Fixed-effect and Random-effect variable selection in linear mixed models using R2 statistics

吳書恆

June, 2022

摘要

本研究旨在探討...。

目次

圖次

表次

第一章　緒論

在迴歸模型中，R^2 是常見用來解釋變異 (explained variance) 比例的指標，它描述模型的依變數被解釋變數詮釋的程度。通常模型變數越多解釋力也會越高，但當過多變數放入模型中，可能會導致共線性 (Collinearity) 的問題，或是資料樣本數不夠導致估計值無法收斂 (**hjort1991estimation**)。不過，在進行迴歸分析之前，能夠發現這些問題是可以提早解決，像是對變數做轉換來消除之間的相關性，又或是再補充樣本解決變數維度過高的問題，但模型變數的平衡與抉擇還是要歸咎回分析的目的與成本，倘落研究的目的著重在預測 (forcasting) 新樣本，那精簡就會是個不錯的選擇，已經有許多研究發現過多解釋變數會導致預測效果的邊際效應遞減 (**helland2000model**, **sundberg1999multivariate**)，同時在解釋模型時也較容易；另外，當個體追蹤的時間點不夠多或中斷追蹤的時間點過多，隨機項係數的推估就有可能會估計不出來，因此追求較精簡在大部分實務上是其有必要性的。

第二章　文獻回顧

第一節　混合模型

縱向資料為數筆隨著時間重複觀測特定個體的數據資料，這種與時間有關的資料主要為了觀察特定變數隨時間的變動與找出可能影響的因子。由於縱向資料的觀測值具有某種相關性，將模型納入與時間有關的共變異數結構 (covariance structure) 是有必要的。當共變異數結構是正確的，模型估計值的標準誤也會被準確地計算，在進行統計推論時就會恰當。因此，在縱向分析會考量兩個部份，除了所有個體平均的反應效果之外，與個體內重複測量的共變異數結構也會納入考慮。線性混和模型 (linear mixed model, LMM) 就是其中一種處理具有相關性資料的迴歸模型。

線性混合模型的係數可拆成兩個部分，平均的反應效果對應固定效果 (fixed-effects)，共變異數結構對應隨機效果 (random-effects)。定義特定觀測值型式 (observation-specific) 的混合模型 (**harville1977maximum**; **laird1982random**) 為

$$Y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i + e_{ij}, \quad i = 1,\ldots,m, \quad j = 1,\ldots,n_i, \tag{1}$$

其中，Y_{ij} 為第 i 個個體 (subject) 之下第 j 個時間點 (occasion) 的觀測值，$\boldsymbol{x}_{ij}$ 為維度 $(p \times 1)$ 的固定效果共變數矩陣，$\boldsymbol{\beta}$ 為維度 $(p \times 1)$ 的固定效果參數向量，$\boldsymbol{z}_{ij}$ 為維度 $(q \times 1)$ 的隨機效果共變數矩陣，$\boldsymbol{z}_{ij}$ 的元素必須是 $\boldsymbol{x}_{ij}$ 的子集合，$\boldsymbol{b}_i$ 為維度 $(q \times 1)$ 且不可觀察到的隨機效果向量，e_{ij} 為隨機誤差值。將個體所有時間點的觀測值合併成矩陣，即 $\boldsymbol{Y}_i = (Y_{i1},\ldots,Y_{in_i})'$、$\boldsymbol{X}_i = (\boldsymbol{x}_{i1}',\ldots,\boldsymbol{x}_{in_i}')$、$\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}',\ldots,\boldsymbol{z}_{in_i}')$ 且 $\boldsymbol{e}_i = (e_{i1},\ldots,e_{in_i})'$，則定義特定個體型式 (subject-specific, matrix notation) 為

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i。$$

混合模型假設 $\boldsymbol{b}_i$ 服從平均數 0 和共變異數矩陣 $\boldsymbol{G}$ 的多維常態分配，即 $\mathrm{E}(\boldsymbol{b}_i) = \boldsymbol{0}$ 且 $\mathrm{Cov}(\boldsymbol{b}_i) = \boldsymbol{G}$，$\boldsymbol{G}$ 的維度為 $(q \times q)$。由於 $\boldsymbol{b}_i$ 會因人而異，此模型特定

