##用R來嘗試一下計算期貨正逆價差

使用的資料為TEJ資料庫，期間為2000/01/01到2022/03/16 將每期期貨價格減到當天現貨價格(加權指數)來做價差的統計。 價差計算為: 期貨價格-現貨價格 舉例: 將每天加權指數價格-當天在期貨市場可以交易的各種期貨價格，以及統計一下正逆價差的天數，最後將正負價差轉換為虛擬變數，看看正價差出現的機率。 首先我們要先載入模組與設定工作路徑:

setwd("C:/Users/Neil/Desktop/期貨與選擇權") # 設定工作目錄  
library(tidyr) #整理數據使用  
library(lubridate) #日期模組  
library(data.table) #把df格式轉換為data.table格式，基本上操作方式一樣，但在處理大量資料(萬筆)的時候，速度有明顯差異  
library(ggplot2) #畫圖

載入檔案之後，因為資料量稍多，先將檔案轉換為data.table格式，以及發現年月日為文字格式，因此把她轉換為日期。以及修改columns欄位名稱。

future\_gap = read.csv("期貨正逆價差.csv",stringsAsFactors = FALSE)   
future\_gap = as.data.table(future\_gap)  
future\_gap$年月日 = ymd(future\_gap$年月日) # 將年月日從字串轉換為日期格式  
colnames(future\_gap) = c("證券代碼","年月日","加權指數收盤價","台指期收盤價") #重新命名欄位名稱

接下來我們發現在原始的資料中，數字為文字格式，以及有小豆點。

head(future\_gap)

將逗點取代之後轉換為數字格式。

#取代逗號再把數據從文字轉換成數字  
future\_gap$加權指數收盤價 = gsub(",","",future\_gap$加權指數收盤價)   
future\_gap$台指期收盤價 = gsub(",","",future\_gap$台指期收盤價)  
future\_gap$加權指數收盤價 = as.numeric(future\_gap$加權指數收盤價)  
future\_gap$台指期收盤價 = as.numeric(future\_gap$台指期收盤價)

在整理資料時，先將加權指數以及期貨指數分成兩個data.table，再利用merge方法合併起來，形成根據日期的兩個欄位，加權指數在相同日期會合併與複製。 以下先將數據拆分:

#很白癡的方法，因為抓下來的資料很酷，把數據拆成兩部分再用merge 方法合併，  
market\_closed = future\_gap[!(future\_gap$加權指數收盤價 == ""),] #df[row,col]  
market\_closed = market\_closed[,-4]  
  
future\_index = future\_gap[!(future\_gap$台指期收盤價 == ""),]  
future\_index = future\_index[,-3]

再做合併

New\_future\_gap = merge(market\_closed,future\_index , by = "年月日" ,all = TRUE) #根據年月日合併資料，保留所有數據  
New\_future\_gap$價差 = New\_future\_gap$台指期收盤價 - New\_future\_gap$加權指數收盤價 #新增一個欄位做數據加減

這樣就完成以下表格惹

head(New\_future\_gap)

接下來就是畫正逆價差圖以及跑虛擬變數。

ggplot(New\_future\_gap, aes(x = 年月日, y = 價差))+ #畫圖  
 geom\_line(aes(y = 價差))  
New\_future\_gap$正價差 = ifelse(New\_future\_gap$價差>0 ,1,0) #將價差轉換成01虛擬變數

最後我們可以將正逆價差的統計值設計成一個函式，方便我們只要輸入日期就可以得知這個日期到2022/3/16出現正逆價差的次數與機率。

gap\_prb = function(date){ #設計一個函式，輸入日期之後可以吐出正逆價差次數，以及出現正價差的機率  
date = as.character(date)  
df = New\_future\_gap[(New\_future\_gap$年月日 >= ymd(date)),]  
tmp = table(df$正價差)  
p = round( mean(df$正價差), digits = 3) #出現正價差的機率  
cat("逆價差的次數為: ",tmp[1],"正價差的次數為: ",tmp[2],"正價差的機率為: ",p\*100 , "%")  
}  
  
gap(20050101)

##結論與可能的原因

經過統計我們發現在2000/0101到2022/3/31中，有22195筆資料呈現逆價差走勢，僅4820筆資料為正價差，出現正價差的機率為17.8%，此外05年後多呈現逆價差走勢，出現正價差的機率下降到10.6%，會有這樣的原因可能是因為在期貨價格理論中，持有期貨領不到股息，持有成本提高，因此有折價表現，以及許多大戶與法人都以期貨當作避險標的，進一步壓低期貨價格，導致多呈現為逆價差的狀態。