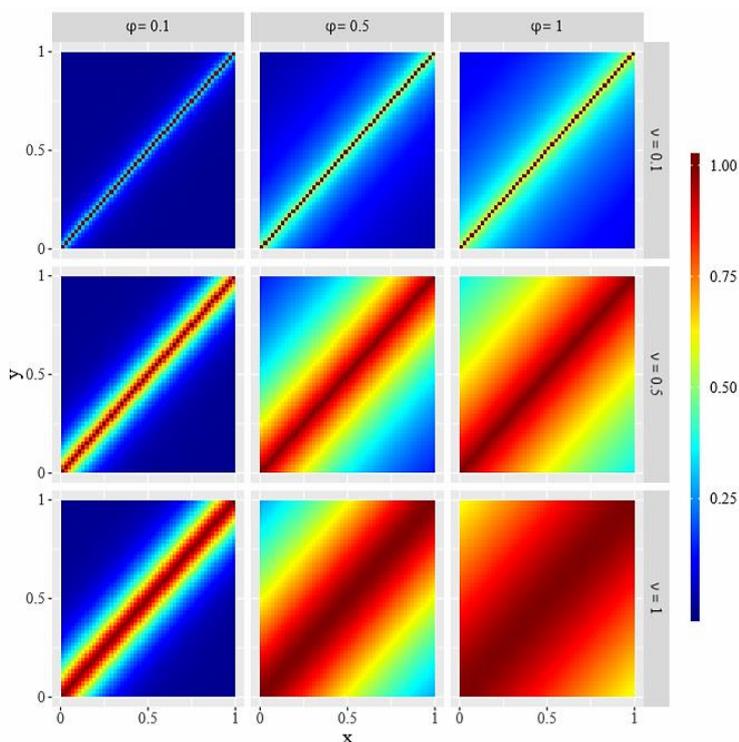


## 空間迴歸模型中隨機效應與解釋變數存在相關性時 迴歸係數估計的探討

國立彰化師範大學數學系暨統計資訊研究所

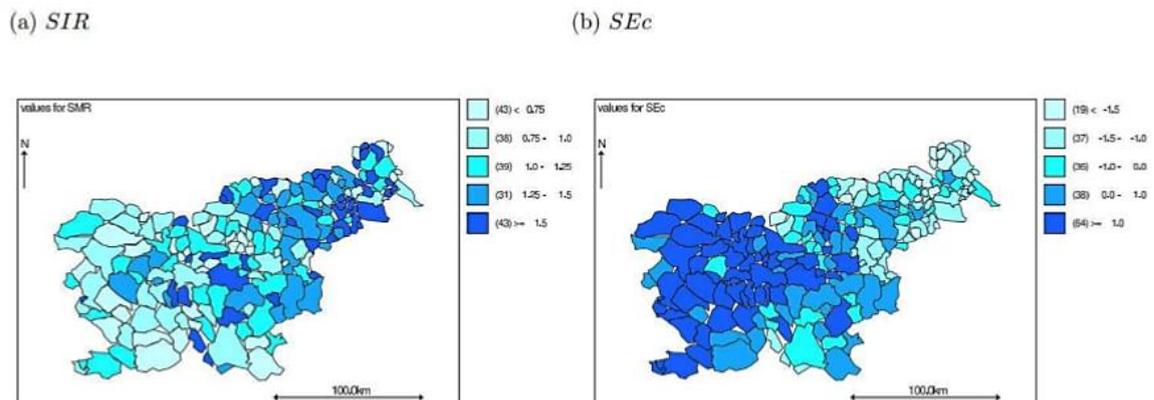
邱詠惠、陳春樹

空間迴歸模型應用甚廣，舉凡具空間位置相關的資料皆可使用它來做分析，例如：空氣中 PM2.5 的濃度預測、各縣市登革熱患病人數的探討、模糊影像資料的重建…等。由於資料在空間上的分布隱含不可觀測的相關性結構，因此空間迴歸模型相較於一般的線性迴歸模型多考慮了資料彼此間的空間相關性於模型中，即所謂的空間隨機效應項(spatial random effect)，因此也增添空間迴歸模型的應用廣度。隨著資料型態的複雜與多樣性，空間隨機效應項的分布設定也有所不同，但大體上皆以距離相近的觀測值相較於距離較遠的觀測值有較強的相關性為主要核心思想。以地震為例，與震央較近的地方其芮氏規模震度較大，隨著距離震央的距離越遠則震度規模會有遞減的趨勢，亦即相關性越小。在實務的應用層面，Matérn 相關性函數(Matérn correlation function) (Matérn 2013)常使用於描述空間隨機效應項的空間相關結構，函數中的  $\nu$  與  $\phi$  分別為平滑參數與尺度參數，( $\nu$ ,  $\phi$ ) 的值越大表示背後的相關性結構越強，此函數亦反應距離相近的觀測值具有較強的相關性(圖一)，隨著觀測值之間的距離越遠，其相關性有遞減的趨勢。



圖一：在不同參數值( $\nu$ ,  $\phi$ )之下，距離為  $|x - y|$  時的 Matérn 相關性函數值。  
圖中函數值越大(紅色)，代表資料空間相關性越強。

