

出國報告（出國類別：其他）

參加 SEACEN 研訓中心與日本央行
合辦之「模型建構與預測方法」
中級課程研習報告書

服務機關：中央銀行

姓名職稱：朱浩榜/經濟研究處四等專員

派赴國家：柬埔寨

出國期間：107 年 6 月 30 日至 7 月 7 日

報告日期：107 年 10 月

目 錄

壹、前言.....	1
貳、狀態空間模型	2
一、模型簡介.....	2
二、Kalman Filter	3
三、應用範例：估計台灣的月頻 GDP	5
參、貝氏計量方法	8
一、貝氏方法的概念.....	8
二、貝氏估計：線性迴歸模型	9
三、貝氏 VAR.....	16
四、應用範例：預測台灣的 CPI 年增率	24
肆、結論與建議	27
參考資料.....	28

參加 SEACEN 研訓中心與日本央行合辦之 「模型建構與預測方法」中級課程研習報告書

壹、前言

東南亞國家中央銀行聯合會（South East Asian Central Banks, SEACEN）所屬研訓中心與日本央行（Bank of Japan）於 2018 年 7 月在柬埔寨暹粒合辦「模型建構與預測方法」中級課程（SEACEN-BOJ Course on Intermediate Modelling and Forecasting Techniques），課程內容涵蓋計量模型建構、估計方法等。

本次研習參加學員共 24 人，包括本行及來自柬埔寨、印度、印尼、馬來西亞、尼泊爾、菲律賓、斯里蘭卡、泰國等共 9 國中央銀行。課程講師均來自 SEACEN 研訓中心，講授常見於實務應用的計量方法，包括狀態空間模型（state space model）、貝氏（Bayesian）估計，以及向量自我迴歸（vector autoregression, VAR）模型等；並輔以統計軟體之操作。課程另邀請日本央行與亞洲開發銀行（Asian Development Bank）的官員，分別說明日本央行使用的經濟分析及預測模型，以及亞洲開發銀行的即時預報（nowcast）模型等，並分享相關經驗。

本報告共分為 4 個部分。除前言外，第 2 部分簡介狀態空間模型與「Kalman filter」之概念，並以月頻實質國內生產毛額（gross domestic product, GDP）之估計為例，說明狀態空間模型之應用。第 3 部分則簡介貝氏估計，以及如何將貝氏方法應用於線性迴歸、VAR 等計量模型，再以貝氏 VAR 模型預測台灣的消費者物價指數（consumer price index, CPI）年增率，作為貝氏估計的應用範例。最後，第 4 部分為結論與建議。

貳、狀態空間模型

在經濟理論或模型中，往往有許多無法觀察到的變數或概念，如潛在產出、自然失業率、中性利率等。而應用狀態空間模型，將有助估計連續（continuous）且不可觀察（unobservable）的變數。此外，狀態空間模型亦可用於處理參數隨時間變動（time-varying）、資料漏失（missing data），或混合頻率資料（mixed-frequency data）等問題¹。

一、模型簡介

狀態空間模型由「測量方程式」（measurement equation）與「狀態方程式」（state equation）等兩部分組成。測量方程式描繪可觀察到的變數、不可觀察的狀態變數（state variable）、其他解釋變數與測量誤差（measurement error）之間的線性關係：

$$y_t = H_t \beta_t + A_t Z_t + e_t. \quad (2.1.1)$$

其中， y_t 為維度 $n \times 1$ 且可觀察到的變數向量， Z_t 為維度 $n \times k$ 的外生變數矩陣； H_t 、 A_t 分別為維度 $n \times m$ 、 $n \times n$ 的系統矩陣（system matrix），通常為稀疏矩陣（sparse matrix）； e_t 則為維度 $n \times 1$ 的誤差項向量。

而狀態方程式則描繪狀態變數本身的動態過程： (2.1.2)

$$\beta_t = \mu_t + F_t \beta_{t-1} + v_t.$$

其中， F_t 為維度 $m \times m$ 的移轉矩陣（transition matrix）， μ_t 為維度 $m \times 1$ 的向量， v_t 為維度 $n \times 1$ 的誤差項向量。同樣地， F_t 為系統矩陣，且通常為稀疏矩陣。

從(2.1.1)式與(2.1.2)式可知，狀態空間模型含有兩個誤差項。

¹ 本節內容主要參考 SEACEN 研習講義（2018）。

可進一步假設其獨立且分配一致 (independent and identically distributed, i.i.d.)、符合常態分配：

$$e_t \sim_{i.i.d.} \mathcal{N}(0, R_t),$$

$$v_t \sim_{i.i.d.} \mathcal{N}(0, Q_t),$$

且 $E(e_t v_t) = 0$ 。此外，在大部分的情況下，通常可再假設 e_t 與 v_t 彼此獨立，但此假設並非一定要成立。

狀態空間模型的目的，即在給定觀察值 y_t (以及 Z_t) 的情況下，估計 H_t 、 A_t 、 μ_t 、 F_t 、 R_t 、 Q_t 等參數，以及不可觀察的狀態變數 β_t 。

二、Kalman Filter

若欲求解狀態空間模型，則需使用 Kalman filter。其為遞迴式 (recursive) 的演算法 (algorithm)，可反覆透過「預測」(prediction) 及「更新」(updating) 等兩步驟進行估計。為便於說明，首先假設模型不含其他外生變數 ($Z_t = 0$)，且 H_t 、 A_t 、 μ 、 F_t 、 R_t 、 Q_t 等參數為已知，則 Kalman filter 的具體做法如下：

1. 給定狀態變數 (β_t) 與其變異數-共變異數矩陣 (P_t) 的起始值 (initial value)。若符合定態 (stationary) 條件²，則可得

$$\beta_{0|0} = (1 - F)^{-1} \mu,$$

$$P_{0|0} = (1 - F \otimes F)^{-1} \text{vec}(Q).$$

其中， vec 為向量化符號。

2. 預測：使用在 $t = t - 1$ 時擁有的全部資訊，推得 $t = t$ 時的狀態變數

$$\beta_{t|t-1} = \mu + F_t \beta_{t-1|t-1},$$

² 若不為定態，則 $\beta_{0|0}$ 可為任意給定 (arbitrary)，而 $P_{0|0}$ 可為某一較大的數值。

以及其變異數-共變異數矩陣

$$P_{t|t-1} = \text{Var}(\beta_{t|t-1}) = F P_{t-1|t-1} F' + Q.$$

此時，可得到可觀察變數的預測值 $y_{t|t-1}$ ，以及其預測誤差

$$\eta_{t|t-1} = y_t - y_{t|t-1} = y_t - H_t \beta_{t|t-1}.$$

而預測誤差之變異數則為

$$f_{t|t-1} = \text{Var}(y_t - y_{t|t-1}) = H_t P_{t|t-1} H_t' + R.$$

3. 更新：利用預測誤差蘊含的資訊更新狀態變數及其變異數-共變異數矩陣，可得

$$\beta_{t|t} = \beta_{t|t-1} + K_t \eta_{t|t-1},$$

$$P_{t|t} = P_{t|t-1} - K_t H_t P_{t|t-1} = (1 - K_t H_t) P_{t|t-1},$$

其中

$$K_t = P_{t|t-1} H_t' (f_{t|t-1})^{-1}$$

稱為「Kalman 增幅」(Kalman gain)，可視為新的資訊所占的權重。

4. 重複上述「預測」及「更新」的步驟，直到最後一個樣本點 $t = T$ 。

值得注意的是，在上述步驟中，已假設 H_t 、 A_t 、 μ 、 F_t 、 R_t 、 Q_t 等參數為已知；惟在大部分的情況下，該參數均為未知。儘管如此，由於模型假設誤差項為常態分配，故可知樣本資料的分配為

$$y_t | H_t \sim \mathcal{N}(H_t \beta_{t|t-1}, H_t P_{t|t-1} H_t' + R),$$

因此，藉由對數概似函數 (log-likelihood function)

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{t=1}^T \ln(2\pi^n [\det(H_t P_{t|t-1} H_t' + R)])$$

$$= -\frac{1}{2} \sum_{t=1}^T (y_t - H_t \beta_{t|t-1})' (H_t P_{t|t-1} H_t' + R)^{-1} (y_t - H_t \beta_{t|t-1}),$$

