經濟論文

# 數學模型的經濟計量方法

## 股票指數預測例子

數學模型是經濟學家常用來預測經濟現象的工具，方法是根據過往經驗（數據），用數學方法建立模型，然後用統計學評估模型的適用性，加以適當的修正，最後將模型用作預測未來之用。

作者：夏玉泉



2021 年 3 月 22 日

# 作者簡歷

夏玉泉，中學畢業後在某大華資銀行的行政部及按揭部工作 40 年。擁有香港理工大學企業管理文憑，澳門東亞大學(現為澳門大學)工商管理學士學位，及亞洲(澳門)國際公開大學(現為澳門城市大學)工商管理碩士學位。

工餘以研習經濟學為樂，對博奕論、公司治理及經濟史等研究頗具興趣。2003 年建立“現實經濟學”網站（http://economicessay.angelfire.com/）討論經濟問題，善用經濟理論解釋社會現象。曾在亞洲(澳門)國際公開大學網上學術期刊“亞洲國際工商資訊”發表經濟文章 40 餘篇，當中多篇被轉載或被引用。

# 目　錄

# 第一章 緒　論

任何現象都不會無聲無色或無緣無故地出現，下雨前必然漫天污雲；生病前必會感到不適，這是由於一切事情總有它的規律，假如我們能夠事前知道規律的變化過程，所有事情就可以在還未發生前得知，對人們趨吉避凶總有一點幫助。經濟現象一樣有其發生的規律，因此經濟學家試圖利用統計學及數學，通過經濟理論來找尋經濟現象的規律，目的是要知道經濟現象的因果關係，及預測發生的時間或結果。

數學模型是經濟學家常用來預測經濟現象的工具，方法是根據過往經驗（數據），用數學方法建立模型，然後用統計學評估模型的適用性，加以適當的修正，最後將模型用作預測未來之用。數學模型用在經濟學上的例子很多，譬如商品生產數量的決策，從經濟理論可知影響生產數量的因素有商品價格、消費者所得、代替品或互補品的價格等；又如國民所得的決定受消費、投資、政府支出、及出入口等因素所影響。故在構建數學模型前有必要先獲得經濟現象的因果關係不可。

本文旨在對股票市場走勢作一合理預測，期望透過建立模型預測恒生指數來年的走勢。首先介紹經濟計量模型的意義及建立，接著通過搜尋經濟數據構建模型，然後對模型作有效性評估及修正，

最後重要的是用作預測。當然所有學術研究工具都有其限制，經濟數學模型亦不例外，本文在結束前會略述使用計量方法預測經濟現象所面對的局限。

# 第二章 數學模型的功能及解法

所謂模型是現實的縮影，作為預測經濟現象的數學模型，有需要將對所考察的經濟情況有影響的因素放進模型內。一般數學模型形式如下：

$$\mathrm{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon\,, \quad \varepsilon \sim N(0, \delta^2) \cdots\cdots\cdots (1)$$

上式是典型的線性迴歸模型（Multiple Liner Regression Model），屬於數學模型一種。其含義是模型內"y"的現象受$x_1, x_2, \cdots, x_m$等因素所影響，y稱為因變量（Controlled Variable），是要研究的經濟現象；$x_1$至$x_m$稱為自變量( Independent Variable )或解釋變量，是獨立於y的變量；$\beta_0, \beta_1, \beta_2, \cdots, \beta_m$為待定常數，稱為參數（Parameter）；$\varepsilon$為隨機誤差（Random Error）。

是否需要把所有影響經濟情況的因素全都放進模型裡，它的效用才達致最大呢?這當然不是，首先，由於我們能力所限，不可能將一切相關因素盡錄；其次，某些因素可能會各自相關，會削弱模型的有效性，此點下文會提及。因此只要選擇適當數量及有效的自變量便足夠。

擬定好模型所需要的變量後，最重要是求解模型之參數β，方法是搜集有關解釋變量的歷史數據，稱為觀察值。為免歷史數據過少失卻其代表性，對經濟數據而言會採用 10 年或以上數字。假設

搜集到 n 次觀察值所得如下：

$$x_{i1}, x_{i2}, \cdots, x_{im}, y_i \qquad i = 1, 2, \cdots, n$$

將觀察值用數學形式表達：

$$\begin{cases} y_1 = a + b_1 x_{11} + b_2 x_{12} + \cdots + b_m x_{1m} + \varepsilon_1 \\ y_2 = a + b_1 x_{21} + b_2 x_{22} + \cdots + b_m x_{2m} + \varepsilon_2 \\ \vdots \\ y_n = a + b_1 x_{n1} + b_2 x_{n2} + \cdots + b_m x_{nm} + \varepsilon_n \end{cases}$$

其中$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$互相獨立，且$\varepsilon_i \sim N(0, \delta^2)$， $i = 1, 2, \cdots, n$。上式可作聯立方程組看待，並利用矩陣式表述，令：

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}, \quad \beta = \begin{pmatrix} a \\ b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

