國 立 交 通 大 學

統計學研究所

碩 士 論 文

LDV 模型參數估計

LDV Model Parameter Estimate of Transaction Costs

研 究 生：黃國瑋

指 導 教 授：洪慧念 教授

中華民國一百零二年六月

LDV Model Parameter Estimate

of Transaction Costs

研 究 生：黃國瑋　　Student： GUO–WEI HUANG

指 導 教 授：洪慧念　　Advisor： HUI – NIEN HUNG

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics

College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistic

June 2012

Hsinchu, Taiwan, Republic of China

LDV Model Parameter Estimate

of Transaction Costs

研 究 生：黃國瑋　　　　　　指 導 教 授：洪慧念

國立交通大學統計學研究所

摘要

在股票交易上交易成本對於報酬率影響是顯著的，在 David A. Lesmond 著作的

A New Estimate of Transaction Costs 這篇文章中作者提出一個模型，這個模型是用來評估交易成本和報酬率之間的關係，但是並沒有說明模型參數的估計方式，我將會在這篇碩士論文中利用一些統計方法來估計這個模型的參數，並利用模擬資料來驗證，我估計方法的準確性，最後我將會使用一些真實料，利用我所推出的估計方法找出最佳參數的估計結果。

誌謝

首先這篇論文能完成，要感謝的是我的指導老師洪慧念教授，在碩士的求學過程中，覺得自己很幸運可以在洪老師的教導下學到很多新知，他告訴了我很多人生道理，並鼓勵我多參與學術活動，使得我這兩年的碩士生活多采多姿，遇到問題不管是論文上的還是程式上的，老師都很樂意提供我想法，也因此能做出這篇論文。最重要的還是要感謝洪老師的栽培與教誨，再來我要感謝碩士生涯中陪伴我的同學他們所給予的幫助與幫忙，給了我想法給了我動力。

最後謝謝口試委員王維菁教授，黃信誠老師，徐南蓉教授在百忙之中聽我口試。

目錄

一、Introduction

在股票市場裡，股票交易的交易成本對於投資報酬率的影響是顯著的，原因是在股票交易的過程中會產生各種費用，以台股交易為例可簡單舉出兩種交易成本：

現股交易交易成本

1.手續費：手續費為買賣金額的千分之 1.425，不管是在進行買進或者是賣出的動作都要收取一次手續費，不同的卷商對手續費打的折扣也不一定一樣，一般卷商會在六折上下，小卷商會有比較優惠的折數最好可以到 3 折左右。

舉例來說：你買了一張 50 元的股票，手續費 $= 50 * 1000 * 0.001425 = 71.25$

也就是 71 元，之後根據卷商打折的折數去計算手續費金額；必須注意的是有些卷商會有最低手續費每筆最低 20 元的限制，假設打完折後是 5 元但因最低手續費限制就必須付 20 元 。

2.證卷交易稅：每一次的"賣出"，都會收一次證卷交易稅，稅率是千分之 3

信用交易交易成本

3.融資(看漲)：付出一筆"融資利息"，跟卷商借部分的錢來買股票，至於可以借多少錢就要看融資成數，上市證卷原則上融資 6 成，上櫃證卷原則上融資 5 成。

舉例來說：要融資買進一檔 50 元的股票(上市)

融資的金額為：50 元＊ 1000 股＊0.6＝30000 元也就是在買進股票的過程中跟券商借了 30000 元。

其中這 30000 元的利息計算方式為：融資利息 = 融資金額＊融資利率＊天數/365

舉例來說：

6 月 1 日融資買進 2882 國泰金一張，成交價位 50 元，6 月 9 日融資賣出 2882 國泰金一張，成交價位 52 元，利息天數共 8 日，融資利率為 5.975%。

融資利息 $= (50{,}000 * 0.6) * 5.975\% * 8/365 = 39$ 元(小數四捨五入)

4. 融卷(看跌)：付出一筆"借卷費" 以及足夠的保證金，跟卷商借卷來賣，卷商會提供給保證金一點點利息。

借券費 = 成交單價 * 成交股數 * 借券費率；借券費率一般為 0.1%

舉例來說： 66 元 x 1000 股 x0.1% = 66 元

上市上櫃證券原則上融券九成，故融券保證金 = 成交單價 * 成交股數 * 融券成數

※若是作空期間標的物上漲，保證金必須給足成數否則將會提前結束交易。

舉例說明：66.2 元 x1000 股 x0.9＝59,580 元

融券保證金 = 59,600 元 (百元為單位)

在 David A. Lesmond (1999)：A New Estimate of Transaction Costs 這篇論文中提出一個 LDV 模型，這個模型是用來評估交易成本和報酬率短期時間內的關係，LDV 模型中假設：報酬率 = 常數 * 市場報酬率 + 誤差

$$\left(R_{jt}^{*} = \beta R_{mt} + \varepsilon_{jt} \text{ ，} \boldsymbol{\varepsilon} \sim N\left(0\text{，}\sigma^{2}\right)\right)$$

是一個正確的報酬率模型。

其中：R_{jt}^{*}是該家公司真實市場報酬率，由於我們無法完完全全得知一家公司內部的情況，不論是財務狀況或者貨品銷售狀況皆是如此，所以一家公司真正的獲利情況我們是無法得知的，所以一家公司真實市場報酬率R_{jt}^{*}是未知的。

R_{mt} 稱為市場報酬率為無風險報酬率加上市場平均風險報酬率(風險溢酬)，其中無風險利率約等於國庫卷的利率。

由於 R_{jt}^{*}是未知的，我們以R_{jt}作為他的估計值用來衡量這家公司的表現，R_{jt}可以稱為報酬率估計值也可以稱為期望報酬率，R_{jt}產生方式如下：

先將R_{jt}^{*}以α_1，α_2分為三個區塊，其中α_1為負訊息門檻，α_2為正訊息門檻 。若時間點 t 的R_{jt}^{*}小於α_1，我們就會認為這個時間點是虧損的，此時時間點 t 的的R_{jt}取為$R_{jt}^{*} - \alpha_1$，R_{jt}^{*}是落在 Region 1，我們將這個時點歸類於 group 1，R_{jt}計算過程中R_{jt}^{*}減去α_1是因為α_1也是交易成本的估計值，所以期望報酬率

(R_{jt})等於真實報酬率(R_{jt}^{*})減掉交易成本。在(1)page5 會舉例子說明

若時間點 t 的R_{jt}^{*}是大於α_2的，則將時間點 t 的R_{jt}取為$R_{jt}^{*}-\alpha_2$，此時R_{jt}^{*}是落在 Region 2 的，且我們將這個時點歸類於 group 2。

最後一個區塊代表時間點 t 的R_{jt}^{*}是介於α_1和α_2之間此時我們無法判斷出它獲利的情形，故將這個時間點 t 的R_{jt}令為 0，0 的意義是指無法判斷獲利表現，並且把這個時點歸類於 group 0。

也就是：

$R_{jt} = R_{jt}^{*} - \alpha_{1j}$ if $R_{jt}^{*} < \alpha_{1j}$ note by group 1

$R_{jt} = 0$ if $\alpha_{1j} < R_{jt}^{*} < \alpha_{2j}$ note by group 0

$R_{jt} = R_{jt}^{*} - \alpha_{2j}$ if $R_{jt}^{*} > \alpha_{2j}$ note by group 2

在此模型中

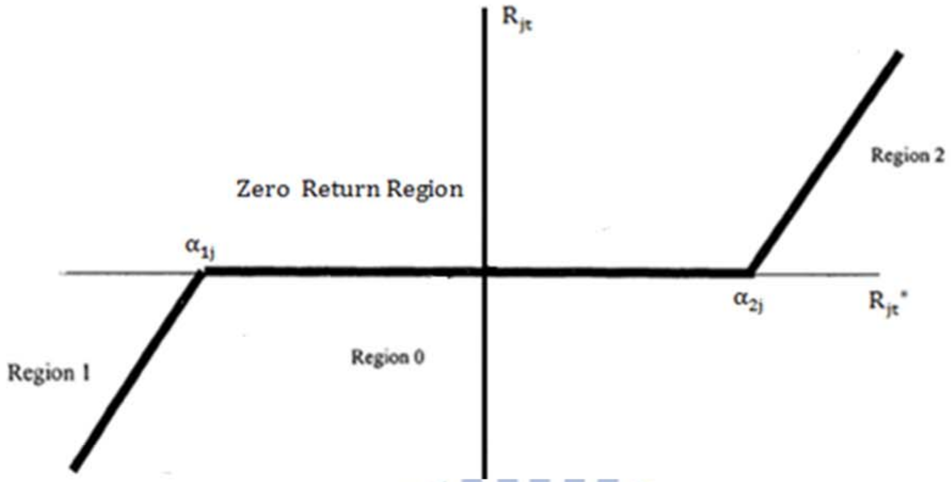
R_{mt}：為 t 時間點市場的報酬率 R_{jt}：某家公司時間點 t 報酬率估計值

R_{jt}^{*}：某家公司 t 時間真實報酬率

α_{1j} ： the threshold for trades on negative information(負訊息門檻)

α_{2j} ： the threshold for trades on positive information(正訊息門檻)

圖示：



x 軸就是R_{jt}^*，y 軸就是R_{jt}

我們把所有落在 Region1 中的時點收集起來計為 group 1

也就是說 group 1 $= \{\, t \mid R_{jt} \in \text{Region1} \}$ 同理得出 group 2，group 0

則在此模型下他的 likelihood function

$$L(\alpha_1, \alpha_2, \beta, \sigma \mid R_{jt}, R_{mt}) = \prod_1 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt} + \alpha_1 - \beta R_{mt}}{\sigma}\right)\right] * \prod_2 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt} + \alpha_2 - \beta R_{mt}}{\sigma}\right)\right]$$
$$* \prod_0 \left[\Phi\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - \Phi\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)\right]$$

其中 $\varphi(.)$是標準常態的 PDF，$\Phi(.)$是標準常態的 CDF，1，2，0 代表的則是他們各自的 group。

我們的目標：尋求每組資料最佳的四個參數$(\alpha_1, \alpha_2, \beta, \sigma)$

我將會利用 Bayesian 和 frequentist 兩種不同的統計推演方法去估計這個 Model 的參數，當中的 Bayesian 我們會先給定 prior 再利用 Gibbs Sampling，frequentist 利用的是 EM，並透過模擬資料來驗證方法的準確性，且我們會將這些方法運用在一些真實的資料並加以分析，詳細流程之後章節會說明。

在 LDV 模型中我們關注的重點在於α_2和α_1之間的差異也就是$\alpha_2 - \alpha_1$這個值的大小，這是因為 David A. Lesmond 在他著作的論文中將這個差值稱之為 round trip transaction costs 代表的就是所謂的短線交易交易成本，短線交易是指短期間內買進

及賣出同一證券，買進時付出成本α_2賣出時得到成本α_1，所以短線交易交易成本就是付出的α_2扣除得到的α_1，所以我們能透過這個差值的大小評估短線交易對報酬率的影響。