變數的係數會隨著個體而改變。然而在推論上不會直接描述係數 $\boldsymbol{b}_i$，研究者大多對它的共變異數矩陣 $\boldsymbol{G}$ 較感興趣，因此 $\boldsymbol{G}$ 才是隨機效果中感興趣的參數。另一方面，$\boldsymbol{b}_i$ 可以接受服從任一多維度分配，為了方便估計與推論，實務上採用多維常態分佈居多。$\boldsymbol{e}_i$ 假設服從平均數 0 和共變異數矩陣 $\boldsymbol{R}_i$ 的多維常態分配，$\boldsymbol{R}_i$ 維度為 $(n_i \times n_i)$，且假設任意 $\boldsymbol{b}_i$ 與 $\boldsymbol{e}_i$ 之間皆須獨立。通常研究者還會再去假設 $\boldsymbol{R}_i$ 為 $\sigma^2 \boldsymbol{I}_{n_i}$ 的對角矩陣。

在這些假設成立之下，給定 $\boldsymbol{X}_i$ 與 $\boldsymbol{b}_i$，$\boldsymbol{Y}_i$ 的條件期望值為

$$\mathrm{E}\,(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i\,,$$

邊際期望值為

$$\mathrm{E}\,(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \boldsymbol{X}_i\boldsymbol{\beta}\,,$$

邊際期望值為條件期望值對 $\boldsymbol{b}_i$ 的分佈取期望值。同樣地，共變異數結構也可以拆成條件與邊際兩個部分，給定 $\boldsymbol{X}_i$ 與 $\boldsymbol{b}_i$，$\boldsymbol{Y}_i$ 的條件共變異數為

$$\mathrm{Cov}\,(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i) = \mathrm{Cov}\,(\boldsymbol{e}_i) = \boldsymbol{R}_i\,,$$

邊際共變異數為

$$\mathrm{Cov}\,(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \boldsymbol{Z}_i\mathrm{Cov}\,(\boldsymbol{b}_i)\boldsymbol{Z}_i' + \mathrm{Cov}\,(\boldsymbol{e}_i) = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i' + \boldsymbol{R}_i\,。$$

邊際共變異數反映模型的兩個變異來源，$\boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i'$ 代表個體與個體之間 (between-subject) 的變異（簡稱個體間），$\boldsymbol{R}_i$ 則代表個體內重複測量之間 (within-subject) 的變異（簡稱個體內）。然而，當共變數會隨時間改變時，這種好用的性質將會消失，$\boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i'$ 因為受到 $\boldsymbol{Z}$ 的影響，導致其混雜著兩種變異來源，這在單獨選擇隨機效果的變數 $\boldsymbol{Z}_i$ 時，會因無法清楚地分辨出變異的類型而無法選出重要的變數 (**orelien2008fixed**)。另外，混淆的共變數也會導致再解釋固定效果的參數 $\boldsymbol{\beta}$ 時做出誤導的推論 (**neuhaus1998between**)。因此，在建立混和模型之前，會先進行時間相依共變數的拆解以解決這個問題。

為了分離個體間與個體內兩種變異來源，將原始資料拆成個體間 ($\bar{\boldsymbol{x}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}$) 與個體內 ($\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i$)，並重新表示式 (1)，定義時間相依共變數拆解

的混合模型為

$$Y_{ij} = \bar{\boldsymbol{x}}_i' \boldsymbol{\beta}_B + \left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i\right)' \boldsymbol{\beta}_W + \left(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_i\right)' \boldsymbol{b}_i + e_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i \text{，} \qquad (2)$$

其中，$\bar{\boldsymbol{x}}_{i\cdot}$ 為維度 $(p \times 1)$ 的個體間固定效果共變數矩陣，$\boldsymbol{\beta}_B$ 為維度 $(p \times 1)$ 的個體間固定效果參數向量，$\left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i\right)$ 為維度 $(r \times 1)$ 的個體內固定效果共變數矩陣（不包含截距項），且 $r < p$，$\boldsymbol{\beta}_W$ 為維度 $(r \times 1)$ 的個體內固定效果參數向量，$\left(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_i\right)$ 為維度 $(q \times 1)$ 的隨機效果設計矩陣，$\boldsymbol{b}_i$ 為維度 $(q \times 1)$ 的隨機效果向量。