即可以最大概似估計法 (maximum likelihood estimation, MLE) 進行估計。此外，對數概似函數亦可以預測誤差的形式表示³：

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{t=1}^T \ln(2\pi^n \det f_{t|t-1}) - \frac{1}{2} \eta_{t|t-1}' f_{t|t-1}^{-1} \eta_{t|t-1}.$$

最後，上述 Kalman filter 每次遞迴均應用新增的資訊進行估計。若直接使用全樣本的資訊估計，則為「Kalman smoother」，估計結果將較為平滑，詳見 Harvey (1993) 之介紹。

三、應用範例：估計台灣的月頻 GDP

在總體經濟變數中，GDP 係衡量一定期間內，某一國家或地區常住居民的生產成果。若與其過往數值相比，亦可衡量經濟成長，廣受各界關注。然而，GDP 通常係按季發布，資料頻率相較大部分的總體變數為低。因此，在實際應用上，使用依時拆分 (temporally disaggregate) 方法將季頻的 GDP 資料插補 (interpolate) 為月頻資料，係為常見的做法。其中，Harvey and Pierse (1984) 將欲插補的高頻資料視為無法直接觀察到的變數，再透過高頻資料與可觀察到的低頻資料之間的關係，以狀態空間模型進行估計。

在國內文獻方面，劉瑞文 (2007) 以 9 種不同的統計方法拆分台灣工業部門的實質 GDP，且發現 Chow and Lin (1971) 方法的拆分結果與製造業生產指數最為吻合；Huang (2010) 則以狀態空間模型拆分台灣的季調後實質 GDP，並發現其樣本內 (in-sample) 配適與樣本外 (out-of-sample) 預測均表現甚佳。

³ 此即「預測誤差分解」(prediction error decomposition) (Schweppe, 1965)。

本報告將以簡單的狀態空間模型估計台灣月頻率之實質 GDP。首先，假設 Y_t 為主計總處公布之季頻 GDP， Y_t^* 為無法直接觀察之月頻 GDP，則季頻 GDP 為當季各月 GDP 之加總： $Y_t = \sum_{i=0}^2 Y_{t-i}^*$ 且 $t = 3, 6, 9, \dots, 3T$ 。其次，假設 Y_t^* 受其前期項 Y_{t-1}^* 影響，且與月頻參考指標 x_t 相關，並令其誤差項 u_t 具有一階自我相關（如 Chow and Lin, 1971）： $u_t = \rho u_{t-1} + w_t$ 且 $\text{Var}(w_t) = \sigma^2$ 。

根據上述假設，(2.1.1) 式之測量方程式可表為

$$Y_t = h_t' \begin{bmatrix} Y_t^* \\ Y_{t-1}^* \\ Y_{t-2}^* \\ u_t \end{bmatrix} = [1 \quad 1 \quad 1 \quad 0] \begin{bmatrix} Y_t^* \\ Y_{t-1}^* \\ Y_{t-2}^* \\ u_t \end{bmatrix}, \quad t = 3, 6, 9, \dots, 3T,$$

且若 $t \neq 3, 6, 9, \dots, 3T$ ，則 $h_t = 0$ ；而 (2.1.2) 式之狀態方程式則為

$$\begin{bmatrix} Y_t^* \\ Y_{t-1}^* \\ Y_{t-2}^* \\ u_t \end{bmatrix} = \begin{bmatrix} x_t b \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \theta & 0 & 0 & \rho \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} Y_{t-1}^* \\ Y_{t-2}^* \\ Y_{t-3}^* \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ 0 \\ w_t \end{bmatrix}, \quad \text{Var} \left(\begin{bmatrix} w_t \\ 0 \\ 0 \\ w_t \end{bmatrix} \right) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}.$$

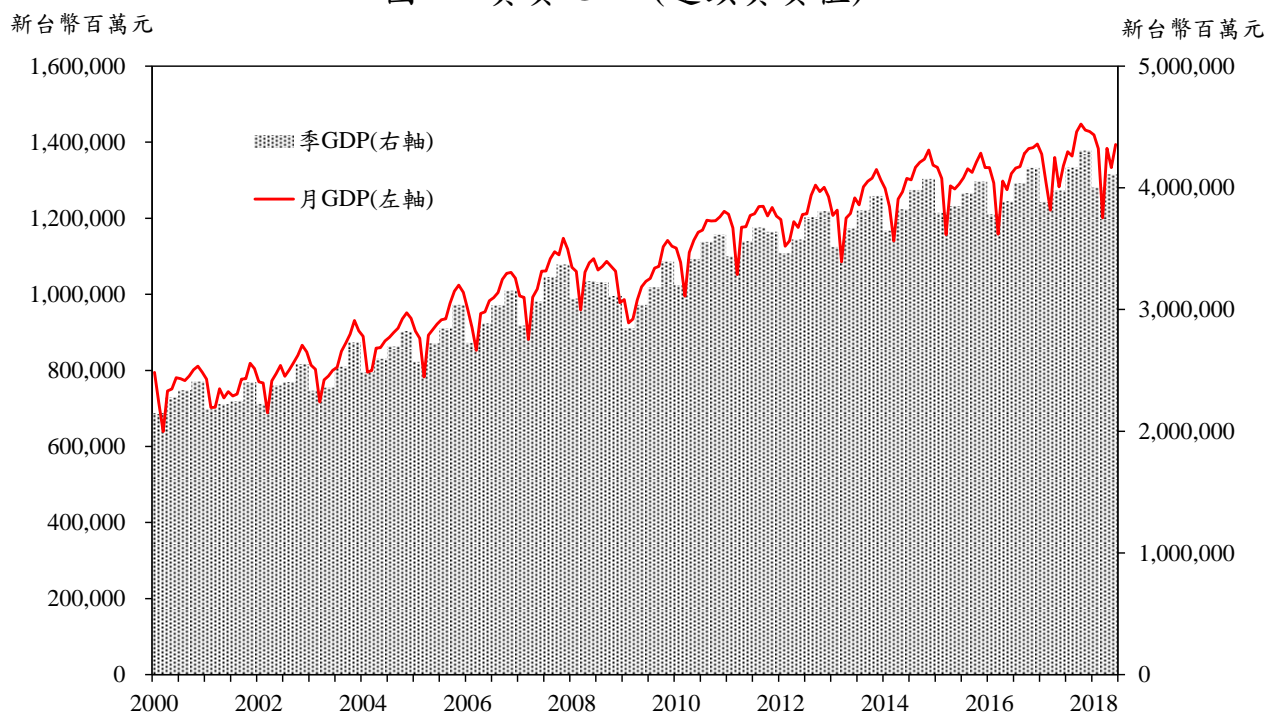
至於月頻參考指標 x_t ，此處選擇以工業生產指數作為月頻 GDP 的參考指標，其資料來源為經濟部統計處。

資料期間涵蓋 2000 年第 1 季至 2018 年第 2 季，且因使用全樣本資訊，故採「Kalman smoother」進行估計。圖 1 繪示月頻 GDP 的估計結果，另亦繪示季頻 GDP，以利比較。顯然，狀態空間模型估得之月頻 GDP 與季頻 GDP 走勢極為接近。接著，圖 2 繪示月頻 GDP 與季頻 GDP 的年增率，可知月頻 GDP 年增率與季 GDP 年增率走勢亦相仿，惟月頻 GDP 年增率波動較大，且易受農曆春節等因素影響。

