Final Project: Guide to Minimize Spending on Gas

Benjamin Chu

2/10/2017

Introduction

Taiwan imports more than 97% of its oil. As such, international price fluctuations in crude oil is almost directly reflected in consumer's receipt. From 2007 to 2008, crude oil price nearly doubled, while from 2014 to 2015 it decreased by more than half, and this trend is very similarly reflected in the prices we see in gas stations. How can citizens of a country such as Taiwan save as much money as possible when they don't produce oils themselves?

Fortunately, the Taiwanese people are not completely hopeless. The CPC corporation is a state owned gasoline company distributing approximately 70% of the gas nationally. This company employs a unique business strategy: any changes in gas price will be announced exactly 1 week in advance. Because of this, the Taiwanese people have time to react to price changes and devise some kind of strategy that will potentially help them save money. I wish to determine a master strategy that can exploit this extra piece of information as much as possible. In the long run, how much money can be saved?

Method Description

I will answer this question through a combination of simulation, surveying, and data analysis on gas price. From CPC's website, I obtained a table listing the dates of price adjustments, and prices after adjustment for 12 different kinds of purchasable gas from 1999 to 2016. I plan to focus on gas prices for the 3 major automobile gases (92, 95, 98). First I will conduct a survey to understand what is the typical refueling process. Based on this data, I will write R scripts that simulate the refueling process according to different refueling strategies. Essentially these scripts will simulate hypothetical people reliving in 1999, each with some kind of refueling strategy, and their expenditures over the course of 17 years will be noted. Then I will analyze the result and try to find the best strategy to conserve fuel.

Simulation

My strategy for simulating the refueling process can be accomplished by modeling groups of people driving independently, employing different refueling strategy. Each day fuels will be lost according a random variable $X \sim N(10, 5^2)$ where negative entries will be set to 0 to reflect a person "not driving" in a particular day.The mean and variance of this normal distribution will be estimated based on a survey I will conduct.

A **control group** will refuel whenever their fuel tank drop below a certain number. A **test group** will start considering refueling whenever their tank drop below a different number, and depending on whether they have information on the following week's gas price, choose a strategy that will (1) not run out of fuel at any point, and (2) minimize their spending on gas. Ultimately I can run these simulations 100 times each and compare their average spending. Hopefully that will give me useful information to answer my master question.

Survey

I plan to release a google survey form into the internet. Main questions will be:

- How many days per month do you drive?
- How ofter do you refuel?
- At what percentage of remaining fuel do you start considering refueling?
- Age
- Sex

The purpose for the last two questions is for me to account for non-response biases, as they will allow me to perform post-stratification.

Data Analysis

I will run the control group and numerous test groups with different strategies 100 times each. The strategies will be partly determined by the online survey result, with minor adjustments as I see fit. This will produce data on average spendings ordered by strategy, and in principle, I can calculate the mean, variance, p-values…etc, which I can use as evidence for finding the "master strategy." Afterwards, depending on simulation results, I will determine whether analytical solutions will be interesting (e.g. perhaps you can't save money no matter what you do) and leave that for future work.

Simulation Code

Control group

Now I have two datasets CPC_data_short.xlsx and CPC_data.xlsx, representing Taiwan's gas price fluctuations of 2009 and from 1999~2017, respectively. Below are descriptions of major functions:

`control`: Simulates the refueling process of the control group. The input is a data set of interest and the fuel type (currently limited to 92, 95, or 98). The output is expense_list, which is a list element constructed the following way: everytime we refuel, we will increase the total_expense by some amount, and the new total expense will be added to the expense_list. This way, we could keep track of the total money we spent on gas as a function of time.

`test-group` Have not been implemented yet. It will essentially be the same as the control group, but they will start considering refueling earlier, and decide if they want to refuel immediately or wait until the following week.

