristics of the series.

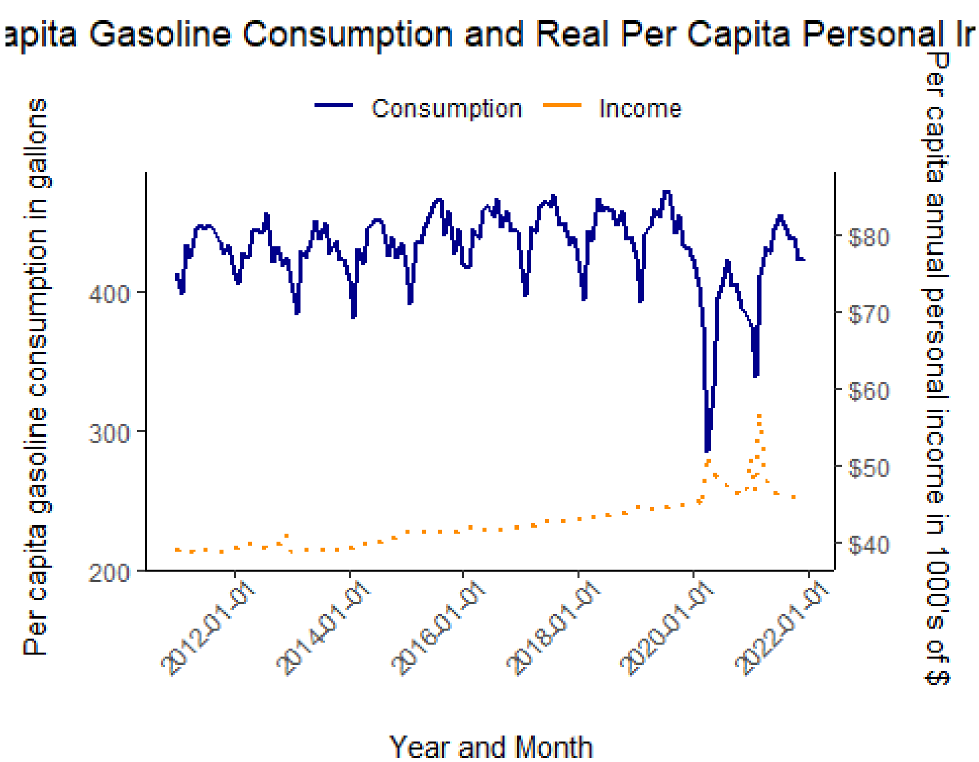
Based on the above graph, there appears to be seasonality in the per capita gasoline consumption in gallons. Every year at roughly the beginning of the year, the consumption decreases. Additionally, from 2012 to 2020, the real price of a gallon of gas using 2012 dollars tended to decrease. However, from 2020 and beyond, the price of gas appears to go up.

(b) Create a time series plot of Y and Z where Y is on the primary y-axis and Z is on the secondary y-axis. Time is on the x-axis.

Run the code chunk below.

```
# create the plot
ggplot() +
```

```
  geom_line(data=df1,aes(x=ym,y=Y,colour="Consumption"),
            linewidth=1) +
  geom_line(data=df1,aes(x=ym,y=Z*scale2,colour="Income"),
            linetype = "dotted",
            linewidth=1) +
  scale_y_continuous(
    # Features of the first axis
    name = "Per capita gasoline consumption in gallons \n",
    # Add a second axis and specify its features
    sec.axis = sec_axis(~ ./scale2,name="Per capita annual personal income in 
1000's of $ \n",
                        labels=scales::dollar_format()),) +
  scale_x_date(date_breaks = "2 year") +
  scale_color_manual(name = "",
                     values = c("Consumption" = "darkblue",
                                "Income" = "darkorange"))+
  theme_classic() +
  ggtitle("Per Capita Gasoline Consumption and Real Per Capita Personal 
Income") +
  theme(legend.position="top") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 0.9)) +
  xlab("\n Year and Month")
```



Identify and describe a few characteristics of the series.

Based on the above graph, the trend of the orange line is to increase over time which means the per capita income continued to increase from 2012 to 2022. We continue to see the seasonality of gas consumption with the blue line.

