iphone analysis

Boyie Chen

12/30/2020

iPhone Second Hand Market Analysis

Enviroment Setting

```
library(rvest)
```

```
## Loading required package: xml2
```

```
library(httr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()         masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()            masks stats::lag()
## x purrr::pluck()          masks rvest::pluck()
```

```
library(stringr)
library(readr)
library(dplyr)
library(sjmisc) #dummies
```

```
##
## Attaching package: 'sjmisc'
```

```
## The following object is masked from 'package:purrr':
##
##     is_empty
```

```
## The following object is masked from 'package:tidyr':
##
##     replace_na
```

```
## The following object is masked from 'package:tibble':
##
##     add_case
```

```
#robust s.e.
#https://www.brodrigues.co/blog/2018-07-08-rob_stderr/
```

```r
library(robustbase)
library(tidyverse)
library(sandwich)
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(modelr)
#library(broom)
library("ivreg")
options(stringsAsFactors = F)
httr::set_config(httr::config(http_version = 0))
```

Loading Data

```r
# environment
load("/Users/Andy 1/Google 雲端硬碟 (r08323004@g.ntu.edu.tw)/0 Semesters/109-1/— 234 資料科學/0_Final_Projec
```

Data Cleaning

```r
# all iphone prices
tidy_df <- Posts %>%
    mutate(IsSell = ifelse(str_detect(ptitle, "\\[販售\\]|\\[賣 | 出售 | 販售 | 售") |
                           str_detect(string = pcontent, pattern = " 售價"), 1, 0),
           IsBuy = ifelse(str_detect(ptitle, "\\[收購\\]|\\[買") |
                          str_detect(string = pcontent, pattern = " 我想要買"), 1, 0),
           IsChange = ifelse(str_detect(ptitle, "\\[交換\\]|\\[換"), 1, 0))%>%
    mutate(IsSold = ifelse(str_detect(string = ptitle, pattern = " 售出 | 已售出 | 已售 | 已出售"), 1, 0),
           IsBought = ifelse(str_detect(string = ptitle, pattern = " 徵到 | 徵得 | 已徵得 | 已徵到 | 收到 | 已購
    filter(!(IsBuy==0 & IsSell ==0 & IsChange ==0 & IsSold==0 & IsBought==0))
#extract prices
tidy_df <- tidy_df %>%
    mutate(price_text = str_extract(pcontent, " 價.{1,12}(\\d{2,5}){1,2}| 價格.{1,8}(\\d{2,5}){1,2}")) %>
    mutate(price_text = gsub(",", "", price_text)) %>%
    mutate(price = str_extract_all(price_text, "\\d{3,4}0")) %>%
    filter(lengths(price)<=2) %>%
    drop_na() %>%
    unnest(price) %>%
    mutate(price = as.numeric(price)) %>%
    group_by(url) %>%
    mutate(avg_price = mean(price)) %>%
    ungroup() %>%
    filter(!duplicated(url)) %>%
    select(-price,-price_text)
```

```
## Warning: Problem with `mutate()` input `price`.
## i 強制變更過程中產生了 NA
## i Input `price` is `as.numeric(price)`.
```

```
## Warning in mask$eval_all_mutate(dots[[i]]): 強制變更過程中產生了 NA
```

```r
#hist(tidy_df$avg_price)

color_pattern <- regex("銀|金|灰|玫瑰|白|綠|紅|藍|石墨")

tidy_df <- tidy_df %>%
    drop_na() %>%
    # ROM
    mutate(ROM_text = str_extract(pcontent, "[1-2]\\d{2}G|[1-2]\\d{2}g|\\d{2}G|\\d{2}g")) %>%
    mutate(ROM = str_extract(ROM_text, "\\d{2,3}")) %>%
    mutate(ROM = as.numeric(ROM)) %>%
    select(-ROM_text) %>%
    # Color
    mutate(color_c = str_extract(pcontent, color_pattern)) %>%
    mutate(color_t = str_extract(ptitle, color_pattern)) %>%
    mutate(color = ifelse(!is.na(color_c), color_c, color_t)) %>%
    select(-color_t, -color_c) %>%
    mutate(color = ifelse(!is.na(color), color, "No_Color")) %>%

    # Usage
    mutate(ptime = parse_datetime(ptime, format = "%a %b %d %H:%M:%S %Y")) %>%
    #mutate(ptime = as.POSIXct(ptime)) %>%
    #mutate(referenceTime = as.POSIXct(parse_datetime("2015-09-25", "%Y-%m-%d"))) %>%
    #mutate(TimeUsed = as.numeric(difftime(ptime-referenceTime, units='days'))/365) %>%
    #filter(TimeUsed>=0) %>%
    mutate(TimeUsed = as.numeric(ptime)) %>%

    # Gender
    mutate(IsFemale = ifelse(str_detect(string = pcontent, pattern = "女用機|女用|女生用"), 1, 0))%>%
    mutate(IsMale = ifelse(str_detect(string = pcontent, pattern = "男用機|男用|男生用"), 1, 0))
```

```
## Warning: Problem with `mutate()` input `ROM`.
## i 強制變更過程中產生了 NA
## i Input `ROM` is `as.numeric(ROM)`.

## Warning: 強制變更過程中產生了 NA

## Warning: Problem with `mutate()` input `ptime`.
## i 39 parsing failures.
##    row col                       expected                         actual
##   9501  -- date like %a %b %d %H:%M:%S %Y Fri Sep 11 10:53:17
##  15503  -- date like %a %b %d %H:%M:%S %Y Tue Feb 18 11:522:15 2020
##  19037  -- date like %a %b %d %H:%M:%S %Y Wed Jan  1 22:33:50
##  22943  -- date like %a %b %d %H:%M:%S %Y Mon Nov 25 11:00:12
##  27178  -- date like %a %b %d %H:%M:%S %Y Tue Oct 22 20:06:26 2019已徵到
## ..... ... .............................. ..............................
## See problems(...) for more details.
##
## i Input `ptime` is `parse_datetime(ptime, format = "%a %b %d %H:%M:%S %Y")`.

## Warning: 39 parsing failures.
##    row col                       expected                         actual
##   9501  -- date like %a %b %d %H:%M:%S %Y Fri Sep 11 10:53:17
##  15503  -- date like %a %b %d %H:%M:%S %Y Tue Feb 18 11:522:15 2020
##  19037  -- date like %a %b %d %H:%M:%S %Y Wed Jan  1 22:33:50
```

```
## 22943  -- date like %a %b %d %H:%M:%S %Y Mon Nov 25 11:00:12
## 27178  -- date like %a %b %d %H:%M:%S %Y Tue Oct 22 20:06:26 2019已 徵 到
## ..... ... .............................. ..............................
## See problems(...) for more details.
```

```r
# get dummies
tidy_df <- tidy_df %>%
    to_dummy(color, suffix = "label") %>%
    bind_cols(tidy_df) %>%
    select(everything())
```