`lose-fuel`: Everyday people will lose $x\%$ of fuels, where $x \sim N(\mu, \sigma^2)$. If $x < 0$, we set it equal to zero (people can't gain fuel other than refueling). Setting equal to zero actually reflects reality better, because people don't always drive.

```
library(readxl) # parsing excel
library(lubridate) # dates/months/years addition made easy
```

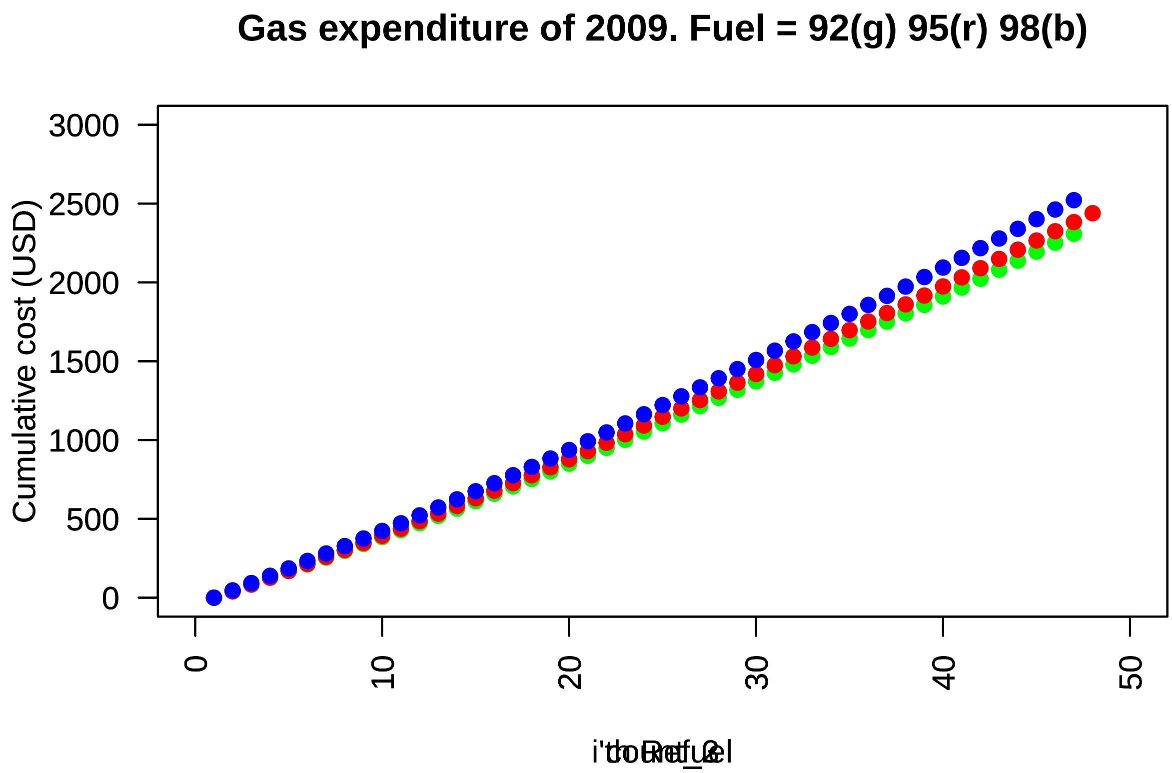
```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```
mydata <- read_excel("CPC_data_short.xlsx")

X_VALUE = 10
Y_VALUE = 30
```

```
Z_VALUE = 40

assign_fuel_type <- function(fuel_type) { # identity fuel type (92/95/98)
  dummy_var = 0
  if (fuel_type == 92) {dummy_var = 2}
  else if (fuel_type == 95) {dummy_var = 3}
  else if (fuel_type == 98) {dummy_var = 4}
  else {
    print("please enter either 92, 95, or 98")
    return()
  }
  return(dummy_var)
}

lose_fuel <- function() { # everyday lose fuel ~ normal
  SD = 5
  lost = rnorm(1, X_VALUE, SD) #generate x ~ N(mean, SD)

  if (lost < 0) {
    lost = 0
  }
  return(lost)
}

control <- function(fuel_type, table) {
  # fuel_type = 92, 95, or 98

  which_column = assign_fuel_type(fuel_type)
  my_price_list = table[[which_column]]
  my_date_list = table[[1]]

  current_date = ymd(my_date_list[1])
  current_price = my_price_list[1]
  my_price_list <- my_price_list[-1] #delete first element
  my_date_list <- my_date_list[-1] #delete first element
  last_day = ymd(my_date_list[length(my_date_list)])

  total_expense = 0 # keep track of total money spent
  expense_list <- list(total_expense) # keeps track of marginal expense
  current_fuel = 100

  while (current_date != last_day) {
    if(current_fuel < Y_VALUE) { # refuel when necessary
      total_expense = total_expense + 57.0 * current_price #57 liters ~ 15 gallons ~ approx gas tank si
      expense_list <- append(expense_list, total_expense / 30.0)
      current_fuel = 100
    }

    current_fuel = current_fuel - lose_fuel()
    current_date = current_date + ddays(1)

    #check if today needs price adjustments. Remember to watch out for NA.
    if (current_date == ymd(my_date_list[1])) {
```

```
      my_date_list <- my_date_list[-1]
      if (!is.na(my_price_list[1])) {
        current_price = my_price_list[1]
      }
      my_price_list <- my_price_list[-1]
