第三章 研究方法

第一節 單根檢定

時間序列資料通常可分為兩種類型：非定態*(Non-Stationary)*與定態*(Stationary)*序列。非定態時間序列資料對於外來衝擊會逐漸累積，並帶來持續且長期性的影響，使其在時間變動中逐漸偏離平均值；而定態的時間序列資料對於外來衝擊只會留下暫時性的影響，在經過干擾後，仍然會回到其平均值。我們在此的定態定義指的是弱式定態*(weakly stationary)*，根據 *Tsay(2005)*，假設$\{r_t\}$為一時間序列資料，則符合以下兩個條件即為弱式定態：

（a）$E(r_t)=\mu,\ \mu$ 為一固定常數。

（b）$Cov(r_t, r_{t-l})=\gamma_l$, 不同期的相關係數只跟落後期數 l 相關。

而一般的時間序列模型皆需要此定態的假設才能得以進行，也因此，我們必須先針對資料做定態的檢定。*Granger and Newbold (1974)*指出，如果直接進行非定態變數的迴歸分析將造成所謂的虛假迴歸*(Spurious Regression)*：兩個獨立且不相關的變數，只因為具有單根*(unitroot)*，就會讓我們估計出一個不存在的相關性。

單根與隨機趨勢*(Stochastic trend)*通常被視為相同的概念，單根也是時間序列資料之所以呈現非定態的原因。假設時間序列 y_t 具有單根，那麼一般我們會說 y_t 具有隨機趨勢。我們在這邊所稱的隨機趨勢，意思通常指的就是時間序列資料持續而長期性的隨機移動，而以總體經濟學來解釋的話，也就是指經濟體系中的外生衝擊對於總體經濟變數的影響是恆久的，任何一次的隨機衝擊都會造成時間序列資料持續而長期性的改變。過去研究總體經濟的學者處理時間序列資料時，都只考慮固定趨勢，認為去除固定趨勢，時間序列資料就可成為定態，但就具有單根的時間序列資料來說，僅僅去掉時間序列的固定趨勢是不夠的。*Nelson and Plosser*（*1982*）指出，大部分經濟變數之時間序列皆具有隨機趨勢。不解決單根

的問題，就會造成我們前面所提到的虛假迴歸問題，而無法利用時間序列模型繼續進行分析。因此，在我們進行分析之前，必須先利用單根檢定，檢查資料屬性是否為非定態，才能繼續後續的研究。

在這邊我們檢定單根的方法選定為 *ADF* 單根檢定。檢定單根的方法最早由 *Dickey* 與 *Fuller(1979)*在其文章中提出，其中假設殘差符合白噪音*(white noise)*，但因為 *DF* 檢定之殘差項通常存在明顯的序列相關*(autocorrelation)*，因此 *Said ＆ Dickey (1984)* 發展出 *Augmented Dickey-Fully Test (ADF 檢定)*，加入額外差分落後期調整，解決以上殘差項序列相關的缺失。*ADF* 檢定另外修正為以 *AR(p)* 的形式進行單根檢定，稱為「修正後 *DF* 檢定」，此檢定的虛無假設為時間序列資料具有單根，對立假設則是時間序列資料為定態資料，因此，在檢定過程中，我們必須得到足以拒絕虛無假設的證據才能套用後續的時間序列模型。

模型如下所示，若 Y_t 為一時間序列資料，β 為自我迴歸係數，b_t 為時間趨勢項，P 則為最適落後期數，e_t 表示干擾項：

模式 1：無截距項且無時間趨勢項之模型 ：

$$\Delta Y_t = \beta Y_{t-1} + \sum_{i=1}^{p} c_j \Delta Y_{t-1} + e_t \qquad (3.1)$$

模式 2：包含截距項之模型：

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \sum_{i=1}^{p} c_j \Delta Y_{t-1} + e_t \qquad (3.2)$$

模式 3：包含截距項及時間趨勢項之模型：

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \sum_{i=1}^{p} c_j \Delta Y_{t-1} + b_t + e_t \qquad (3.3)$$

ADF 檢定允許同質的殘差和白噪音的性質。

另外，在 *ADF* 檢定之外，我們也試著用 *Phillips and Perron(1988)*所提出的 *PP 檢定(Phillips-Perron Test)*。提供兩種單根檢定的方法，一方面是為了更加確認定態時間序列的存在，另一方面也是為了解決 *ADF* 檢定力較低的問題。如果時間序列資料不如我們所預期的為定態，則必須要採取差分*(differencing)*的動作，再重複的進行單根檢定，直到我們的資料呈現定態為止。