使用迴歸模型分析資料主要是想了解感興趣的解釋變數與反應變數之間的關係，進一步也可以針對未抽樣的位置進行預測，使得研究者能依據建模與分析結果執行相對應的因應措施或擬定改善計畫。然而如 Zadnik 與 Reich (2006) 的文章所提及，斯洛維尼亞國家(Slovenia)的人民之胃癌標準化發生比(SIR)與社會經濟地位(SEc)應呈現負相關的趨勢(如圖二所示)，同時透過線性迴歸模型的配適結果亦得到顯著的負值迴歸係數(亦即 SIR 與 SEc 呈現顯著負相關)。但是透過空間迴歸模型並藉由貝氏方法估計模型參數所得到的社會經濟地位之迴歸係數估計值為-0.02，且在 95%的信心水準下可得此係數的可信區間為(-0.10, 0.06)。此結果表示此迴歸係數估計值-0.02 並不顯著，甚至有可能為正值，這似乎與實務上的認知違背且可能讓研究者做出錯誤的推論。



圖二：斯洛維尼亞國家每個城市的胃癌標準化發生比(SIR)與社會經濟地位(SEc)分布圖。  
此圖摘錄自 Reich et al. (2006)。

此現象的發生可能是因為社會經濟地位觀測值是隨著此國家的城市地理位置而收集，背後可能存在某種未知的空間相關性結構，此時可能與空間迴歸模型中的隨機效應項產生共線性，在空間統計中亦稱之為空間混淆效應(Spatial confounding effect)，此現象將導致不準確的迴歸係數估計量(e.g., Page et al. 2017)。實務上若假設觀測值間彼此相互獨立，線性迴歸模型之迴歸係數可藉由最小平方方法(LS)得之，然而存在隨機效應項的空間迴歸模型則因觀測值間非獨立，此時迴歸係數的估計值可使用加權最小平方方法(WLS)得之。但是當空間混淆效應存在時，直接使用最小平方方法或加權最小平方方法估計迴歸係數均會產生偏誤的估計結果(Page et al. 2017; Chiou 2019)。假如空間迴歸模型中的解釋變數與隨機效應項存在共線性，並將它們之間的相關性記作  $\rho$ ，我們藉由模擬實驗得知，當  $\rho > 0$  時，使用 LS 或 WLS 估計迴歸係數會有高估的現象(表一)。反之，當  $\rho < 0$  時，則迴歸係數會有低估的現象(表二)。實務上，空間混淆效應並不容易發現，因此導致分析上的疏忽甚至可能造成結論的誤判。為了減輕空間混淆效

應所造成的影響，我們介紹一個迴歸係數估計的修正方法，稱之為 Adj-GLS。

假設迴歸模型中只有兩個解釋變數且只有第二個解釋變數  $x_2$  與隨機效應項  $W$  存在共線性(空間混淆)，並假設  $x_2$  與  $W$  皆來自平均數為零，變異數分別為  $\sigma_x^2 R_x$  與  $\sigma_w^2 R_w$  的多維常態分佈，其中空間相關矩陣  $R_x$  與  $R_w$  皆由 Matérn 相關函數構造。則 Adj-GLS 方法的迴歸係數估計式為

$$\hat{\beta}_{Adj} = (X'X)^{-1}X'Y - \rho \frac{\sigma_w}{\sigma_x} M_{Adj} (X'X)^{-1}X'R_w^{1/2}R_x^{-1/2}x_2$$

上式中  $X$  為解釋變數矩陣， $Y$  為反應變數向量。 $\hat{\beta}_{Adj}$  可視為 LS 迴歸係數估計式  $(X'X)^{-1}X'Y$  與修正量  $\rho \frac{\sigma_w}{\sigma_x} M_{Adj} (X'X)^{-1}X'R_w^{1/2}R_x^{-1/2}x_2$  的組合。此修正量的修正方向將隨著  $\rho$  值的正負而自動調整。亦即當  $\rho > 0$  時，修正量為正，此時將能下修被 LS 估計法高估的部分，使得迴歸係數估計值  $\hat{\beta}_{Adj}$  較為準確，反之亦然。其中  $M_{Adj}$  是對角元素為(0,0,1)的對角矩陣，表示此方法只修正與隨機效應項有相關性的解釋變數  $x_2$  之迴歸係數。下列表一與表二的模擬數據是在不同  $R_x$  的參數值( $\nu_x, \varphi_x$ )設定下，分別模擬  $\rho = 0.5$  及  $\rho = -0.5$  時的迴歸係數估計結果，其中迴歸係數真值設定為  $\beta_0 = \beta_1 = \beta_2 = 1$ 。

表一：模擬當  $x_2$  與隨機效應項存在空間混淆且相關性  $\rho = 0.5$  時，使用三種方法(LS、WLS、Adj-GLS)對迴歸係數做估計的 400 次模擬平均值，括號內為 400 次估計結果的標準差。此外， $\sigma_x^2 = \sigma_w^2 = 1$ ，以及  $R_w$  中參數( $\nu_w, \varphi_w$ ) = (0.5, 1)。