上式 X 是一個 m 行 n 列矩陣。上文模型(1)可以用矩陣表示為 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$，其中$\varepsilon \sim N(0, \delta^2 \boldsymbol{I}_n)$，$\boldsymbol{I}_n$為 n 階個單位矩陣( Unit Matrix )。

線性迴歸方程是利用最小二乘法( Least Squares Method )求解，原理是用各期之自變量觀察值數據代入迴歸方程式求得 Y 的預測值，用以與實際產生之數值作比較，得到各期差異，然後令各觀察期的總差異最小時，求取迴歸式的各參數，這裡不打算將求解方法列出，但模型參數的結果可用下列公式獲得：

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad \cdots\cdots\cdots (2)$$

其中，$X^T$為X的倒置矩陣( Transpose Matrix )，指的是矩陣元素不變，

經過行轉列，列轉行後所形成的一個新矩陣，本來 m 行 n 列矩陣，現在變為 n 行 m 列，所以

$$X^T = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{32} & \cdots & x_{n1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1m} & x_{2m} & x_{3m} & \cdots & x_{nm} \end{pmatrix}$$

並且$(X^TX)^{-1}$是矩陣$X^TX$的逆矩陣（Inverse Matrix），$\hat{\beta}$就是參數 β 的最小二乘法估計。(2)式將於下節作詳細應用。

# 第三章 股票指數實例應用

股票市場內股價之漲落，受很多因素影響。從市場基本分析看，股市對地區內整體經濟的表現，有如寒暑表般效應，因此某些經濟數據，可以用來分析及預測股票市場表現的指標。在香港，影響股市價格的數據包括：本地生產總值（Gross Domestic Product，GDP）、通貨膨脹（Inflation）、利率水平（Interest Rate）、政府財政（Fiscal Policy）、及貨幣供應量（Money Supply）等。

本地生產總值是一個用來評估經濟情況好壞的最普通及最佳的工具，根據它的升跌幅度，可以了解經濟環境之興衰。投資者從本地生產總值數據，得知各行業受惠或受影響程度，據此評估各種股票所受影響之情況，採取相應之投資抉擇。因此 GDP 理想股票價值高，反之價值則低。

通貨膨脹是購買力的侵蝕來源，所以對股票所產生之未來股息收入及回報，都具有侵蝕的作用。在投資觀點看，高通脹期不是一個理想的投資時期，結果是股票價值在高通脹期相對較低，在低通脹期則相對較高。

利率在投資上是一個非常重要的元素，若利率處於低水平，企業與個人都會傾向借貸，以從事生產及消費，增加經濟活動能力；而且利率也決定了企業的借貸成本，若利率處於高水平，則會蠶食

其利潤，降低盈利能力；高利率亦提高了持有股票的成本，直接導致要求較高之預期回報，從而影響股票需求及股價。

財政政策是政府直接地以財政預算來平衡經濟過熱或衰退的情況。過熱時，政府會減少公共開支、增加稅收，以遏止企業及私人的消費；衰退時，政府增加開支、減低稅收，以增加社會貨品及勞務之需求，並增加就業需求。兩者均會對企業的盈利能力有影響，前者對企業有負面影響，對股價有壓力；後者對企業有正面影響，對股價有幫助。

貨幣供給量對通貨膨脹有直接影響，當 GDP 之增加幅度遠低於貨幣供給量時，便會出現通貨膨脹，通貨膨脹對股票價值的關係已於上文論述，此處不贅。

表 1 列出影響恒生指數變化之各種因素$x_1, x_2, x_3, x_4, x_5$，並搜集從 2010 年至 2019 年 10 年的經濟數據，作為建立迴歸模型之用。

| 年份 | 恒生指數 (年底) | 本地生產總值 (百萬港元) | 綜合消費物價指數 | 最優惠利率 (%) $x^3$ | 政府收支淨額 (百萬港元) | 貨幣供應量 M1 (百萬港元) $x^5$ |
|---|---|---|---|---|---|---|
| 2010 | 23,035.45 | 1,633,535 | 98.4 | 5 | 25,917 | 901,819 |
| 2011 | 18,434.39 | 1,748,090 | 100.7 | 5 | 75,121 | 1,017,227 |
| 2012 | 22,656.92 | 1,934,430 | 106 | 5 | 73,686 | 1,127,320 |
| 2013 | 23,306.39 | 2,037,059 | 110.3 | 5 | 64,826 | 1,377,359 |
| 2014 | 23,605.04 | 2,138,305 | 115.1 | 5 | 21,803 | 1,510,895 |
| 2015 | 21,914.40 | 2,260,005 | 120.2 | 5 | 72,797 | 1,708,724 |
| 2016 | 22,000.56 | 2,398,280 | 123.8 | 5 | 14,374 | 1,971,146 |
| 2017 | 29,919.15 | 2,490,438 | 103 | 5 | 111,072 | 2,213,970 |
| 2018 | 25,845.70 | 2,659,384 | 104.5 | 5 | 148,974 | 2,431,461 |
| 2019 | 28,189.75 | 2,835,119 | 107 | 5.03 | 67,949 | 2,421,598 |

