



國立政治大學

日經指數之預測模型

學生

103300325 Jan Rusnak

103351016 張佳誠

103351010 李杰恩

指導教授

莊皓鈞

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。

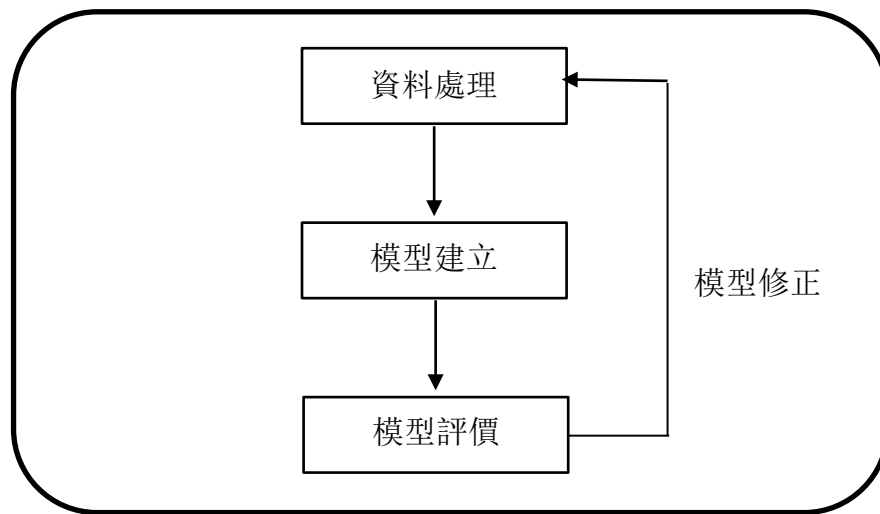
目 錄

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。



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 $y_t = x1_t + x2_{t-1} + u_t$ 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 y_t 同期的自變數 $x1_t$ ，故我們改成 $y_{t+1} = x1_t + x2_t + u_t$ 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，其一為 $t + 1$ 期的日經漲幅 $\Delta nikkei$ ，其二為 $JP_NKY_FutureExpect_t = \frac{1}{3} \sum_{i=1}^3 \Delta nikkei_{t+i}$ ，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()` 函式的演算法外，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：

$$\text{Nikkei}_{\text{exp3}} = \log(\text{GDP}) + \text{CPI} + \text{Interest} + \text{CC} + \text{Unemploy} + \text{Exchange} + t$$

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：

$$\text{Nikkei_exp3} = \log(\text{GDP}) + \text{Interest} + \text{CC} + \text{Interest_fall2} + \text{CPI_fall2} + \text{Interest: CC}$$

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：

$$\begin{aligned} \text{Nikkei_exp3} = & \log(\text{GDP}) + \text{Exchange} + \text{CPI} + \text{Interest} + \text{CC} + \text{Unemploy} \\ & + \text{Unemploy_fall2} + \text{CC_rise2} + \text{Interest_fall2} + \text{CPI_rise2} + \text{CPI_fall2} \end{aligned}$$

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 0，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：

```
Call:
lm(formula = y ~ log(GDP) + CPI + Interest + CC + Unemploy +
    Exchange + t)

Residuals:
    Min       1Q   Median       3Q      Max
-5511.5  -489.5    11.2   485.2  2195.8

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  55538.794   21533.190    2.579  0.011109 *
log(GDP)     -3246.494    1764.983   -1.839  0.068330 .
CPI           -72.732     53.286   -1.365  0.174830
Interest     -655.635    133.785   -4.901  3.02e-06 ***
CC             85.693     24.329    3.522  0.000606 ***
Unemploy     -246.147    181.617   -1.355  0.177865
Exchange     -27.450     13.071   -2.100  0.037820 *
t             -11.081      6.041   -1.834  0.069072 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 964.9 on 120 degrees of freedom
Multiple R-squared:  0.3709, Adjusted R-squared:  0.3342
F-statistic: 10.11 on 7 and 120 DF, p-value: 6.772e-10

> AIC(reg1)
[1] 2132.215
```

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：

```
Call:
lm(formula = y ~ log(GDP) + Exchange + Unemploy_fall2 + CC_rise2 + Interest_fall2 + CPI_rise2 + CPI_fall2 + t)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-5492.6  -539.3  -118.1   714.8  2200.6
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  61862.811   16430.458   3.765 0.000260 ***
log(GDP)     -4222.984   1111.688  -3.799 0.000231 ***
Exchange      -37.281     10.439  -3.571 0.000513 ***
Unemploy_fall2  677.607    231.594   2.926 0.004115 **
CC_rise2       503.331    213.054   2.362 0.019777 *
Interest_fall2  571.014    245.830   2.323 0.021889 *
CPI_rise2       25.102    248.928   0.101 0.919848
CPI_fall2       98.754    272.001   0.363 0.717202
t               2.771      4.579   0.605 0.546218
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1059 on 119 degrees of freedom
Multiple R-squared:  0.2481, Adjusted R-squared:  0.1976
F-statistic: 4.909 on 8 and 119 DF,  p-value: 2.945e-05
```

```
> AIC(reg2)
[1] 2157.036
```

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：

```
Call:
lm(formula = y ~ CPI + log(GDP) + Interest + CC + Interest_fall2 +
    Interest:CC - 1)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-5092.2  -442.7    -7.1   528.7  2381.3
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
CPI           -154.69     36.23   -4.270 3.89e-05 ***
log(GDP)       526.73    304.54    1.730 0.08623 .
Interest      1199.20    529.51    2.265 0.02529 *
CC             203.04     37.97    5.347 4.24e-07 ***
Interest_fall2  567.45    221.88    2.557 0.01177 *
Interest:CC    -34.32     11.94   -2.875 0.00476 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 940.9 on 122 degrees of freedom
Multiple R-squared:  0.3936,
Adjusted R-squared:  0.3638
F-statistic: 13.2 on 6 and 122 DF, p-value: 1.785e-11
```

```
> AIC(fwd.model)
[1] 2123.904
```

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