值得注意的是，此處的模型設定較為簡便，實際上或仍有進一步改善的空間。關於狀態空間模型於依時拆分方法之應用，可參見前述

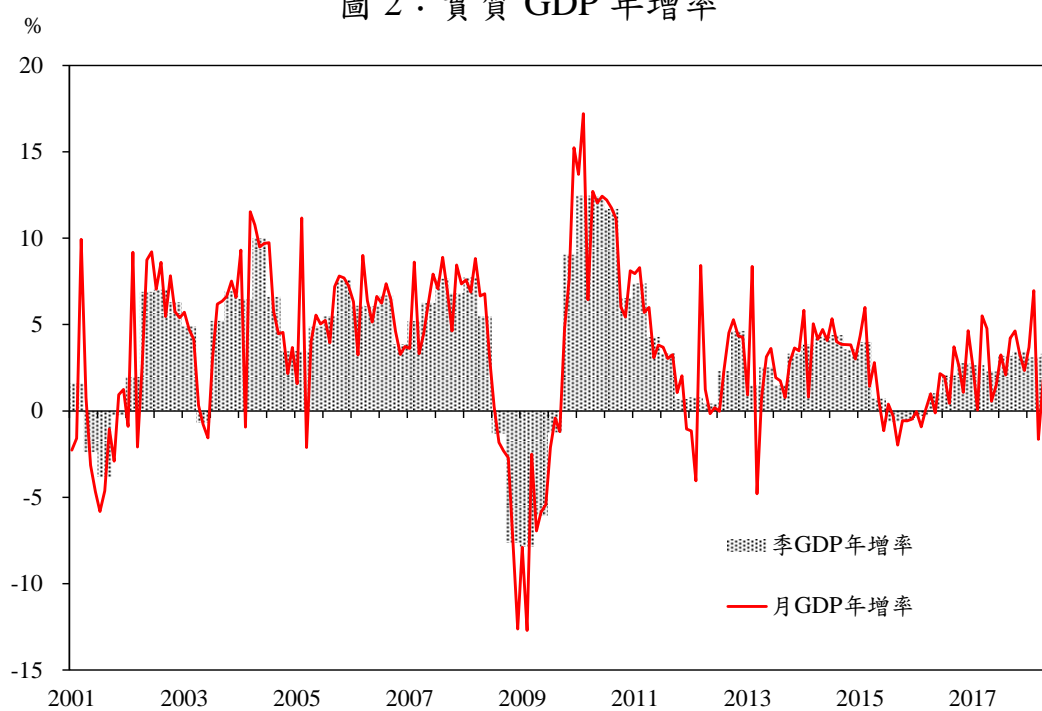
之 Harvey and Pierse (1984) 或 Huang (2010) 等；而若欲拆分多變量之低頻變數，則可參見 Moauro and Savio (2005)。

圖 1：實質 GDP (連鎖實質值)



說明：連鎖實質值之參考年為 2011 年。

圖 2：實質 GDP 年增率



參、貝氏計量方法

一般而言，資訊愈多將愈有助於估計。除了樣本資料所能提供的資訊外，也有一些非樣本（non-sample）資訊，如經濟理論、過去文獻的研究成果等。貝氏方法的目的即是將這些非樣本資訊應用於計量分析中，協助估計或統計推論，通常用於模型參數眾多、過於複雜或求解不易之情況⁴。

一、貝氏方法的概念

傳統的計量方法視樣本資料來自某一資料產生過程（data generating process, DGP），研究者可利用樣本資料推估 DGP 的參數，且資料樣本具有隨機（stochastic）性質，但 DGP 的參數則否。貝氏方法不然，其視待估計參數為隨機、但樣本資料則否，須結合研究者對參數本身的主觀信念（beliefs）加以估計。此外，參數的統計推論係基於樣本資料，DGP 較不重要。事實上，在貝氏估計中，待估計參數的「真實值」（true value）既不存在、也不重要。

具體而言，研究者會先給予待估計參數某些先驗（prior）上的機率分配；接著，再加入樣本資料的訊息加以求解（可藉模型的概似函數捕捉），以求得參數的事後（posterior）分配。事後分配已涵蓋研究者在觀察樣本資料後，得知的所有資訊。說明如下。

首先，根據貝氏定理，在給定B的情況下，A的條件機率為

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \quad (3.1.1)$$

由於 $\mathbb{P}(A|B)$ 是在已知B的條件下之A的機率，故為A的事後機率；相對地， $\mathbb{P}(A)$ 為A的先驗機率，因其未考慮任何B的因素。

⁴ 本節內容主要參考 SEACEN 研習講義（2018）。

其次，假設 θ 為待估計參數， y 為產生自聯合（joint）密度函數 $f(y|\theta)$ 的樣本觀察值，則 $f(y|\theta)$ 等同於 θ 的概似函數 $\mathcal{L}(\theta|y)$ ，其包含所有蘊含在 y 中，關於 θ 的資訊。由於研究者的主觀機率分配反映在 θ 上，故 θ 為隨機變數，而 $f(y|\theta)$ 則為給定 θ 的條件下， y 的條件密度函數。據此，即可將（3.1.1）式改寫為

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)}, \quad (3.1.2)$$

其中， $f(\theta|y)$ 即為給定樣本觀察值 y 下， θ 的事後分配； $f(y|\theta)$ 可表示成給定樣本觀察值 y 下， θ 的概似函數 $\mathcal{L}(\theta|y)$ ； $f(\theta)$ 為 θ 的先驗分配；而 $f(y) = \int f(y|\theta)f(\theta)d\theta$ 為邊際（marginal）概似函數，因與 θ 無關，故視為常數，可將事後分配機率值標準化為1。（3.1.2）式可再簡化：

$$f(\theta|y) \propto \mathcal{L}(\theta|y)f(\theta), \quad (3.1.3)$$

表示「事後分配」與「樣本資訊（概似函數）乘以先驗分配」成正比。亦即，研究者可將其對於 θ 的主觀認知反映在先驗分配 $f(\theta)$ ，並藉由樣本觀察值 y 的資訊，予以更新及修正，再得到事後分配 $f(\theta|y)$ 。

先驗分配 $f(\theta)$ 應符合 θ 本身應有的特性，如若 θ 為變異數，則 $f(\theta)$ 不應為標準常態分配，因為 θ 不可能為負值。而若先驗資訊較為缺乏，研究者為減少先驗資訊對事後分配的干擾程度，則可採用較為「平坦」（flat）的先驗分配⁵。此外，為確保事後分配在先驗分配加入概似函數的訊息後，仍與先驗分配屬於同樣類型的機率分配，通常會假設「共軛先驗」（conjugate prior）。

二、貝氏估計：線性迴歸模型

此處將以線性迴歸模型為範例，說明如何以貝氏方法進行估計。

⁵ 即先驗資訊精確性低、變異數大，或稱「分散」（diffuse）、「未含資訊」（non-informative）等。

假設線性迴歸模型如下：

$$y = X\beta + \varepsilon,$$

其中， y 為維度 $T \times 1$ 的被解釋變數向量、 X 為維度 $T \times k$ 的解釋變數矩陣； β 為維度 $k \times 1$ 的參數向量； ε 為維度 $T \times 1$ 的模型誤差項，且 $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ 。

在傳統計量模型中，係利用最小平方法（ordinary Least squares, OLS）或MLE，直接以 y 、 X 等資料估計未知參數 β 、 σ^2 。貝氏估計則不然，以下將分別假設 β 未知、 σ^2 已知，以及 β 與 σ^2 均未知等兩種情形，說明如何以貝氏方法估計未知參數。

（一） β 未知、 σ^2 已知

假設 σ^2 為已知，僅有 β 為待估計參數。首先，可給定 β 的先驗分配為

$$\beta \sim \mathcal{N}(\beta_0, \Omega_0),$$

其中 β_0 、 Ω_0 為已知， β_0 代表研究者對於 β 的主觀信念， Ω_0 則反映主觀信念的強烈程度。若 Ω_0 越大，則表示研究者對於 β_0 越不確定（即越「平坦」）；若 Ω_0 越小，則表示研究者對於 β_0 越有信心。因此， β 先驗分配之機率密度函數為

$$\begin{aligned} f(\beta) &= (2\pi)^{-\frac{k}{2}} |\Omega_0|^{-\frac{1}{2}} \exp \left[-\frac{(\beta - \beta_0)' \Omega_0^{-1} (\beta - \beta_0)}{2} \right] \\ &\propto \exp \left[-\frac{(\beta - \beta_0)' \Omega_0^{-1} (\beta - \beta_0)}{2} \right]. \end{aligned} \quad (3.2.1)$$

其次，從 y 、 X 等實際資料得到線性迴歸模型的概似函數：

$$\mathcal{L}(\beta|y) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[-\frac{(y - X\beta)' (y - X\beta)}{2\sigma^2} \right]. \quad (3.2.2)$$

由於 (3.2.2) 式等同於 MLE 的做法，故可得到 MLE 估計值（亦等同於 OLS 估計式）： $\hat{\beta} = (X'X)^{-1}X'y$ 與 $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ 。再以

$$(y - X\beta)'(y - X\beta) = (y - X\hat{\beta})'(y - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})$$

