* 國立政治大學

**日經指數之預測模型**

**學生**

**103300325** [Jan Rusnak](https://www.facebook.com/jan.rusnak.1?fref=nf)

**103351016 張佳誠**

**103351010 李杰恩**

**指導教授**

[**莊皓鈞**](http://www.mis2.nccu.edu.tw/news/news.php?Sn=1445)

**摘　要**

日經指數之預測模型

學生：張佳誠

李杰恩

利用時間序列模型為基礎，對日經指數的趨勢做預測，目的在於了解哪一個經濟因子對未來的日經指數有確切的影響，研究方法為利用R的快速運算環境，對經濟數據做前處理以及建立非傳統統計分析的模型評價，此外，雖然預測下跌就只是預測上漲一樣的方法反向操作一遍，但此研究主要是以預測上漲為主。

**目 錄**

第一章 研究方法 4

1.1研究方法 4

第二章 資料與前處理 5

2.1資料介紹 5

2.2資料前處理 5

第三章 模型建置 9

3.1應變數設定 9

3.2自變數設定 9

3.3最佳化演算法 9

3.4模型 10

第四章 模型評價 12

4.1 預測勝率 12

4.2 模型評價分析 13

第五章 結論 16

**第一章 研究方法**

**1.1 研究方法**

研究過程中最會費時為蒐集資料，蒐集的資料除了原始資料外，可能會需要製作延伸的資料，將資料統整後便可以利用線性迴歸演算法作為預測的基礎建置模型，最後再評價該模型的優劣進而回頭修正變數。

資料處理

模型修正

模型建立

模型評價

研究流程圖

**第二章 資料與前處理**

**2.1資料介紹**

資料的蒐集主要來源以Bloomberg為主，蒐集了從1980年第一季到2014年的第四季經濟數據季資料，其中有日本GDP、CPI、利率、失業率、消費者信心、日經指數、匯率(消費者信心指數遺失5季的資料、匯率為日資料)。

變數列表

|  |  |  |  |
| --- | --- | --- | --- |
| **變數** | **符號** | **變數** | **符號** |
| **國內生產毛額** | **GDP** | **物價連兩降** | **CPI\_fall2** |
| **物價指數** | **CPI** | **物價連兩升** | **CPI\_rise2** |
| **利率** | **Interest** | **消費信心連兩升** | **CC\_rise2** |
| **失業率** | **Unemploy** | **日經未來三季平均變化量** | **Nikkei\_exp3** |
| **消費者信心** | **CC** | **日經** | **Nikkei** |
| **匯率** | **Exchange** | **日經下一期** | **Nikkei\_plus1** |
| **失業連兩降** | **Unemploy\_fall2** | **日經下一期漲** | **Nikkei\_rise\_plus1** |
| **利率連兩降** | **Interest\_fall2** | **時間** | **t** |

**2.2資料前處理**

由於原始的資料有遺失值、格式不同或是需要原始資料的延伸作為分析，故我們必須對資料做前處理。以下將介紹在笨研究中利用R程式做的資料處理。

* 將匯率的日資料轉換成季資料；方法為利用grep()函數以及for迴圈將日期為1979+i年的資料抓出來，再將當年度3月、6月、9月、12月的抓出的資料中取出最大值存進新的向量即為季資料。
* 相關程式碼：

season=as.Date("1979-12-31") #set any date that let season become a time variable

for(i in 1:35){

a=Code[grep(eval(1979+i),Code)] #select first quarter of years

b=a[grep(eval("-03-"),a)]

season=c(season,max(b)) #max(b) will find the last date of March

a=Code[grep(eval(1979+i),Code)] #select second quarter of years

b=a[grep(eval("-06-"),a)]

season=c(season,max(b))

a=Code[grep(eval(1979+i),Code)] #select third quarter of years

b=a[grep(eval("-09-"),a)]

season=c(season,max(b))

a=Code[grep(eval(1979+i),Code)] #select forth quarter of years

b=a[grep(eval("-12-"),a)]

season=c(season,max(b))

}

season=season[2:length(season)]

#eliminate first element "1979-12-31" from our data

* 數據連兩期上升；由於我們除了關心原始的經濟數據外我們還需要製作一些延伸的的數據，譬如消費者信心連續兩次上升就為1否則為0就為一個我們關心的延伸資料。
* 相關程式碼：