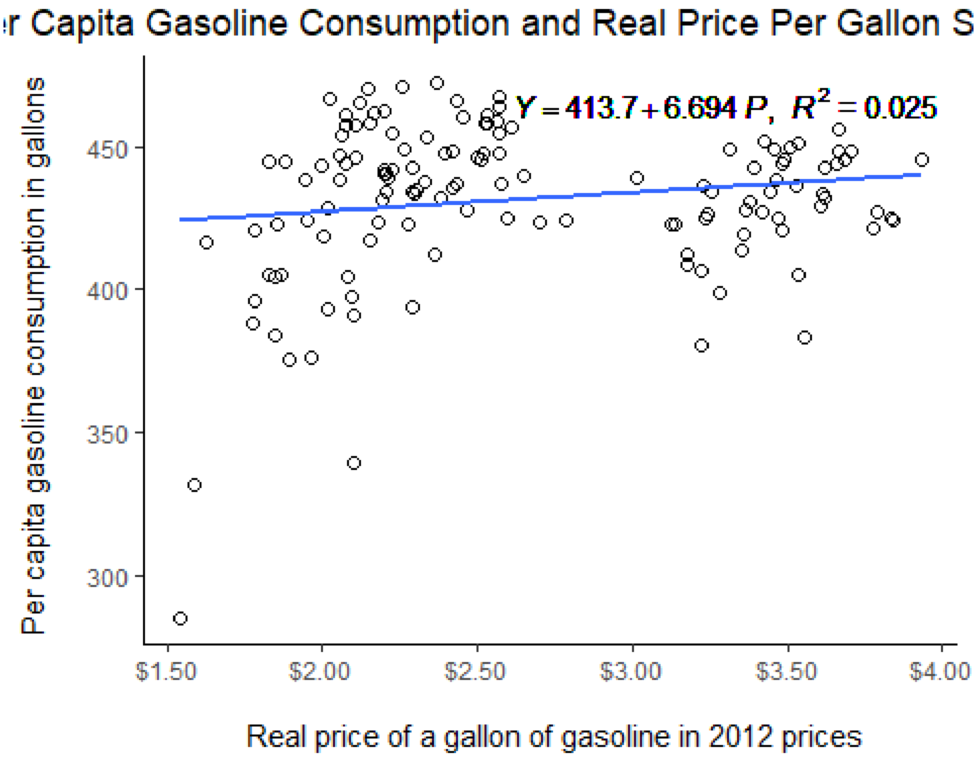
(c) Create a scatter plot of Y and P where Y is on the y-axis and P is on the x-axis. Include a least-squares regression line in the plot. Ensure the estimated coefficients and R2 are provided in the illustration.

Run the code chunk below.

```
# function to obtain regression output
biv_reg1c <- function(dep,ind,df){
  m <- lm(dep ~ ind, df);
  eq <- substitute(italic(Y) == a +
                     b~italic(P)*","~~italic(R)^2~"="~R2,
                   list(a = format(unname(coef(m)[1]),digits=4),
                        b = format(unname(coef(m)[2]),digits=4),
                        R2 = format(summary(m)$r.squared,digit=2)))
  as.character(as.expression(eq));
}

# create the plot
ggplot(df1, aes(x=P, y=Y)) +
  geom_point(size=2, shape=1,fill="black",color="black") +
  geom_smooth(method=lm,se=FALSE, fullrange=TRUE) +
  theme_classic() +
  ggtitle("Per Capita Gasoline Consumption and Real Price Per Gallon Scatter
Plot") +
  theme(legend.position="top") +
  theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "\n Real price of a gallon of gasoline in 2012 prices") +
  labs(y = "Per capita gasoline consumption in gallons \n") +
  geom_text(x = 3.3, y = 465,
            label = biv_reg1c(df1$Y,df1$P,df1), parse = TRUE) +
  scale_x_continuous(labels=scales::dollar_format())

## `geom_smooth()` using formula = 'y ~ x'
```



Identify and describe a few characteristics of the series.

The above graph shows per capita gasoline consumption vs. the real price of a gallon of gasoline in 2012 prices. The blue line shows the overall trend that consumption increases very softly as price increases which intuitively doesn't make sense but the demand for gas is relatively inelastic as there aren't good and readily available alternatives to consumers using gas. However, the low R2 means that this correlation isn't that strong.

(d) Create a scatter plot of Y and Z where Y is on the y-axis and Z is on the x-axis. Include a least-squares regression line in the plot. Ensure the estimated coefficients and R2 are provided in the illustration.