第二節 *AR(P)*－*GARCH(1,1)*－*M* 模型

在處理價量關係中價格波動的異質變異時，我們通常會使用一般化自我迴歸條件異質變異數模型*(GARCH)*，而在這裡採取的是 *GARCH(1,1)*這個國內外文獻在股票市場、金融市場都普遍被使用的模型來配適，不僅因為其變數較少，在文獻中通常也都能達到改善殘差項平方自我相關函數*(AutoCorrelation Function, ACF)*的顯著效果。而在我們實際測試模型的結果當中，的確也獲得不錯的實證結論，*GARCH(1,1)*的模型假設仍然受到青睞。

這邊所採用的一般化自我迴歸條件異質變異數模型（*GARCH*）並非普通的 *GARCH* 模型，而是為了將報酬率波動與交易量連結起來所採用的 *GARCH-M* 模型，當然也考慮了市場報酬中風險溢酬的部分。這個模型跟標準的 *GARCH(1,1)* 主要有幾點不同，除了在條件變異數方程式中，參考 GJR-GARCH 的設計，多了不對稱的條件異質變異數(*ARCH-term*, 通常也稱為"*leverage effect* ",亦即槓桿效果項)以外，本文在條件報酬率的部分更多了跟條件變異數連結的風險溢酬項(*in-mean term)*，而這可以幫助我們了解交易量變數加入條件變異數方程式後，所能夠解釋報酬波動的程度。但在一開始的條件變異數方程式*(3.5)*中，我們並未加入交易量變數，在後面的兩個條件變異數方程式*(3.6)*、*(3.7)*才分別加入兩個不同的交易量變數來加以解釋。

一、條件報酬率方程式*(Conditional Return Equation)*

$$R_t = \mu + \sum_{i=1}^{p} \rho_i R_{t-i} + \lambda \sigma^2_{t,F_t} + a_t , \qquad a_t \sim \left(0, \sigma^2_{t,F_t}\right), \quad t = 1,\ldots,T \qquad (3.4)$$

R_t：條件報酬率　σ^2_{t,F_t}:母體分配的條件變異數

我們所設定的條件報酬率方程式中，除了風險溢酬項 σ^2_{t,F_t} 以外，就是簡單的 *AR(p)*模型。模型中包含了一個常數項 μ，以及不同落後期數的自我相關項

$\sum_{i=1}^{p}\rho_i R_{t-i}$，至於干擾項 a_t 的設定，則是獨立抽樣自一平均數為零，條件變異數為 σ^2_{t,F_t} 的厚尾分配。在 *AR(p)*設定下的意義，不外乎就是認為在台灣股票市場中，實際上存在著不同期股票指數的前後期影響，而這也是相關文獻中財務金融時間序列資料最為人知的特性，往往具有與前期的自我相關。至於風險溢酬項的設計，源自於 *Genotte and Marsh(1993)*在美國股票市場的研究，而之後也陸續有研究報酬變動與風險溢酬相互關係的實證結果，在大多數的結果中，此項係數的符號顯著為正，代表市場中明顯存在著風險溢酬。

二、條件變異數方程式*(Conditional variance equation)*

$$\sigma^2_{t,F_t} = h_t^2 \equiv \omega_0 + \omega_1 a^2_{t-1} + \omega_2 a^2_{t-1} I_{\{a_{t-1}<0\}} + \omega_3 \sigma^2_{t-1,F_{t-1}} \quad (3.5)$$

$$\sigma^2_{t,F_t} = h_t^2 + \omega_4 U_t \quad (3.6)$$

$$\sigma^2_{t,F_t} = h_t^2 + \omega_4 U_t I_{\{U_t>0\}} \quad (3.7)$$

條件變異數方程式的設計，我們總共分為三種模式，第一個模式*(3.5)*中我們並未加入交易量變數，而只是普通的 *GARCH(1,1)*模型。在這個模式當中，除了常數項(ω_0)與 *ARCH* 項(a^2_{t-1})、*GARCH* 項($\sigma^2_{t-1,F_{t-1}}$)以外，還有一個不對稱的 *ARCH* 項($a^2_{t-1}I_{\{a_{t-1}<0\}}$)，也就是我們俗稱的槓桿效果項。這一項的主要效果，就是為了要檢驗不同正負面的衝擊，所帶給報酬波動的影響程度是否有所不同，一般來說，一般文獻檢定後的結果多半此項係數的符號都會顯著為正，也就是負面衝擊所帶來的效果，往往比正面衝擊來得影響深遠。