α_1和α_2如何影響交易成本 (1)

在 LDV 模型中α_2代表的是當真實報酬率大於α_2時才算是獲益，這個現象是因為股票交易方式造成的

例子：

假設：某檔股票在 1 個月後會有 10%的報酬率(R_t^*)，假設現值 50 元 1 個月後會漲到 55，在訊息出來時掛買單變多，最後能成交的價格可能是 52 成交。也就是說必須要用 52 元買了一個現在最低價值為 50 元股票，2 元就是交易成本買的是一個未來，未來指的是股價之後的趨勢，2 元換算成報酬率的話是 4%(α_2)只有在 A 公司股價報酬率大於 4%時投資才獲利，若 1 個月後股價報酬率真的達到 10%，實際收益在不考慮折現的情形下為 6%(R_t)。

這意味著某檔股票就算未來表現會不錯，當有意願投資這檔股票的人很多時，我們的投資也會付出較高的交易成本，也因此會影響我們的報酬率所以我們在評估他的報酬率時才會以$R_t^* - \alpha_2$作為我們的估計值R_t。

α_1代表的是當真實報酬率小於α_1時才算是虧損，這也是因為股票交易方式造成的只是這時候我們對於未來趨勢是看跌 不是看漲。

舉例說明

假設：某檔股票在 1 個月後會有負 20%的報酬率(R_t^*)，也就是現值 50 元 1 個月後會跌到 40 在訊息出來時掛賣單變多 ，最後能成交的價格可能是 45 成交。也就是說現在是用 45 元價格買了一個現在價值為 50 元股票，當中的差值負 5 元就是交易成本買的是一個未來趨勢，負 5 元換算成報酬率的話是負 10%(α_1)，只有 A 公司報酬率小於負 10%這次的交易才算是虧損。若 1 個月後股價真的跌到 40，實際虧損在不考慮折現的情形下為負 10%。

其中交易成本是負 10%是合理的，這是因為如果我是股票持有人當我認定未來趨勢會跌這時我必然會想拋售我的股票 ，但是因為股票成交方式我必然必須壓低我拋售的金額，且對於買方而言我不但會賣給他股票還會同時把"未來趨勢的風險"一起賣出去，也因此為了讓買方願意承擔"未來趨勢的風險" 壓低的價格會使得買方的交易成本為負值。

如果我是投資人當我認定某支股票未來會漲這時我必然會想買進股票 ，但是因為股票成交方式我必須開出比較高的金額才有可能買到，對於買方而言我不但會買到這張股票還會同時把"未來趨勢的希望"一起買進來，所以交易成本為正值。

這這些例子中我們能看到交易成本大小對於報酬率的影響性以及α_2和α_1的意義在接下來的章節中我們會介紹方法以及驗證這些方法的準確性並且將這些方法運用在真實資料中。

二、介紹方法&方法流程

二、1 Bayesian

我將利用 data augmentation 以及 Gibbs sampling 來幫助我估計參數。

介紹 Gibbs sampling

以兩變數為例子：若 $P(X，Y)$聯合分配未知，但是我們知道他的條件分配 $P(X|Y)$ 和 $P(Y|X)$，這種情況之下我們可以用亂數產生樣本來推論聯合機率分配的長相。

Gibbs sampling 演算法： (2)

Given initial value y_0

for $i = 1 \sim N$

x_i由 $P(X|Y = y_i)$生成

y_i由 $P(Y|X = x_i)$生成

生滿 N 組樣本後推論 $P(X，Y)$

Example：assume real distribution of $P(X，Y)$

$$\text{is BN}\left(\mu = \begin{pmatrix} 2 \\ 3 \end{pmatrix}，\Sigma = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}\right)$$

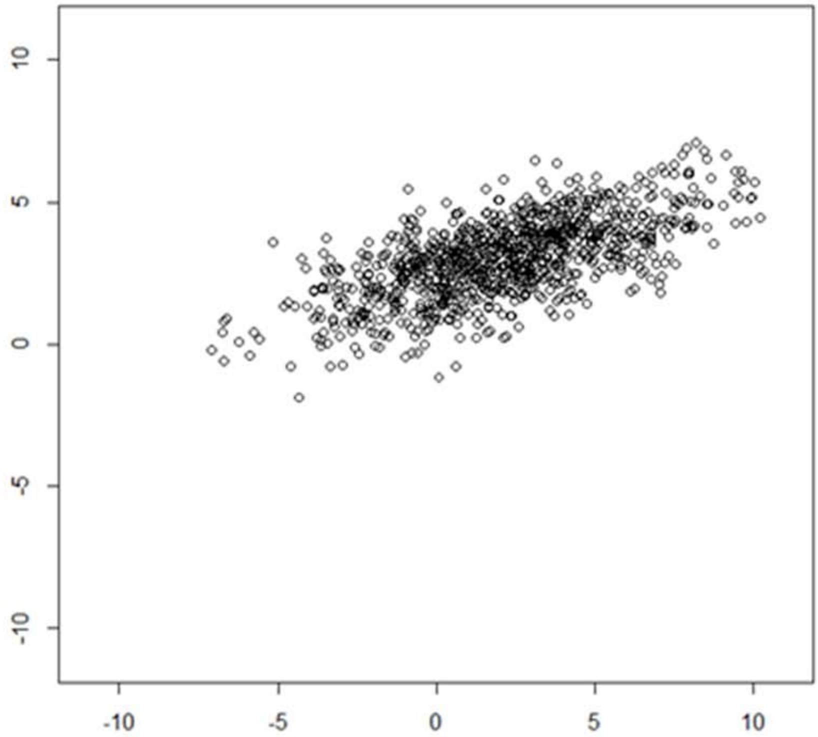
we know that

$$P(X|Y = y) = N(2 + \frac{3}{\sqrt{10 * 2}} * \sqrt{\frac{10}{2}} * (y - 3)，10 * (1 - \frac{3^2}{10 * 2}))$$

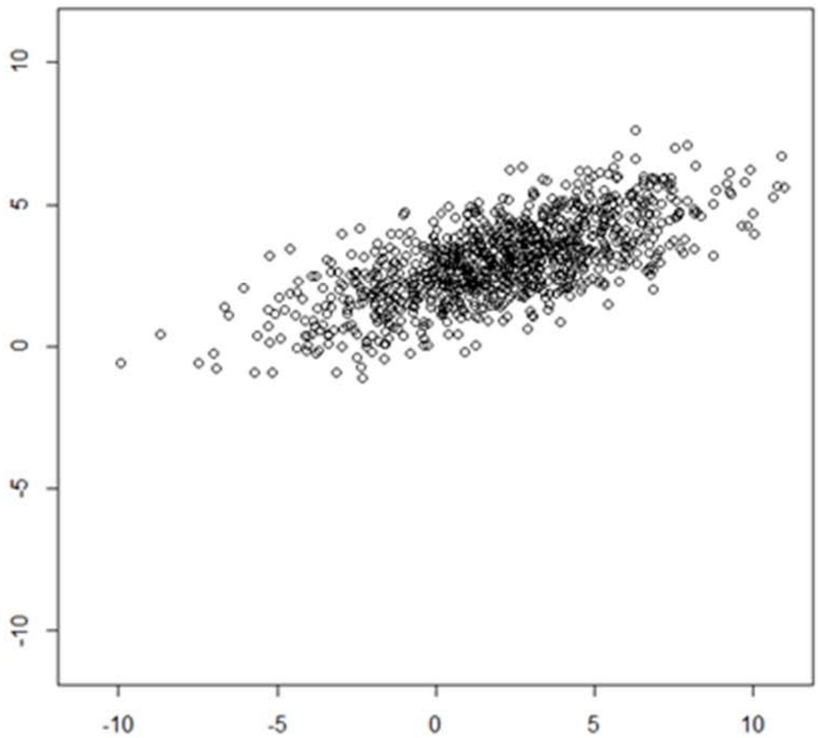
$$P(Y|X = x) = N(3 + \frac{3}{\sqrt{10 * 2}} * \sqrt{\frac{2}{10}} * (x - 2)，2 * (1 - \frac{3^2}{10 * 2}))$$

透過(2)的演算法去生成資料：

圖一是以 Gibbs sampling 推論出來的分配形狀 Sample size 為 1000



圖二 real distribution BN $\left(\mu = \begin{pmatrix} 2 \\ 3 \end{pmatrix}，\Sigma = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}\right)$ Sample size 1000



可以很明顯看出 Gibbs sampling 生成結果和真實分配外型很相似，在 Sample size 1000 下平均和變異數和真實值差距並不大。

```
> mean(x)            > var(x)
[1] 2.072584         [1,] 10.64462
> mean(y)            > var(y)
[1] 2.990774         [1,] 2.269077
```

也就是說對於未知的分配 $P(X，Y)$如果我們知道 X 和 Y 個別的條件分配我們就能對 $P(X，Y)$做推論。

因為 Gibbs sampling 的特性，我們嘗試著解決這個模型參數估計的問題

觀察 Model：

$$R_{jt}^{*} = \beta R_{mt} + \varepsilon_{jt} \qquad \boldsymbol{\varepsilon} \sim N(0，\sigma^2)$$

$$R_{jt} = R_{jt}^{*} - \alpha_{1j} \qquad \text{if } R_{jt}^{*} < \alpha_{1j} \qquad \text{note by group 1}$$

$$R_{jt} = 0 \qquad \text{if } \alpha_{1j} < R_{jt}^{*} < \alpha_{2j} \qquad \text{note by group 0}$$

$$R_{jt} = R_{jt}^{*} - \alpha_{2j} \qquad \text{if } R_{jt}^{*} > \alpha_{2j} \qquad \text{note by group 2}$$

已知：likelihood function $L(\alpha_1, \alpha_2, \beta, \sigma | R_{jt}, R_{mt})$ (3)

$$= \underset{\text{group 1}}{\Pi}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right] * \underset{\text{group 2}}{\Pi}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right] * \underset{\text{group 0}}{\Pi}\left[\Phi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)-\Phi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)\right]$$

其中 $\varphi(.)$是標準常態的 PDF，$\Phi(.)$是標準常態的 CDF

prior distribution ：$p(\alpha_1, \alpha_2, \beta, \sigma | R_{jt}, R_{mt}) = 1/\sigma$

二、1－1　Gibbs Sampling Estimator

為簡化問題假設：R_{jt}^{*} 已知

式子(3)可簡化為：

$$p(\alpha_1,\alpha_2,\beta,\sigma)=\pi\left(\alpha_1,\alpha_2,\beta,\sigma\middle|R_{jt},R_{mt},R_{jt}^{*}\right)$$

$$=\prod_{group\ 1}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_{group\ 2}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_{group\ 0}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^{*}-\beta R_{mt}}{\sigma}\right)\right]$$