代入 (3.2.2) 式⁶，則可得

$$\begin{aligned} \mathcal{L}(\beta|y) &= (2\pi\sigma^2)^{-\frac{T}{2}} \exp\left[-\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{2\sigma^2}\right] \times \exp\left[-\frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{2\sigma^2}\right] \\ &\propto \exp\left[-\frac{(\beta - \hat{\beta})'(\sigma^2(X'X)^{-1})^{-1}(\beta - \hat{\beta})}{2}\right], \end{aligned} \quad (3.2.3)$$

亦即，概似函數與多變量常態分配 $\mathcal{N}(\hat{\beta}, \hat{\Omega})$ 成正比。

最後，根據 (3.1.3) 式，可知 β 之事後分配 $f(\beta|y) \propto \mathcal{L}(\beta|y)f(\beta)$ ；因此，代入 (3.2.3) 式之概似函數 $\mathcal{L}(\theta|y)$ 與 (3.2.1) 式之先驗分配 $f(\theta)$ ，即可得事後分配

$$f(\beta|y) \propto \mathcal{L}(\beta|y)f(\beta)$$

$$\propto \exp\left[-\frac{(\beta - \hat{\beta})'(\sigma^2(X'X)^{-1})^{-1}(\beta - \hat{\beta})}{2}\right] \times \exp\left[-\frac{(\beta - \beta_0)'\Omega_0^{-1}(\beta - \beta_0)}{2}\right].$$

由於共軛先驗與概似函數均屬常態，故事後分配亦屬常態。因此，可推得事後分配⁷：

$$f(\beta|y) \sim \mathcal{N}(\beta_{\text{post}}, \Omega_{\text{post}}).$$

其中，

$$\beta_{\text{post}} = [\Omega_0^{-1} + (\sigma^2(X'X)^{-1})^{-1}]^{-1}[\Omega_0^{-1}\beta_0 + (\sigma^2(X'X)^{-1})^{-1}\hat{\beta}], \quad (3.2.4)$$

$$\Omega_{\text{post}} = [\Omega_0^{-1} + (\sigma^2(X'X)^{-1})^{-1}]^{-1}. \quad (3.2.5)$$

⁶ 參見 Judge et al. (1988)。

⁷ 參見 Judge et al. (1988)。

可知參數 β 的事後分配係結合資料(X)與先驗資訊(β_0 、 Ω_0)而得。

從(3.2.4)式可知，事後分配的條件平均值 β_{post} 為先驗平均 β_0 與OLS估計式 $\hat{\beta}$ 之加權平均。此時，若使用較為「分散」或「未含資訊」的先驗資訊($\Omega_0 \rightarrow \infty$ ，即 $\Omega_0^{-1} \rightarrow 0$)，則 β_{post} 將越為接近OLS估計式 $\hat{\beta}$ ；亦即， β_{post} 將完全由資料決定。反之，若 Ω_0 越小，則先驗資訊的比重越大。同樣地，(3.2.5)式之 Ω_{post} 亦然。若沒有任何樣本資訊，則事後分配將等同於先驗分配。

(二) β 與 σ^2 均為未知

接著，探討 β 與 σ^2 均為待估計參數的情形。同樣地，可給定 β 的先驗分配為常態分配

$$\beta \sim \mathcal{N}(\beta_0, \Omega_0).$$

由於 σ^2 亦為未知，故亦須給定先驗分配。而 σ^2 為變異數，不得為負值。為方便起見，可視 σ^2 的倒數 σ^{-2} 為「準確度」(precision)，並給定 σ^{-2} 的先驗分配為「Gamma分配」(Gamma distribution)⁸：

$$\sigma^{-2} \sim \mathcal{G}\left(\frac{T_0}{2}, \frac{v_0}{2}\right).$$

若 σ^{-2} 的先驗資訊係來自某一樣本群體，則 T_0 為該樣本群體的大小，而 v_0 為該樣本群體的殘差平方和(sum of squared residual)。因此，可得到 σ^{-2} 的先驗機率密度函數為

$$f(\sigma^{-2}) = \frac{\left(\frac{v_0}{2}\right)^{\frac{T_0}{2}} (\sigma^{-2})^{\frac{T_0}{2}-1} \exp\left(-\frac{v_0}{2} \sigma^{-2}\right)}{\Gamma\left(\frac{T_0}{2}\right)} \propto (\sigma^{-2})^{\frac{T_0}{2}-1} \exp\left(-\frac{v_0}{2} \sigma^{-2}\right), \quad (3.2.6)$$

⁸ 若Y服從Gamma分配，即 $Y \sim \mathcal{G}(a, b)$ ， $a > 0$ 、 $b > 0$ ，則其機率分配函數為 $\Gamma(a)^{-1} b^a Y^{a-1} \exp(-bY)$ ，($\Gamma(\cdot)$ 為Gamma函數)，期望值 $E(Y) = a/b$ ，變異數 $\text{Var}(Y) = a/b^2$ 。不同的 a 會使機率分配產生不同的型態(shape)，故 a 稱為「型態參數」(shape parameter)；不同的 b 會改變機率分配的尺度(scale)，故 b 稱為「尺度參數」(scale parameter)。

其中 $\Gamma(\cdot)$ 為 Gamma 函數⁹。

若假設 β 與 σ^2 的先驗分配不互相獨立 (dependent)¹⁰，則先驗資訊將為常態分配與 Gamma 分配的聯合分配

$$f(\beta, \sigma^{-2}) = f(\beta|\sigma^{-2})f(\sigma^{-2}),$$

其中 $f(\beta|\sigma^{-2}) \sim \mathcal{N}(\beta_0, \sigma^2 \Omega_0)$ 。再從 (3.2.1) 式，可知

$$f(\beta|\sigma^{-2}) \propto \sigma^{-k} \exp\left[-\frac{(\beta - \beta_0)'(\sigma^2 \Omega_0)^{-1}(\beta - \beta_0)}{2}\right]. \quad (3.2.7)$$

其次，使用 y 、 X 等實際資料得到線性迴歸模型的概似函數：

$$\mathcal{L}(\beta, \sigma^{-2}|y) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp\left[-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right],$$

仿照 (3.2.2) 式至 (3.2.3) 式的做法，可進一步改寫為

$$\begin{aligned} \mathcal{L}(\beta, \sigma^{-2}|y) &= (2\pi\sigma^2)^{-\frac{T}{2}} \exp\left[-\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{2\sigma^2}\right] \\ &\times \exp\left[-\frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{2\sigma^2}\right]. \end{aligned} \quad (3.2.8)$$

最後，由於 β 與 σ^{-2} 的聯合事後分配為

$$f(\beta, \sigma^{-2}|y) \propto \mathcal{L}(\beta, \sigma^{-2}|y)f(\beta|\sigma^{-2})f(\sigma^{-2}),$$

將 (3.2.6) 式、(3.2.7) 式與 (3.2.8) 式分別代入，即可得

⁹ 或也可以直接給定 σ^2 的先驗分配為「反 Gamma 分配」(inverse-Gamma distribution)：

$$\sigma^2 \sim \mathcal{IG}\left(\frac{T_0}{2}, \frac{v_0}{2}\right).$$

¹⁰ 亦可假設 β 與 σ^2 的先驗分配互相獨立 (independent)，其後續過程仍大致相同，惟 β 與 σ^2 事後的邊際機率分配將沒有解析解。

$$\begin{aligned}
f(\beta, \sigma^{-2} | y) &\propto (\sigma^2)^{-\frac{T}{2}} \exp \left[-\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{2\sigma^2} \right] \\
&\times \exp \left[-\frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{2\sigma^2} \right] \\
&\times \sigma^{-k} \exp \left[-\frac{(\beta - \beta_0)'(\sigma^2\Omega_0)^{-1}(\beta - \beta_0)}{2} \right] \\
&\times (\sigma^{-2})^{\frac{T_0}{2}-1} \exp \left(-\frac{\nu_0}{2} \sigma^{-2} \right).
\end{aligned}$$

由於共軛先驗 $f(\beta, \sigma^{-2})$ 為常態分配與 Gamma 分配的聯合分配，故 β 與 σ^{-2} 的聯合事後分配亦為常態分配與 Gamma 分配的聯合分配，並包括 β_{post} 、 Ω_{post} 、 $\frac{T_{\text{post}}}{2}$ 、 $\frac{\nu_{\text{post}}}{2}$ 等參數：

$$(\beta, \sigma^{-2}) \sim \mathcal{NG} \left(\beta_{\text{post}}, \Omega_{\text{post}}, \frac{T_{\text{post}}}{2}, \frac{\nu_{\text{post}}}{2} \right). \quad (3.2.9)$$