表 1：過去 10 年恒生指數及香港主要經濟指標的歷史數據
資料來源：恒生指數—hk.investing.com
其他數據—香港政府統計處

需要指出的，是這裡的經濟指標（自變量）對恒生指數（因變量）的影響在時間上有滯後作用，因為通常人們是消化了去年的經濟因素，才會調整對股市的預期，即$\mathrm{Y_t}$受$\mathrm{x_{t-1}}$影響，並非受$\mathrm{x_t}$影響。所以表 1 中在 2010 年的經濟指標，實為 2009 年發生的數字，餘此類推。

設所要建立的模型是：

$$\mathrm{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon\,, \quad \varepsilon \sim N(0, \delta^2)$$

將表 1 數據配合上式用矩陣式表示，得：

$$Y = \begin{pmatrix} 23035.45 \\ 18434.39 \\ 22656.92 \\ 23306.39 \\ 23605.04 \\ 21914.4 \\ 22000.56 \\ 29919.15 \\ 25845.7 \\ 28189.75 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1633535 & 98.4 & 5 & 25917 & 901819 \\ 1 & 1748090 & 100.7 & 5 & 75121 & 1017227 \\ 1 & 1934430 & 106 & 5 & 73686 & 1127320 \\ 1 & 2037059 & 110.3 & 5 & 64826 & 1377359 \\ 1 & 2138305 & 115.1 & 5 & 21803 & 1510895 \\ 1 & 2260005 & 120.2 & 5 & 72797 & 1708724 \\ 1 & 2398280 & 123.8 & 5 & 14374 & 1971146 \\ 1 & 2490438 & 103 & 5 & 111072 & 2213970 \\ 1 & 2659384 & 104.5 & 5 & 148974 & 2431461 \\ 1 & 2835119 & 107 & 5.03 & 67949 & 2421598 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix},$$

由(2) 式：

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= \left[\left(\begin{pmatrix} 1 & 1633535 & 98.4 & 5 & 25917 & 901819 \\ 1 & 1748090 & 100.7 & 5 & 75121 & 1017227 \\ 1 & 1934430 & 106 & 5 & 73686 & 1127320 \\ 1 & 2037059 & 110.3 & 5 & 64826 & 1377359 \\ 1 & 2138305 & 115.1 & 5 & 21803 & 1510895 \\ 1 & 2260005 & 120.2 & 5 & 72797 & 1708724 \\ 1 & 2398280 & 123.8 & 5 & 14374 & 1971146 \\ 1 & 2490438 & 103 & 5 & 111072 & 2213970 \\ 1 & 2659384 & 104.5 & 5 & 148974 & 2431461 \\ 1 & 2835119 & 107 & 5.03 & 67949 & 2421598 \end{pmatrix}^T \begin{pmatrix} 1 & 1633535 & 98.4 & 5 & 25917 & 901819 \\ 1 & 1748090 & 100.7 & 5 & 75121 & 1017227 \\ 1 & 1934430 & 106 & 5 & 73686 & 1127320 \\ 1 & 2037059 & 110.3 & 5 & 64826 & 1377359 \\ 1 & 2138305 & 115.1 & 5 & 21803 & 1510895 \\ 1 & 2260005 & 120.2 & 5 & 72797 & 1708724 \\ 1 & 2398280 & 123.8 & 5 & 14374 & 1971146 \\ 1 & 2490438 & 103 & 5 & 111072 & 2213970 \\ 1 & 2659384 & 104.5 & 5 & 148974 & 2431461 \\ 1 & 2835119 & 107 & 5.03 & 67949 & 2421598 \end{pmatrix}\right)\right]^{-1} \left[\left(\begin{pmatrix} 1 & 1633535 & 98.4 & 5 & 25917 & 901819 \\ 1 & 1748090 & 100.7 & 5 & 75121 & 1017227 \\ 1 & 1934430 & 106 & 5 & 73686 & 1127320 \\ 1 & 2037059 & 110.3 & 5 & 64826 & 1377359 \\ 1 & 2138305 & 115.1 & 5 & 21803 & 1510895 \\ 1 & 2260005 & 120.2 & 5 & 72797 & 1708724 \\ 1 & 2398280 & 123.8 & 5 & 14374 & 1971146 \\ 1 & 2490438 & 103 & 5 & 111072 & 2213970 \\ 1 & 2659384 & 104.5 & 5 & 148974 & 2431461 \\ 1 & 2835119 & 107 & 5.03 & 67949 & 2421598 \end{pmatrix}^T \begin{pmatrix} 23035.45 \\ 18434.39 \\ 22656.92 \\ 23306.39 \\ 23605.04 \\ 21914.4 \\ 22000.56 \\ 29919.15 \\ 25845.7 \\ 28189.75 \end{pmatrix}\right)\right]$$

由於運算過程繁複，此處從略，結果

$$= \begin{pmatrix} 1388480.802 \\ 0.029103173 \\ -444.6246335 \\ -271965.7498 \\ -0.056625479 \\ -0.009659658 \end{pmatrix}$$