其中 $\varphi(.)$是標準常態的 PDF 。

推出個別條件分配：

$$f\left(\alpha_1\middle|\alpha_2,\beta,\sigma,R_{jt},R_{mt},R_{jt}^{*}\right)=\frac{\pi\left(\alpha_1,\alpha_2,\beta,\sigma\middle|R_{jt},R_{mt},R_{jt}^{*}\right)}{\int_{-1}^{\min(R_{jt}^{*})}\pi\left(\alpha_1,\alpha_2,\beta,\sigma\middle|R_{jt},R_{mt},R_{jt}^{*}\right)d\alpha_1}$$

$$=\frac{\prod_{1}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]}{\int_{-1}^{\min(R_{jt}^{*})}\prod_{1}\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]d\alpha_1}$$

$$=\frac{\frac{1}{\sqrt{2\pi}\left(\frac{\sigma}{\sqrt{\#1}}\right)}*\exp\left[\frac{-1}{2}\left(\frac{\alpha_1+\frac{\sum_1\left(R_{jt}-\beta R_{mt}\right)}{\#1}}{\frac{\sigma}{\sqrt{\#1}}}\right)^2\right]}{\left[\Phi\left(\frac{\min\left(R_{jt}^{*}\right)+\frac{\sum_1\left(R_{jt}-\beta R_{mt}\right)}{\#1}}{\frac{\sigma}{\sqrt{\#1}}}\right)-\Phi\left(\frac{-1+\frac{\sum_1\left(R_{jt}-\beta R_{mt}\right)}{\#1}}{\frac{\sigma}{\sqrt{\#1}}}\right)\right]}$$

k is number of group k

二、1－2 Alpha1 & Alpha2 Distribution

得出：

$$(\alpha_1|\alpha_2,\beta,\sigma,R_{jt},R_{mt},R_{jt}^*)\sim N\left(-\frac{\sum_1(R_{jt}-\beta R_{mt})}{\#1}\ ,\ \frac{\sigma^2}{\#1}\right)\Bigg|_{-1}^{\min(R_{jt}^*)}$$

同理：

$$(\alpha_2|\alpha_1,\beta,\sigma,R_{jt},R_{mt},R_{jt}^*)\sim N\left(-\frac{\sum_2(R_{jt}-\beta R_{mt})}{\#2}\ ,\ \frac{\sigma^2}{\#2}\right)\Bigg|_{\max(R_{jt}^*)}^{1}$$

k is number of group k

其中 $\min(R_{jt}^*)$ 是 minmum R_{jt}^* of group 0

其中 $\max(R_{jt}^*)$ 是 maxmum R_{jt}^* of group 0

$$f(\beta|\alpha_1,\alpha_2,\sigma,R_{jt},R_{mt},R_{jt}^*)=\frac{\pi(\alpha_1,\alpha_2,\beta,\sigma|R_{jt},R_{mt},R_{jt}^*)}{\int_{-\infty}^{\infty}\pi(\alpha_1,\alpha_2,\beta,\sigma|R_{jt},R_{mt},R_{jt}^*)\,d\beta}$$

$$=\frac{\prod_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_0\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^*-\beta R_{mt}}{\sigma}\right)\right]}{\int_{-\infty}^{\infty}\prod_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_0\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^*-\beta R_{mt}}{\sigma}\right)\right]d\beta}$$

$$=\frac{\prod_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_0\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^*-\beta R_{mt}}{\sigma}\right)\right]}{\left(\frac{1}{2\pi\sigma}\right)^{\frac{n}{2}}\exp\left[\left(\frac{\sum_1(R_{jt}+\alpha_1)*R_{mt}+\sum_2(R_{jt}+\alpha_2)*R_{mt}+\sum_1(R_{mt}*R_{jt}^*)}{\sqrt{\sum_{0\,1\,2}(R_{mt})^2}}\right)^2\right]*\frac{\sqrt{\pi}}{\sqrt{W}}}$$

$$=\frac{\exp\left(\frac{-1}{2\sigma^2}\left[\sum_1(R_{jt}+\alpha_1-\beta R_{mt})^2+\sum_2(R_{jt}+\alpha_2-\beta R_{mt})^2+\sum_0(R_{jt}^*-\beta R_{mt})^2\right]\right)}{\left(\frac{1}{2\pi\sigma}\right)^{\frac{n}{2}}\exp\left[\left(\frac{\sum_1(R_{jt}+\alpha_1)*R_{mt}+\sum_2(R_{jt}+\alpha_2)*R_{mt}+\sum_1(R_{mt}*R_{jt}^*)}{\sigma^2\sqrt{\sum_{0\,1\,2}(R_{mt})^2}}\right)^2\right]*\frac{\sqrt{\pi}}{\sqrt{W}}}$$

當中：$W=\frac{1}{2\sigma^2}\sum_{0\,1\,2}(R_{mt})^2$

$$=\frac{\exp\left[\frac{-1}{2}\left(\frac{\beta-\frac{\sum_1\left((R_{jt}+\alpha_1)*R_{mt}\right)+\sum_2\left((R_{jt}+\alpha_2)*R_{mt}\right)+\sum_0(R_{mt}*R_{jt}^*)}{\sum_{1\,2\,0}\{R_{mt}^2\}}}{\frac{\sigma}{\sqrt{\sum_{1\,2\,0}\{R_{mt}^2\}}}}\right)^2\right]}{\sqrt{2\pi}\frac{\sigma}{\sqrt{\sum_{1\,2\,0}\{R_{mt}^2\}}}}$$

二、1－3 Beta Distribution

得出：

$\left(\beta \middle| \alpha_1, \alpha_2, \sigma, R_{jt}, R_{mt}, R_{jt}^{*}\right) \sim N(M，\Sigma^2)$

其中

$$M = \frac{\sum_1\left((R_{jt}+\alpha_1)*R_{mt}\right)+\sum_2\left((R_{jt}+\alpha_2)*R_{mt}\right)+\sum_0\left(R_{mt}*R_{jt}^{*}\right)}{\sum_{1\,2\,0}\left\{R_{mt}^{\ 2}\right\}}$$

$$\Sigma = \frac{\sigma}{\sqrt{\sum_{1\,2\,0}\left\{R_{mt}^{\ 2}\right\}}}$$

$$f\left(\sigma \middle| \alpha_1, \alpha_2, \beta, R_{jt}, R_{mt}, R_{jt}^{*}\right) = \frac{\pi\left(\alpha_1,\alpha_2,\beta,\sigma \middle| R_{jt}, R_{mt}, R_{jt}^{*}\right)}{\int_0^\infty \pi\left(\alpha_1,\alpha_2,\beta,\sigma \middle| R_{jt}, R_{mt}, R_{jt}^{*}\right) d\sigma}$$

$$= \frac{\prod_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_0\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^{*}-\beta R_{mt}}{\sigma}\right)\right]}{\int_0^\infty \prod_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_0\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^{*}-\beta R_{mt}}{\sigma}\right)\right] d\sigma}$$

$$= \frac{\prod_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\prod_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right]*\prod_0\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}^{*}-\beta R_{mt}}{\sigma}\right)\right]}{\frac{1}{2}*\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}*\left(\frac{1}{W}\right)^{\frac{n-1}{2}}*\Gamma\left(\frac{n-1}{2}\right)}$$

$$= \frac{2*\left(\frac{1}{\sigma}\right)^n \exp\left(W\frac{-1}{\sigma^2}\right)}{\left(\frac{1}{W}\right)^{\frac{n-1}{2}}*\Gamma\left(\frac{n-1}{2}\right)}$$

當中 $W = \frac{1}{2}*\left[\sum_1\left(R_{jt}+\alpha_1-\beta R_{mt}\right)^2+\sum_2\left(R_{jt}+\alpha_2-\beta R_{mt}\right)^2+\sum_0\left(R_{jt}^{*}-\beta R_{mt}\right)^2\right]$

則 $Y = \left(\frac{1}{\sigma^2} \middle| \alpha_1, \alpha_2, \beta, R_{jt}, R_{mt}, R_{jt}^{*}\right)$ 的分配為

$$f_Y(y) = f_\sigma\left(y^{\frac{-1}{2}}\right)*\frac{1}{2}*y^{\frac{-3}{2}}$$

$$= \frac{2*y^{\frac{n}{2}}*\exp(-y*W)}{\left(\frac{1}{W}\right)^{\frac{n-1}{2}}*\Gamma\left(\frac{n-1}{2}\right)}*\frac{1}{2}*y^{\frac{-3}{2}}$$

$$=\frac{y^{\frac{n-1}{2}-1}*\exp(-y*W)}{\left(\frac{1}{W}\right)^{\frac{n-1}{2}}*\Gamma\left(\frac{n-1}{2}\right)}$$

二、1－4 Sigma Distribution

得出：

$$\left(\frac{1}{\sigma^2}\middle|\alpha_1,\alpha_2,\beta,R_{jt},R_{mt},R_{jt}{}^*\right)\sim\text{Gamma}\left(\alpha=\frac{n-1}{2}\ ，\beta=\frac{1}{W}\right)$$

其中

n ：sample size

$$W=\frac{1}{2}\left[\sum_1\left(R_{jt}+\alpha_1-\beta R_{mt}\right)^2+\sum_2\left(R_{jt}+\alpha_2-\beta R_{mt}\right)^2+\sum_0\left(R_{jt}{}^*-\beta R_{mt}\right)^2\right]$$

由於當初我們為了簡化問題我們假設了 $R_{jt}{}^*$是已知的，所以我們必須設法去模擬 $R_{jt}{}^*$的資料且這筆模擬的$R_{jt}{}^*$ 必須合乎於當前參數估計的結果，也因此我們必須推導 $R_{jt}{}^*$的分配。

觀察 Model

$R_{jt}{}^*=R_{jt}+\alpha_1$ note by group 1

$\alpha_1<\ R_{jt}{}^*<\alpha_2$ note by group 0

$R_{jt}{}^*=R_{jt}+\alpha_2$ note by group 2

不難發現 group 1 的 $R_{jt}{}^*$ 是可以直接以 $R_{jt}+\alpha_1$ 來做為估計值同理 group 2 的 $R_{jt}{}^*$也可以直接以 $R_{jt}+\alpha_2$作為估計值 唯有 group 0 的 $R_{jt}{}^*$ 是需要生成的，也因此我們針對 group 0 推導$R_{jt}{}^*$的分配。

$$f\left(R_{jt_0}{}^*(\text{some } t_0\ ,t_0\in\text{group }0)\middle|\alpha_1,\alpha_2,\beta,\sigma,R_{jt},R_{mt},(\text{given all t})\right)$$

$$=\frac{\frac{1}{\sigma}\varphi\left(\frac{R_{jt_0}{}^*-\beta R_{mt_0}}{\sigma}\right)}{\int_{\alpha_1}^{\alpha_2}\frac{1}{\sigma}\varphi\left(\frac{R_{jt_0}{}^*-\beta R_{mt_0}}{\sigma}\right)dR_{jt_0}{}^*}$$

$$=\frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-1}{2}\left(\frac{R_{jt}^{*}-\beta R_{mt_0}}{\sigma}\right)^2\right]}{\Phi\left(\frac{\alpha_2-\beta R_{mt_0}}{\sigma}\right)-\Phi\left(\frac{\alpha_1-\beta R_{mt_0}}{\sigma}\right)}$$