其中，

$$\begin{aligned}
\beta_{\text{post}} &= [\Omega_0^{-1} + X'X]^{-1} [\Omega_0^{-1}\beta_0 + X'X\hat{\beta}], \\
\Omega_{\text{post}} &= [\Omega_0^{-1} + X'X]^{-1}, \\
T_{\text{post}} &= T_0 + T, \\
\nu_{\text{post}} &= \nu_0 + (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \beta_0)'[\Omega_0^{-1} + X'X]^{-1}(\hat{\beta} - \beta_0).
\end{aligned}$$

(三) Gibbs 抽樣法

經由 (3.2.9) 式得到聯合事後分配 $f(\beta, \sigma^{-2} | y)$ 後，可進一步計算 β 、 σ^{-2} 的邊際機率密度函數 $f(\beta | y) = \int f(\beta, \sigma^{-2} | y) d\sigma^{-2}$ 、 $f(\sigma^{-2} | y) = \int f(\beta, \sigma^{-2} | y) d\beta$ ，以進行 β 與 σ^{-2} 的統計推論。然而，邊際密度函數之積分往往求解不易，甚至沒有解析解 (analytical solution)。此時，可借助馬可夫鏈蒙地卡羅 (Markov chain Monte Carlo, MCMC) 方法，在條件分配下進行抽樣，模擬參數的邊際事後分配。Gibbs 抽樣法 (Gibbs Sampling) 為 MCMC 方法的一種，可在已知條

件分配的函數形式下，用於模擬聯合或邊際事後分配¹¹。

假設有 $\theta_1, \theta_2, \dots, \theta_k$ 等 k 個隨機變數， y 為樣本資料，而 $\theta_1, \theta_2, \dots, \theta_k$ 的聯合密度函數 $f(\theta_1, \theta_2, \dots, \theta_k | y)$ 未知，但已知條件分配 $f(\theta_1 | \theta_2, \dots, \theta_k, y), f(\theta_2 | \theta_1, \theta_3, \dots, \theta_k, y), \dots, f(\theta_k | \theta_1, \theta_2, \dots, \theta_{k-1}, y)$ 的函數形式，則 Gibbs 抽樣法的抽樣步驟如下：

1. 給定 $(\theta_1, \theta_2, \dots, \theta_k)$ 一組起始值

$$(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}).$$

2. 在給定 $(\theta_2, \dots, \theta_k)$ 現有數值的情況下，從 θ_1 的條件分配抽取 $\theta_1^{(1)}$ ：

$$\theta_1^{(1)} \sim f(\theta_1 | \theta_2^{(0)}, \dots, \theta_k^{(0)}, y).$$

3. 在給定 $(\theta_1, \theta_3, \dots, \theta_k)$ 現有數值下，從 θ_2 的條件分配抽取 $\theta_2^{(1)}$ ：

$$\theta_2^{(1)} \sim f(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, y).$$

4. 依上述步驟，繼續抽取 $\theta_3^{(1)}, \theta_4^{(1)}, \dots$ ，直到第 k 個變數 $\theta_k^{(1)}$ ：

$$\theta_k^{(1)} \sim f(\theta_k | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)}, y).$$

此時，Gibbs 抽樣法即完成一次的疊代 (iteration)，可得到一筆 $(\theta_1, \theta_2, \dots, \theta_k)$ 的抽樣結果 $(\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)})$ 。若不斷持續重複上述抽樣步驟，則可利用抽樣結果建構機率分配。當疊代次數趨近無限大，則從條件分配取得的抽樣結果，將以指數速度收斂至 $(\theta_1, \theta_2, \dots, \theta_k)$ 的聯合分配或邊際分配¹²。此外，由於每次的抽樣結果 $(\theta_1^{(n)}, \theta_2^{(n)}, \dots, \theta_k^{(n)})$ 僅與上一次抽樣結果 $(\theta_1^{(n-1)}, \theta_2^{(n-1)}, \dots, \theta_k^{(n-1)})$ 有關，故屬馬可夫鏈。

¹¹ 若條件分配的函數形式未知，則須採更為一般化的 Metropolis–Hastings 演算法模擬事後分配。事實上，Gibbs 抽樣法可視為 Metropolis–Hastings 演算法的一種特例。

¹² 參見 Casella and George (1992)。

實證上，若進行 M 次疊代，為確保任意給定的起始值不會影響模擬結果，通常僅會保留最後 H 次疊代獲得的抽樣結果，而將前 $(M - H)$ 筆抽樣結果予以捨棄 (burn-in)。保留下來的 H 筆 $\theta_1, \theta_2, \dots, \theta_k$ ，其直方圖 (histogram) 即可分別用於近似 $\theta_1, \theta_2, \dots, \theta_k$ 邊際密度函數的圖形。

最後，回到 (3.2.9) 式，可知 β 與 σ^{-2} 的事後分配分別為

$$f(\beta|\sigma^{-2}, y) \sim \mathcal{N}(\beta_{\text{post}}, \Omega_{\text{post}}), \quad (3.3.10)$$

$$f(\sigma^{-2}|\beta, y) \sim \mathcal{G}\left(\frac{T_{\text{post}}}{2}, \frac{\nu_{\text{post}}}{2}\right). \quad (3.3.11)$$

式中符號皆與 (3.2.9) 式相同。接著，藉由 Gibbs 抽樣法，從 (3.3.10) 式與 (3.3.11) 式等 2 個條件分配不斷抽樣，抽樣結果將會逐漸收斂至聯合密度函數 $f(\beta, \sigma^{-2}|y)$ ，此時即可據此進行 β 與 σ^{-2} 的統計推論。例如，若抽樣 N 次，則可得到抽樣結果 $(\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(N)})$ ，而 $\beta|y$ 的平均值的近似值即為

$$\mathbb{E}(\beta|y) = \frac{1}{N} \sum_{n=1}^N \beta^{(n)}.$$

三、貝氏 VAR

若將貝氏方法應用在 VAR 模型上，即為貝氏 VAR。由於 VAR 係估計許多不同變數彼此間的關係，故只要考慮的變數或落後期數較多，即可能面臨樣本自由度不足的問題。因此，若以貝氏方法對 VAR 模型的係數與變異數給予先驗假設，再加入實際資料的訊息，即可改善估計結果。

假設落後期數為 p 的 VAR(p)模型如下：

$$y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + \varepsilon_t, \quad (3.4.1)$$

其中， y_t 為維度 $n \times 1$ 的變數向量，且 $t = 1, \dots, T$ ； B_i 為維度 $n \times n$ 的係數矩陣，且 $i = 1, \dots, p$ ； c 為維度 $n \times 1$ 的截距項； ε_t 為維度 $n \times 1$ 的誤差項，且 $\varepsilon_t \sim i.i.d. \mathcal{N}(0, \Sigma)$ 、 $\Sigma = \mathbb{E}(\varepsilon_t \varepsilon_t')$ 為 ε_t 的變異數-共變異數矩陣。

接著，令

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}; Y = [y_1 \quad y_2 \quad \cdots \quad y_T]'; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix};$$

$$x_t = [1 \quad y'_{t-1} \quad \cdots \quad y'_{t-p}]; X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}; B = \begin{bmatrix} c \\ B_1 \\ \vdots \\ B_p \end{bmatrix}; b = \text{vec}(B),$$

以及 $k = np + 1$ ，為 VAR(p) 模型中每條方程式的係數個數。因此， y 與 ε 的維度皆為 $nT \times 1$ ，而 x_t 、 X 、 B 、 Y 的維度分別為 $1 \times k$ 、 $T \times k$ 、 $k \times n$ 、 $T \times n$ ， b 則為維度 $kn \times 1$ 的 VAR 係數向量。據此，即可進一步將 (3.4.1) 式堆疊為

$$y = (I_n \otimes X)b + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T). \quad (3.4.2)$$

應用 (3.4.2) 式，有助將 VAR(p) 的概似函數

$$\mathcal{L}(b, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr}[(Y - XB)' \Sigma^{-1} (Y - XB)] \right\}$$