#Two consecutive rise, binary start from 1980 Q3

transform\_rise=function(index){

bi=c()

for(i in 1:(length(index)-2)){

if((index[i+2]>index[i+1])&&(index[i+1]>index[i])){

bi[i]=1}

else{

bi[i]=0}

}

return(bi)

}

只要將資料丟入此函數便可產生連兩升的向量，但是CC缺少5季資料有NA值，故CC的做法須先加入以下程式碼才能套用。

CI=CI[!is.na(CI)]

* 數據連兩期下降；不同於連兩升的信號，有些數據連兩降對於日經的預期影響為正（本研究主要探討影響上漲的因素），如利率或失業率下降，因此我們也將轉換該數據。
* 相關程式碼：

#Two consecutive rise, binary start from 1980 Q3

transform\_fall=function(index){

bi=c()

for(i in 1:(length(index)-2)){

if((index[i+2]>index[i+1])&&(index[i+1]>index[i])){

bi[i]=1}

else{

bi[i]=0}

}

return(bi)

}

**第三章 模型建製**

**3.1 應變數建置**

在傳統時間序列分析中，迴歸式的設定如自變數可能為同期也可能為落後項，但是在以預測第t期作為首要目標的情況下，我們永遠無法事先知道與同期的自變數故我們改成的模式，相當於自變數只放入落後項。

應變數為線性迴歸中用放置於等號左邊，為欲估計之目標式，我們總共設定了兩種，其一，其二為，以未來三季的日經漲幅平均作為應變數。

* 相關程式碼：

result=c()

for(i in 1:length(JP\_index)-3){

result[i]=((JP\_index[i+1]-JP\_index[i])+(JP\_index[i+2]-JP\_index[i+1])+(JP\_index[i+3]-JP\_index[i+2]))/3

}

**3.2 自變數建置**

除了日本GDP、CPI、利率、失業率、消費者信心、日經指數、匯率外，由於報章雜誌常常在經濟數據的描述中，形容已連續幾次的上升或下降，故我們以第一章資料前處理的方法特別建置出是否為連兩升或連兩降的虛擬變數。

**3.3 最佳化演算法**

除了相關的經濟理論作為基礎，並使用lm()涵式的演算法外，我還利用step()函數作為輔助，其為R中內建的函數，原理為自動檢定每一變數及n階交互項的顯著性，汰弱留強後以AIC最小的模型作為最後的模型。

* 相關程式碼：

rdc.model=lm(JP\_NKY\_FutureExpect~0)

fwd.model=step(rdc.model,direction = "both",scope=(~(log(GDP)+Exchange+Unemploy+CC+Interest+CPI + Unemploy\_fall2 + CC\_rise2+Interest\_fall2 + CPI\_rise2+CPI\_fall2 )^5))

在此模型中，direction參數為選擇檢定汰弱留強的方向，scope為欲檢定之所有變數，5次方則代表自動檢驗最多5階的交互項，故只要先將應變數隨意與常數跑回歸作為一個初始模型rdc.model，再將其丟入fwd.model後即可得到最佳化AIC的模型。

**3.4 模型**

最後完成之三個模型如下：

* 模型一；為detrend過後的日經趨勢模型：

此模型為我們最原始的模型，我們認為以上的自變項皆會影響日經未來三季平均變化量，而GDP上升的係數應當為正，也就是經濟發展狀況越好，股市應該會上漲；物價指數對於股市的影響我們認為可正可負，因適度的通貨膨脹對經濟發展是有好處的，太高或太低的通膨都會對經濟體系造成負面的影響；而我們認為利率越低股市越蓬勃發展，因融資成本的降低會使人們有誘因借錢投資股市，其斜率應為負值；消費者信心應與股市成同向變動，當人民有消費信心時，對於景氣態度應是樂觀的，此時也應對股市感到樂觀；失業率某種程度上是利用勞動力市場的狀況來衡量經濟發展的情形，當失業率越低時股市應該表現會很好；匯率對於股市我們認為是負向關係，因貨幣貶值是有利於國家產品的出口，對經濟發展有利。