二、1－5 R_{jt}^{*} Distribution

得出：

$$\left(R_{jt_0}^{*}(\text{some } t_0\ , t_0 \in \text{group } 0)\middle|\alpha_1,\alpha_2,\beta,\sigma,R_{jt},R_{mt},(\text{given all } t)\right)\sim N\left(\beta R_{mt_0}，\sigma^2\right)\Big|_{\alpha_1}^{\alpha_2}$$

二、1－6 Gibbs sampling 方法過程：

已知：R_{mt}為 t 時間時市場的報酬率，R_{jt}為某家(第 j 家)公司的報酬率

1\.

determining the number of sample size n

given initial parameter $\alpha_1,\alpha_2,\beta,\sigma$

PS: $-1<\alpha_1<\alpha_2<1$ And $\sigma>0$

2\. 觀察R_{jt}的正負將資料分組

$R_{jt}>0$ 為 group 2

$R_{jt}<0$ 為 group 1

$R_{jt}=0$ 為 group 0

3\. Use

$$\left(R_{jt_0}^{*}(\text{some } t_0\ , t_0 \in \text{group } 0)\middle|\alpha_1,\alpha_2,\beta,\sigma,R_{jt},R_{mt},(\text{given all } t)\right)$$

$$\sim N\left(\beta R_{mt_0}，\sigma^2\right)\Big|_{\alpha_1}^{\alpha_2}$$

And relation

$R_{jt}^{*}=R_{jt}+\alpha_1$ note by group 1

$R_{jt}^{*}=R_{jt}+\alpha_2$ note by group 2

generator corresponding R_{jt}^{*}

4.

Calculate：

minmum R_{jt}^* of group 0 denote by $\min(R_{jt}^*)$ And

maxmum R_{jt}^* of group 0 denote by $\max(R_{jt}^*)$

Use：

$$(\alpha_1|\alpha_2,\beta,\sigma,R_{jt},R_{mt},R_{jt}^*)\sim N\left(-\frac{\Sigma_1(R_{jt}-\beta R_{mt})}{\#1}\ ,\ \frac{\sigma^2}{\#1}\right)\Big|_{-1}^{\min(R_{jt}^*)}$$

$$(\alpha_2|\alpha_1,\beta,\sigma,R_{jt},R_{mt},R_{jt}^*)\sim N\left(-\frac{\Sigma_2(R_{jt}-\beta R_{mt})}{\#2}\ ,\ \frac{\sigma^2}{\#2}\right)\Big|_{\max(R_{jt}^*)}^{1}$$

generator：α_1 and α_2

storage： α_1 and α_2 (Each time)

replace： old α_1 and old α_2

5.

Calculate：

$$M=\frac{\Sigma_{group\ 1}\left((R_{jt}+\alpha_1)*R_{mt}\right)+\Sigma_{group\ 2}\left((R_{jt}+\alpha_2)*R_{mt}\right)+\Sigma_{group\ 0}(R_{mt}*R_{jt}^*)}{\Sigma_{group\ 1\ 2\ 0}\{R_{mt}^2\}}$$

$$\Sigma=\frac{\sigma}{\sqrt{\Sigma_{group\ 1\ 2\ 0}\{R_{mt}^2\}}}$$

Use：

$(\beta|\alpha_1,\alpha_2,\sigma,R_{jt},R_{mt},R_{jt}^*)\sim N(M，\Sigma^2)$

generator：β

storage： β (Each time)

replace： old β

6.

Calculate：

$$W=\frac{1}{2}\left[\sum_{group\ 1}(R_{jt}+\alpha_1-\beta R_{mt})^2+\sum_{group\ 2}(R_{jt}+\alpha_2-\beta R_{mt})^2+\sum_{group\ 0}(R_{jt}^*-\beta R_{mt})^2\right]$$

Use：

$\left(\frac{1}{\sigma^2}\middle|\alpha_1, \alpha_2, \beta, R_{jt}, R_{mt}, R_{jt}^{*}\right) \sim \text{Gamma}(\alpha = \frac{n-1}{2}$ ，$\beta = \frac{1}{W})$

generator：σ

storage： σ (Each time)

replace： old σ

Above 3~6 Called one cycle

7.

do100,000 cycle

8. Calculate four groups parameter median ，As a result。

二、2 Frequentist

EM(Estimation Maximization)

介紹

假定：有一筆 data 分為兩個部分

Observed data $\quad X=(x_1，x_2，\ldots，x_n)$

Unobserved data $Z=(z_1，z_2，\ldots，z_n)$

h are the parameters

定義：$P(X，Z|h)$是參數的機率分配函數；其中(X，Z) is the full data

EM Algorithm：

給定參數起始值h_0

Estimation (E)step：

計算 $E[\,P(X，Z|h)\,]$ 其中：利用 observed X 和 當前參數 h 去估計 Z

Maximization (M)step：

找出使得 E[P(Y|h)]為最大值的h'並取代掉先前的參數 h

重複地去做(E)step和(M)step直到h'和先前的參數差異不大時，

h'就是我們參數 h 的估計結果。

二、2－1 EM Method

Model：

$$
\begin{aligned}
L &= L(\alpha_1,\alpha_2,\beta,\sigma|R_{jt},R_{mt}) \\
&= \pi_1\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_1-\beta R_{mt}}{\sigma}\right)\right]*\pi_2\left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}+\alpha_2-\beta R_{mt}}{\sigma}\right)\right] \\
&\quad *\pi_0\left[\Phi_2\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)-\Phi_1\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)\right]
\end{aligned}
$$

其中 $\varphi(.)$是標準常態的 PDF，$\Phi(.)$是標準常態的 CDF

Analysis ：

L 為一個 4 參數的概似函數，若使用 MLE 我們必須計算出個別參數的估計值在計算上是一個很大的困難條件也非常多，為了簡化計算我將α_1, α_2先固定起來也就是我們的 Model 由 $L(\alpha_1, \alpha_2, \beta, \sigma | R_{jt}, R_{mt})$簡化為 $L(\beta, \sigma | \alpha_1, \alpha_2, R_{jt}, R_{mt})$ 再求解 MLE 找出 β, σ 的估計值。

觀察：

$$L(\beta, \sigma | \alpha_1, \alpha_2, R_{jt}, R_{mt}) = \pi_1 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt} + \alpha_1 - \beta R_{mt}}{\sigma}\right)\right] * \pi_2 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt} + \alpha_2 - \beta R_{mt}}{\sigma}\right)\right] * \pi_0 \left[\Phi_2\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - \Phi_1\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)\right]$$

可以看出在求解 β, σ 的 MLE 時$\pi_0 \left[\Phi_2\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - \Phi_1\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)\right]$

計算上有些許複雜也因此我們在對這區域做些簡化

Analysis ：

π_0這個區域為 group 0 代表：$R_{jt}{}^* \in (\alpha_{1j}，\alpha_{2j})$

其中$R_{jt}{}^* = \beta R_{mt} + \varepsilon_{jt}$ ，$\boldsymbol{\varepsilon} \sim N(0，\sigma^2)$

$\Rightarrow$ in (group 0) $R_{jt}{}^* \sim N(\beta R_m, \sigma^2)|_{\alpha_1}^{\alpha_2}$

$$\therefore \pi_0 \left[\Phi_2\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - \Phi_1\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)\right] \Rightarrow \pi_0 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}{}^* - \beta R_{mt}}{\sigma}\right)\right]$$

$\varphi(.)$是標準常態的 PDF

Model：

$$L = L(\beta, \sigma | \alpha_1, \alpha_2, R_{jt}, R_{mt}) = \pi_1 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt} + \alpha_1 - \beta R_{mt}}{\sigma}\right)\right] * \pi_2 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt} + \alpha_2 - \beta R_{mt}}{\sigma}\right)\right] * \pi_0 \left[\frac{1}{\sigma}\varphi\left(\frac{R_{jt}{}^* - \beta R_{mt}}{\sigma}\right)\right]$$

$$\Rightarrow \ln L = \ln L(\beta, \sigma | \alpha_1, \alpha_2, R_{jt}, R_{mt})$$

$$= \frac{-1}{2}\left[\sum_{0,1,2} \ln(2\pi\sigma^2) + \sum_{1} \frac{\left(R_{jt} + \alpha_1 - \beta R_{mt}\right)^2}{\sigma^2} + \sum_{2} \frac{\left(R_{jt} + \alpha_2 - \beta R_{mt}\right)^2}{\sigma^2} + \sum_{0} \frac{(R_t{}^* - \beta R_{mt})^2}{\sigma^2}\right]$$

此時

Observed data R_{jt}, R_{mt}

Unobserved data $R_{jt}{}^*$(group 0)

則在α_1, α_2固定下 也就是$\ln L\left(\beta, \sigma \middle| R_{jt}, R_{mt}, \alpha_1, \alpha_2\right)$我們有意願尋找 β, σ 的最佳解，所以我們去尋找 β, σ 的 MLE

二、2 − 2 Beta MLE

$$\frac{\partial \ln L\left(\beta, \sigma \middle| \alpha_1, \alpha_2, R_{jt}, R_{mt}\right)}{\partial \beta}$$

$$= \frac{1}{\sigma^2}\left[\sum_{1} (R_{jt} + \alpha_1 - \beta R_{mt}) R_{mt} + \sum_{2} (R_{jt} + \alpha_2 - \beta R_{mt}) R_{mt} + \sum_{0} (R_t{}^* - \beta R_{mt}) R_{mt}\right] = 0$$

$$\Rightarrow \sum_{1} (R_{jt} + \alpha_1) * R_{mt} - \sum_{2} \left(R_{jt} + \alpha_2\right) * R_{mt} + \sum_{0} R_{mt} * R_t{}^* = \beta \sum_{0,1,2} (R_{mt})^2$$

$$\Rightarrow \hat{\beta}_{MLE} = \frac{\sum_1 (R_{jt} + \alpha_1) * R_{mt} - \sum_2 \left(R_{jt} + \alpha_2\right) * R_{mt} + \sum_0 R_{mt} * R_t{}^*}{\sum_{012} (R_{mt})^2}$$

其中$R_t{}^*$以 E(X) 估計 $X \sim N(\beta R_{mt}, \sigma^2)|_{\alpha_1}^{\alpha_2}$

$$E(X) = \beta R_{mt} + \sigma * \left[\frac{\varphi\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right) - \varphi\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right)}{\Phi\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - \Phi\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)}\right]$$