iPhone 6s Specific Data Cleaning

```r
# Conditional on iPhone 6s
# 型號、RAM/ROM 容量、color、發文時間（作為使用幾年的依據）
type_pattern <- regex("plus|\\+", ignore_case=TRUE)

iphone6s_df <- tidy_df %>%
    filter(str_detect(string = ptitle, pattern = "6s")) %>%
    # Type
    mutate(Is_6s_plus = (ifelse(str_detect(ptitle, type_pattern), 1, 0))) %>%
    # ROM
    mutate(TimeUsed = as.numeric(ptime-parse_datetime("2015-09-25", "%Y-%m-%d"))/24) %>%
    filter(TimeUsed>=0)
```

Explorative Data Analysis

Inference Statistics

1. 「女用機」比較值錢？

在 Macshop 版中，有時可以見到欲出售 iphone 的貼文內標注「女用機」或「女生用」等字眼，通常被認為是要顯示「原主人為細心使用的女性，因此手機保養得很好，品項優良」等隱含訊息。然而，「女用機」真的可以賣得比較貴嗎？亦或是其實並沒有呢？若有，我們暫且將此 “premium” 稱為 “gender rent”

```r
tidy_df %>%
  filter(IsFemale==1) %>%
  nrow
```

```
## [1] 2922
```

```r
tidy_df %>%
  filter(IsMale==1) %>%
  nrow
```

```
## [1] 83
```