* 模型二：

模型二基本上也是依循模型一的經濟邏輯，不同的地方在於我們將幾種指標改成虛擬變數來操作，也因我們主要預測的是日經漲幅，確定影響方向的指標我們將選擇做多股市的變項，如失業、利率連二降或是消費者信心連二升，由於物價指數的影響不明確，兩二升或降的虛擬變項我們都有放入模型來分析。

* 模型三：

此模型主要是利用step涵式來找最佳AIC之回歸模型，我們將可能影響日經未來三季平均的變項入（模型一、二包含的所有變項），最後找出最有影響力的自變項，並檢視其是否符合我們的經濟直覺以及有經濟上的顯著意義。

**第四章 模型評價**

**4.1 預測勝率**

除了傳統統計分析中最常用到的模型評價指標R-square、Adjusted-R-square、AIC外，作為預測的目的，我們有興趣知道其預測的勝率，但除了預測的勝率外我們還關心其評價次數，若是預測的勝率很高，但是一百多季中只出手預測一次，並不足以讓人取信其勝率。

* 相關程式碼1：

#建製向量rise，下一期日經有上漲這期rise就為1否則為0。

rise=c()

for(i in 1:(length(日經)-1)){

if(日經[i+1]-日經[i]>0){

rise[i]=1

}else{

rise[i]=0

}

}

* 相關程式碼：

k=c()

#取A到B，因為summary(predict(fwd.model))中顯示平均數約為A

#最大值約B。

for(j in A:B){

Comment=(predict(fwd.model))>j

Success=(( Comment +rise)==2)

Fail=(( Comment -rise)==1)

if(sum(Comment,na.rm = T)>30){

k[j]=sum(Success)/sum(Success+Fail)

}

}

max(k,na.rm = T)

Comment =(predict(fwd.model))>which.max(k)

#計算出首次數

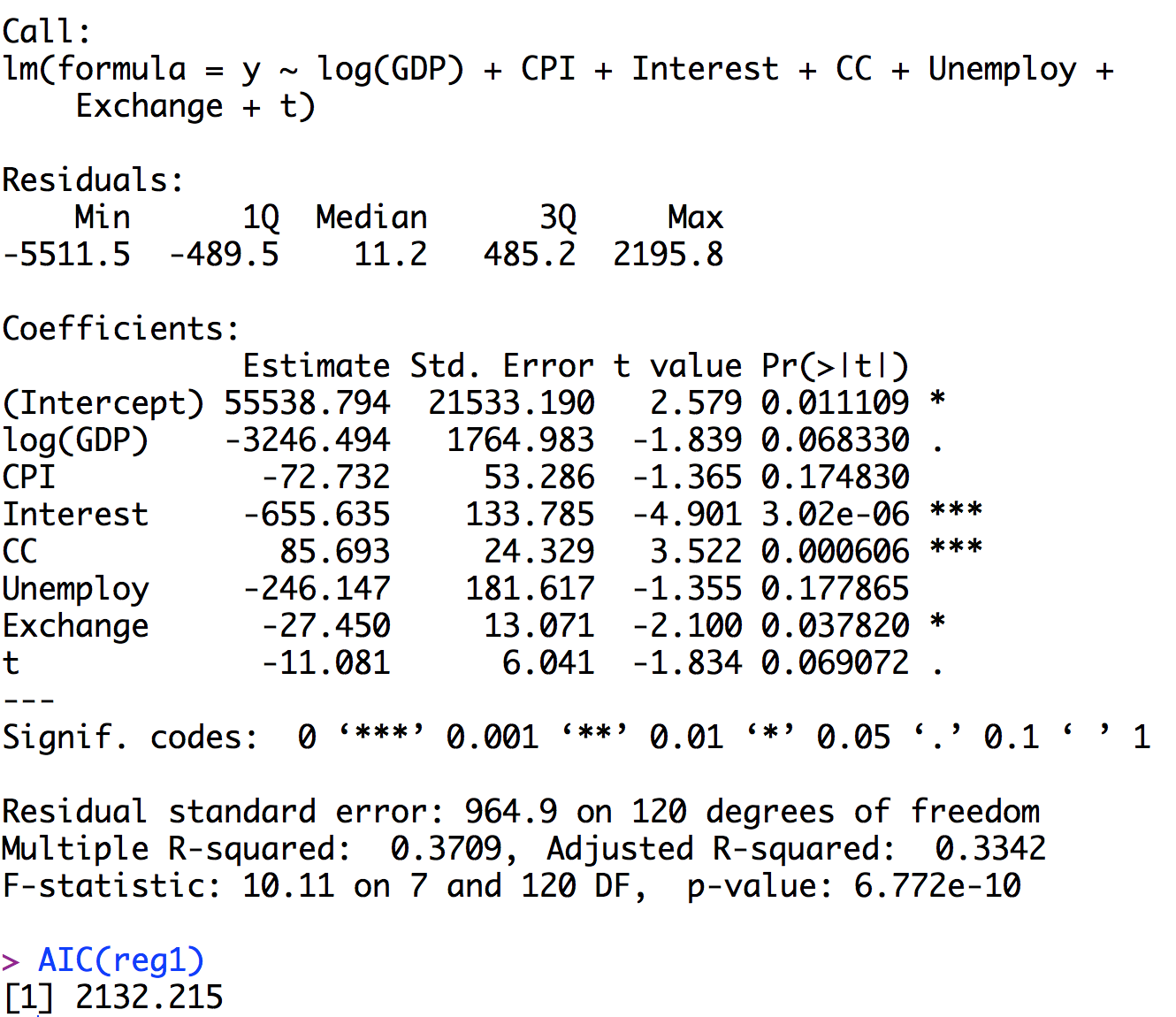
times=sum(Buy,na.rm = T)