二、2－3 Sigma MLE

$$\frac{\partial \ln L(\beta,\sigma|\alpha_1,\alpha_2,R_{jt},R_{mt})}{\partial\sigma}$$

$$=\frac{-1}{2}*\frac{\partial}{\partial\sigma}\left[\sum_{012}\ln(2\pi\sigma^2)+\sum_{1}\frac{\left(R_{jt}+\alpha_1-\beta R_{mt}\right)^2}{\sigma^2}+\sum_{2}\frac{(R_{jt}+\alpha_2-\beta R_{mt})^2}{\sigma^2}\right.$$

$$\left.+\sum_{0}\frac{(R_t{}^*-\beta R_{mt})^2}{\sigma^2}\right]$$

$$=-\sum_{012}\frac{1}{\sigma}+\sum_{1}\frac{\left(R_{jt}+\alpha_1-\beta R_{mt}\right)^2}{\sigma^3}+\sum_{2}\frac{\left(R_{jt}+\alpha_2-\beta R_{mt}\right)^2}{\sigma^3}+\sum_{0}\frac{(R_t{}^*-\beta R_{mt})^2}{\sigma^3}$$

$$=0$$

$$\Rightarrow\sum_{1}\frac{\left(R_{jt}+\alpha_1-\beta R_{mt}\right)^2}{\sigma^3}+\sum_{2}\frac{\left(R_{jt}+\alpha_2-\beta R_{mt}\right)^2}{\sigma^3}+\sum_{0}\frac{(R_t{}^*-\beta R_{mt})^2}{\sigma^3}=\frac{n}{\sigma}$$

$$\Rightarrow\hat{\sigma}^2{}_{MLE}$$

$$=\frac{\sum_1\left(R_{jt}+\alpha_1-\beta R_{mt}\right)^2+\sum_2\left(R_{jt}+\alpha_2-\beta R_{mt}\right)^2+\sum_0\left[[R_t{}^*]^2-2\beta R_{mt}R_t{}^*+\beta^2 R_{mt}{}^2\right]}{n}$$

其中$[R_t{}^*]^2$以 $E(X^2)$估計，$R_t{}^*$以 $E(X)$ 估計　$X\sim N(\beta R_{mt},\sigma^2)|_{\alpha_1}^{\alpha_2}$

$$E(X^2)=[E(X)]^2$$

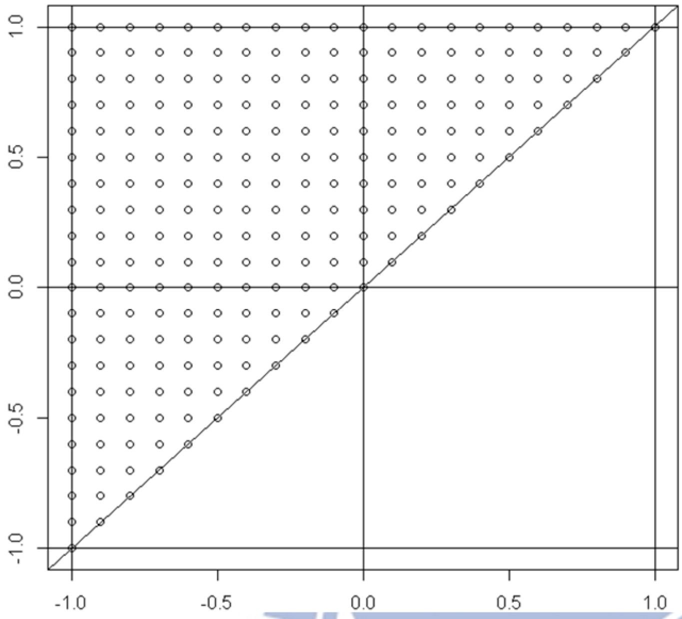
$$+\sigma^2\left[1+\frac{\frac{\alpha_1-\beta R_{mt}}{\sigma}\varphi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)-\frac{\alpha_2-\beta R_{mt}}{\sigma}\varphi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)}{\Phi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)-\Phi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)}\right.$$

$$\left.-\left(\frac{\varphi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)-\varphi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)}{\Phi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)-\Phi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)}\right)^2\right]$$

二、2－4 EM 方法過程

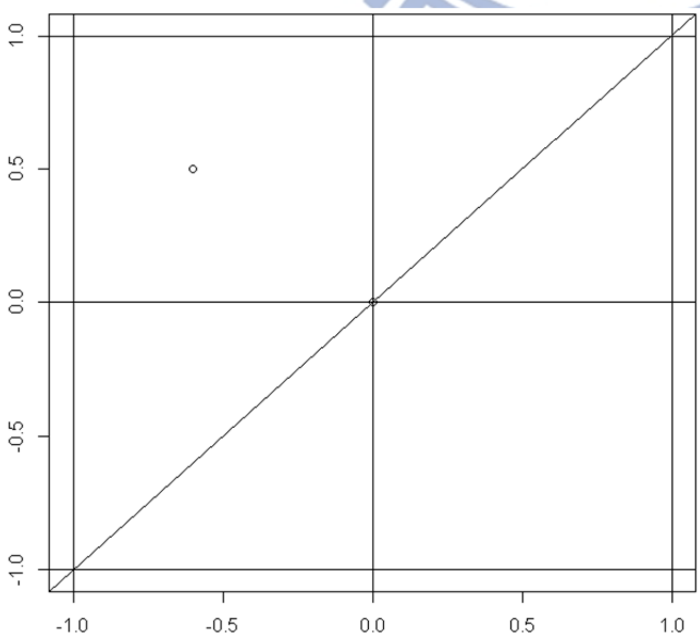
Given initial $\alpha_1, \alpha_2, \beta, \sigma$

其中 $-1 < \alpha_1 < \alpha_2 < 1$　　在間距 0.1 下 我們考慮下面這些點

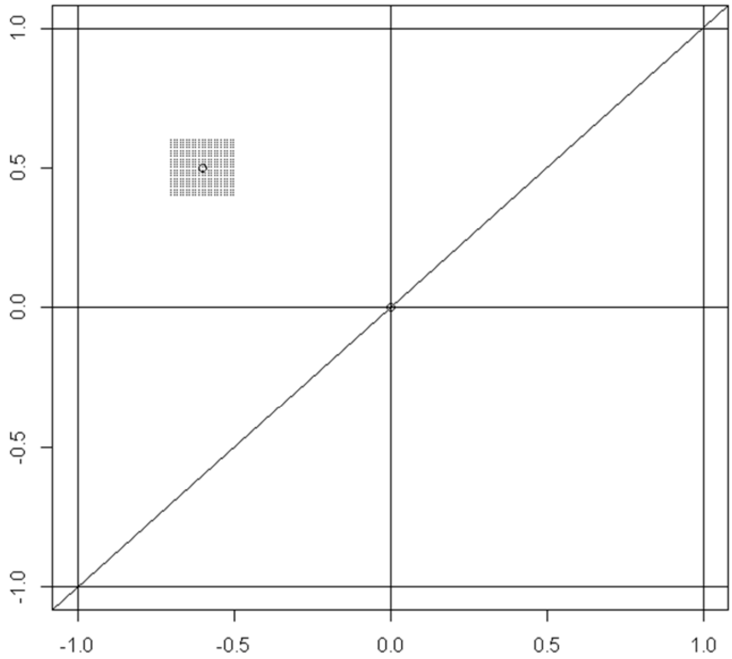


其中 X 軸是α_1，Y 軸是α_2

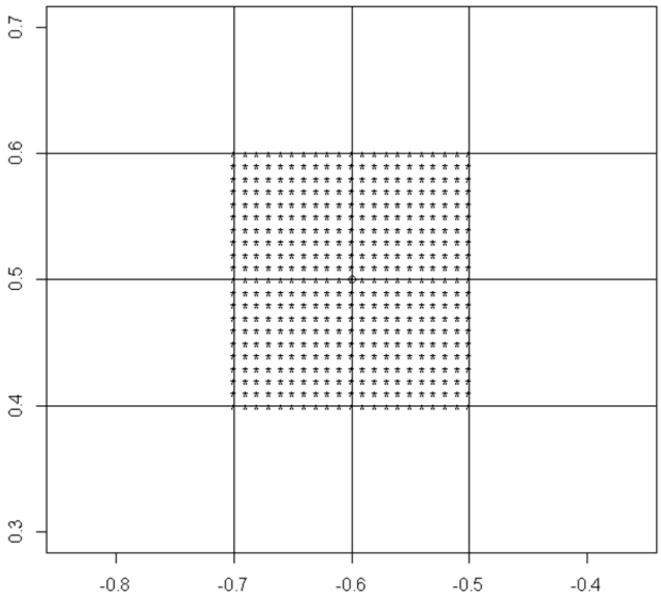
透過先前推導出的公式去估計每個點最有可能的 β 和 σ 的估計值(重複疊帶 5 次而產生)，並且記錄這點在引入 data 下的概似值$\left(\ln L\left(\beta, \sigma \middle| \alpha_1, \alpha_2, R_{jt}, R_{mt}\right)\right)$，從中找出最大概似值的點並且記為$\alpha_1^*, \alpha_2^*$如下圖所示：



假定上圖中的點是最大概似值出現的地方，我們以這點上下 0.1 的範圍，間距 0.01 從中再找一次α_1, α_2如下圖所示：



放大顯示



透過上面推導出的公式去估計每個點最有可能的 β 和 σ 的估計值(重複疊帶 5 次而產生)，並且記錄這點在引入 data 下的概似值$\left(\ln L\left(\beta, \sigma \middle| \alpha_1, \alpha_2, R_{jt}, R_{mt}\right)\right)$，從中找出最大概似值的點並且記為$\alpha_1^{**}, \alpha_2^{**}$，在從這點上下 0.01 的範圍間距 0.001 從中再找一次α_1, α_2做到間距10^{-5}為止。

※其中 β 和 σ 的估計值是透過上述公式重複疊帶 5 次而產生。

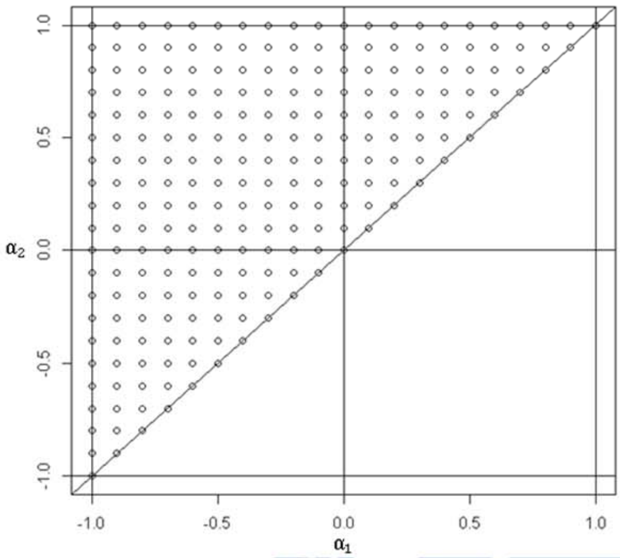
疊代 5 次是因為參數估計值和疊代 4 次時值已經很相近(已經收斂)。

二、2－5 EM 詳細流程

1. given initial max 概似值

2. 將 α_1，α_2以間隔 0.1 做切割其中 $-1 < \alpha_1 < \alpha_2 < 1$

   在間距 0.1 下 我們考慮下面這些點如圖：



3. 每一組$(\alpha_1，\alpha_2)$(點) given initial β_0，σ_0

   透過

$$\hat{\beta} = \frac{\sum_1 (R_{jt} + \alpha_1) * R_{mt} - \sum_2 (R_{jt} + \alpha_2) * R_{mt} + \sum_0 R_{mt} * {R_t}^*}{\sum_{0,1,2} (R_{mt})^2}$$

$$\hat{\sigma}^2 = \frac{\sum_1 (R_{jt} + \alpha_1 - \beta R_{mt})^2 + \sum_2 (R_{jt} + \alpha_2 - \beta R_{mt})^2 + \sum_0 \left[[{R_t}^*]^2 - 2\beta R_{mt} {R_t}^* + \beta^2 {R_{mt}}^2 \right]}{n}$$