所求的模型（迴歸方程式）如下：

$$= 1388481 + 0.0291x_1 - 444.6246x_2 - 271965.7498x_3 - 0.05663x_4 - 0.0097x_5 \cdots\cdots\cdots (3)$$

其中： Y 是恒生指數

$x_1$是本地生產總值

$x_2$是綜合消費物價指數

$x_3$是最優惠利率

$x_4$是政府收支淨額

$x_5$是貨幣供應量 M1

# 第四章 數學模型的評估與修正

模型是否能夠反映真實情況，日後作出有效預測，對模型的實用性有很重要作用。幸虧是在初建之模型中，其實可以找到端倪。因此對模型的評估工作，是建構模型時極重要的工作。評估前需要回答以下問題：

1. 模型是否符合有關經濟理論?
2. 迴歸模型的線性假設$Y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_mx_m+\varepsilon$是否正確?
3. 模型中每個自變量$x_{i,}\ i=1,2,\cdots,m$是否對因變量 Y 都產生了顯著影響?
4. 隨機項$\varepsilon$是否確實服從正態分佈$N(0,\delta^2)$?
5. 模型中的自變量$x_1,x_2,\cdots,x_m$之間是否相互獨立或線性不相關?

上述問題可以利用構建模型之觀察資料，計算一些統計量或編製指標來檢測這些問題，坊間很多計量模型電腦軟件都可以製作出各種統計量/指標。表 2 詳列各統計量之意義及其用法：

| 統計量/指標 | 意義 | 評估方法 | 解決之問題 |
|---|---|---|---|
| 參數的正/負號 | 參數之正與負關係著自變量與因變量的相關方向 | 正數表示兩者呈正相關；負數表示負相關，假如現實呈正相關，但參數是負數，則自變量與合理論相椁 | 問題 1 |
| 判定係數（$R^2$） | 檢驗樣本數據的擬合優度 | 0<$R^2$<1<br>$R^2$ 越大，數據的擬合度越高，迴歸模型的線性假設越準確 | 問題 2 |
| 自變量的顯著性（t 檢驗） | 檢驗每個自變量對因變量影響力的顯著程度 | $\|t\|$值（t 的絕對值）越大，顯著程度越高，參數的適用性越強 | 問題 3 |
| 殘差分析圖 | 模型預測誤差的分佈情況 | 殘差點散亂及密集地分佈在以直線 $e=0$ 為中心的水平區域時，隨機項確實服從正態分佈 | 問題 4 |
| 多重共線性識別（自變量間的相關係數$r_{mn}$） | 分別每兩個自變量檢驗其相關係數 | 0<$r_{mn}$<1<br>係數越大，彼此越高度相關，多重共線性存在，自變量可取性越弱 | 問題 5 |

表 2：各統計量的意義及用法

上文求得的迴歸方程式(3)，使用電腦軟件可得表 3 的附加資料：

| 摘要輸出 | |
|---|---|
| | |
| 迴歸統計 | |
| R 的倍數 | 0.887889 |
| $R^2$ | 0.788346 |
| 調整的 $R^2$ | 0.523779 |
| 標準誤 | 2283.815 |
| 觀察值個數 | 10 |

| ANOVA | | | | | |
|---|---|---|---|---|---|
| | 自由度 | SS | MS | F | 顯著值 |
| 迴歸 | 5 | 77709218.08 | 15541844 | 2.979756 | 0.1562028 |
| 殘差 | 4 | 20863242.36 | 5215811 | | |
| 總和 | 9 | 98572460.44 | | | |

| | 係數 | 標準誤 | t 統計 | P-值 | 下限 95% | 上限 95% | 下限 95.0% | 上限 95.0% |
|---|---|---|---|---|---|---|---|---|
| 截距 | 1388481 | 1204729.681 | 1.152525 | 0.313296 | -1956385 | 4733347 | -1956385 | 4733346.6 |
| $X_1$ | 0.029103 | 0.026416516 | 1.101704 | 0.332424 | -0.0442408 | 0.102447 | -0.0442408 | 0.1024472 |
| $X_2$ | -444.625 | 230.7990908 | -1.92646 | 0.126329 | -1085.4256 | 196.1764 | -1085.4256 | 196.17637 |
| $X_3$ | -271966 | 242428.3788 | -1.12184 | 0.324721 | -945054.84 | 401123.3 | -945054.84 | 401123.34 |
| $X_4$ | -0.05663 | 0.040761106 | -1.3892 | 0.237108 | -0.1697965 | 0.056545 | -0.1697965 | 0.0565455 |
| $X_5$ | -0.00966 | 0.014738248 | -0.65541 | 0.547982 | -0.0505796 | 0.03126 | -0.0505796 | 0.0312603 |

表 3：迴歸方程式電腦計算結果—附加資料

根據表 3 資料，迴歸方程式(3)的效用並不大。首先，所有自變量的 t 值都甚低（從-0.66 至 1.15），不足以確定自變量對因變量影響力的顯著程度，而且更可斷言自變量甚不適用；其次，雖然判定係數$R^2$數值中規中矩（等於 0.788），由於 F 統計量是檢驗$R^2$數據擬合度的顯著性，但因 F 統計量太低，只有 2.98，所以也不能接受迴歸模型(3)的線性假設是否準確；重要是從表 1 得知，最優惠利率從 2010 至 2018 年都維持在 5%水平，亦即自變量$x_4$與因變量 Y 不相關，因此可以肯定將最優惠利率這經濟因素（$x_4$）剔除在模型(3)之外。