k為一向量，若模型的估計值大於一定值(j)，就預測期未來會漲，否則就不加以評論，若評論會漲後確實也上漲了Success就為1，但評論會漲後沒有漲Fail就為1，最後再加以計算勝率並存放在k向量中，存放的過程中可以限定出手評論的次數要大於一定值才存放，在此例子中是30。

**4.2 模型評價**

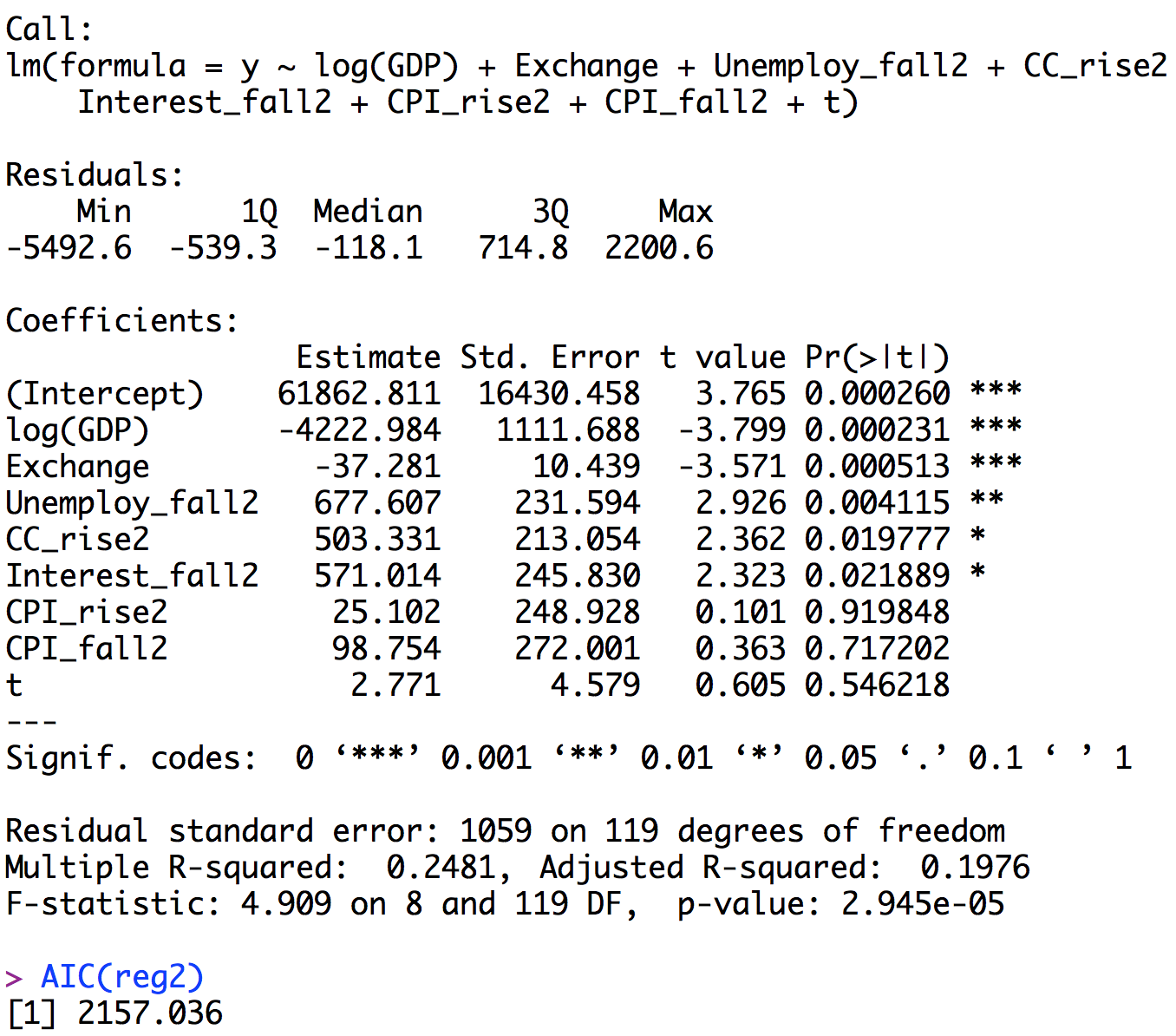
**由於習慣將依變項令做y，以下回歸模型的y變項都代表是日經未來三期平均變化，以下分析只會選擇統計上有顯著的變項來做解釋。**