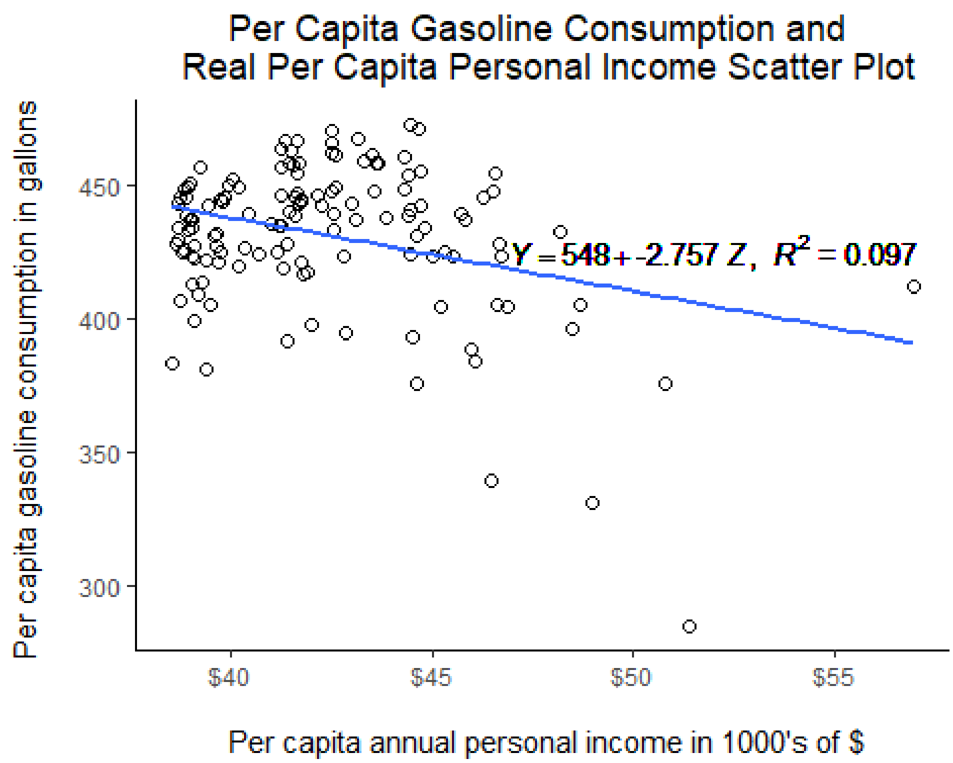
Run the code chunk below.

```
# function to obtain regression output
biv_reg1d <- function(dep,ind,df){
  m <- lm(dep ~ ind, df);
  eq <- substitute(italic(Y) == a + b~italic(Z)*","~~italic(R)^2~"="~R2,
                   list(a = format(unname(coef(m)[1]),digits=4),
                        b = format(unname(coef(m)[2]),digits=4),
                        R2 = format(summary(m)$r.squared,digit=2)))
  as.character(as.expression(eq));
}

# create plot
ggplot(df1, aes(x=Z, y=Y)) +
```

```
  geom_point(size=2, shape=1,fill="black",color="black") +
  geom_smooth(method=lm,se=FALSE, fullrange=TRUE) +
  theme_classic() +
  ggtitle("Per Capita Gasoline Consumption and \n Real Per Capita Personal 
Income Scatter Plot") +
  theme(legend.position="top") +
  theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "\n Per capita annual personal income in 1000's of $") +
  labs(y = "Per capita gasoline consumption in gallons \n") +
  geom_text(x = 52, y = 425,
            label = biv_reg1d(df1$Y,df1$Z,df1), parse = TRUE) +
  scale_x_continuous(labels=scales::dollar_format())

## `geom_smooth()` using formula = 'y ~ x'
```



Identify and describe a few characteristics of the series.

The above graph shows per capita gasoline consumption vs. per capita annual personal income. The general trend is that as per capita income increases, the per capita gasoline consumption decreases. However, the low R2 means that this correlation isn't that strong.

(e) Using the detect outliers function of tsrobprep with default settings except for the value S, which should be set to 12 since we are modeling monthly data, identify the month-year combinations where Y is detected as an outlier, the values of Y, and the probability that the observation is an outlier.