使用混合模型還有其他好處。首先，資料不需要是平衡 (balanced) 的縱向設計，也就是說每個個體觀測的時間不用完全一樣，原因是時間可以當作模型的共變數而量化其效果，這在實務上的研究提供了許多彈性的空間。第二，共變異數矩陣 $\boldsymbol{G}$ 的參數並不會因某些人測量時間點的數量而影響，這個特性接納遺失值隨時間增加的問題，儘管有些人中途退出研究。最後，不同時間點變異程度要相同的假設是不需要成立的，只要隨機項納入時間的共變數，共變異數矩陣就能作為與時間有關的矩陣，這也放寬了變異數的型式並非要複合對稱 (Compound symmetry) 或一階自迴歸模型 (First-order autoregressive, AR(1)) 等。

第二節　混合模型的 R^2 統計量

目前已有許多 R^2 統計量將一般的迴歸中的定義衍生至混合模型上，儘管這些統計量有陸續提出一些修正版本，但都存在一些缺失和嚴苛的先前條件，直至今日混合模型 R^2 才發展的較為完善。不管是何種模型，R^2 都應該能夠反應 X 與 Y 的關係，並提供除了顯著性以外的資訊。然而更重要的是，當選擇固定效果變數時，AIC、BIC 或 LRT 這些選模準則和方式直接套用至受限最大概似函數 (REML) 有時候是不恰當的，這也增加了不需要概似函數為基礎的 R^2 被使用的機會。由於過去有許多不同但卻類似的 R^2，這一章節將簡單探討較具代表性的混合模型 R^2，以及這些統計量表現的相關研究。

對於式 (1)，主流 R^2 的定義為

$$R^2 = 1 - \frac{\text{Var}(\text{提出模型的} Y_{ij})}{\text{Var}(\text{虛無模型的} Y_{ij})},$$

提出模型 (proposed model) 和虛無模型 (null model) 對 R^2 的解釋扮演重要的角色。當分子為 $\text{Var}(Y_{ij}|\boldsymbol{x}_{ij})$，則稱作邊際 R^2 (*marginal* R^2)，只反應模型固定的未解釋量；當分子為 $\text{Var}(Y_{ij}|\boldsymbol{x}_{ij}, \hat{\boldsymbol{b}}_i)$，則 R^2 統計量稱作條件 R^2 (*conditional* R^2)，反應模型固定和隨機效果的未解釋量。這邊注意，邊際 R^2 與條件 R^2 皆從同個模型計算出來，只是計算的成分不同，邊際 R^2 僅考慮 $Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$，條件 R^2 考慮的是 $Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i$。

而虛無模型大致可以分為兩種，一是僅含固定效果截距項的 $\text{Var}(Y_{ij})$，也就是 $Y_{ij} = \mathbf{1}'_{n_i}\beta_0$，$\mathbf{1}_{n_i}$ 為維度 $(n_i \times 1)1$ 的向量，描述只放固定效果截距項（沒有任何解釋變數放入模型）的未解釋量；二是含有兩種截距項的 $\text{Var}(Y_{ij}|b_{0i})$，即 $Y_{ij} = \mathbf{1}'_{n_i}\beta_0 + b_{0i}$，大多是為了看兩種效果斜率項增加的解釋量。要採用哪種型式並沒有一套準則，甚至並不一定要以上兩種型式，選用何種虛無模型只會影響它的性質與解釋。

在時間相依共變數拆解的想法提出之前，條件 R^2 和邊際 R^2 是普遍用來當作混合模型的解釋統計量。最早由 Vonesh 和 Chinchilli (1996) 提出的統計量 R^2_{VC}，他沿用一般的迴歸的想法至廣義線性混合模型，並假設 $\boldsymbol{R}_i$ 為 $\sigma^2\boldsymbol{I}_{n_i}$，表示

為

$$R^2_{VC} = 1 - \frac{\sum_{i=1}^{m}(\boldsymbol{Y}_i - \hat{\boldsymbol{Y}}_i)'(\boldsymbol{Y}_i - \hat{\boldsymbol{Y}}_i)}{\sum_{i=1}^{m}(\boldsymbol{Y}_i - \bar{Y}\mathbf{1}_{n_i})'(\boldsymbol{Y}_i - \bar{Y}\mathbf{1}_{n_i})},$$