據此，我們將自變量$x_4$剔除，然後不斷將餘下 4 個自變量不停組合求得不同的迴歸方程，再比較它們的統計量/指標，得到下到表 4：

| 行次 | 參數 t 值 | | | | | | $R^2$ | F 值 |
|---|---|---|---|---|---|---|---|---|
| | 截距 | 本地生產總值 | 綜合消費物價 | 最優惠利率 | 政府收支淨額 | 貨幣供應量 | | |
| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | | |
| 1 | 1.1525 | 1.1017 | -1.9265 | -1.1218 | -1.3892 | -0.6554 | 0.7883 | 2.9798 |
| 2 | 2.0893 | 0.0641 | -1.7172 | | -0.4075 | 0.5248 | 0.7187 | 3.8315 |
| 3 | 2.7013 | 3.9628 | -1.8123 | | -0.1994 | | 0.7057 | 5.5961 |
| 4 | 2.4581 | 0.2777 | -1.7964 | | | 0.4056 | 0.7109 | 5.7367 |
| 5 | 1.1552 | -0.0181 | | | 0.2167 | 0.4606 | 0.5804 | 3.2273 |
| 6 | 3.4367 | | -1.8530 | | -0.5177 | 4.0902 | 0.7185 | 5.9544 |
| 7 | 3.1010 | 4.2730 | -2.0262 | | | | 0.7041 | 9.5166 |
| 8 | 2.4601 | 3.1428 | | | 0.5349 | | 0.5677 | 5.2520 |
| 9 | 1.5669 | -0.1464 | | | | 0.69310 | 0.5776 | 5.4690 |
| 10 | 1.6471 | | -0.3132 | | 0.3767 | | 0.0456 | 0.1910 |
| 11 | 3.8000 | | -1.8952 | | | 4.3108 | 0.7077 | 9.6836 |
| 12 | 6.9314 | | | | 0.2736 | 3.2277 | 0.5804 | 5.5320 |
| 13 | 2.6869 | 3.3314 | | | | | 0.5522 | 11.0984 |
| 14 | 2.1429 | | -0.5153 | | | | 0.0287 | 0.2655 |
| 15 | 13.262 | | | | 0.56177 | | 0.0339 | 0.3156 |
| 16 | 7.6847 | | | | | 3.4998 | 0.5764 | 12.2483 |

表 4：不同自變量擬合之迴歸方程的統計量/指標一覽表

首先，自變量$x_1, x_2, x_3, x_4, x_5$兩兩的相關係數分別是$r_{1\cdot 2} = 0.2762$、$r_{1\cdot 4} = 0.081$、$r_{1\cdot 5} = 0.986$、$r_{2\cdot 4} = -0.375$、$r_{2\cdot 5} = 0.2402$、及$r_{4\cdot 5} = 0$，除了$x_1$及$x_5$的係數較大外，即有多重共線性問題，其餘都微不足道。

其次，比較表 4 各項，發覺第 7 項之數字似乎較為理想，雖然$R^2$不算太高，但都有 0.70，即有 70%機會自變量能夠解釋因變量；F 值則較優異，有 9.52 算是頗高，足夠接受判定係數$R^2$有顯著程度；截距及自變數$x_1$的 t 值也可通過接受存在顯著程度，唯獨是$x_2$的 t 值較低，在較低的顯著水平下，有被確定對因變量有顯著的可能。

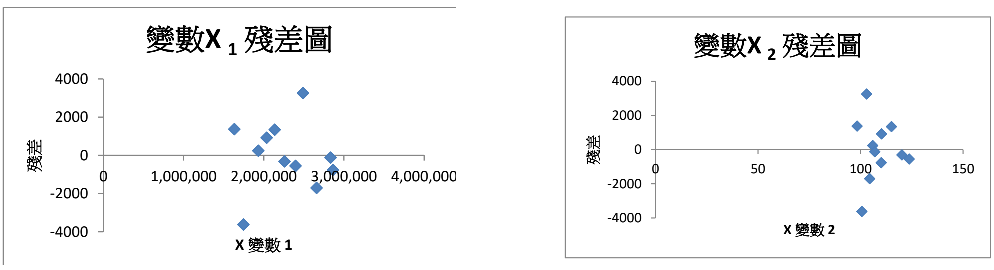


圖 1：兩自變量迴歸方程殘差圖

最後，採用$x_1$及$x_2$為自變量的迴歸方程殘差圖，散佈點是散亂地分佈，見圖 1。因此與表 4 眾多情況作比較以後，第 7 項是最佳選擇。因此最後鎖定的計量迴歸方程式是：

$$Y = 27175.1508 + 0.0067x_1 - 167.7614x_2 \qquad \cdots\cdots\cdots (4)$$

# 第五章 數學模型用作預測工具

建立數學模型後，最後一步工作便是進行預測。前文曾說經濟變量對股票價格的反應在時間上有滯後，因此我們將上年的經濟數據與是年恒生指數作迴歸，即自變量所發生的時間跟因變量發生時間是呈 t-1 與 t 之問關係。現在我們拿 2019 年當年本地生產總值及綜合消費物價指數來預測 2020 年的恒生指數，由此順道檢驗一下迴歸方程(4)的準確度。