分解為 b 與 Σ 條件密度函數之乘積¹³，其中 b 為常態分配：

$$f(b|\Sigma, y) \sim \mathcal{N}(\hat{b}, \Sigma \otimes (X'X)^{-1}).$$

而 Σ 則為「反 Wishart 分配」(inverse-Wishart distribution)¹⁴：

$$f(\Sigma|y) \sim \mathcal{IW}(T - k - n - 1, S).$$

¹³ 參見 Kadiyala and Karlsson (1997)。

¹⁴ 在貝氏估計中，若有一個多變量常態分配的隨機向量，則可用反 Wishart 分配作為其變異數-共變異數矩陣的共軛先驗（另如前述，若為單變量常態隨機變數，則可用反 Gamma 分配作為其變異數的共軛先驗，見註 8）。

其中， $\hat{b} = \text{vec}(\hat{B})$ 且 \hat{B} 為B的 OLS 估計式； $S = (Y - X\hat{B})'(Y - X\hat{B}) = E'E$ （維度為 $T \times n$ ），為 VAR 模型以 OLS 估計之殘差平方和的矩陣。

（一）Minnesota 先驗

在貝氏 VAR 中，「Minnesota 先驗」(Minnesota prior) 是常見的先驗分配之一，係由 Litterman (1980, 1986) 提出。其認為，總體經濟變數多具有隨機趨勢或隨機漫步 (random walk) 的性質；對變數本身而言，越接近當期的落後項含有越多當期變數的資訊；相較其他變數，變數本身的落後項含有較多有關當期變數本身的資訊。

首先，可給定 b 的先驗分配為

$$b|\Sigma \sim \mathcal{N}(b_0, V_0),$$

其中， $b_0 = \text{vec}(B_0)$ 且

$$B_0 = \begin{bmatrix} 0_{1 \times n} \\ I_n \\ 0_{(k-n-1) \times n} \end{bmatrix}, \quad (3.4.3)$$

亦即，各變數的 AR(1)係數為 1，其餘為 0； V_0 為對角矩陣，其對角線上之元素 $v_{ij,l}$ 為

$$v_{ij,l} = \begin{cases} \left(\frac{\lambda_1}{l\lambda_3}\right)^2, & i = j \\ \left(\frac{\lambda_1\lambda_2\sigma_i}{l\lambda_3\sigma_j}\right)^2, & i \neq j \\ (\sigma_i\lambda_4)^2, & \text{for the constant} \end{cases} \quad (3.4.4)$$

為係數 $b_{ij,l}$ 先驗分配的變異數； $l = 1, \dots, p$ ，為變數之落後期數； $i、j$ 分別代表 VAR 模型中，第 i 條方程式的被解釋變數、解釋變數，故 $\sigma_i、\sigma_j$ 分別為其誤差項之標準差¹⁵。 $\lambda_1、\lambda_2、\lambda_3、\lambda_4$ 為超參數 (hyperparameter)，

¹⁵ 例如， σ_1 即為第 1 條方程式誤差項 ε_{1t} 之標準差。

係由研究者給定，可簡化變異數-共變異數矩陣之設定，減少待估計參數。 λ_1 主要控制被解釋變數落後期係數之先驗分配的標準差， λ_1 越大分配將越平坦； λ_2 ($0 < \lambda_2 \leq 1$) 控制其他解釋變數的落後期係數之先驗分配的標準差，若 $\lambda_1 > \lambda_2$ ，則代表被解釋變數本身落後項的重要性較其他解釋變數為高； λ_3 ($\lambda_3 > 0$) 使落後越多期的係數越傾向為 0，並控制其趨向 0 速度；而 λ_4 則僅用於常數項。換句話說， λ_1 、 λ_2 係控制整體的緊度 (tightness) 以及變數間互相影響的程度， λ_3 則控制落後項衰退的速度。

因此，若以雙變量的 VAR(2) 模型為例，則可從 (3.4.3) 式、(3.4.4) 式得到

$$b_0 = [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]',$$

且

$$V_0 = \begin{bmatrix} (\sigma_1 \lambda_4)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \left(\frac{\lambda_1 \lambda_2 \sigma_1}{1^{\lambda_3} \sigma_2}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\frac{\lambda_1}{2^{\lambda_3}}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \left(\frac{\lambda_1 \lambda_2 \sigma_1}{2^{\lambda_3} \sigma_2}\right)^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (\sigma_2 \lambda_4)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_1 \lambda_2 \sigma_2}{1^{\lambda_3} \sigma_1}\right)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_1 \lambda_2 \sigma_2}{2^{\lambda_3} \sigma_1}\right)^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_1}{2^{\lambda_3}}\right)^2 \end{bmatrix}.$$

至於 Minnesota 先驗的超參數應如何設定，許多文獻上皆有相關探討。例如，Canova(2007)建議可設定 $\lambda_1 = 0.2$ 、 $\lambda_2 = 0.5$ 、 $\lambda_3 = 1$ 或2、 $\lambda_4 = 10^5$ 。

其次，關於 Σ ，可令其為已知以簡化計算，並以 VAR 模型的變異數-共變異數矩陣估計式 $\hat{\Sigma}$ 代替（有時甚至簡化至對角矩陣），故無須假設其先驗分配。

最後，在給定 b 的先驗分配後，即可再加入樣本資料的訊息，得到 b 的事後分配

$$g(b|\hat{\Sigma}, y) \sim \mathcal{N}(b_{\text{post}}, V_{\text{post}}),$$

其中，

$$\begin{aligned} V_{\text{post}} &= \left(V_0^{-1} + (\hat{\Sigma}^{-1} \otimes X'X) \right)^{-1}, \\ b_{\text{post}} &= V_{\text{post}} \left(V_0^{-1} b_0 + (\hat{\Sigma}^{-1} \otimes X)'Y \right) \\ &= \left(V_0^{-1} + (\hat{\Sigma}^{-1} \otimes X'X) \right)^{-1} \left(V_0^{-1} b_0 + (\hat{\Sigma}^{-1} \otimes X)'Y \right). \end{aligned}$$

若無任何先驗資訊（ $V_0 = 0$ 且 $b_0 = 0$ ），則 $V_{\text{post}} = (X'X)^{-1} \otimes \hat{\Sigma}$ 且 $b_{\text{post}} = (X'X)^{-1}(X'Y)$ ，即與 OLS 相同。此外，由於 Σ 係以殘差的變異數-共變異數矩陣 $\hat{\Sigma}$ 代替，未必能代表參數的不確定性，故亦可使用 Gibbs 抽樣法，從反 Wishart 分配中抽樣並模擬 Σ 。

(二) Normal-Wishart 先驗

若 Σ 為未知，則除了 b 外，亦須給定 Σ 的先驗分配。對常態分配的資料而言，其自然共軛先驗為 Normal-Wishart 分配：

$$f(b|\Sigma) \sim \mathcal{N}(b_0, \Sigma \otimes V_0), \quad f(\Sigma) \sim \mathcal{IW}(v_0, S_0). \quad (3.4.5)$$

自然共軛代表先驗分配、樣本的概似函數、事後分配皆來自同一個分配族 (distribution family)，在加入樣本資料的訊息後，可得到 b 與 Σ 的事後分配為

$$g(b|\Sigma, y) \sim \mathcal{N}(b_{\text{post}}, \Sigma \otimes V_{\text{post}}), \quad g(\Sigma|y) \sim \mathcal{IW}(v_{\text{post}}, S_{\text{post}}).$$

其中，

$$V_{\text{post}} = (V_0^{-1} + X'X)^{-1},$$

$$b_{\text{post}} = \text{vec}[V_{\text{post}}(V_0^{-1}B_0 + X'X\hat{B})],$$

$$v_{\text{post}} = v_0 + T,$$

$$S_{\text{post}} = S + S_0 - \hat{B}X'X\hat{B} + B_0'V_0^{-1}B_0 - B_{\text{post}}'(V_0^{-1} + X'X)B_{\text{post}}.$$

接著，有關 VAR 模型係數 b 的統計推論，則可藉積分的方式推得 b 的邊際事後分配；該分配將為多變量 t -分配，其平均數為 \bar{b} 、自由度為 v_{post} ，且變異數-共變異數矩陣為

$$\text{Var}(b|y) = \frac{1}{v_{\text{post}} - n - 1} (S_{\text{post}} \otimes V_{\text{post}}).$$