其中${R_t}^*$以 $E(X)$估計，$[{R_t}^*]^2$以 $E(X^2)$估計 $X \sim N(\beta R_m, \sigma^2)|_{\alpha_1}^{\alpha_2}$

$$E(X) = \beta R_m + \sigma * \left[\frac{\varphi\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right) - \varphi\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right)}{\Phi\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - \Phi\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)} \right]$$

$$E(X^2) = [E(X)]^2$$

$$+\sigma^2\left[1+\frac{\frac{\alpha_1-\beta R_{mt}}{\sigma}\varphi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)-\frac{\alpha_2-\beta R_{mt}}{\sigma}\varphi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)}{\Phi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)-\Phi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)}\right.$$

$$\left.-\left(\frac{\varphi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)-\varphi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)}{\Phi\left(\frac{\alpha_2-\beta R_{mt}}{\sigma}\right)-\Phi\left(\frac{\alpha_1-\beta R_{mt}}{\sigma}\right)}\right)^2\right]$$

估計 β，σ $\rightarrow$ 取代掉舊的 β_0，σ_0 結果計為β_1，σ_1

重新估計 β，$\sigma \rightarrow$ 取代先前估出的β_1，σ_1

重新估計 β，$\sigma \rightarrow$ 取代先前估出的β_2，σ_2 …(做到收斂為止)

收斂代表的是前一次估計結果和這次估計結果差不多，在此公式之下 5 次就能得到不錯的結果。

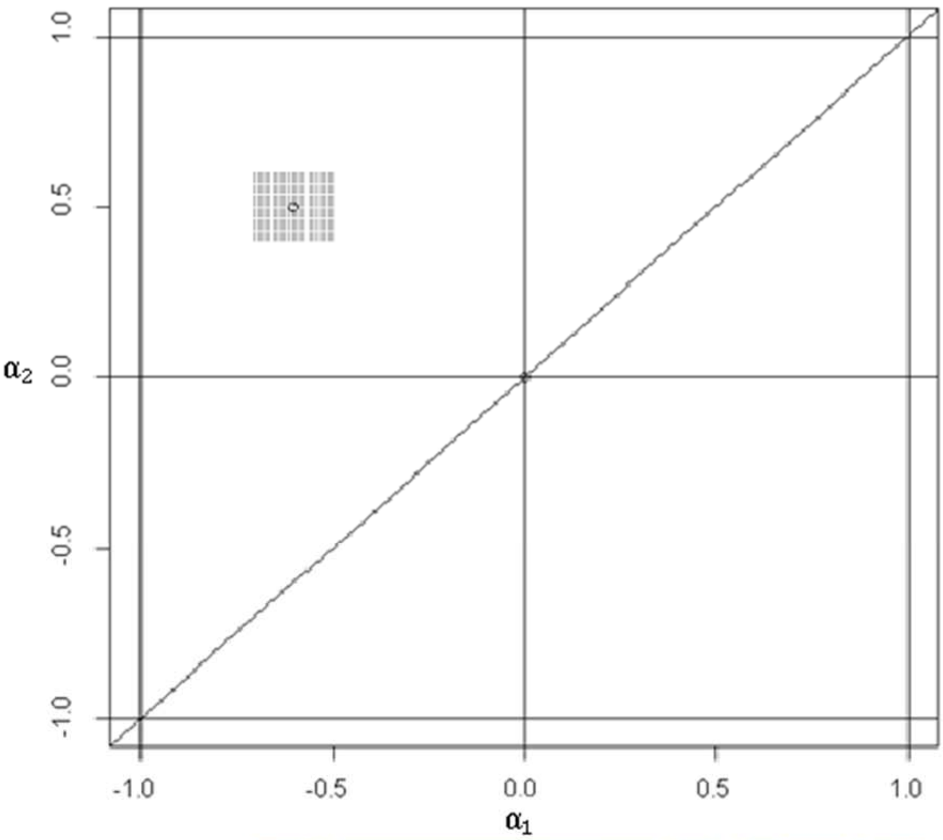
假定收斂在第 k 次計為 β_k，σ_k

計算 $\ln L(\beta,\sigma|\alpha_1,\alpha_2,R_{jt},R_{mt})$ 並和 max 概似值比較

若 $\ln L(\beta,\sigma|\alpha_1,\alpha_2,R_{jt},R_{mt}) >$ max 概似值

紀錄(α_1，α_2 ，β_k，σ_k)並將(α_1，α_2)計為($\alpha_1{}^*$，$\alpha_2{}^*$)

4. 將($\alpha_1{}^*$，$\alpha_2{}^*$)作為這中心點考慮前次間距內的所有點(以前次間隔的$\frac{1}{10}$做切割)如圖：



5. 會到步驟 3. 做到點和點距離小於10^{-5}

三、1 Gibbs Sampling Simulation Result

Simulation data

Gibbs sampling

Sample size 1000 run time 1,000 true parameter：

```
> true
            [,1]       [,2]      [,3]       [,4]
[1,] -0.0151977 0.01871786 -1.548321 0.04971042
```

α_1 α_2 β σ

Result：

```
> mean(note[,1])
[1] -0.01502501
> median(note[,1])
[1] -0.01511472
```

α_1

```
> mean(note[,2])
[1] 0.01649408
> median(note[,2])
[1] 0.01636214
```

α_2

```
> mean(note[,3])
[1] -1.547073
> median(note[,3])
[1] -1.546946
```

β

```
> mean(note[,4])
[1] 0.04952052
> median(note[,4])
[1] 0.04949694
```

σ

Sample size 1000 run time 100,000 true parameter：

```
> true
          [,1]      [,2]      [,3]       [,4]
[1,] -0.1317812 0.1117481 0.8080516 0.03848179
```

α_1 α_2 β σ

Result：

```
> mean(note[,1])       > mean(note[,2])       > mean(note[,3])       > mean(note[,4])
[1] -0.1309725         [1] 0.1123192          [1] 0.8066624          [1] 0.0389362
> median(note[,1])     > median(note[,2])     > median(note[,3])     > median(note[,4])
[1] -0.1309366         [1] 0.1122961          [1] 0.8066226          [1] 0.03890593
```

α_1 α_2 β σ

三、2 EM Simulation Result

Simulation data

EM

Sample size 200

```
> alpha1_ans  > alpha2_ans  > beta_ans   > sigma_ans
[1] -0.05311  [1] 0.00457   [1] 0.53368  [1] 0.03372031

> true
          [,1]   [,2]      [,3]       [,4]
[1,] -0.05328 0.0121 0.5523969 0.03241055
```

Size 1000

```
> alpha1_ans  > alpha2_ans  > beta_ans     > sigma_ans
[1] -0.05461  [1] 0.01498   [1] 0.5562606  [1] 0.03317146

> true
          [,1]   [,2]      [,3]       [,4]
[1,] -0.05328 0.0121 0.5523969 0.03241055
```

四、1 & 四、2 Gibbs Sampling Result V. S EM Result

True Data

data1 size 100

data1 use EM Result

```
> alpha1_ans   > alpha2_ans   > beta_ans     > sigma_ans
[1] -0.03231   [1] 0.02947    [1] 1.133122   [1] 0.02555444
```

data1 time 2000 Gibbs Sampling Result

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.03258749       [1] 0.02549582        [1] 0.6182785         [1] 0.02506359
> median(note[,1])    > median(note[,2])    > median(note[,3])    > median(note[,4])
[1] -0.03252269       [1] 0.02549222        [1] 0.5949665         [1] 0.02491645
```

α_1 α_2 β σ

data1 time 4000 Gibbs sampling

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.03301035       [1] 0.02642665        [1] 0.7964052         [1] 0.02527237
> median(note[,1])    > median(note[,2])    > median(note[,3])    > median(note[,4])
[1] -0.03236322       [1] 0.02622001        [1] 0.7933636         [1] 0.02511396
```

α_1 α_2 β σ

data1 time 10，000 Gibbs sampling

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.03245519       [1] 0.02887423        [1] 1.409818          [1] 0.02576819
> median(note[,1])    > median(note[,2])    > median(note[,3])    > median(note[,4])
[1] -0.03210413       [1] 0.02858063        [1] 1.410577          [1] 0.02559092
```

α_1 α_2 β σ

data1 time 100，000 Gibbs sampling

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.03280459       [1] 0.02939318        [1] 1.146148          [1] 0.02507361
> median(note[,1])    > median(note[,2])    > median(note[,3]     > median(note[,4])
[1] -0.03246302       [1] 0.02911841        [1] 1.138823          [1] 0.02489669
```

α_1 α_2 β σ

True Data

data2 size 1000

data2 use EM Result

```
> alpha1_ans    > alpha2_ans   > beta_ans      > sigma_ans
[1] -0.02687    [1] 0.02394    [1] 0.1082602   [1] 0.05151914
```

data2 time 1，000 Gibbs Sampling Result

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.02765413       [1] 0.02258955        [1] 0.1840856         [1] 0.05135991
> median(note[,1])    > median(note[,2])    > median(note[,3])    > median(note[,4])
[1] -0.02777365       [1] 0.02257664        [1] 0.1881298         [1] 0.05128735
```

data2 time 10，000 Gibbs Sampling Result

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.02772204       [1] 0.02410697        [1] 0.201663          [1] 0.05218017
> median(note[,1])    > median(note[,2])    > median(note[,3])    > median(note[,4])
[1] -0.02764434       [1] 0.02412101        [1] 0.2006337         [1] 0.05213576
```