* 模型一：



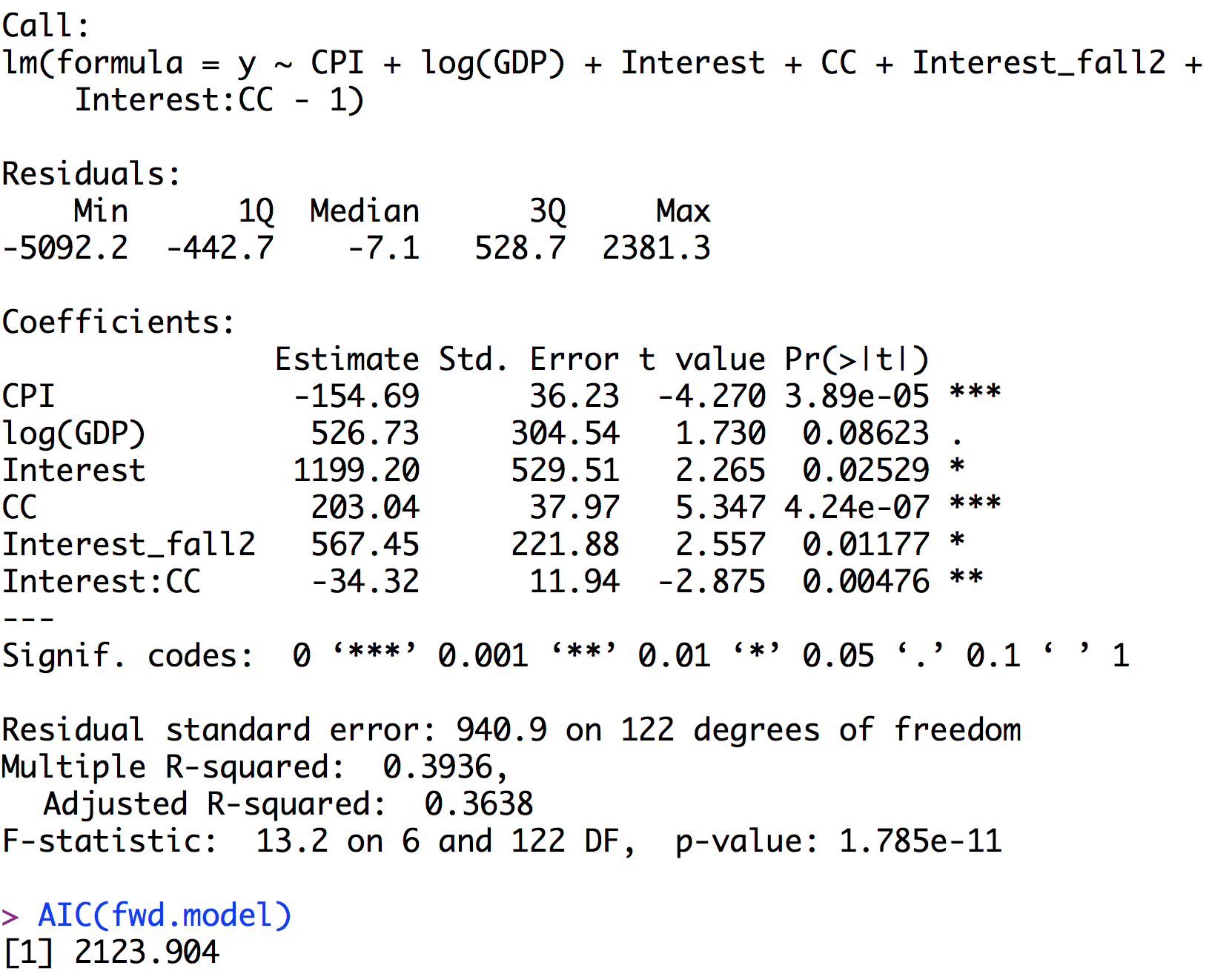
此模型跑出有顯著影響力的變項有利率、消費者信心以及匯率。當利率上升一單位，預期日經走低656點；消費者信心上升一單位，日經預期漲86點；日圓匯率上升一單位（代表匯率貶值），日經預期下跌27點。此模型結果除了匯率的影響方向不符合我們的假設之外，利率與消費者信心的影響皆符合預期。而用我們所建構的勝率模型來估計，根據此預測來操作看多或看空未來日經，其勝率有89.9%（出手36次）。