值得一提的是，Sims and Zha (1998) 曾基於 Litterman (1986) 的概念，提出其先驗假設。茲舉例說明如下。

令 $\mathcal{N}(b_0, V_0^{\text{SZ}})$ 為 b 的先驗分配¹⁶，其中 $b_0 = \text{vec}(B_0)$ ；若為隨機漫步模型，則

$$B_0 = \begin{bmatrix} 0_{1 \times n} \\ I_n \\ 0_{(k-n-1) \times n} \end{bmatrix}, \quad (3.4.6)$$

¹⁶ 同樣地，此處先假設 b 與 Σ 不互相獨立。關於 b 與 Σ 互相獨立的情形，可參見 Sims and Zha (1998) 之探討。

與 Minnesota 先驗相同。 V_0^{SZ} 為對角矩陣，其對角線上之元素 $v_{ij,l}^{SZ}$ 為

$$v_{ij,l}^{SZ} = \begin{cases} \left(\frac{\lambda_0 \lambda_1}{l^{\lambda_3} \sigma_j}\right)^2, & j = 1, \dots, n \\ (\lambda_0 \lambda_4)^2, & \text{for the constant} \end{cases} \quad (3.4.7)$$

且 $l = 1, \dots, p$ ，為變數之落後期數； σ_j 為第 j 條方程式誤差項之標準差。同樣地， λ_0 、 λ_1 、 λ_3 、 λ_4 係由研究者所給定之超參數。 λ_0 控制整個變異數-共變異數矩陣的緊度，若 $\lambda_0 \rightarrow 0$ 則資料將不會影響估計結果，但若 $\lambda_0 \rightarrow \infty$ 則先驗資訊將極為寬鬆 (loose)，此時估計結果將完全由資料決定，即與 OLS 相同。 λ_1 主要控制被解釋變數落後期係數之先驗分配的標準差， λ_1 越大分配將越平坦； λ_3 使落後越多期的係數越傾向為 0，並控制其趨向 0 的速度¹⁷，故若 λ_3 越大，則落後越多期的係數將越快趨於 0； λ_4 僅用於常數項。

若比較 (3.4.4) 式、(3.4.7) 式，則可知除了加入 λ_0 ， Sims and Zha (1998) 逕給定 $\lambda_2 = 1$ ，故無論被解釋變數的落後項或其他解釋變數，只要落後期數相同，其係數在先驗分配上的條件變異數皆相同；且 Sims and Zha (1998) 將干擾項變異數標準化為 1，故 (3.4.7) 式中為 $(1/\sigma_j)$ 而非 (σ_i/σ_j) 。

因此，若以雙變量的 VAR(2) 模型為例，則可從 (3.4.6) 式、(3.4.7) 式得到

$$b_0 = [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]',$$

且

¹⁷ 在隨機漫步的假設中，除了 AR(1) 係數為 1 外，其餘皆為 0。

$$V_0^{SZ} = \begin{bmatrix} (\lambda_0\lambda_4)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \left(\frac{\lambda_0\lambda_1}{\sigma_1}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \left(\frac{\lambda_0\lambda_1}{\sigma_2}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\frac{\lambda_0\lambda_1}{2^{\lambda_3}\sigma_1}\right)^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \left(\frac{\lambda_0\lambda_1}{2^{\lambda_3}\sigma_2}\right)^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (\lambda_0\lambda_4)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_0\lambda_1}{\sigma_1}\right)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_0\lambda_1}{\sigma_2}\right)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_0\lambda_1}{2^{\lambda_3}\sigma_1}\right)^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \left(\frac{\lambda_0\lambda_1}{2^{\lambda_3}\sigma_2}\right)^2 \end{bmatrix}$$

另外，在貝氏 VAR 模型中，對於變數取差分是否能改善模型的預測結果尚無定論。因此，Sims and Zha (1998) 使用 μ_5 (≥ 0) 及 μ_6 (≥ 0) 等 2 個超參數。其中， μ_5 控制「係數總和虛擬觀察值」(sum-of-coefficients dummy observation) 之先驗，若 $\mu_5 \rightarrow \infty$ ，則模型傾向於將變數取差分，此時各變數均含有單根 (unit root) 且無共整合 (cointegration) 關係。 μ_6 控制「虛擬初始觀察值」(dummy initial observation) 之先驗，若 $\mu_6 \rightarrow \infty$ ，則模型將傾向於變數皆屬定態且其平均為初始條件之樣本平均值，抑或是變數具有不含飄移項 (drift) 之單根的情形。詳見 Sims and Zha (1998)。

綜上所述，關於 (3.4.5) 式中先驗分配的參數， b_0 與 V_0 可參照前述 Minnesota 先驗或 Sims and Zha (1998) 的方式給定；而 S_0 可設為

$(v_0 - n - 1)\text{diag}[\sigma_1^2, \dots, \sigma_n^2]$ ， v_0 則可設為 $n + 2$ （如 Carriero et al., 2011；Karlsson, 2013）。除此之外，Koop and Korobilis（2010）則是設定 $b_0 = 0$ 、 $V_0 = 10 \times I_n$ 、 $v_0 = n$ 、 $S_0 = I_n$ ，端視研究者之主觀認知而定。

四、應用範例：預測台灣的 CPI 年增率

物價穩定是貨幣政策的重要目標，各國的中央銀行均極為關注物價情勢的變化。然而，貨幣政策自擬定、施行至影響經濟活動，往往具有時間上的落差；因此，物價預測對央行而言相當重要。在國內文獻方面，陳佩玗（2013）以時間序列模型預測台灣 CPI 的短期走勢，並發現分別預測 CPI 的各組成項目後，再將其加總為 CPI 的做法，預測績效將較直接預測 CPI 本身為佳。

本報告擬依照台灣 CPI 的組成項目，以貝氏 VAR 預測各組成項目的年增率後，依權數加總得到 CPI 年增率的預測值，再比較其與直接預測 CPI 年增率的預測績效。參考陳佩玗（2013），貝氏 VAR 考量的變數包括食物類、衣著類、房租、住宅維修費、家庭用品、家庭管理費用、水電燃氣、交通及通訊設備、油料費、交通服務及維修零件、醫藥保健類、教養費用、娛樂費用及雜項類等，共 14 項組成項目之年增率。而 CPI 及其組成項目的資料來源均為主計總處。

關於貝氏 VAR 之先驗資訊，此處將使用 Normal–Wishart 先驗。首先，如 Carriero et al.（2011）所述，Minnesota 先驗假設變數的 AR(1) 係數為 1，以考量總體資料的持續性；故若變數為差分形式，則應將 AR(1) 係數的先驗假設為 0。由於此處係直接預測 CPI 各組成項目的年增率，故假設為 0。其次，Carriero et al.（2011）發現當 $\lambda_0 = 0.1$ 、

模型的落後期數為 12 期時，模型的表現將難以擊敗 (hard to beat)¹⁸；故此處亦設定 $\lambda_0 = 0.1$ 、落後期數為 12 期。至於其他超參數，則參照 Carriero et al. (2011) 的基準設定，給定 $\lambda_2 = 1$ 、 $\lambda_3 = 2$ 。

另一方面，在陳佩玗 (2013) 中，AR 模型係表現較佳的時間序列模型，故此處亦使用 AR 模型直接預測 CPI 年增率。而 AR 模型的落後期數，則是在最大落後期數為 12 的前提下，依據赤池資訊評選準則 (Akaike information criterion, AIC) 選定為 12。