在 163950 則貼文中，標注「男用機」的貼文有 83 則，標注「女用機」的貼文有 2922 則，是男用機的 35 倍。

```r
# Gender Rent
gender_df <- tidy_df %>%
  filter(IsBought==1 | IsSold==1)
reg1 <- lm(formula = avg_price ~ IsFemale + IsMale + TimeUsed + ROM +
           color_ 紅 + color_ 灰 + color_ 金 + color_ 藍 + color_ 綠 + color_ 玫瑰 + color_ 銀,
           data = gender_df)
summary(reg1)
```

```
##
## Call:
## lm(formula = avg_price ~ IsFemale + IsMale + TimeUsed + ROM +
##     color_紅 + color_灰 + color_金 + color_藍 + color_綠 +
##     color_玫瑰 + color_銀, data = gender_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -33551  -6566  -1410   6351  39696
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.677e+04  3.826e+03  17.450  < 2e-16 ***
## IsFemale    -5.205e+03  8.737e+02  -5.957 2.82e-09 ***
## IsMale      -6.481e+03  4.150e+03  -1.562   0.1185
## TimeUsed    -3.974e-05  2.607e-06 -15.244  < 2e-16 ***
## ROM          8.748e+01  2.235e+00  39.133  < 2e-16 ***
## color_紅     3.956e+03  7.457e+02   5.305 1.20e-07 ***
## color_灰     2.694e+03  4.609e+02   5.845 5.51e-09 ***
## color_金     2.587e+03  3.787e+02   6.831 9.86e-12 ***
## color_藍     9.708e+03  1.063e+03   9.132  < 2e-16 ***
## color_綠     9.449e+03  9.492e+02   9.954  < 2e-16 ***
## color_玫瑰   1.051e+03  5.109e+02   2.058   0.0397 *
## color_銀     2.078e+03  4.547e+02   4.570 5.05e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8288 on 3559 degrees of freedom
##   (113 observations deleted due to missingness)
## Multiple R-squared:  0.3611, Adjusted R-squared:  0.3592
## F-statistic: 182.9 on 11 and 3559 DF,  p-value: < 2.2e-16
```

```r
coeftest(reg1, vcov = vcovHC(reg1, type = "HC3"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)  6.6765e+04  3.8878e+03  17.1730 < 2.2e-16 ***
## IsFemale    -5.2049e+03  5.8177e+02  -8.9466 < 2.2e-16 ***
## IsMale      -6.4808e+03  2.8031e+03  -2.3120   0.02084 *
## TimeUsed    -3.9740e-05  2.6729e-06 -14.8675 < 2.2e-16 ***
## ROM          8.7480e+01  2.4696e+00  35.4230 < 2.2e-16 ***
## color_紅     3.9561e+03  6.8944e+02   5.7382 1.037e-08 ***
## color_灰     2.6941e+03  4.6818e+02   5.7544 9.428e-09 ***
## color_金     2.5873e+03  3.6616e+02   7.0661 1.909e-12 ***
## color_藍     9.7082e+03  1.5156e+03   6.4054 1.696e-10 ***
## color_綠     9.4491e+03  1.1865e+03   7.9641 2.220e-15 ***
## color_玫瑰   1.0511e+03  4.8286e+02   2.1768   0.02956 *
## color_銀     2.0776e+03  4.6752e+02   4.4440 9.099e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

事實上，與預期相反，在所有型號中且是「已售出」或「已徵得」的 iPhone 中，標注有「女用機」的貼文的成交價格是比較低的。

```r
gender_df <- iphone6s_df %>%
  filter(IsBought==1 | IsSold==1)

reg1 <- lm(formula = avg_price ~ IsFemale + IsMale + TimeUsed + ROM +
           color_ 灰 + color_ 金 + color_ 玫瑰 + color_ 銀,
           data = gender_df)
summary(reg1)
```

```
##
## Call:
## lm(formula = avg_price ~ IsFemale + IsMale + TimeUsed + ROM +
##     color_灰 + color_金 + color_玫瑰 + color_銀, data = gender_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20150.9  -2055.3   -348.9   1901.0  11235.0
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21343.6835   885.7408  24.097  < 2e-16 ***
## IsFemale     -457.1488  1187.1237  -0.385    0.700
## IsMale              NA         NA      NA       NA
## TimeUsed      -15.2216     0.4118 -36.966  < 2e-16 ***
## ROM            50.2456     7.4092   6.782 5.67e-11 ***
## color_灰     -572.0821   892.9311  -0.641    0.522
## color_金      941.8240   851.0533   1.107    0.269
## color_玫瑰   1200.3281   789.4811   1.520    0.129
## color_銀     -294.1154   959.3444  -0.307    0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3452 on 323 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.8214, Adjusted R-squared:  0.8175
## F-statistic: 212.2 on 7 and 323 DF,  p-value: < 2.2e-16
```

```r
coeftest(reg1, vcov = vcovHC(reg1, type = "HC3"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 21343.68352  1480.37045  14.4178 < 2.2e-16 ***
## IsFemale     -457.14882  1492.68059  -0.3063    0.7596
## TimeUsed      -15.22156     0.55273 -27.5388 < 2.2e-16 ***
## ROM            50.24563     8.26413   6.0800  3.39e-09 ***
## color_灰     -572.08213  1215.11728  -0.4708    0.6381
## color_金      941.82397  1267.18374   0.7432    0.4579
## color_玫瑰   1200.32805  1205.09409   0.9960    0.3200
## color_銀     -294.11541  1313.01866  -0.2240    0.8229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

若是限制在 iPhone 6s 這個機型的話，則沒有顯著地異於零，但大致的方向仍是負的，意味著標示「女用機」並沒有辦法「提升價格」。事實上，在 PTT 的文化中，更常見的是嘲諷標注女用機的貼文者，或許這可以部分地解釋為何標示了「女用機」反而會有比較低的成交價格。

2. Estimating the Willingness To Pay & the Willingness To Be Paid

由於 iPhone 6s 此一機型非常地暢銷，自 2015 年發售以來便持續活躍於二手交易市場，以致於在關鍵字為 iPhone 的 19 萬則貼文中，儘管 iPhone 有從 3GS 到 iPhone 12 等數十個型號，但仍有近 10% 的貼文是此單一機型，因此我們將所估計的財貨限制在此一型號

首先，由於二手交易版需要使用者在貼文時標注自己希望「出售」還是「購買」(徵求)。在使用者打算「徵求」購買一支二手的手機時，我們可以將帶有「購買」分類標籤的貼文簡單地視為使用者標注了自己的 "Willingness To Pay"(願付價格)；而對於打算出售二手手機的使用者，文內的「希望價格」我們則以 "Willingness To Be Paid" 來稱之。