Run the code chunk below.

```
# find outliers
outlier1 <- detect_outliers(df1$Y,S=12)

# get row names
df1$id <- as.numeric(rownames(df1))

# get all outlier probabilities
df1$prob_out <- outlier1$outlier.probs

# get observations with outlier probability over 0.5
df1$outlier <-ifelse(df1$prob_out>.5,1,0)
df1_out <- subset(df1, outlier == 1)
df2_out <- df1_out[,c('ym','Y','prob_out')]

# display results
df2_out

##             ym        Y  prob_out
## 110 2020-02-01 404.4602 0.6020660
## 111 2020-03-01 375.7291 1.0000000
## 112 2020-04-01 284.9080 1.0000000
## 113 2020-05-01 331.2385 0.9012528
## 124 2021-04-01 432.3017 0.7004913
```

Given the outliers detected, we will only use data before January 2020 for the remainder of the project.

Monthly indicators, which will be used to model seasonality. These are created in the code chunk below.

Run the code chunk below.

```
# Creating additional explanatory variables
df1$M1 <- ifelse(month(df1$ym)==1,1,0)
df1$M2 <- ifelse(month(df1$ym)==2,1,0)
df1$M3 <- ifelse(month(df1$ym)==3,1,0)
df1$M4 <- ifelse(month(df1$ym)==4,1,0)
df1$M5 <- ifelse(month(df1$ym)==5,1,0)
df1$M6 <- ifelse(month(df1$ym)==6,1,0)
df1$M7 <- ifelse(month(df1$ym)==7,1,0)
df1$M8 <- ifelse(month(df1$ym)==8,1,0)
df1$M9 <- ifelse(month(df1$ym)==9,1,0)
df1$M10 <- ifelse(month(df1$ym)==10,1,0)
df1$M11 <- ifelse(month(df1$ym)==11,1,0)

rownames(df1) <- NULL
df1$trend <- log(as.numeric(rownames(df1))+1)
df2 <- df1[df1$ym < "2020-01-01",]
```

2. Consider the following simple equilibrium model for the demand of gasoline (Equation 1):

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 p_t + \gamma_0 w_t + \sum_{i=1}^{11} \gamma_i \, x_{it} + \epsilon_t$$

where the ϵ_t are assumed to be independent and identically distributed. Note we are using the natural logarithm of consumption, income and price. Since the model is specified as a log-log model in consumption, income and price, β_1 is the long-run income elasticity of demand for gasoline and β_2 is the long-run price elasticity of demand for gasoline for small changes in income and price, respectively.

(a) Estimate the model given by Equation (1) using the "lm" function Report the coefficients, coefficient standard errors, R^2 and Adjusted R^2 of the model.

In the data frame called "df2",

y is the variable y_t,

z is the variable z_t,

p is the variable p_t,

the w_t variable is called "trend",

the $\sum_{i=1}^{11} \gamma_i \, x_{it}$ will be the sum of the terms labeled M1, M2 . . ., M11.

Enter your code in the chunk below. Store the regression output in an object called "reg2".

```
reg2 <- lm(y ~ z + p + trend + M1 +M2 + M3 + M4 + M5 + M6 + M7 + M8 + M9 +
M10 + M11, data=df2)
summary(reg2)
```

```
##
## Call:
## lm(formula = y ~ z + p + trend + M1 + M2 + M3 + M4 + M5 + M6 +
##     M7 + M8 + M9 + M10 + M11, data = df2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026546 -0.008086 -0.001191  0.007757  0.043656
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.514919   0.193360  28.521  < 2e-16 ***
## z            0.173673   0.052087   3.334  0.00123 **
## p           -0.061654   0.009575  -6.439 5.22e-09 ***
## trend       -0.008236   0.002698  -3.053  0.00296 **
## M1          -0.043412   0.006104  -7.112 2.32e-10 ***
## M2          -0.092616   0.006072 -15.253  < 2e-16 ***
## M3           0.009892   0.006061   1.632  0.10605
## M4           0.003857   0.006091   0.633  0.52811
```

```
## M5          0.047504   0.006128   7.752 1.11e-11 ***
## M6          0.053482   0.006121   8.738 9.54e-14 ***
## M7          0.052569   0.006101   8.617 1.72e-13 ***
## M8          0.063026   0.006099  10.333  < 2e-16 ***
## M9          0.015509   0.006105   2.540  0.01274 *
## M10         0.032990   0.006076   5.430 4.48e-07 ***
## M11        -0.007506   0.006049  -1.241  0.21780
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01281 on 93 degrees of freedom
## Multiple R-squared:  0.9306, Adjusted R-squared:  0.9201
## F-statistic: 89.04 on 14 and 93 DF,  p-value: < 2.2e-16
```

(b) Test the joint hypothesis that the coefficients of β1 and β2 are zero.