$\nu_x$	$\varphi_x$	LS			WLS			Adj-GLS		
		$\beta_0 = 1$	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_0 = 1$	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_0 = 1$	$\beta_1 = 1$	$\beta_2 = 1$
0.1	0.5	1.0201 (0.0777)	0.9969 (0.0077)	1.4170 (0.0121)	1.0189 (0.0729)	1.0010 (0.0057)	1.2756 (0.0069)	1.0201 (0.0777)	0.9969 (0.0077)	0.9412 (0.0121)
0.5	1	1.0740 (0.0727)	1.0015 (0.0070)	1.9625 (0.0311)	1.0757 (0.0656)	1.0012 (0.0056)	1.9960 (0.0174)	1.0740 (0.0727)	1.0015 (0.0070)	0.9625 (0.0311)
1	0.5	1.0771 (0.0739)	1.0024 (0.0071)	1.8923 (0.0370)	1.0838 (0.0660)	1.0015 (0.0057)	2.1757 (0.0240)	1.0771 (0.0739)	1.0024 (0.0071)	0.9327 (0.0369)

如表一所示，在  $\rho = 0.5$  時，沒有空間混淆效應影響的迴歸係數  $\beta_0$  和  $\beta_1$ ，在三種估計方法(LS、WLS、Adj-GLS)下都有不錯的估計結果，但有空間混淆效應影響的迴歸係數  $\beta_2$ ，在 LS 和 WLS 兩種常用的估計法下皆呈現高估的現象，而 Adj-GLS 除了能下修 LS 估計值外，亦使估計結果相當靠近真值 1。

表二呈現  $\rho = -0.5$  時的估計結果，數據亦顯示對於沒有空間混淆效應影響的迴歸係數  $\beta_0$  和  $\beta_1$ ，在三種估計方法下仍然有不錯的估計表現，但是對於有空間混淆效應影響的迴歸係數  $\beta_2$ ，使用 LS 和 WLS 估計法皆有低估的現象，甚

至如前述 SIR 與 SEC 所遇到的問題一樣，LS 和 WLS 將真實係數  $\beta_2 = 1$  估計為負值，此現象將嚴重影響後續的分析與推論。而使用 Adj-GLS 估計法除了上修 LS 低估的估計值外，同時也提供較為接近真值 1 的估計結果。

**表二：** 模擬當  $\mathbf{x}_2$  與隨機效應項存在空間混淆且相關性  $\rho = -0.5$  時，使用三種方法 (LS、WLS、Adj-GLS) 對迴歸係數做估計的 400 次模擬平均值，括號內為 400 次估計結果的標準差。此外， $\sigma_x^2 = \sigma_w^2 = 1$ ，以及  $\mathbf{R}_w$  中參數  $(\nu_w, \phi_w) = (0.5, 1)$ 。

$\nu_x$	$\phi_x$	LS			WLS			Adj-GLS		
		$\beta_0 = 1$	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_0 = 1$	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_0 = 1$	$\beta_1 = 1$	$\beta_2 = 1$
0.1	0.5	1.1249 (0.0756)	0.9990 (0.0081)	0.5538 (0.0126)	1.1310 (0.0704)	1.0011 (0.0057)	0.7136 (0.0067)	1.1249 (0.0756)	0.9990 (0.0081)	1.0297 (0.0122)
0.5	1	1.0740 (0.0727)	1.0015 (0.0070)	-0.0375 (0.0311)	1.0757 (0.0656)	1.0012 (0.0056)	-0.0040 (0.0174)	1.0740 (0.0727)	1.0015 (0.0070)	0.9625 (0.0311)
1	0.5	1.0839 (0.0734)	1.0023 (0.0069)	0.0056 (0.0375)	1.0627 (0.0661)	1.0013 (0.0056)	-0.1557 (0.0239)	1.0839 (0.0734)	1.0023 (0.0069)	0.9653 (0.0380)

藉由上述的討論與模擬結果，我們可以清楚了解，當分析空間型態的資料時，若存在空間混淆效應於解釋變數和隨機效應項之間時，常用的 LS 和 WLS 估計法將對迴歸係數估計值產生嚴重程度不等的偏誤，因而可能影響後續的分析、推論與決策。因此對估計式做適當的調整是有其必要性，此時建議使用 Adj-GLS 的方法來估計迴歸係數，將有利於結果的分析與判斷。此部分的相關議題仍相當多且重要，而其所蘊含的理論及應用問題相當廣泛且有趣，期待更多有興趣的研究者與學生投入解決。

## 參考文獻

1. Chiou, Y. H. (2019). On fixed effects estimation for spatial regression under the presence of spatial confounding. Master Thesis.
2. Matérn, B. (2013). Spatial Variation. Springer Science & Business Media.
3. Page, G. L., Liu, Y., He, Z., and Sun, D. (2017). Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scandinavian Journal of Statistics*, 44, 780-797.
4. Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, 62, 1197-1206.
5. Zadnik, V. and Reich, B. J. (2006). Analysis of the relationship between socioeconomic factors and stomach cancer incidence in Slovenia. *Neoplasma*, 53, 103-110.