2019 年本地生產總值及綜合消費物價指數分別是 2,865,659 百萬元及 110.1（資料來源與表 1 同），將$x_1$ =2,865,659 及$x_2$ =110.1 代入方程式(4)，得：

$$\hat{\mathrm{Y}} = 27175.1508 + 0.0067 \times 2{,}865{,}659 - 167.7614 \times 110.1$$
$$= 28004.59$$

2020 年年底恒生指數實際是 27231.13，較預測數字相差不到 3%。

以上的預測方法稱為點估計（Point Estimation），現在再用區間估計（interval estimation）預測一次。2020 年恒生指數在 95%的信任區間內，其預測值

$$\mathrm{Y}_{2020} = \hat{\mathrm{Y}} \pm t_{7,0.025} S_F \qquad \cdots\cdots\cdots(5)$$

由於 $S_F = 2865.276$及

$t_{7,0.025} = 2.365$

由式(5)，得 $\mathrm{Y}_{2020} = 28004.59 \pm 2.365 \times 2865.276$

$\because 21228.21 < \mathrm{Y}_{2020} < 34780.97$

根據計算結果得到結論：2020 年底恒生指數收市點數有 95%的可能在 21228.21 至 34780.97 之間。果然實際指數確實落在這區間內。

有興趣的讀者可找 2020 年本地生產總值及綜合消費物價指數的資料，用方程式(4)來預測 2021 年底恒生指數的收市點。

# 第六章 結論

隨著電腦普及與運作速度增加，計量經濟模型亦有很多應用程式軟件供使用，如 EViews、LIMDEP、SHAZAM，甚至 Microsoft Excel 內也有迴歸方程的運算程式，這有助於數學模型廣泛應用。然而數學模型有其本身限制，使用前務必充分了解。這些限制包括：

1. 模型基於歷史數據而建立，它反映的是過去發生的事情，包括因果關係。今天發生的事由某些因素所造成，但同樣事情未來發生的原因，或許未必與今天所有這些因素有關，所以不可能完全參照過去情況來解釋未來。一旦將來發生變化，模型的效用便會有商榷。

2. 事情發生的因素眾多，要將所有因素都放進模型內，現代人使用電腦程式，這目標總可以達到的，我們無法解決的問題，是怎麼知道所有影響事情發生的因素?只要當中缺少一項因素，結果都產生偏差，模型的質素便遜色起來。

3. 假如事情的因果，彼此間關係微妙，例如某因變量受多個自變量影響，但在這多個自變量中，某個自變量亦受其它變量及該因變量之影響，其關係如用數學式表示，即：$y = \mathrm{f}\,(a,\ b,\ c)$，及 $b = \mathrm{f}\,(d,\ e,\ y)$。這時需要使用聯立方程式，才能將所有變量囊括於模型方程內。現代經濟社會層層相扣，一道數學模型用上 10 多條

方程已是司空見慣，不用說擁有百條以上方程的模型，可想而知在處理時會頗感繁複。

4. 最後，股票市場變幻莫測，價格時脹時跌，難以論斷。有人說：股票價格受消息影響（如上文指的基本分析），假如作了基本分析掌握了市場消息，股價便可測定。也有人說：市場訊息流轉速度甚快，瞬間已轉化在市價內，根本沒有時候對市場作預測。持此觀點者首推 2013 年諾貝爾經濟學獎得主尤金·法瑪（E. Fama），他的有效市場理論（Efficient Markets Theory）指市場無時無刻都已經把市場訊息有效地反映在市價之中。這樣似乎硬要去預測股價，甚至用無比高效的計量模型，總是徒勞無功。

# 附錄　股票價格變化的隨機過程

上次嘗試用迴歸分析預測恒生指數走勢，除了迴歸分析可以預測股價外，這裡介紹另一種方法，名叫「隨機過程」（Stochastic Processes），由俄國數學家馬科夫（A. Markov，1856-1922）發展而成，故又稱為「馬科夫過程」（Markov Process）。

「馬科夫過程」用在股價預測上，原理是將股價的變化分成多種狀態，假設 S1、S2、及 S3 三種，而且每種狀態轉變成他種狀態的機率可從過往股價走勢獲得，如表 A-1 所示。

| | | 到狀態 | | |
|---|---|---|---|---|
| 由狀態 | | S1 | S2 | S3 |
| | S1 | P11 | P12 | P13 |
| | S2 | P21 | P22 | P23 |
| | S3 | P31 | P32 | P33 |