若分別以 inverse supply function 以及 inverse demand function 的角度視之，我們可以分別寫下：

$$p_t = \alpha_0 + \alpha_1 q_t^s + \alpha_2 W_t + u_t$$
$$p_t = \alpha_0 + \alpha_1 q_t^d + \alpha_2 W_t + u_t$$

其中，q_t^s, q_t^d 分別代表時間 t 時，iPhone 6s 的供給以及需求

在此，由於我們有區別是在 supply side 或 demand side 的分類標籤，因此無須處理 simultaneous equation 的問題。當然，我們所估計的上面兩條迴歸式尚且不能稱之為供給及需求函數，但可以作為此二函數的近似。在此，我們尚且稱呼此二式為 Willingness To Be Paid (WTBP) 及 Willingness To Pay (WTP)

WTBP

```
# iphone 6s : 15873 obs
# supply
i6s_WTBP_df <- iphone6s_df %>%
    filter(IsSell == 1) %>%
    mutate(week = as.Date(cut(ptime, "week"))) %>%
    group_by(week, ROM) %>%
    mutate(quantity = n()) %>%
    ungroup() %>%
    group_by(week) %>%
    summarise(price = mean(avg_price),
              quantity = quantity,
              ROM = ROM,
              TimeUsed = mean(TimeUsed),
              Is_6s_plus = Is_6s_plus) %>%
    filter(!duplicated(week, ROM))
```

```
## `summarise()` regrouping output by 'week' (override with `.groups` argument)
```

```
reg3 <- lm(formula = price ~ quantity + TimeUsed + ROM + Is_6s_plus,
           data = i6s_WTBP_df)
summary(reg3)
```

```
##
## Call:
## lm(formula = price ~ quantity + TimeUsed + ROM + Is_6s_plus,
##     data = i6s_WTBP_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3875.8 -1865.2  -423.1  1750.0  9013.6
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22346.2630   441.2814  50.639  < 2e-16 ***
```

```
## quantity       20.5398     4.5111   4.553 8.19e-06 ***
## TimeUsed      -12.5391     0.2896 -43.301  < 2e-16 ***
## ROM             3.2760     4.9763   0.658    0.511
## Is_6s_plus    160.4526   292.7805   0.548    0.584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2259 on 255 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.9135, Adjusted R-squared:  0.9122
## F-statistic: 673.7 on 4 and 255 DF,  p-value: < 2.2e-16
```

```r
coeftest(reg3, vcov = vcovHC(reg3, type = "HC3"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 22346.26298   543.25356  41.1341 < 2.2e-16 ***
## quantity       20.53982     5.76329   3.5639  0.000436 ***
## TimeUsed      -12.53909     0.35069 -35.7552 < 2.2e-16 ***
## ROM             3.27603     4.90388   0.6680  0.504706
## Is_6s_plus    160.45261   289.86773   0.5535  0.580381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

也就是，WTBP:

$$p = 22346.26 + 20.54q^s$$

WTP

```r
# demand
i6s_WTP_df <- iphone6s_df %>%
    filter(IsBuy == 1) %>%
    mutate(week = as.Date(cut(ptime, "week"))) %>%
    group_by(week, ROM) %>%
    mutate(quantity = n()) %>%
    ungroup() %>%
    group_by(week) %>%
    summarise(price = mean(avg_price),
              quantity = quantity,
              ROM = ROM,
              TimeUsed = mean(TimeUsed),
              Is_6s_plus = Is_6s_plus) %>%
    filter(!duplicated(week, ROM))
```

```
## `summarise()` regrouping output by 'week' (override with `.groups` argument)
```

```r
reg4 <- lm(formula = price ~ quantity + TimeUsed + ROM + Is_6s_plus,
           data = i6s_WTP_df)
summary(reg4)
```

```
##
## Call:
## lm(formula = price ~ quantity + TimeUsed + ROM + Is_6s_plus,
##     data = i6s_WTP_df)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -5265  -2597  -1053   2074  16395
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23691.7842   944.5196  25.083   <2e-16 ***
## quantity       -8.5358    42.3515  -0.202   0.8405
## TimeUsed      -12.4746     0.5611 -22.232   <2e-16 ***
## ROM           -18.4653     9.9484  -1.856   0.0650 .
## Is_6s_plus   1134.3397   523.7186   2.166   0.0316 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3496 on 184 degrees of freedom
##   (81 observations deleted due to missingness)
## Multiple R-squared:  0.7816, Adjusted R-squared:  0.7769
## F-statistic: 164.7 on 4 and 184 DF,  p-value: < 2.2e-16
```

```r
coeftest(reg4, vcov = vcovHC(reg4, type = "HC3"))
```

```
##
## t test of coefficients:
##
##                Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 23691.7842   954.1712  24.8297  < 2e-16 ***
## quantity       -8.5358    55.3531  -0.1542  0.87762
## TimeUsed      -12.4746     0.7538 -16.5489  < 2e-16 ***
## ROM           -18.4653     8.3369  -2.2149  0.02799 *
## Is_6s_plus   1134.3397   552.4897   2.0531  0.04147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