其中，$\hat{\boldsymbol{Y}}_i$ 可以為 $\boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{b}}_i$ 或是 $\boldsymbol{X}_i\hat{\boldsymbol{\beta}}$。Vonesh 透過有無考慮隨機項來分別描述整體與固定效果的適合度 (goodness of fit)，這樣的想法雖然合理，但是對於選模時卻是不恰當的。在式 (1) 中，$\text{Cov}(\boldsymbol{Y}_i|\boldsymbol{X}_i)$ 的 $\boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i'$ 為個體間的變異來源，說明固定效果的缺適性 (lack of fit) 會被轉移到這裏，並且反應到係數 $\hat{\boldsymbol{b}}_i$。當 $\hat{\boldsymbol{Y}}_i$ 為 $\boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{b}}_i$，$R^2_{VC}$ 的分子將不會根據變數的進入或進出而有明顯的變動，這說明了條件 R^2 無法選擇出正確固定效果變數但邊際 R^2 可以的原因。

R^2_{VC} 的虛無模型僅考慮固定效果截距項，Xu (2003) 修正 R^2_{VC} 的虛無模型，讓模型包含了兩種截距項，試圖去消除隨機項中個體間變異來源的干擾。但這並不管用，當分母納入隨機項後，原本固定效果的解釋變異又會被隨機截距項彌補回來，分子與分母的缺適性還是無法有明顯的差距出現。另一方面，當模型納入了時間相依共變數，那麼固定效果的缺適性 (lack of fit) 多少比例被轉移到 $\boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{b}}_i$ 和 $\boldsymbol{R}_i$ 並無法直接判斷出來，這取決於該變數個體間與個體內解釋變異的比例。

對於這更加複雜的問題，Snijders 和 Bosker (2011) 的多層次模型（multilevel model）變異成分提供可行的解決方案。在式 (2) 的型式下，整體解釋變異可拆解成個體間和個體內兩種變異，推理如下，

$$\begin{aligned}
\text{Var}(Y_{ij}) &= \text{Var}\left(\bar{\boldsymbol{x}}_i'\boldsymbol{\beta}_B + \left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i\right)'\boldsymbol{\beta}_W + \left(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_i\right)'\boldsymbol{b}_i + e_{ij}\right) \\
&= \boldsymbol{\beta}_B'\boldsymbol{\Sigma}_B\boldsymbol{\beta}_B + \boldsymbol{\beta}_W'\boldsymbol{\Sigma}_W\boldsymbol{\beta}_W + g_{00} + tr(\boldsymbol{G}\boldsymbol{\Sigma}_Z) + \sigma^2 \\
&= \underbrace{\boldsymbol{\beta}_B'\boldsymbol{\Sigma}_B\boldsymbol{\beta}_B + g_{00}}_{\text{個體間變異}} + \underbrace{\boldsymbol{\beta}_W'\boldsymbol{\Sigma}_W\boldsymbol{\beta}_W + tr(\boldsymbol{G}\boldsymbol{\Sigma}_Z) + \sigma^2}_{\text{個體內變異}}。
\end{aligned} \tag{3}$$

其中，tr 表示對矩陣對角線上各個元素做加總，$\boldsymbol{\Sigma}_B$、$\boldsymbol{\Sigma}_W$ 與 $\boldsymbol{\Sigma}_Z$ 分別為 $\bar{\boldsymbol{x}}_i$、$\left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i\right)$ 與 $\left(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_i\right)$ 的共變異數矩陣，g_{00} 為斜率截距項變異數，σ^2 為誤差的變異數。因此，$\boldsymbol{\beta}_B'\boldsymbol{\Sigma}_B\boldsymbol{\beta}_B$ 代表個體間固定效果的解釋變異，$\boldsymbol{\beta}_W'\boldsymbol{\Sigma}_W\boldsymbol{\beta}_W$ 代表個體內固定效果的解釋變異，$tr(\boldsymbol{G}\boldsymbol{\Sigma}_Z)$ 代表所有隨機斜率項的解釋變異。

由式(3)中可發現，$\boldsymbol{\beta}'_B\boldsymbol{\Sigma}_B\boldsymbol{\beta}_B+g_{00}$ 反映個體間變異，$\boldsymbol{\beta}'_W\boldsymbol{\Sigma}_W\boldsymbol{\beta}_W+tr(\boldsymbol{G}\boldsymbol{\Sigma}_Z)+\sigma^2$ 則反映個體內變異。將這兩個區別開來的好處是能夠清楚的看到解釋變異的流動，