Run the code below.

```
# joint hypothesis test
linearHypothesis(reg2,c('z = 0','p = 0'))
```

```
## Linear hypothesis test
##
## Hypothesis:
## z = 0
## p = 0
##
## Model 1: restricted model
## Model 2: y ~ z + p + trend + M1 + M2 + M3 + M4 + M5 + M6 + M7 + M8 + M9 +
##     M10 + M11
##
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1     95 0.030047
## 2     93 0.015262  2  0.014785 45.047 2.089e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Is the null hypothesis that both coefficients of z and p are equal to 0 rejected, or not? Is at least one of the coefficients of z and p equal to 0? Use a 5% level of significance.

Based on the above p-value for model 2, we can reject the null hypothesis because the p-value well below the 5% significance level (2.089e-14 < 0.05). We don't have sufficient evidence to say that the regression coefficients for z and p are both equal to zero.

Neither coefficient of z or p are zero.

(c) Test if the income elasticity of demand is unitary elastic. That is, test the following null hypothesis against a two-sided alternative: H0: β1 = 1.

Run the code below.

```r
# conduct hypothesis test
linearHypothesis(reg2, "z = 1")
```

```
## Linear hypothesis test
##
## Hypothesis:
## z = 1
##
## Model 1: restricted model
## Model 2: y ~ z + p + trend + M1 + M2 + M3 + M4 + M5 + M6 + M7 + M8 + M9 +
##     M10 + M11
##
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1     94 0.056564
## 2     93 0.015262  1  0.041302 251.67 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Is the null hypothesis that the coefficient of z is equal to 1 rejected, or not? Use a 5% level of significance.

We can reject the null hypothesis that the coefficient of z is equal to 1 based on the p-value being well below 0.05 (2.2e-16 < 0.05). Therefore, the income elasticity of demand is not unitary elastic.

(d) Test if the price elasticity of demand is unitary elastic. That is, test the following null hypothesis against a two-sided alternative: H0: β2 = −1.

Run the code below.

```r
# conduct hypothesis test
linearHypothesis(reg2, "p = -1")
```

```
## Linear hypothesis test
##
## Hypothesis:
## p = - 1
##
## Model 1: restricted model
## Model 2: y ~ z + p + trend + M1 + M2 + M3 + M4 + M5 + M6 + M7 + M8 + M9 +
##     M10 + M11
##
##   Res.Df     RSS Df Sum of Sq    F    Pr(>F)
## 1     94 1.59119
## 2     93 0.01526  1    1.5759 9603 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Is the null hypothesis that the coefficient of p is equal to -1 rejected, or not? Use a 5% level of significance.

We can reject the null hypothesis that the coefficient of p is equal to -1 and that the price elasticity of demand is unitary elastic. The p-value of the hypothesis test is below 0.05 so we reject the null hypothesis (2.2e-16 < 0.05)

(e) Test for first order autocorrelated error using the DW statistic for a two-sided alternative hypothesis that the errors are first order autocorrelated errors. Ensure you state the conclusion of the test.

Run the code below.

```
# test for autocorrelation
durbinWatsonTest(reg2)

##  lag Autocorrelation D-W Statistic p-value
##    1     -0.01174551      1.989025   0.822
##  Alternative hypothesis: rho != 0
```

Is the null hypothesis that there is no first order autocorrelated error to be rejected? Use a 5% level of significance. What is the implication of this?

We cannot reject the null hypothesis that there isn't any first order autocorrelation error. The p-value of 0.852 is above the 5% significance level, so we can't reject the null hypothesis. Therefore, we will need to adjust our model to account for autocorrelation.

(f) Test for autocorrelated errors of up to order 12 using the Breusch-Godfrey test (Breusch (1978) and Godfrey (1978)).

Run the code below.