也就是，WTP:

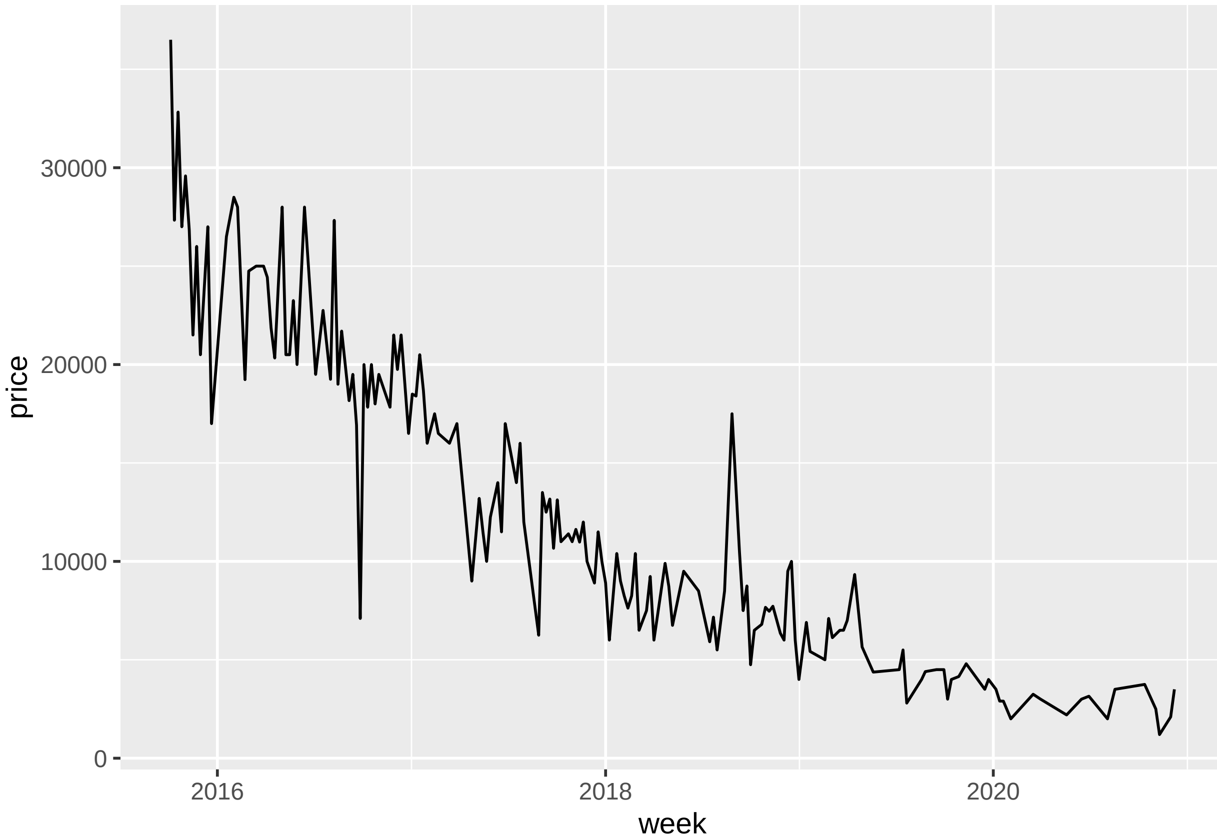
$$p = 23691.78 - 8.54q^d$$

3. Estimating the demand function for iPhone 6s

iPhone 6s 週成交價格走勢圖

```r
iphone6s_df %>%
    filter(IsBought == 1 | IsSold == 1) %>%
    mutate(week = as.Date(cut(ptime, "week"))) %>%
    group_by(week, ROM) %>%
    mutate(quantity = n()) %>%
    ungroup() %>%
    group_by(week) %>%
    summarise(price = mean(avg_price),
              quantity = quantity,
              ROM = ROM,
              TimeUsed = mean(TimeUsed),
              Is_6s_plus = Is_6s_plus) %>%
    filter(!duplicated(week, ROM)) %>%
    ggplot(aes(week, price)) +
    geom_line()
```

```
## `summarise()` regrouping output by 'week' (override with `.groups` argument)
```



有別於 (2.) 中的估計，我們可以針對「已售出」或「已徵得」的貼文（以下稱之為已成交的貼文），將其內的價格視為「均衡價格」，並利用 IV Estimation 來控制住 supply shifter，藉此估計出 demand function。

為何需要 supply shifter?