對於 Var $\left(\boldsymbol{z}_{ij}-\bar{\boldsymbol{z}}_i\right)'\boldsymbol{b}_i$) 能夠，後來由 Rights 和 Sterba (2019) 將其概念到套用到 R^2，先定義整體解釋變異，

$$100\% = R_C^2 + \frac{\sigma^2}{\text{Var}\,(Y_{ij})},\ \text{where}\ R_C^2 = R_M^2 + R_u^2 + R_v^2 = (R_B^2 + R_W^2) + R_u^2 + R_v^2 \qquad (4)$$

第三節　AIC

$$AIC = 2\ln(f(y|\hat{\beta},\hat{\theta})) + 2k \qquad (5)$$

$$AICc = AIC + \frac{2(k+1)(k+2)}{nk2} \qquad (6)$$

where k is the number of parameters (including the intercept) in the model.

$$mAIC = 2\ln(f(y|\hat{\beta},\hat{\theta})) + 2\alpha_n(p+q) \qquad (7)$$

$\alpha_n = 1$ in the infinite sample form or $\alpha_n = n/(npq1)$ in the finite sample form (Sugiura, 1978).

$$cAIC = 2\ln(f(y|\hat{\beta},\hat{b},\hat{\theta}) + 2(\rho+1), \qquad (8)$$

$$\rho = trace\left(\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z+G^{-1} \end{pmatrix}^{-1}\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix}\right)$$

第三章　修正 R^2 成分

前章節已提及條件 R^2 在混合模型中不具有選模的功用，這部分可用隨機斜率 R^2 來替代隨機效果（不含隨機截距項）變數選擇的依據，邊際 R^2 則是繼續作為固定效果變數選擇的依據。然而這些統計量仍會面臨到與傳統回歸同樣的問題－這些未調整的 R^2 容易選到較為複雜的模型，很可惜目前沒有相關研究對這部分提出修正。在接下來的章節將會嘗試探討邊際 R^2 和隨機斜率 R^2 的調整，調整方式分為兩種，第一種會根據傳統調整 R^2 的方式做類似的推廣，第二種會新增精簡指數來使得 R^2 成分可以更有彈性地使用。

第一節　調整 R^2_{adj} 成分

根據 Mordecai Ezekiel 的定義，一般迴歸調整 R^2 的公式為

$$R^2_{adj} = 1-(1-R^2)\frac{N-1}{N-p} = 1-\frac{\widehat{\text{Var}}_{res}}{\widehat{\text{Var}}_{tot}}$$

其中 $\widehat{\text{Var}}_{res}$ 與 $\widehat{\text{Var}}_{tot}$ 分別對應殘差與總誤差變異數的不偏估計值，邊際 R^2 可透過調整 R^2 的概念進行調整，如下。

$$R^2_{adj.M} = 1-\frac{\widehat{\text{Var}}\,(u_i+\boldsymbol{s}'_{ij}\boldsymbol{v}_i+e_{ij})}{\widehat{\text{Var}}\,(Y_{ij})} = 1-\frac{\hat{\sigma}^2_u+\text{tr}(\hat{\boldsymbol{T}}\boldsymbol{\Sigma}_s)+\hat{\sigma}^2}{\widehat{\text{Var}}\,(Y_{ij})} \tag{9}$$

這調整的概念是將隨機效果與誤差皆視為不可解釋的部分，而分子對應的是受限概似函數 (restricted maximum likelihood, REML) 底下計算出來的不偏變異數成分，並假設固定效果及各種變異成分之間皆獨立；同樣地，隨機斜率 R^2 修正如下。

$$R^2_{adj.v} = 1-\frac{\widehat{\text{Var}}\,(\bar{\boldsymbol{x}}'_{i\cdot}\boldsymbol{\beta}_B+\boldsymbol{w}'_{ij}\boldsymbol{\beta}_W+u_i+e_{ij})}{\widehat{\text{Var}}\,(Y_{ij})} = 1-\frac{\boldsymbol{\beta}'_B\boldsymbol{\Sigma}_B\boldsymbol{\beta}_B+\boldsymbol{\beta}'_W\boldsymbol{\Sigma}_W\boldsymbol{\beta}_W+\hat{\sigma}^2_u+\hat{\sigma}^2}{\widehat{\text{Var}}\,(Y_{ij})} \tag{10}$$