```
bgtest(y ~ z + p + trend + M1 + M2 + M3 +
         M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11,
       order=12, data=df2)

##
##  Breusch-Godfrey test for serial correlation of order up to 12
##
## data:  y ~ z + p + trend + M1 + M2 + M3 + M4 + M5 + M6 + M7 + M8 + M9 +
M10 + M11
## LM test = 6.5388, df = 12, p-value = 0.8865
```

For this test the null hypothesis is that there is no serial correlation of any order up to 12. Should it be rejected at a 5% level of significance?

We cannot reject the null hypothesis that there is no serial correlation for any order up to 12 because the p-value if 0.8865 which is great than the 0.05 we need to reject the null hypothesis. Therefore, there is serial correlation.

Why would one expect twelfth order serially correlated errors?

Twelfth order serial correlation is to be expected with certain time series data because it represents the same month but different year. If there is seasonality in the data, it would make sense there is twelfth order serial correlation.

3. Model (1) does not seem adequate since it implies a constant price elasticity for all prices. Model (2) is a more appealing equilibrium model (see the project pdf for details).

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 p_t + \beta_3 p_t^2 + \gamma_0 w_t + \sum_{i=1}^{11} \gamma_i x_{it} + \epsilon_t$$

For this model the price elasticity of demand is

$$\eta = \frac{\partial y_t}{\partial p_t} = \beta_2 + 2\beta_3 p_t$$

β3 is the long-run elasticity for its explanatory variable. Since the end of World War II β3 < 0, and of course β2 < 0.

(a) Goverment revenue, otherwise known as tax collection, is maximized when η = −1. Provide a formula for computing the revenue maximizing price for the model given by Equation (2). (Recall pt is the natural logarithm of Pt.)

P^*, the revenue maximizing price, is

$$P^* = e^{(-1-\beta_2)/2\beta_3}$$

(b) Estimate the model given by Equation (2). Report the coefficients, coefficient standard errors, R^2 and Adjusted R^2 of the model.

Enter your code in the chunk below. Store your regression output in an object called “reg3”.

```
# create the p-squared variable, add it to the df
# do not edit this; use "psquared" in this model
df2$psquared <- df2$p*df2$p

reg3 <- lm(y ~ z + p + psquared + trend + M1 +M2 + M3 + M4 + M5 + M6 + M7 +
M8 + M9 + M10 + M11, data=df2)
summary(reg3)

##
## Call:
## lm(formula = y ~ z + p + psquared + trend + M1 + M2 + M3 + M4 +
##     M5 + M6 + M7 + M8 + M9 + M10 + M11, data = df2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026314 -0.007919 -0.000922  0.007665  0.043666
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.520856   0.196822  28.050  < 2e-16 ***
## z            0.170681   0.054637   3.124  0.00239 **
## p           -0.048659   0.068468  -0.711  0.47908
## psquared    -0.006973   0.036375  -0.192  0.84840
## trend       -0.008372   0.002804  -2.986  0.00362 **
## M1          -0.043413   0.006136  -7.075 2.88e-10 ***
## M2          -0.092483   0.006143 -15.055  < 2e-16 ***
## M3           0.010029   0.006134   1.635  0.10549
## M4           0.003981   0.006157   0.647  0.51949
## M5           0.047599   0.006180   7.702 1.49e-11 ***
## M6           0.053524   0.006156   8.694 1.27e-13 ***
## M7           0.052601   0.006135   8.574 2.28e-13 ***
## M8           0.063090   0.006140  10.275  < 2e-16 ***
## M9           0.015571   0.006146   2.534  0.01298 *
## M10          0.032973   0.006108   5.398 5.22e-07 ***
## M11         -0.007548   0.006085  -1.240  0.21796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01288 on 92 degrees of freedom
## Multiple R-squared:  0.9306, Adjusted R-squared:  0.9193
## F-statistic: 82.25 on 15 and 92 DF,  p-value: < 2.2e-16
```

Calculate the price elasticity of demand for a real price of $3 (use the coefficients from your regression model).

Use the code chunk below to perform and display the calculation.

```
## $$\eta=\frac{\partial y_t}{\partial p_t}=\beta_2 + 2\beta_3 p_t$$

-0.048659+2* -0.006973

## [1] -0.062605
```

Is the estimated price elasticity sensible? Explain your answer.

I expect demand for gasoline to be pretty inelastic because there aren't really any other widely available and easily accessible alternatives to gas. The model above suggests that for a real price of $3, demand would decrease but only slightly which makes sense.

(c) Test for autocorrelated errors of order 1 using the Breusch-Godfrey test.

Enter your code in the chunk below.