由於我們所觀察到的所有的成交的「價格」與「數量」(paired data) 皆是在「均衡」時所觀察到的，然而，均衡是需求等於供給之時，因此我們無法簡單地把這些均衡的價格與數量描繪成 scatter plot，然後知道這些點應該沿著需求還是供給線移動。

若是簡單地將價格對數量跑迴歸，將會面臨內生性問題，這是源自於 simultaneous equation 的問題。

這意味著，若我們想要知道需求曲線，就需要找到「只會影響供給，但不會影響需求的變數」，以此作為 supply shifter，我們便可描繪出需求曲線。符合此條件的 supply shifter 實際上就是一個 IV，因此我們可以簡單地透過 2SLS 或者 IV Regression 來得到對於 demand function 正確地估計（至少是避免內生性問題的估計）

在此，我們 propose 一個 IV，它是「販賣或購買 iPhone 6s 的貼文內是否標注了『台積電/TSMC 晶片』或『三星/Samsung 晶片』」，因此我們的 IV 為兩個 Dummy Variable，分別以 `IsTSMC` 及 `IsSamsung` 稱之。

為何以此作為 IV？

由於 2015.9.25 發表 iPhone 6s 及 6s Plus 之後，隨即發生了「晶片門」事件。事件經過大致是 6s 的 A9 晶片製造廠商分別有台積電以及三星，但三星的晶片有易發熱、速度較慢、電池續航較差等問題，導致拿到三星晶片版本的消費者不滿，希望退貨，在十月開始在香港等其他地區亦陸續有退貨潮。然而，台灣當時尚無蘋果直營店，因此除了透過官網購買的消費者外，並無無條件退貨的門路。再考量到當時主要取得貨源的三大電信業者皆為消費者推出各類針對 iPhone 新機的綁約服務，我們認定「晶片門」事件會影響 iPhone 二手交易市場的「供給」，因為拿到三星版晶片的消費者退貨無門，較可能傾向在二手市場出售手上的手機，而且以較低的價格出售；若是拿到台積電版本晶片的消費者反而可以獲得一些「套利」空間，以較高的價格出售。

至於需求，我們認定「晶片門」事件對「iPhone 6s 二手交易市場的需求」沒有影響，因為 iPhone 6s 在事件爆發時尚屬新品，且在二手市場有需求者大多是對於價格較實惠的舊世代產品有需求，所以我們可以認為晶片門事件對於本來就想買二手 iPhone 6s 的消費者需求沒有影響。

```r
# 晶片門
chip_pattern <- regex("A9 晶片 | 台積電 | 三星 |TSMC|samsung",ignore_case=TRUE)
i6s_IV_df <- iphone6s_df %>%
    #filter(IsBought==1 | IsSold==1) %>%
    mutate(IsTSMC = ifelse(str_detect(pcontent, pattern = regex(" 台積電 |TSMC",ignore_case=TRUE)), 1, 0)
    mutate(IsSamsung = ifelse(str_detect(pcontent, pattern = regex(" 三星 |samsung",ignore_case=TRUE)), 1
    mutate(IsChipGate = ifelse(str_detect(pcontent, pattern = chip_pattern), 1, 0)) %>%

    mutate(week = as.Date(cut(ptime, "week"))) %>%
    group_by(week, ROM) %>%
    mutate(quantity = n()) %>%
    ungroup() %>%
    group_by(week) %>%
    summarise(price = mean(avg_price),
              quantity = quantity,
              ROM = ROM,
              TimeUsed = mean(TimeUsed),
              Is_6s_plus = Is_6s_plus,
              IsTSMC = IsTSMC,
              IsSamsung = IsSamsung,
              IsChipGate = IsChipGate) %>%
    #drop_na() %>%
    filter(!duplicated(week, ROM))
```

```
## `summarise()` regrouping output by 'week' (override with `.groups` argument)
```

IV Regression

我們真實想估計的 demand function（非 inverse demand function）是：

$$q_t^d = \beta_0 + \beta_1 p_t + \beta_3 W_t + \epsilon_t$$

其中 W_t 表示其他外生變數。但由於 p_t 存在內生性問題（$Cov(p_t, \epsilon_t) \neq 0$），因此我們可以透過 $IsTSMC$ 以及 $IsSamsung$ 來此二外生變數來 serve p_t

透過 IV regression 可以得到：

```r
### IV reg
# demand function
m_iv <- ivreg(quantity ~ price + TimeUsed + ROM + Is_6s_plus |
                  IsTSMC + IsSamsung + TimeUsed + ROM + Is_6s_plus,
              data = i6s_IV_df)
summary(m_iv)
```

```
##
## Call:
## ivreg(formula = quantity ~ price + TimeUsed + ROM + Is_6s_plus |
##     IsTSMC + IsSamsung + TimeUsed + ROM + Is_6s_plus, data = i6s_IV_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -745.4 -380.2 -120.3  283.5 2342.3
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4284.6034 18815.6238   0.228    0.820
```

```
## price          -0.1833     0.8156  -0.225    0.822
## TimeUsed       -2.3286    10.1989  -0.228    0.820
## ROM             0.1215     1.1386   0.107    0.915
## Is_6s_plus     36.8174   167.6908   0.220    0.826
##
## Diagnostic tests:
##                   df1 df2 statistic p-value
## Weak instruments    2 251     0.026 0.97427
## Wu-Hausman          1 251    10.771 0.00118 **
## Sargan              1  NA     0.013 0.90755
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 496.2 on 252 degrees of freedom
## Multiple R-Squared: -141.9,  Adjusted R-squared: -144.2
## Wald test: 0.1222 on 4 and 252 DF,  p-value: 0.9744
```

```r
coeftest(m_iv, vcov = vcovHC(m_iv, type = "HC3"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## (Intercept)  4284.60338 18578.79407  0.2306   0.8178
## price          -0.18328     0.80501 -0.2277   0.8201
## TimeUsed       -2.32859    10.07136 -0.2312   0.8173
## ROM             0.12153     1.02909  0.1181   0.9061
## Is_6s_plus     36.81740   159.14455  0.2313   0.8172
```