第二節　精簡指數的 $R^2(\alpha)$ 成分

單純對 R^2 成分做簡單地調整，這個調整是很直白的。令可調控的精簡指數為 α，且 α 介於 0 到 1 之間，則加入精簡指數 R^2 成分為

$$\begin{aligned} R^2(\alpha) &= \alpha R^2 + (1-\alpha)\frac{R^2}{\#(\text{parameter})}, \quad \alpha \in [0,1] \\ &= \left(\alpha + \frac{1-\alpha}{\#(\text{parameter})}\right) R^2 \end{aligned} \tag{11}$$

最不精簡的特性為考量總體參數解釋量，反映在 αR^2，而最精簡模型的特性為考量平均一個參數解釋量，反映在 $(1-\alpha)\frac{R^2}{\#(\text{parameter})}$，其中 #(parameter) 表示該 R^2 成分對應模型的部分參數個數。想要達到多精簡是可以調整的，因此對這兩個解釋量取加權平均，權重由 α 決定，α 越接近 0 表示越精簡，α 越接近 1 則越不精簡。

套用至 R^2_M、R^2_v 如下

$$R^2_M(\alpha) = \left(\alpha + \frac{1-\alpha}{p}\right) \times R^2_M \tag{12}$$

$$R^2_v(\alpha) = \left(\alpha + \frac{1-\alpha}{q}\right) \times R^2_v \tag{13}$$

第三節　使用 R^2_{adj} 與選擇 α

假如要比較固定效果的解釋量，則計算要比較模型的 $R^2_M(\alpha)$，最後選擇最大 $R^2_M(\alpha)$ 的模型；假如比較隨機斜率效果的解釋量，則計算要比較模型的 $R^2_v(\alpha)$，最後選擇最大 $R^2_v(\alpha)$ 的模型。選混合模型時的變數時，通常不會同時比較固定效果與隨機效果，這兩者在估計時會彼此影響這特性已在第三節討論過，因此通常選模會考慮以下三種情形，一為已知 fixed 只選 random 部分；二為已知 random 只選 fixed 部分；三為 random 和 fixed 未知。

第四章　資料模擬與實際資料

第一節　資料模擬

為了方便表示模擬的各種模型，定義模型符號如下：

$$M(fixed, random)$$

下標前面表示隨機部分的模型類型，後面表示固定部分，模型類型符號包含：

$over$ = 超飽和模型，包含全部顯著與部分不顯著變數，

$full$ = 完整模型，包含全部顯著，

$reduce$ = 精簡模型，只包含部分不重要的顯著變數，

$-reduce$ = 過度精簡模型，只包含部分重要的顯著變數，

$other$ = 其他模型，包含部分顯著與部分不顯著變數。

舉例：$M(over, full)$ 表示模型的固定是飽和、隨機部分是超飽和。

模擬的主要流程

1. 產生具有相關性的資料：$m = 100$，$n_i = 10$，組間變數有 13 個，組內有 6 個

   $\beta_B = (-4, 5, -6, 8, -2, 0, 0, 0, 0, 0, 0, 0, 0)^T$,

   $\beta_W = (1, -1, 0, 0, 0, 0)^T$,

$$T = \begin{pmatrix} 31 & 12 & 0 & 0 & 0 & 0 & 0 \\ 12 & 28 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

顯著地的固定效果變數為 Vb1,Vb2,Vb3,Vb4,Vw1,Vw2，重要的顯著變數只有 Vb1,Vb2,Vb3；顯著地的隨機效果變數 Vb1,Vb2,Vb3,Vb4，重要的顯著變數只有 Vb1,Vb2。

2. 挑選出候選模型：決定候選模型就是在決定剔除變數的順序，在所有變數都放入的情況之下，透過向後選變數（backward selection）的概念，陸續剔除最不重要的變數，但是直至模型剩下最後一個變數才停止，並非完全等於向後選變數的方式。根據向後選變數的原理，會先從最不重要的變數開始剔除，也就是該參數的統計量最小的開始移除，最後模型剩下的變數通常是最重要，其參數的統計量也最大。例如，如果變數有 10 個，則候選模型就會有 10 個，其中第一個模型會包含所有 10 個變數，第二個模型會剔除 1 個最不重要的變數剩下 9 個，第三個模型剩下 8 個，以此類推，到第十個模型會剩下 1 個。
3. 算出候選模形的選模準則：根據前個步驟算出的候選模型，分別計算不同精簡程度的 R^2 成分。
4. 根據選模準則挑選最佳模型：不同精簡程度 R^2 成分挑選最佳模型的方式都是依據最大值（AIC 系列則皆是最小值），並記錄下該模型。
5. 重複執行一到四步驟，直到執行完 1000 次，觀察各種模型的比例。