    }
  }
  return(expense_list)
}


test_group <- function(fuel_type, table) {
  # fuel_type = 92, 95, or 98

  which_column = assign_fuel_type(fuel_type)
  my_price_list = table[[which_column]]
  my_date_list = table[[1]]

  current_date = ymd(my_date_list[1])
  current_price = my_price_list[1]
  my_price_list <- my_price_list[-1] #delete first element
  my_date_list <- my_date_list[-1] #delete first element
  last_day = ymd(my_date_list[length(my_date_list)])

  total_expense = 0 # keep track of total money spent
  expense_list <- list(total_expense) # keeps track of marginal expense
  current_fuel = 100

  while (current_date != last_day) {
    if(current_fuel < Y_VALUE) { # refuel when necessary
      total_expense = total_expense + 57.0 * current_price #57 liters ~ 15 gallons ~ approx gas tank si
      expense_list <- append(expense_list, total_expense / 30.0)
      current_fuel = 100
    }

    current_fuel = current_fuel - lose_fuel()
    current_date = current_date + ddays(1)

    #check if today needs price adjustments. Remember to watch out for NA.
    if (current_date == ymd(my_date_list[1])) {
      my_date_list <- my_date_list[-1]
      if (!is.na(my_price_list[1])) {
        current_price = my_price_list[1]
      }
      my_price_list <- my_price_list[-1]
    }
  }
  return(expense_list)
}


test_1 = control(92, mydata)
number_1 = length(test_1)
```

```
count_1 = 1:number_1
plot(count_1, test_1, col="green", las = 2, xlim = c(0, 50), ylim = c(0, 3000), main = "Gas expenditure

test_2 = control(95, mydata)
number_2 = length(test_2)
count_2 = 1:number_2
par(new=TRUE) #plots the second graph on top of first graph
plot(count_2, test_2, col="red", las = 2, xlim = c(0, 50), ylim = c(0, 3000), ylab="Cumulative cost (USD

test_3 = control(98, mydata)
number_3 = length(test_3)
count_3 = 1:number_3
par(new=TRUE) #plots the second graph on top of first graph
plot(count_3, test_3, col="blue", las = 2, xlim = c(0, 50), ylim = c(0, 3000), ylab="Cumulative cost (US
```



Results

none so far…

References

- http://new.cpc.com.tw/division/mb/oil-more4.aspx
- http://www.macrotrends.net/1369/crude-oil-price-history-chart
- http://www.insight-post.tw/insight-report/20161010/16027
- http://www.restore.ac.uk/PEAS/nonresponse.php

- Mant pages of Wiki