也就是 demand function 為：

$$q^d = 4284.60 - 0.18328p$$

移項得到 inverse demand function:

$$p = 23377.34 - 5.4561q^d$$

回顧在 (2.) 中估計的 WTP:

$$p = 23691.78 - 8.54q^d$$

可以發現二者結果相近。當然，IV Estimation 所給我們的估計較為保守，可以見到在絕對值的意義上，IV Estimation 的估計值較小。

4. Probit Model : the probability of successfully selling an iPhone 6s

由於我們的貼文有是否成交的標注，因此可以簡單地透過 Probit Model 來看看哪些因素影響一個賣家成功賣出手機的機率。

LPM

```r
reg_LPM <- lm(IsSold ~ avg_price + ROM + Is_6s_plus+TimeUsed+IsFemale+
                     color_ 灰+color_ 金+color_ 玫瑰+color_ 銀,
                 data=iphone6s_df %>% filter(IsSell==1))
summary(reg_LPM)
```

```
##
## Call:
## lm(formula = IsSold ~ avg_price + ROM + Is_6s_plus + TimeUsed +
##     IsFemale + color_ 灰 + color_ 金 + color_ 玫 瑰 + color_ 銀 ,
```

```
##     data = iphone6s_df %>% filter(IsSell == 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.05492 -0.03073 -0.02601 -0.01873  1.00222
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.730e-02  1.302e-02   2.866  0.00417 **
## avg_price   -1.343e-06  4.526e-07  -2.967  0.00302 **
## ROM          5.673e-06  5.368e-05   0.106  0.91583
## Is_6s_plus   1.019e-02  3.241e-03   3.144  0.00167 **
## TimeUsed    -1.067e-05  8.182e-06  -1.304  0.19232
## IsFemale    -3.420e-03  8.805e-03  -0.388  0.69770
## color_灰     1.022e-02  8.179e-03   1.250  0.21138
## color_金     1.233e-02  8.005e-03   1.540  0.12361
## color_玫瑰   1.362e-02  7.667e-03   1.776  0.07579 .
## color_銀     9.134e-03  8.542e-03   1.069  0.28500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1574 on 11657 degrees of freedom
##   (517 observations deleted due to missingness)
## Multiple R-squared:  0.002656,   Adjusted R-squared:  0.001886
## F-statistic:  3.45 on 9 and 11657 DF,  p-value: 0.0002944
```

```r
coeftest(reg_LPM, vcov = vcovHC(reg_LPM, type = "HC3"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## (Intercept)  3.7299e-02  1.3766e-02  2.7095 0.006748 **
## avg_price   -1.3428e-06  4.7938e-07 -2.8011 0.005101 **
## ROM          5.6730e-06  4.9910e-05  0.1137 0.909505
## Is_6s_plus   1.0190e-02  3.5333e-03  2.8839 0.003935 **
## TimeUsed    -1.0668e-05  9.0650e-06 -1.1768 0.239304
## IsFemale    -3.4201e-03  9.0412e-03 -0.3783 0.705230
## color_灰     1.0223e-02  7.0065e-03  1.4590 0.144582
## color_金     1.2327e-02  6.9306e-03  1.7787 0.075320 .
## color_玫瑰   1.3615e-02  6.5062e-03  2.0927 0.036399 *
## color_銀     9.1335e-03  7.3937e-03  1.2353 0.216739
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Probit

```r
reg_probit <- glm(IsSold ~ avg_price + ROM + Is_6s_plus+TimeUsed+IsFemale+
                      color_灰+color_金+color_玫瑰+color_銀,
                  data=iphone6s_df %>% filter(IsSell==1), family=binomial(probit))
summary(reg_probit)
```

```
##
## Call:
## glm(formula = IsSold ~ avg_price + ROM + Is_6s_plus + TimeUsed +
```

```
##     IsFemale + color_灰 + color_金 + color_玫瑰 + color_銀,
##     family = binomial(probit), data = iphone6s_df %>% filter(IsSell ==
##         1))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3939  -0.2485  -0.2261  -0.1912   3.1601
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.757e+00  2.305e-01  -7.623 2.47e-14 ***
## avg_price   -2.471e-05  7.837e-06  -3.152  0.00162 **
## ROM          1.385e-04  9.022e-04   0.154  0.87795
## Is_6s_plus   1.655e-01  5.397e-02   3.066  0.00217 **
## TimeUsed    -2.033e-04  1.367e-04  -1.487  0.13693
## IsFemale    -5.606e-02  1.465e-01  -0.383  0.70196
## color_灰     2.112e-01  1.567e-01   1.348  0.17774
## color_金     2.396e-01  1.535e-01   1.560  0.11865
## color_玫瑰   2.617e-01  1.486e-01   1.761  0.07827 .
## color_銀     1.869e-01  1.630e-01   1.146  0.25165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2766.8  on 11666  degrees of freedom
## Residual deviance: 2735.0  on 11657  degrees of freedom
##   (517 observations deleted due to missingness)
## AIC: 2755
##
## Number of Fisher Scoring iterations: 6
```