```
bgtest(y ~ z + p + psquared + trend + M1 + M2 + M3 +
         M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11,
       order=1, data=df2)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  y ~ z + p + psquared + trend + M1 + M2 + M3 + M4 + M5 + M6 +     M7
+ M8 + M9 + M10 + M11
## LM test = 0.019318, df = 1, p-value = 0.8895
```

For this test the null hypothesis is that there is no serial correlation of order 1. Should it be rejected at a 5% level of significance?

No, the null hypothesis should not be rejected because the p-value of 0.8895 is greater than 0.05.

(d) Test for autocorrelated errors of up to order 12 using the Breusch-Godfrey test.

Enter your code in the chunk below.

```
bgtest(y ~ z + p + psquared + trend + M1 + M2 + M3 +
         M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11,
       order=12, data=df2)

##
##  Breusch-Godfrey test for serial correlation of order up to 12
##
## data:  y ~ z + p + psquared + trend + M1 + M2 + M3 + M4 + M5 + M6 +     M7
+ M8 + M9 + M10 + M11
## LM test = 6.8584, df = 12, p-value = 0.8668
```

For this test the null hypothesis is that there is no serial correlation of any order up to 12. Should it be rejected at a 5% level of significance?

No, the null hypothesis should not be rejected because the p-value of 0.8668 is greater than 0.05.

(e) Compute the revenue maximizing price using the formula you derived above.

Use the following chunk to compute and display this value.

```
## P^*=e^((-1-??_2 )/2??_3 )

(-1--0.048659)/(2*-0.006973)

## [1] 68.21605

exp(68.21605)

## [1] 4.225266e+29
```

Is this revenue maximizing price sensible? Explain your answer to this question.

No, this price doesn’t seem reasonable because it is so large. Nobody would ever pay this much for gas.

4. A problem with Model (2) is that demand adjusts simultaneously for changes in price and income. A model that incorporates dynamics, we'll call it Model (4), is

$$y_t = \beta_0 + \alpha_0 y_{t-1} + \delta_0 \Delta y_{t-1} + \beta_1 z_t + \beta_2 p_t + \beta_3 p_t^2 + \gamma_0 w_t + \sum_{i=1}^{11} \gamma_i x_{it} + \epsilon_t$$

where $\Delta y_{t-1} = y_{t-1} - y_{t-2}$. The parameters α_0 and δ_0 determine the short-run dynamics of the model. In equilibrium models all differences of variables are 0 (i.e., $\Delta y_{t-1} = 0$ for all t).

Thus the long-run price elasticity of demand of Model (4) is

$$\eta = \frac{\partial y_t}{\partial p_t} = (1 - \alpha_0)^{-1}(\beta_2 + 2\beta_3 p_t)$$

(a) Provide a formula for computing the revenue maximizing price for the model given by Equation (4).

The revenue maximizing price for model 4 is

$$P^* = e^{(\alpha_0 - 1 - \beta_2)/2\beta_3}$$

(b) Estimate model given by Equation (4). Report the coefficients, coefficient standard errors, and R2 and Adjusted R2 of the model.

Enter your code to complete the chunk below.

```r
# Dynamic model - additional explanatory variable
# do not edit this code; enter your code below
df3 <- df2
df3$lagy1 <- lag(df3$y)
df3$lagy2 <- lag(df3$y,2)
df3$deltay1y2 <- df3$lagy1 - df3$lagy2

reg4 <- lm(y ~ lagy1 + deltay1y2 + z + p + psquared + trend + M1 +M2 + M3 +
M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11, data=df3)
summary(reg4)

##
## Call:
## lm(formula = y ~ lagy1 + deltay1y2 + z + p + psquared + trend +
##     M1 + M2 + M3 + M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11, data = df3)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.022807 -0.008500 -0.001119  0.007574  0.043095
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.726443   0.791987   5.968 4.94e-08 ***
## lagy1        0.150304   0.144593   1.039  0.30142
```