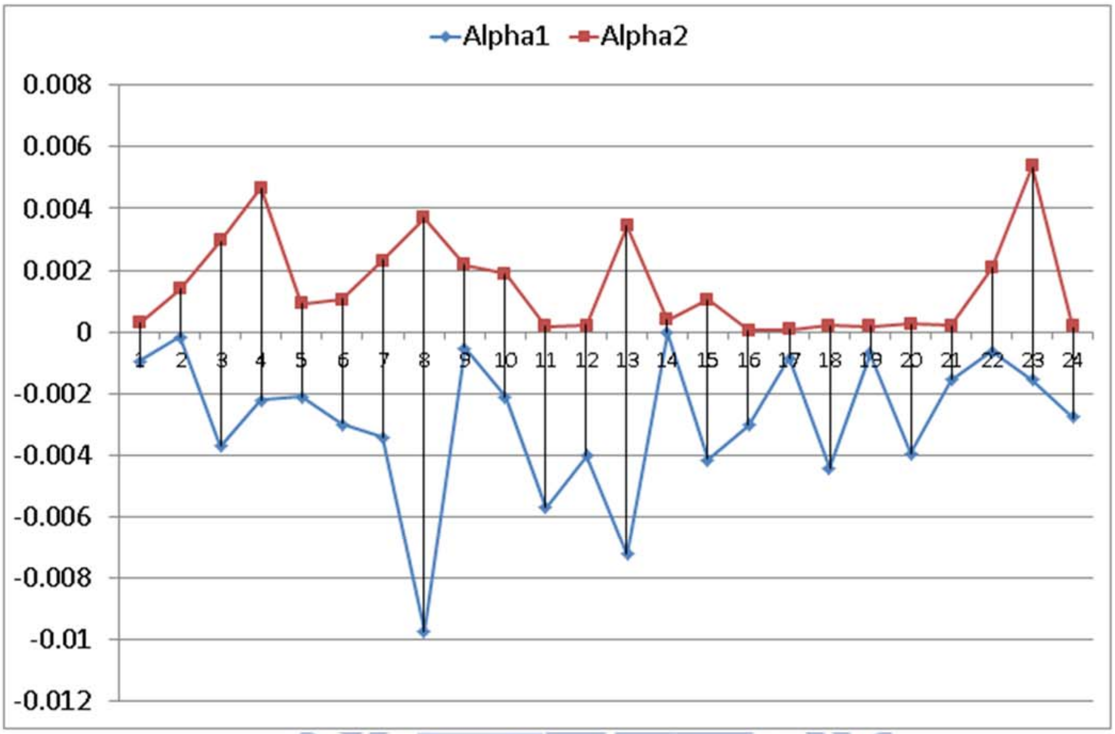
data2 time 100，000 Gibbs Sampling Result

```
> mean(note[,1])      > mean(note[,2])      > mean(note[,3])      > mean(note[,4])
[1] -0.02696358       [1] 0.02341591        [1] 0.0945296         [1] 0.0514199
> median(note[,1])    > median(note[,2])    > median(note[,3])    > median(note[,4])
[1] -0.027014         [1] 0.02335248        [1] 0.09387119        [1] 0.05136821
```

五、驗證

由於真實資料的資料時間點大多由 1986~2011 年，這期間恰巧涵蓋次貸危機(時間點 2007 年 4 月)，所以我有興趣研究這段時間內的交易成本。

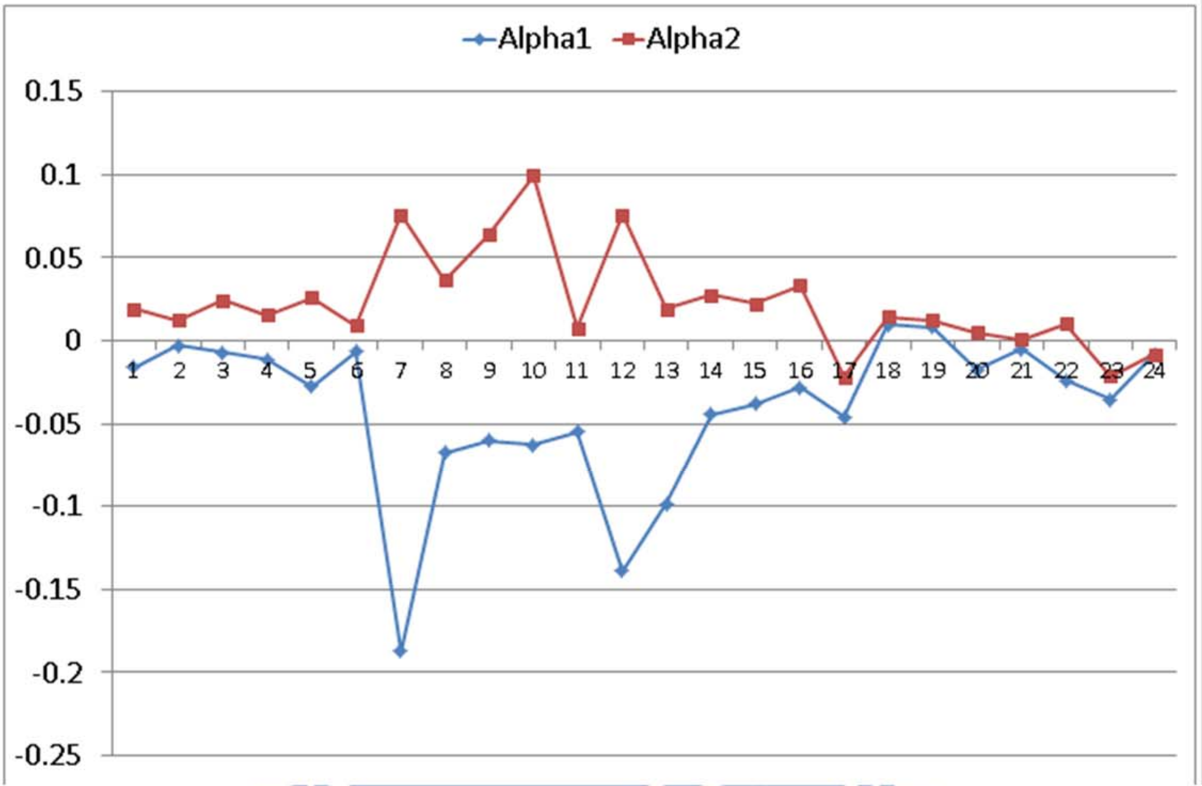
下圖每一個點代表為是用 20 個交易日所得出的估計值：公司代號為 A 公司分析時間大約為 24 個月(簡稱 24 個時點)。(開始為 2007 年 4 月 2 日)



我們可以觀察出α_1和α_2都非常的小，代表的是這段期間內投資 A 公司賺賠情形能很容易的判斷出來，另一方面也能夠說明這期間交易成本比較低。原因是次貸危機這個負面事件帶來的衝擊使得市場趨於保守，此時大部分投資人可能選擇把錢丟入債卷市場或者銀行定存，這時銀行利率必然下降，因為銀行獲利能力降低，此時銀行或者卷商為了活絡市場可能提出比較優惠的借貸利率來刺激市場。

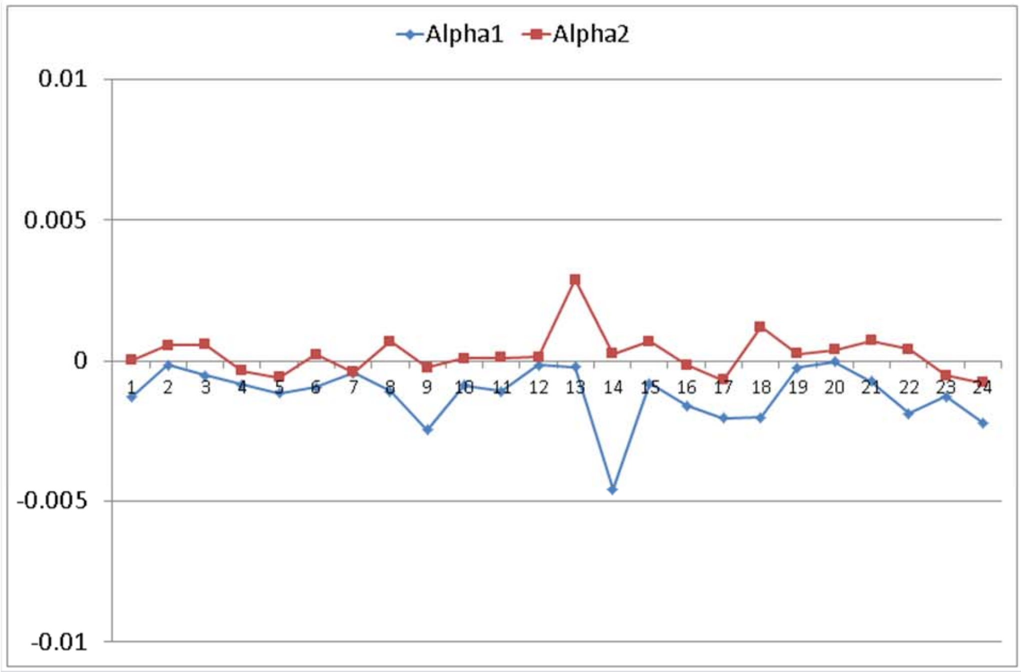
上圖中能看出α_1對交易成本的影響是大於α_2的，原因在於這個事件所帶來的負面效應導致我們對未來市場比較看跌。

公司代號為 B 公司



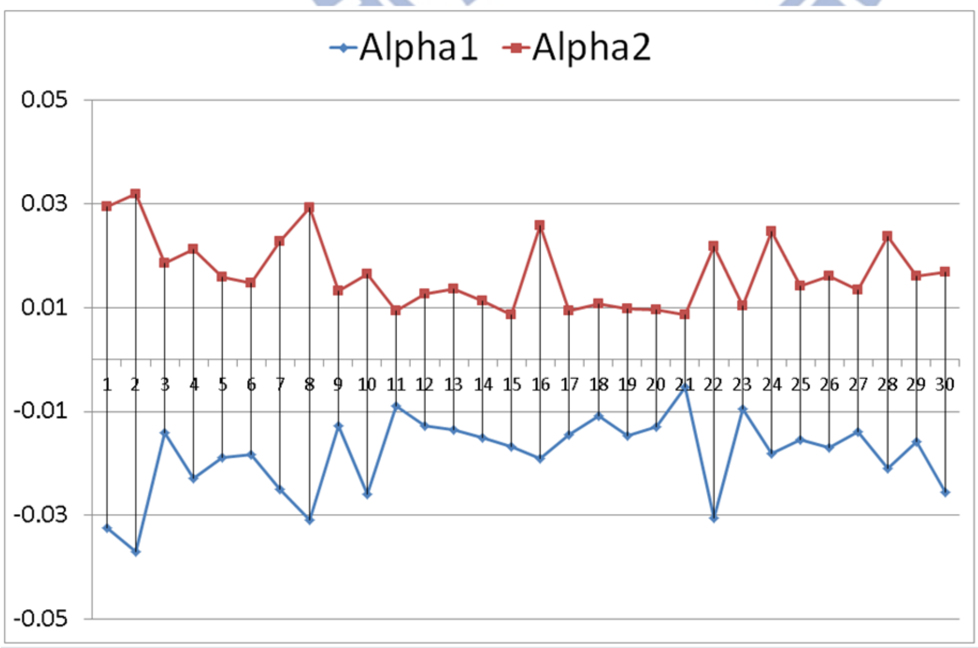
我們能看出在海嘯發生後 B 公司前 6 個月α_1和α_2挺小的，但是 B 公司α_1對交易成本的影響在這 6 個月並沒有特別大於α_2，我推測的原因是就算我們對未來看空想要借卷來賣，但我們借來的卷可能沒人買又或者我們掛的金額必須很低，可能造成借來的卷 10 元，但掛單必須 5 元才能成交的情況所以賣出後價格必須跌得很低才有獲利的能力才會出現接近負 20%的數值。A 和 B 公司有個共同點就是在時點 7 的地方市場漸漸有些起色但是在時點 16 之後α_1和α_2又變的非常小這是因為在 2008 年 9 月 15 日美國第四大投資銀行雷曼兄弟倒閉，市場再度受到嚴重的打擊。