```r
coeftest(reg_probit, vcov = vcovHC(reg_probit, type = "HC3"))
```

```
##
## z test of coefficients:
##
##                Estimate  Std. Error z value  Pr(>|z|)
## (Intercept) -1.7569e+00  2.4610e-01 -7.1389 9.407e-13 ***
## avg_price   -2.4706e-05  8.0708e-06 -3.0611  0.002205 **
## ROM          1.3855e-04  8.2493e-04  0.1679  0.866625
## Is_6s_plus   1.6549e-01  5.6492e-02  2.9294  0.003396 **
## TimeUsed    -2.0327e-04  1.4356e-04 -1.4159  0.156794
## IsFemale    -5.6058e-02  1.4617e-01 -0.3835  0.701349
## color_灰     2.1121e-01  1.5505e-01  1.3622  0.173143
## color_金     2.3960e-01  1.5243e-01  1.5719  0.115985
## color_玫瑰   2.6174e-01  1.4728e-01  1.7772  0.075532 .
## color_銀     1.8686e-01  1.6171e-01  1.1555  0.247867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Logit

```r
reg_logit <- glm(IsSold ~ avg_price + ROM + Is_6s_plus+TimeUsed+IsFemale+
                      color_ 灰+color_ 金+color_ 玫瑰+color_ 銀,
                  data=iphone6s_df %>% filter(IsSell==1), family=binomial(logit))
summary(reg_logit)
```

```
##
## Call:
## glm(formula = IsSold ~ avg_price + ROM + Is_6s_plus + TimeUsed +
##     IsFemale + color_灰 + color_金 + color_玫瑰 + color_銀,
##     family = binomial(logit), data = iphone6s_df %>% filter(IsSell ==
##         1))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4128  -0.2478  -0.2250  -0.1912   3.1338
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.168e+00  5.517e-01  -5.742 9.35e-09 ***
## avg_price   -5.928e-05  1.850e-05  -3.204  0.00135 **
## ROM          2.299e-04  2.136e-03   0.108  0.91426
## Is_6s_plus   4.037e-01  1.270e-01   3.178  0.00148 **
## TimeUsed    -4.857e-04  3.185e-04  -1.525  0.12730
## IsFemale    -1.286e-01  3.461e-01  -0.372  0.71026
## color_灰     4.863e-01  3.877e-01   1.254  0.20974
## color_金     5.620e-01  3.796e-01   1.481  0.13871
## color_玫瑰   6.136e-01  3.685e-01   1.665  0.09587 .
## color_銀     4.392e-01  4.026e-01   1.091  0.27521
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2766.8  on 11666  degrees of freedom
## Residual deviance: 2734.8  on 11657  degrees of freedom
##   (517 observations deleted due to missingness)
## AIC: 2754.8
##
## Number of Fisher Scoring iterations: 6
```

```r
coeftest(reg_logit, vcov = vcovHC(reg_logit, type = "HC3"))
```

```
##
## z test of coefficients:
##
##                Estimate  Std. Error z value  Pr(>|z|)
## (Intercept) -3.1681e+00  5.9295e-01 -5.3430 9.144e-08 ***
## avg_price   -5.9276e-05  1.9055e-05 -3.1107  0.001866 **
## ROM          2.2993e-04  1.9502e-03  0.1179  0.906144
## Is_6s_plus   4.0372e-01  1.3258e-01  3.0451  0.002326 **
## TimeUsed    -4.8572e-04  3.3617e-04 -1.4449  0.148493
## IsFemale    -1.2858e-01  3.4595e-01 -0.3717  0.710136
## color_灰     4.8629e-01  3.8594e-01  1.2600  0.207661
## color_金     5.6200e-01  3.7943e-01  1.4812  0.138555
```

```
## color_玫瑰    6.1361e-01  3.6748e-01  1.6698  0.094962 .
## color_銀      4.3925e-01  4.0170e-01  1.0935  0.274191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

可以發現，無論是在線性機率模型中還是 Pobit 或 Logit Model，最重要的因素仍是價格：價格越低，成交的機率就越高。有趣的是，2015 年推出 iPhone 6s 時，「玫瑰金」是新出的顏色而比較搶手，因而有段時間在市場上無論是平行輸入還是二手 iPhone 6s，只要顏色是玫瑰金就可以賣出比較高的價格。而在此機率模型中，也體現了該顏色的搶手程度，若是顏色是玫瑰金，就有比較高的機會賣出，這個效果在其他顏色中是沒有體現的。