至於第二個模式*(3.6)*，除了跟第一個模式相同的變數設計以外，我們還多加了一項 U_t。而這一項 U_t 事實上就是交易量變數，只是它並不是代表簡單的交易

量而已，它代表的是異常交易量*(abnormal volume)*。本文試著加入此新變數，看看是否會使市場報酬的波動持續性減弱，讓我們用下面簡單的數學式來表達。

$$U_t = V_t - \overline{V_t} \quad ,$$

所以，我們從上面的數學式看出，V_t 表示在 t 時間內的所有交易量，$\overline{V_t}$ 則表示在 t 時間內的平均交易量，而我們所需要的 U_t，基本上表示的就是一般交易量再減去平均交易量的概念，但因為交易量也需要經過定態的檢驗，因此並不單純只是去除平均數的動作，有關交易量的處理，在下一節會有仔細的介紹。利用異常交易量解釋市場波動的想法，其實並不難理解，雖然混和分配假設*(MDH)*所告訴我們的，是一般的交易量變數可以替代*(proxy)*市場資訊的流入，但 *Wagner and Marsh(2003)*認為並不是所有的交易量都帶有豐富的資訊含量。往往只有那些異常的、超過平均數的交易量，才是我們想要拿來觀察報酬波動的變數，通常也只有這樣的交易量，會含有比較多私有資訊的內容，可以被我們拿來當作觀察訊息到來的替代指標，一般的交易量大部分是受到公開資訊的影響，對我們的模型解釋可能不太理想，也沒有辦法較清楚的觀察到市場是否有重大私有資訊流入的情形。而本文預期 ω_4 這項衡量交易量變數解釋能力的係數符號為正，主要也是因為在上述的理論假設前提之下，越大的異常交易量應該會導致股市較大的波動。

最後看到第三個模式*(3.7)*，這邊的交易量變數又跟第二個模式有所不同。在第二個模式中我們雖然把交易量減去平均交易量，包括高於平均的交易量以及低於平均的交易量，但是在*(3.7)*中，我們完全只取用高於平均的交易量，而將低於平均的交易量給屏除了，在這邊將之設定為零。也就是說，我們在這兩個模式的實證結果比較當中，將可以很清楚的看到超過平均的異常交易量所能夠帶來的解釋能力差異。而 ω_4 的係數符號在這邊也被預期顯著為正。

第三節 *Hodrick-Prescott* 分解*(HP* 分解*)*、自我相關與季節性效果調整

在這一節所使用的研究方法，主要是為了定義本文的交易量變數。就像報酬率的時間序列資料一樣，交易量的時間序列資料一樣有可能會存在隨機趨勢的資料非定態問題。而使用混和分配假設的前提，要加入交易量變數來解釋報酬率的波動程度，一定是在時間序列資料為定態之下才可行，所以必須要找出一個定態的交易量數列，才可運用後續的假設與模型去驗證報酬率與交易量的相關性。過去的文獻為了捕捉非定態數列中的隨機趨勢，使用過許多方法，除了最常見的方法-差分*(differencing)*以外，還有各式各樣的理論運用。但本文在這邊使用的則是 *Hodrick and Prescott(1997)*提出*的 HP 分解*。而在得到定態的交易量數列之後，本文又援引 *Box/Jenkins* 的自我相關整合移動平均*(ARIMA)*的時序模型架構，進一步捕捉交易量數列的季節性與自我相關性，並試圖以是否去除相關性的兩種不同交易量在實證結果中做比較。

一、*Hodrick-Prescott* 分解*(HP* 分解*)*

$$V_t = g_t + c_t \qquad for \quad t = 1,...,T$$

*HP 分解*的主要作法如下：如果在給定上式中的交易量時間序列V_t為非定態序列之下，就可以將其分解成非定態部分與定態部分。而這兩部分其實就代表了隨機趨勢項(g_t)與波動項(c_t)，而長期之下波動項的平均為零。非定態的序列資料也就表示了，在受到不同的衝擊時，會產生恆常的衝擊效果以及暫時的衝擊效果，使得此序列資料在一個隨機趨勢上下波動，而在恆常的衝擊效果發生時，正是本文想要從非定態序列中移除的隨機趨勢所導致的，如果可以用原始的序列資料減去從 *HP 分解*得到的隨機趨勢，那麼就可以找到一個定態的異常交易量時間序列，可供我們做為有效的交易量解釋變數。下式*(3.8)*表示我們想要找出隨機趨勢項(g_t)的大小，所以極小化交易量序列跟趨勢項序列間的平方差，且對於過度波動的隨機趨勢項做出懲罰，因此在給定的平滑係數之下，去控制其波動程度。