最後我們在分析公司代號為簡稱 C 公司：

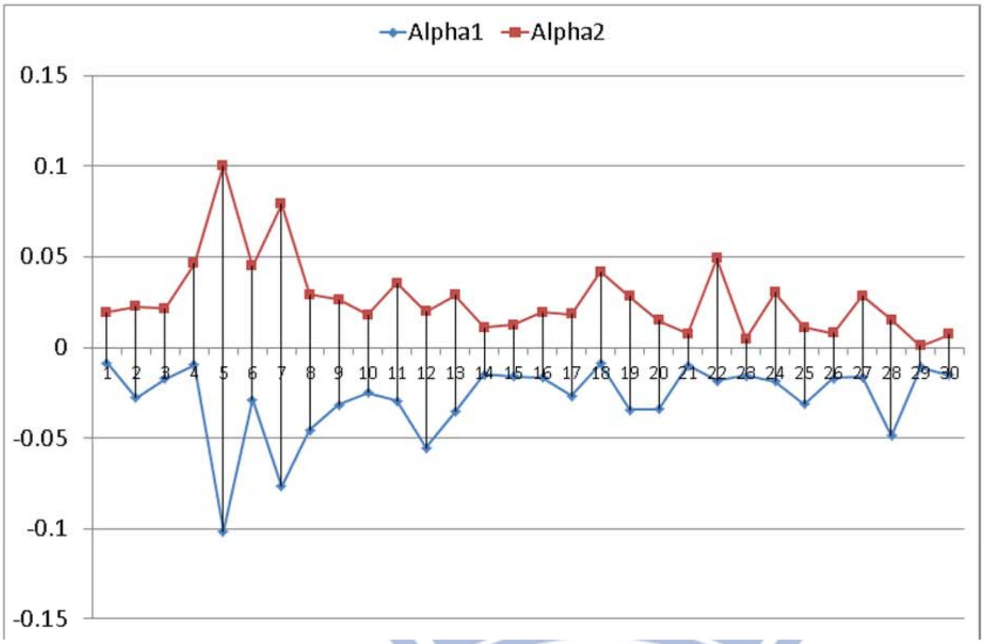


看到 C 公司的估計結果我們能得到和 A，B 兩公司差不多的結論代表模型對於市場概況有著一致性。

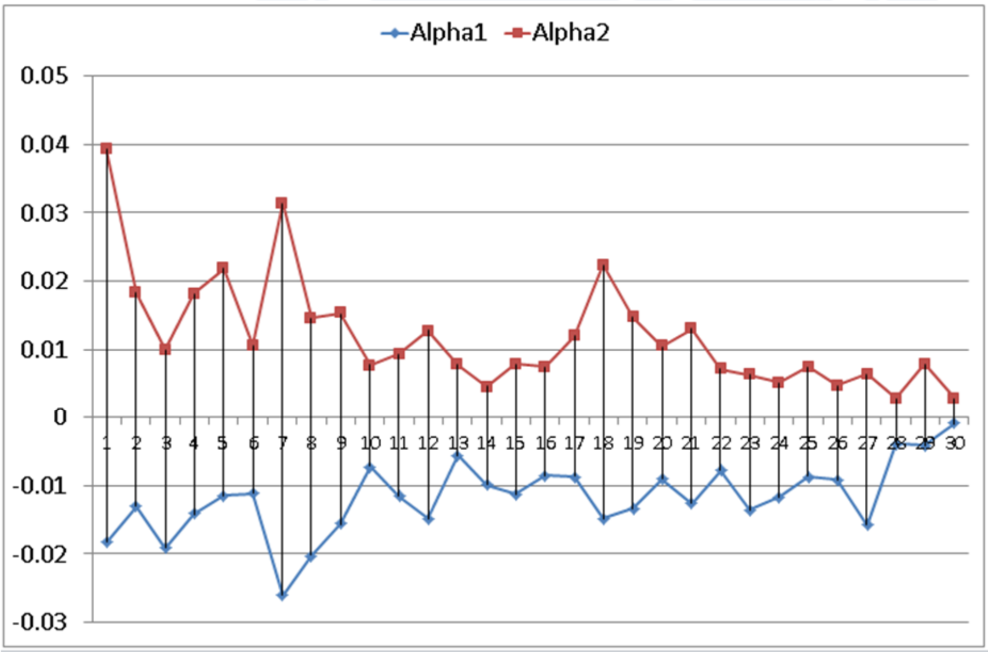
再來我們將分析這幾家公司在 1986 年到 1989 年 6 月的表現，用以說明 LDV 模型對於不同公司有著鑑別性。



A 公司 1986 年 1 月到 1989 年 6 月



B 公司 1986 年 4 月到 1989 年 9 月



C 公司 1986 年 1 月到 1989 年 6 月

結論：

我們能看出針對不同公司同一個時間點，這個模型能估計出相對應的市場趨勢，並且

在市場概況不佳的情況下估計出一致的結果，這代表著 LDV 模型所估計出的α_1和α_2能充分反應出市場狀況。故我們能透過α_1和α_2的值判斷出正確的交易成本並且判斷出影響證卷報酬率的程度。

References

[1] Bhushan, R., 1994, "An Informational Efficiency Perspective on the Post-Earnings Drift," Journal of Accouniting anid Econornics, 18, 46-65.

[2]Benston, G. J., and R. L. Hagerman, 1974, "Determinants of Bid-Ask Spreads in the Over-the-Counter Market," Journal of Finianicial Economics, 1, 353-364.

[3]Berkowitz, S. A., D. E. Logue, and E. A. Noser, Jr., 1988, "The Total Cost of Transaction Costs in the NYSE," Jourtnal of Finiance, 43, 97-112.

[4]Bhardwaj, R. K., and L. D. Brooks, 1992, "The January Anomaly: Effects of Low Share Price, Transaction Costs, and Bid-Ask Bias," Journial of Finiance, 47, 553-574.

[5]Bhushan, R., 1991, "Trading Costs, Liquidity, and Asset Holdings," Review of Finiancial Studies, 4, 343-360.

[6]Copeland, T. E., and D. Galai, 1983, "Information Effects of the Bid-Ask Spread," Journal of Finance, 38, 1457-1469.

[7]Conrad, J., G. Kaul, and M. Nimalendran, 1991, "Components of Short-Horizon Individual Security Returns," Jourlnal of Finiancial Econiomics, 29, 365-384.

[8]Constantinides, G., 1986, "Capital Market Equilibrium with Transaction Costs," Journal of Political Econiomy, 94, 842-862.

[9]Dumas, B., and E. Luciano, 1991, "An Exact Solution to a Dynamic Portfolio Choice Problem Under Transactions Costs," Journal of Finianice, 46, 577-598.

[10]Demsetz, H., 1968, "The Cost of Transacting," Quarterly Journal of Econiomics, 82, 33-53.

[11]Grossman, S., and M. Miller, 1988, "Liquidity and Market Structure," Journal of Finance, 43, 617-636.

[12]Glosten, L., and P. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogenously Informed Traders," Journal of Finiancial Economics, 14, 71-100.

[13]Huang, R., and H. Stoll, 1996, "Dealer Versus Auction Markets: A Paired Comparison of Execution Costs on NASDAQ and the NYSE," Journal of Finiancial Ecotnomics, 41,313-357.

[14]Harris, L., 1990, "Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator," Journal of Finiance, 45, 579-590.

[15]Huang, R., and H. Stoll, 1994, "Market Microstructure and Stock Return Predictions," Review of Finiancial Studies, 7, 179-213.

[16]Johnson, D. B., 1994, "Property Rights to Investment Research: The Agency Costs of Soft Dollar Brokerage." Yale Journal on Regulation, 11 , 75-113.

[17]Kyle, A. S., 1985, "Continuous Auctions and Insider Trading," Econometrica, 53, 1315-1335.

[18]Karpoff, M. K., and R. A. Walkling, 1988, "Short Term Trading Around Ex-Dividend Days: Addition Evidence," Journal of Finiancial Economics, 21, 291-298.

[19]Kato, K., and U. Loewenstein, 1995, "The Ex-Dividend-Day Behavior of Stock Prices: The Case of Japan," Review of Finanicial Studies, 8, 817-848.

[20]Keim, D. B., and A. Madhavan, 1995, "Execution Costs and Investment Performance: An Empirical Analysis of Institutional Equity Trades," working paper, University of Pennsylvania.

[21]Knez, P., and M. Ready, 1996, "Estimating the Profits from Trading Strategies," Review of Financial Studies, 9, 1121-1163.

[22]Lesmond, D.; J. Ogden; AND C. Trzcinka. A New Estimate of Transaction Costs. Review of Financial Studies **12** (1999): 1113–41.

[23] Lee, C., and M. Ready, 1991, "Inferring Trade Direction from Intraday Data," Journial of Finiance, 46, 733-746.

[24]Lesmond, D., 1995, "Transaction Costs and Security Return Behavior: The Effect on Systematic Risk and Firm Size," unpublished dissertation, University at Buffalo.

[25]Maddala, G., 1983, "Limited Dependent and Qualitative Variables in Econometrics," Cambridge University Press, Cambridge, Mass.

[26] Maddala, G., 1991, "A Perspective on the Use of Limited Dependent and Qualitative Variables Models in Accounting Research," Accounting Review, 66, 788-807.

[27]Merton, R., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information Exchange," Jolurnalo f Finianice,4 2, 483-510.

[28]Petersen, M. A., and D. Fialkowski, 1994, "Posted versus Effective Spreads: Good Prices or Bad Quotes?" Jolurnial of Financial Economics, 35, 269-292.

[29]Plexus Group, 1996, "Quality of Trade Execution in Comparative Perspective: AMEX vs. NASDAQ vs. NYSE," August, 1-18.

[30]Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," Journial of Finianice, 39, 1127-1140.

[31]Rosett, R., 1959, "A Statistical Model of Friction in Economics," Ecotnotetrica, 27, 263-267.

[32]Scott, D., 1983, The Inivestor's Guide to Discount Brokers, Prager, New York.

[33]Stoll, H. R., 1989, "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," Journal of Finianicial Economics, 12, 57-79.

[34]Stoll, H. R., and R. Whaley, 1983, "Transactions Costs and the Small Firm Effect," Joutrnal of Finianicial Econiomics, 12, 57-79.

[35]Tobin, J., 1958, "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26, 24-36.

[36]Wood, R. A., T. H. Mclnish, and J. K. Ord, 1985, "An Investigation of Transactions Data for NYSE Stocks," Joturtnal of Finanice, 40, 723-739. 1141