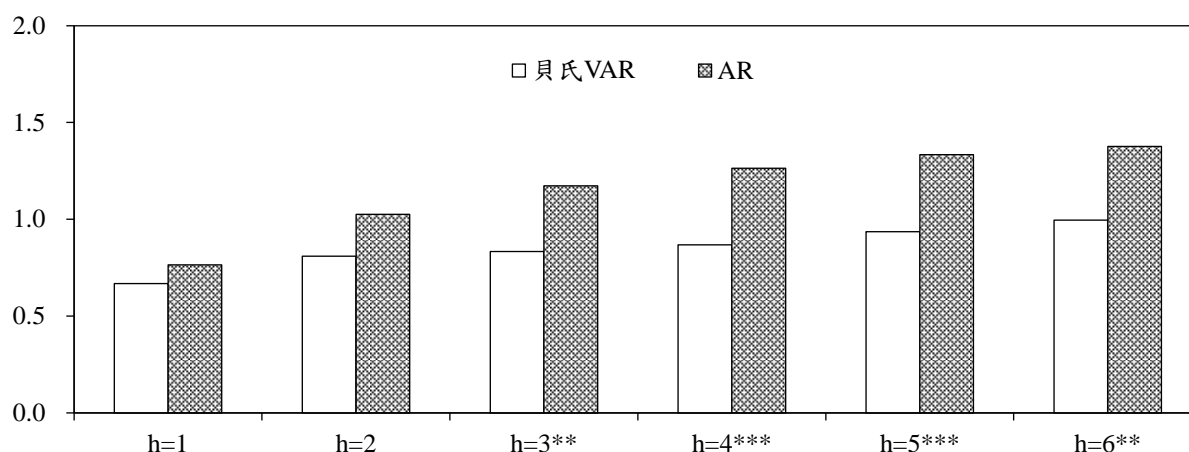
資料期間涵蓋 1991 年 1 月至 2018 年 6 月。為評估樣本外預測績效，首先以 1991 年 1 月至 2012 年 12 月的資料進行樣本內估計，並以疊代的方式間接預測未來半年 (共 6 期) 的 CPI 年增率；接著，再以遞迴的方式，逐月增加模型的樣本數，直到 2017 年 12 月為止。在得到未來半年各月的預測值後，即可計算預測誤差在各預測期間下的均方根差 (root mean square error, RMSE)，以比較貝氏 VAR 與 AR 模型的預測績效。

圖 3 為 CPI 預測值之 RMSE。從圖中可知，在未來 1 至 6 期之預測期間內，貝氏 VAR 的 RMSE 皆小於 AR 模型，故貝氏 VAR 的預測績效較佳。為了進一步瞭解兩者的預測績效是否顯著不同，另以 Diebold–Mariano 檢定檢驗兩者的均方差 (mean square error, MSE) 是否具有顯著差異 (Diebold and Mariano, 1995)。結果顯示，在預測期間為未來 1、2 期時，雖然貝氏 VAR 的 MSE 較小，但兩者差異仍未達顯著水準；至未來 3 期以上時，貝氏 VAR 的預測績效均能顯著優於 AR 模型。因此，如同陳佩玗 (2013) 之實證結果，分別預測 CPI

¹⁸ 惟在 Carriero et al. (2011) 中，這並非最適設定。該文基於美國的資料，以最大化邊際概似函數的方式，尋找最適的 λ_0 與落後期數，並發現最適的 λ_0 為 0.25 或 0.3，最適落後期數則為 1 或 2 期。詳見該文。

各組成項目的年增率後，再將其加總為 CPI 年增率的做法，的確可能有較佳的預測績效。

圖 3：不同預測期間下 CPI 預測值之 RMSE



說明：1. h=預測期間。

2. ***, **與*分別表示 Diebold–Mariano 檢定統計量在 1%、5%與 10%的水準下顯著，損失函數設定為 MSE 的形式。

同樣地，此處模型設定較為簡便，多僅參考 Carriero et al.(2011)，且未考慮農曆春節、國際油價等影響物價的重要因素，故實際上應仍有改善空間。關於貝氏 VAR 之介紹，以及估計、預測等說明，可參見前述之 Canova (2007)、Carriero et al. (2011)、Koop and Korobilis (2010) 或 Karlsson (2013) 等。

肆、結論與建議

本次研習內容主要涵蓋狀態空間模型、貝氏估計，以及 VAR 模型等常見於總體實證之計量模型；並輔以統計軟體 EViews 與 R 之操作，俾助理論學習及實際運用。課程講師均由 SEACEN 研訓中心之經濟學者擔任，皆具豐富的學識及經驗，熟稔相關主題。

主辦單位亦邀請日本央行與亞洲開發銀行的官員，說明其運用之經濟預測模型，並分享相關經驗。日本央行除了設定詳盡的大型總體計量模型，尚有用於估計長期通膨趨勢與通膨預期的模型，以及即時預報模型等；亞洲開發銀行則使用動態因子模型 (dynamic factor model)，應用大量的經濟指標，即時預報區域內經濟體當季或當年的經濟成長率。

本報告除了簡介狀態空間模型與貝氏估計等主要研習內容，並將其應用於相關的範例。首先，藉由狀態空間模型，可將月頻 GDP 視為不可觀察到的變數，再透過其與季頻 GDP 及工業生產指數之間的關係加以估計；其次，以貝氏 VAR 分別預測 CPI 各組成項目的年增率後，再將其加總為 CPI 年增率的做法，其預測績效將可能較直接預測 CPI 本身為佳。

近年總體經濟環境變遷快速，而本次研習涵蓋之內容，應用層面相當廣泛；故參與本次研習，不僅可學習這些常見的分析工具，亦可將其應用於未來研究。其次，透過與各國學員交流討論，發現多數央行均建置即時預報模型，以隨時加入新的資訊，更新預測結果。有鑑於此，未來亦可持續精進或發展相關模型，提高短期經濟預測能力。此外，與其他學員進行交流的過程中，尚可瞭解其他國家之經濟估測及研究經驗。故參與本次研習，當有助增進估測及研究分析能力。

參考資料

- 陳佩玗 (2013), 「台灣地區短期通貨膨脹率之預測」, 中央銀行季刊, 35 (1), 63–89。
- 劉瑞文 (2007), 「由靜態到動態之依時拆分—臺灣工業部門實質 GDP 之按月推估」, 臺灣經濟預測與政策, 38 (1), 75–125。
- Canova, F. (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton.
- Carriero, A., T. E. Clark, and M. Marcellino (2011), “Bayesian VARs Specification Choices and Forecast Accuracy,” Federal Reserve Bank of Cleveland Working Papers 11-12.
- Casella, G., and E. I. George (1992), “Explaining the Gibbs Sampler,” *American Statistician*, 46(3), 167–174.
- Chow, G. C. and A.-L. Lin (1971), “Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series,” *Review of Economics and Statistics*, 53(4), 372–375.
- Diebold, F. X., and R. S. Mariano (1995), “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics*, 13(3), 253–263.
- Harvey, A. C. (1993), *Time Series Models*, 2nd ed., Harvester, New York.
- Harvey, A. C., and R. G. Pierse (1984), “Estimating Missing Observations in Economic Time Series,” *Journal of the American Statistical Association*, 79(385), 125–131.
- Huang, Y.-L. (2010), “Estimating Taiwan's Monthly GDP in an Exact Kalman Filter Framework: A Research Note,” *Taiwan Economic Review*, 38(1), 147–160.
- Judge, G. G., R. C. Hill, W. Griffiths, H. Lütkepohl, and T.-C. Lee (1988), *Introduction to the Theory and Practice of Econometrics*, 2nd ed.,

- Wiley, New York.
- Kadiyala, K. R., and S. Karlsson (1997), “Numerical Methods for Estimation and Inference in Bayesian VAR-Models,” *Journal of Applied Econometrics*, 12(2), 99–132.
- Karlsson, S. (2013), Forecasting with Bayesian Vector Autoregression, in: G. Elliott, and A. Timmermann (ed.), *Handbook of Economic Forecasting*, Volume 2B. Elsevier, Amsterdam, pp.791–897.
- Koop, G., and D. Korobilis (2010), “Bayesian Multivariate Time Series Methods for Empirical Macroeconomics,” *Foundations and Trends in Econometrics*, 3(4), 267–358.
- Litterman, R. B. (1980), “A Bayesian Procedure for Forecasting with Vector Autoregressions,” *Journal of Business and Economic Statistics*, 4(1), 25–38.
- Litterman, R. B. (1986), “Forecasting with Bayesian Vector Autoregressions—Five Years of Experience,” *Journal of Business and Economic Statistics*, 4(1), 25–38.
- Moauero, F., and G. Savio (2005), “Temporal Disaggregation Using Multivariate Structural Time Series Models,” *Econometrics Journal*, 8(2), 214–234.
- Schweppe, F. (1965), “Evaluation of Likelihood Functions for Gaussian Signals,” *IEEE transactions on Information Theory*, 11(1), 61–70.
- SEACEN(2018),「SEACEN-BOJ Course on Intermediate Modelling and Forecasting Techniques」 研習課程講義。
- Sims, C. A., and T. Zha (1998), “Bayesian Methods for Dynamic Multivariate Models,” *International Economic Review*, 39(4), 949–968.