表 A-1：每種狀態轉變狀況表

表 A-1 可以用「矩陣」（Matrix）表示，並稱為轉移矩陣（Transition Matrix），其型式如下。

$$M = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}$$

「馬科夫過程」之目的是在知道轉移矩陣，及初始狀態矩陣下，如何計算（預測）一定時間後狀態矩陣的結果。初始狀態矩陣是指開始時社會狀態之矩陣分佈。我們把 Q 及 M 分別表示社會狀態矩

陣及狀態轉移矩陣，矩陣的上標代表在不同時間下的矩陣分佈，例如 Q^n 是第 n 期的社會狀態矩陣、Q^(n-1)是第 n-1 期時的社會狀態矩陣，Q^0 就是初始狀態矩陣。

$$Q^n = Q^0 M^n \qquad \cdots\cdots\cdots(1)$$

上式稱為「馬科夫第一定律」，其推導過程是每期的狀態矩陣是由上一期的狀態矩陣乘以轉移矩陣所求得，即：

$$Q^1 = Q^0 M$$

$$Q^2 = Q^1 M$$

$$\vdots$$

$$Q^{n-1} = Q^{n-2} M$$

$$Q^n = Q^{n-1} M$$

$$\because Q^n = Q^0 \overbrace{MMM\cdots\cdots\cdots}^{n}$$

$$\therefore Q^n = Q^0 M^n$$

現在我們將利用「馬科夫過程」預測滙豐控股的價格。表 A-2 是滙豐控股由 2020 年 10 月 29 日至 2021 年 3 月 22（100 天）在港交所每天的交易收市價：

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 日期 | 2021年3月23日 | 2021年3月22日 | 2021年3月19日 | 2021年3月18日 | 2021年3月17日 | 2021年3月16日 | 2021年3月15日 | 2021年3月12日 | 2021年3月11日 | 2021年3月10日 | 2021年3月09日 | 2021年3月08日 | 2021年3月05日 | 2021年3月04日 | 2021年3月03日 | 2021年3月02日 | 2021年3月01日 | 2021年2月26日 | 2021年2月25日 | 2021年2月24日 |
| 收市價 | 45.35 | 46 | 46.8 | 46.8 | 45.75 | 45.8 | 46.5 | 45.75 | 46.9 | 47.6 | 48.597 | 47.6 | 45.955 | 46.404 | 46.105 | 44.908 | 45.955 | 47.151 | 47.5 | 45.955 |
| 日期 | 2021年2月23日 | 2021年2月22日 | 2021年2月19日 | 2021年2月18日 | 2021年2月17日 | 2021年2月16日 | 2021年2月11日 | 2021年2月10日 | 2021年2月09日 | 2021年2月08日 | 2021年2月05日 | 2021年2月04日 | 2021年2月03日 | 2021年2月02日 | 2021年2月01日 | 2021年1月29日 | 2021年1月28日 | 2021年1月27日 | 2021年1月26日 | 2021年1月25日 |
| 收市價 | 46.553 | 46.354 | 45.407 | 46.254 | 45.905 | 45.606 | 42.516 | 42.466 | 41.569 | 41.22 | 41.569 | 40.871 | 41.22 | 40.821 | 40.622 | 40.821 | 41.419 | 42.616 | 42.067 | 42.666 |
| 日期 | 2021年1月22日 | 2021年1月21日 | 2021年1月20日 | 2021年1月19日 | 2021年1月18日 | 2021年1月15日 | 2021年1月14日 | 2021年1月13日 | 2021年1月12日 | 2021年1月11日 | 2021年1月08日 | 2021年1月07日 | 2021年1月06日 | 2021年1月05日 | 2021年1月04日 | 2020年12月31日 | 2020年12月30日 | 2020年12月29日 | 2020年12月28日 | 2020年12月24日 |
| 收市價 | 43.064 | 44.211 | 43.114 | 43.662 | 41.918 | 42.815 | 42.865 | 42.616 | 43.114 | 42.865 | 43.264 | 43.264 | 41.37 | 40.024 | 40.323 | 40.522 | 40.472 | 40.124 | 39.874 | 40.074 |
| 日期 | 2020年12月23日 | 2020年12月22日 | 2020年12月21日 | 2020年12月18日 | 2020年12月17日 | 2020年12月16日 | 2020年12月15日 | 2020年12月14日 | 2020年12月11日 | 2020年12月10日 | 2020年12月09日 | 2020年12月08日 | 2020年12月07日 | 2020年12月04日 | 2020年12月03日 | 2020年12月02日 | 2020年12月01日 | 2020年11月30日 | 2020年11月27日 | 2020年11月26日 |
| 收市價 | 39.426 | 39.326 | 39.376 | 41.021 | 41.569 | 41.17 | 40.871 | 41.22 | 41.519 | 41.768 | 41.619 | 41.22 | 42.765 | 43.613 | 43.264 | 42.366 | 40.871 | 41.37 | 41.868 | 41.868 |
| 日期 | 2020年11月25日 | 2020年11月24日 | 2020年11月23日 | 2020年11月20日 | 2020年11月19日 | 2020年11月18日 | 2020年11月17日 | 2020年11月16日 | 2020年11月13日 | 2020年11月12日 | 2020年11月11日 | 2020年11月10日 | 2020年11月09日 | 2020年11月06日 | 2020年11月05日 | 2020年11月04日 | 2020年11月03日 | 2020年11月02日 | 2020年10月30日 | 2020年10月29日 |
| 收市價 | 42.217 | 39.625 | 39.226 | 38.778 | 38.977 | 39.077 | 39.127 | 38.678 | 37.133 | 38.23 | 39.127 | 37.582 | 34.741 | 34.591 | 34.043 | 33.694 | 33.694 | 32.697 | 31.999 | 32.199 |