```
## deltay1y2   -0.128150   0.103843  -1.234  0.22046
## z            0.133525   0.066131   2.019  0.04652 *
## p           -0.031730   0.076986  -0.412  0.68123
## psquared    -0.011727   0.040912  -0.287  0.77507
## trend       -0.006951   0.004100  -1.695  0.09353 .
## M1          -0.037696   0.007616  -4.950 3.55e-06 ***
## M2          -0.090220   0.007531 -11.979  < 2e-16 ***
## M3           0.020163   0.012116   1.664  0.09964 .
## M4           0.018067   0.013686   1.320  0.19025
## M5           0.049022   0.006627   7.398 7.73e-11 ***
## M6           0.055128   0.009091   6.064 3.24e-08 ***
## M7           0.048785   0.009254   5.272 9.53e-07 ***
## M8           0.058483   0.009398   6.223 1.61e-08 ***
## M9           0.010719   0.010222   1.049  0.29723
## M10          0.028017   0.007514   3.729  0.00034 ***
## M11         -0.006457   0.007727  -0.836  0.40559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01294 on 88 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.9292, Adjusted R-squared:  0.9155
## F-statistic: 67.92 on 17 and 88 DF,  p-value: < 2.2e-16
```

Calculate the price elasticity of demand for a real price of $3.

Use the code chunk below to perform and display the calculation.

```
(1-0.150304)^(-1)*(-0.031730+2*-0.011727)

## [1] -0.06494558
```

Is the estimated price elasticity sensible? Explain your answer.

I expect demand for gasoline to be pretty inelastic because there aren’t really any other widely available and easily accessible alternatives to gas. The model above suggests that for a real price of $3, demand would decrease but only slightly which makes sense.

(c) Test for autocorrelated errors of order 1 using the Breusch-Godfrey test.

Enter your code in the chunk below.

```
bgtest(y ~ lagy1 + deltay1y2 + z + p + psquared + trend + M1 +M2 + M3 + M4 +
M5 + M6 + M7 + M8 + M9 + M10 + M11, order = 1, data=df3)

##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  y ~ lagy1 + deltay1y2 + z + p + psquared + trend + M1 + M2 +     M3
+ M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11
## LM test = 1.7986, df = 1, p-value = 0.1799
```

For this test the null hypothesis is that there are no autocorrelated errors of order 1. Should it be rejected at a 5% level of significance?

No, the null hypothesis that no autocorrelation errors of order 1 exist should not be rejected because the p-value of 0.1799 is greater than 0.05.

(d) Test for autocorrelated errors of up to order 12 using the Breusch-Godfrey test.

Enter your code in the chunk below.

```
bgtest(y ~ lagy1 + deltay1y2 + z + p + psquared + trend + M1 +M2 + M3 + M4 +
M5 + M6 + M7 + M8 + M9 + M10 + M11, order = 12, data=df3)

##
##  Breusch-Godfrey test for serial correlation of order up to 12
##
## data:  y ~ lagy1 + deltay1y2 + z + p + psquared + trend + M1 + M2 +     M3
+ M4 + M5 + M6 + M7 + M8 + M9 + M10 + M11
## LM test = 12.045, df = 12, p-value = 0.4421
```

For this test the null hypothesis is that there is no serial correlation of any order up to 12. Should it be rejected at a 5% level of significance?

No, the null hypothesis that no autocorrelation errors of order 1 exist should not be rejected because the p-value of 0.4421 is greater than 0.05.

(e) Compute the revenue maximizing price using the formula you derived above.

Use the following chunk to compute and display this value.

```
(0.150304-1--0.031730)/(2*-0.011727)

## [1] 34.87533

exp((0.150304-1--0.031730)/(2*-0.011727))

## [1] 1.400115e+15
```

Is this revenue maximizing price sensible? Explain your answer to this question.

No, this price isn't sensible either because it is way too costly to be for a gallon of gas.

(f) The long-run estimates of the elasticity are βi divided by $1 - \alpha 0$ for $i = \{1, 2, 3\}$. Calculate these elasticities.

Use the code chunk below to compute and display the three results.

```
1/(1-0.150304)

## [1] 1.176892

2/(1-0.150304)
```

```
## [1] 2.353783
```

```r
3/(1-0.150304)
```

```
## [1] 3.530675
```

(g) What suggestions do you have that may further improve the model fit and its usability to determine tax policy (by changing prices)?

I would want to further dive into the autocorrelation by looking at different cuts of the data (daily, weekly) to see if the autocorrelation still exists with different cuts of the data. Additionally, additional variables could be included in the model such as crude oil supply. Lastly, I'd want to look into and adjust the model for seasonality by using differencing.