這邊要注意，候選模型可能不包含以上指定的完整模型或精簡模型，因為何種變數會納入候選模型是交給向後選模的方式決定，不直接指定候選模形而

是透過選變數的方式的原因，是考量實務上變數量多的情形，因此模擬結果也會收到向後選變數的影響，這也比較貼近實務的情形。

以下將不同精簡模型的類型區分為三種，每一種都與要比較的完整模型的參數進行比較。首先先比較當隨機已知時，只選固定；再來是已知固定只選隨機部分；最後同時選固定和隨機。

表 1: Table to type1 for R_M.

	-R	R	F	O	M
$\alpha = 0$	NA	0.004	0.002	0.750	0.244
$\alpha = 0.1$	NA	0.089	0.053	0.556	0.301
$\alpha = 0.2$	NA	0.402	0.103	0.257	0.238
$\alpha = 0.3$	NA	0.711	0.071	0.101	0.117
$\alpha = 0.4$	NA	0.915	0.040	0.012	0.034
$\alpha = 0.5$	NA	0.990	0.008	NA	0.002
$\alpha = 0.6$	0.006	0.994	NA	NA	NA
$\alpha = 0.7$	0.034	0.966	NA	NA	NA
$\alpha = 0.8$	0.150	0.850	NA	NA	NA
$\alpha = 0.9$	0.495	0.505	NA	NA	NA
$\alpha = 1$	0.905	0.095	NA	NA	NA

表 2: Table to type1 for AIC.

	-R	R	F	O	M
AIC	NA	0.055	0.050	0.483	0.412
AICc	NA	0.059	0.053	0.463	0.424
mAIC	NA	0.891	0.034	0.002	0.073
cAIC	0.103	0.198	0.030	0.352	0.317

第二節　實際資料

表 3: Table to type2 for R_v.

	-R	R	F	O	M
$\alpha = 0$	NA	0.002	0.212	0.786	NA
$\alpha = 0.1$	NA	0.004	0.358	0.638	NA
$\alpha = 0.2$	NA	0.004	0.544	0.452	NA
$\alpha = 0.3$	NA	0.012	0.738	0.250	NA
$\alpha = 0.4$	NA	0.042	0.826	0.132	NA
$\alpha = 0.5$	NA	0.176	0.784	0.040	NA
$\alpha = 0.6$	NA	0.484	0.510	0.006	NA
$\alpha = 0.7$	NA	0.924	0.076	NA	NA
$\alpha = 0.8$	0.006	0.990	0.004	NA	NA
$\alpha = 0.9$	0.082	0.918	NA	NA	NA
$\alpha = 1$	0.794	0.206	NA	NA	NA

表 4: Table to type2 for AIC.

	-R	R	F	O	M
AIC	NA	NA	0.914	0.086	NA
AICc	NA	NA	0.932	0.068	NA
mAIC	0.100	0.900	NA	NA	NA
cAIC	NA	NA	0.680	0.320	NA

第五章　結論

過往有許多研究，...。

另外，這邊建議選變數的方式為向後或逐步向後（stepwise backward selection），比起向前（forward selection）和逐步向前（stepwise forward selection），向後比較能夠保留可能顯著的變數，但當樣本數遠小於變數個數時，向後的方式可能會估計不出來，此時向前和逐步向前的方式就會比較好，另一個解決這個情形的方式是納入懲罰項進去來解決資料維度較高的情形，也就是 Lasso 迴歸，但 Lasso…(待補)。倘落變數不多或是沒有時間成本的考量，也可考慮所有集合選變數的方式（Best subset selection）。

附錄一、R2 轉變

表 5: type1Rchange

	Mean(SD，Min，Max)		
	V3	V2	V1
1	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)
2	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.56 (0.05,0.39,0.70)
3	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.55 (0.05,0.38,0.70)
4	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.01 (0.01,0.00,0.04)
5	0.58 (0.05,0.42,0.73)	0.15 (0.03,0.07,0.25)	0.13 (0.02,0.07,0.23)
6	0.27 (0.03,0.17,0.38)	0.27 (0.03,0.18,0.38)	0.26 (0.03,0.18,0.38)