表 A-2：滙豐控股由 2020 年 10 月 29 日至 2021 年 3 月 22 每天收市價

解法：

1. 根據表 A-2 數據將股票價格劃分為 10 種狀態，如表 A-3

| 狀態 | 股價範圍 |
|---|---|
| S1 | 29.9<股價≤31.9 |
| S2 | 31.9<股價≤33.9 |
| S3 | 33.9<股價≤35.9 |
| S4 | 35.9<股價≤37.9 |
| S5 | 37.9<股價≤39.9 |
| S6 | 39.9<股價≤41.9 |
| S7 | 41.9<股價≤43.9 |
| S8 | 43.9<股價≤45.9 |
| S9 | 45.9<股價≤47.9 |
| S10 | 47.9<股價≤49.9 |

表 A-3：滙豐控股 10 種價格狀態

2. 接著計算轉移矩陣機率。圖 A-1 是滙豐控股 100 天的股價散佈圖，根據圖 A-1 逐個數據點分析其去向，獲得表 A-4 之轉向數字及每種價格狀態的總點數。

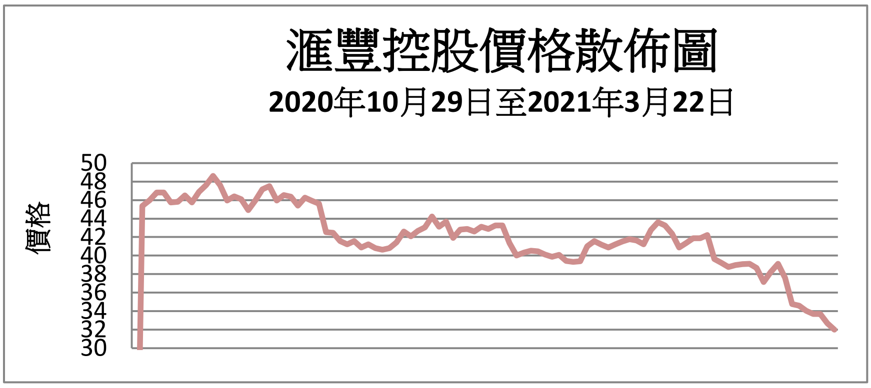


圖 A-1：滙豐控股 100 天價格散佈圖

| | | 到狀態 | | | | | | | | | | 總點數 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ | |
| 由狀態 | $S_1$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | $S_2$ | 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| | $S_3$ | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| | $S_4$ | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 3 |
| | $S_5$ | 0 | 0 | 0 | 1 | 8 | 3 | 1 | 0 | 0 | 0 | 13 |
| | $S_6$ | 0 | 0 | 0 | 0 | 3 | 22 | 4 | 0 | 0 | 0 | 29 |
| | $S_7$ | 0 | 0 | 0 | 0 | 0 | 4 | 14 | 2 | 0 | 0 | 20 |
| | $S_8$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 7 | 0 | 12 |
| | $S_9$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 5 | 1 | 13 |
| | $S_{10}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |

表 4：10 種狀態的轉向數字

3. 將每種狀態的總點數分別除以轉向數字，得到轉移矩陣如下：

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0.6 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0.7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0.6 & 0.2 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0.8 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.7 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.3 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

4. 由於最後 1 天的股價是 45.35，在狀態 $S_8$ 範圍，故初始狀態矩陣

$$Q^0 = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0)$$

由公式(1)得：

下一個交易日股價的狀態矩陣 $Q^1 = Q^0 \times M$

$$= (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0.08 \quad 0.33 \quad 0.58 \quad 0)$$

$$\because 0.58 > 0.33 > 0.08$$

$\therefore$ 下一交易日股價出現在狀態 $S_9$的機會最大，

即股價預測會在 45.9 至 47.9 之水平。

「馬科夫過程」與迴歸分析同屬數學模型的預測方法，而且都是利用歷史數據作為模型建立的根據。不同之處，是迴歸分析注重事情發生的因果關係;「馬科夫過程」則只針對事情發生的機會率，由事件過往不同情況的轉變機率作預測，因此往往予人缺少實在的感覺。

# 參考文獻


- 林華德(1978)。計量經濟學導論。三民書局印行。
- 王生喜主編(2018)。應用統計學。科學出版社。
- 謝識予(2004)。計量經濟學教程。復旦大學出版社。
- 蔡國強(1995)。香港証劵投資ー理論與實務。中華書局。