* 模型二：



放入虛擬變項後的回歸結果與模型一稍有不同，但我們發現在匯率、利率和消費者信心上的影響皆與模型一的預測方向相同，但匯率的方向還是與我們預設的不同。而此GDP的係數顯著為負，代表GDP上升1%，預期日經下跌4223點，此與我們的假設有很大的出入，我們猜測可能是因為GDP對於日經指數的影響在當下比在未來三期大，若其影響在當期都已經反應完了，對於未來三期日經平均變化的預測能力便很低。最後，在失業率的部分的影響係數符合我們的假設，當失業訊號連兩降的信號出現時，預期日經未來三季平均上漲678點。而用我們所建構的勝率模型來估計，根據此預測來操作看多或看空未來日經，其勝率有82.4%（出手37次）。

* 模型三：



**利用step模型所選出的變項為log(GDP)、Interest、CC、Interest\_fall2與Interest:CC，而我們發現此模型雖為AIC最佳化模型，但有些係數仍偏離我們的經濟直覺。像是利率上升，由模型的預測會看多日經，當利率上升一單位，未來三季日經預期漲1199點；利率連兩降對日經也是正向的，這樣的結果不太合理，因利率上升與利率兩兩降是兩種相對的概念。而此模型與其它兩個模型不同的地方還有CPI呈現非常的顯著，此模型預測當物價上升時，股市會下跌。另外，**我們所建構的勝率模型來估計，根據此預測來操作看多或看空未來日經，其勝率有80.6%（出手37次），是三個模型中最低的。

**第五章 結論**

在三個最後完成的計量模型中，都有一共同的現象，可以發現利率與消費者信心不管在哪個模型都扮演著舉足輕重的腳色，這兩個變量的改變可以對日經做跨期的影響，故我等在觀察財經數據時需特別關注，其他如匯率與GDP等就無法在各模型中都保持一定的顯著，但這也有其經濟上的道理，匯率對股市的影響本來就是不一定的，GDP則建議未來可以取年增成長率做測試。雖然三個模型在評價中無非常明顯的優劣之分，但做較少資料前處理的第一個模型反而稍微勝出，不管是在36次評論次數中高達89%的勝率，或是其變數的係數也都符合經濟理論的方向，故在未來日經做預測時可以帶入的一個模型，以預測值的大小來決定對未來日經漲幅的細心。

參考資料:

Bloomberg數據庫