而這樣設定的λ係數(λ=14400)控制了隨機趨勢項的平滑程度,也就是說,當λ越大，隨機趨勢項也就越平滑，而一般來說，只要透過這一個式子，解出其一階條件的值，我們就可以獲取隨機趨勢的估計值。

$$\underset{\{g_t\}_{t=-1}^{T}}{Min}\left\{\sum_{t=1}^{T}c_t^2+\lambda\sum_{t=1}^{T}\left[\left(g_t-g_{t-1}\right)-\left(g_{t-1}-g_{t-2}\right)\right]^2\right\} \tag{3.8}$$

但是，一旦資料的期間過長，使得資料筆數過多，直接利用 *HP 分解*必須要使用到矩陣的轉化，太高階的的矩陣轉化不僅難以計算、相當耗時，而且容易產生計算上的困難或是數值上的誤差問題，所以，*Hodrick and Prescott(1997)*建議使用卡爾曼濾波(*Kalman Filtering)*的方法來做後續的運算處理，而為了這樣的方法使用，必須要設定一些基本的假設以及變數的基本轉換，以符合卡爾曼濾波使用時所需要的狀態空間模型(*State Space Model)*型態。

$$c_t \overset{iid}{\sim} N\left(0,\sigma_1^2\right)$$

$$\left(g_t-g_{t-1}\right)-\left(g_{t-1}-g_{t-2}\right)\overset{iid}{\sim} N\left(0,\sigma_2^2\right)$$

首先，我們根據過去學者的做法及前面對波動項的假設，做了上式兩個常態分配的假設。上面的兩個式子說明了波動項和隨機趨勢序列的二階差分項在長期之下會服從均數為零的常態分配。另外，為了轉換成狀態空間模型，再令：

$$x_t=V_t-V_{t-1}$$

$$s_t=g_t-g_{t-1}$$

x_t是V_t的一階差分，而s_t則是g_t的一階差分，這樣一來就可以得到下面的狀態空間模型：

$$x_t=s_t+e_t\ ,\qquad e_t=c_t-c_{t-1}\sim N\left(0,2\sigma_1^2\right) \tag{3.9}$$

$$s_t=s_{t-1}+\eta_t\ ,\qquad \eta_t\sim N\left(0,\sigma_2^2\right) \tag{3.10}$$

*(3.9)*式中表現的就是觀察方程式*(observation equation)*，而*(3.10)*式中表現的則是狀態方程式*(state equation)*，透過狀態空間模型的建立，利用遞歸的方式不斷的更新，這樣的好處是不需要利用到太高階的矩陣計算，在電腦軟體的使用上也較

為便利可行。但在求解 *HP* 分解的時候，我們必須利用交易量的先行處理 ($x_t = V_t - V_{t-1}$)，才能使用統計軟體求得狀態值(*smoothed state*, s_t)，接著再利用相同的方法轉回我們想要的隨機趨勢序列(g_t)，最後取原始對數交易量跟此趨勢序列的差為本文所訂的異常交易量，也就是$u_{t,1}$為我們的交易量變數。$u_{t,1}$就是經過 *HP* 分解過後所得到的異常交易量變數。

二、自我相關與季節性效果調整

$$\Phi\left(B^s\right)\phi\left(B\right)\nabla_s^D\nabla^d x_t = \Theta\left(B^s\right)\theta\left(B\right)\varepsilon_t \tag{3.11}$$

解決了隨機趨勢的問題，接下來要克服的就是交易量資料的自我相關性與季節性。在同時考慮這兩個問題之下，我們引用文獻常用的 $ARIMA\left((1,0,1)\times(1,0,1)_5\right)$。這個模型的好處在於，使用的自我相關與移動平均期數都只有一期，不必利用過多的變數來加以解釋，而事實上，經過我們實際測試更高階的期數，對實證結果並沒有太大的影響，在經過取捨後，還是使用較低的期數；另外，在季節性的考量之下，我們使用的是一週五天的季節效果，同時，為了觀察我們所設定的週季節效果所帶來的影響，我們並未多加考慮月的季節效果，而上一小段獲得的異常交易量變數$u_{t,1}$，在配適過這邊所用的 *ARIMA* 模型後，我們得到無相關的異常交易量變數$u_{t,2}$作為我們的第二個交易量變數。$u_{t,2}$就是經過自我相關、季節性因子等調整的異常交易量變數。