表 6: type2Rchange

	Mean(SD，Min，Max)		
	V3	V2	V1
1	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)
2	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.56 (0.05,0.39,0.70)
3	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.55 (0.05,0.38,0.70)
4	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.01 (0.01,0.00,0.04)
5	0.58 (0.05,0.42,0.73)	0.15 (0.03,0.07,0.25)	0.13 (0.02,0.07,0.23)
6	0.27 (0.03,0.17,0.38)	0.27 (0.03,0.18,0.38)	0.26 (0.03,0.18,0.38)

表 7: type3Rchange

	Mean(SD，Min，Max)			
	V3	V2	V1	V4
1	0.81 (0.03,0.73,0.88)	0.92 (0.01,0.89,0.95)	0.81 (0.03,0.72,0.88)	0.92 (0.01,0.88,0.95)
2	0.11 (0.05,0.01,0.31)	0.11 (0.05,0.01,0.31)	0.53 (0.05,0.34,0.68)	0.53 (0.05,0.35,0.67)
3	0.11 (0.05,0.01,0.31)	0.11 (0.05,0.01,0.31)	0.53 (0.05,0.34,0.68)	0.53 (0.05,0.35,0.67)
4	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)
5	0.58 (0.05,0.41,0.70)	0.58 (0.05,0.43,0.71)	0.15 (0.03,0.08,0.25)	0.15 (0.02,0.09,0.24)
6	0.13 (0.02,0.07,0.20)	0.23 (0.03,0.15,0.34)	0.13 (0.02,0.07,0.20)	0.24 (0.03,0.15,0.33)

附錄二、精簡模型與完整模型模擬參數比較

表 8: type1beta

	Mean(SE，SD，CP(95%))		
	V3	V2	V1
1	0.01 (1.14,1.20,0.94)	-0.02 (0.56,0.59,0.93)	-0.02 (0.53,0.55,0.94)
2	NA (NA,NA,NA)	-0.00 (0.56,0.58,0.94)	-0.01 (0.53,0.54,0.94)
3	NA (NA,NA,NA)	-0.00 (0.56,0.59,0.93)	-0.01 (0.53,0.56,0.94)
4	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.01 (0.53,0.57,0.93)
5	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.00 (0.54,0.56,0.93)
6	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.00 (0.51,0.52,0.94)

表 9: type2T

	Mean(SD，CP(95%))		
	V3	V2	V1
1	0.57 (5.18,0.97)	0.32 (4.57,0.96)	0.35 (4.42,0.95)
2	0.52 (4.13,0.94)	0.27 (3.66,0.96)	0.31 (3.62,0.96)
3	0.64 (4.41,0.97)	0.84 (4.07,0.98)	0.75 (4.06,0.96)
4	NA (NA,NA)	-0.09 (3.73,0.97)	-0.13 (3.67,0.95)
5	NA (NA,NA)	NA (NA,NA)	0.06 (0.90,0.97)
6	NA (NA,NA)	NA (NA,NA)	0.02 (0.63,0.95)

表 10: type3betaT

	Mean(SE，SD，CP(95%))			
	V3	V2	V1	
1	0.03 (1.17,1.16,0.95)	0.04 (1.15,1.15,0.95)	0.04 (0.60,0.61,0.94)	0.05 (0.57,0.60,0.93)
2	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.02 (0.60,0.62,0.94)	-0.01 (0.57,0.61,0.93)
3	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.00 (0.61,0.63,0.94)	0.01 (0.58,0.61,0.94)
4	104.72 (NA,19.82,0.00)	104.77 (NA,19.32,0.00)	4.43 (NA,5.79,0.90)	4.41 (NA,5.38,0.90)
5	0.44 (NA,7.24,0.49)	0.51 (NA,6.88,0.46)	0.66 (NA,4.10,0.94)	0.68 (NA,3.82,0.94)
6	1.32 (NA,4.98,0.93)	1.25 (NA,4.50,0.93)	1.33 (NA,4.94,0.94)	1.25 (NA,4.49,0.94)
7	NA (NA,NA,NA)	0.63 (NA,3.93,0.95)	NA (NA,NA,NA)	0.63 (NA,3.93,0.95)