



## 行政院國家科學委員會專題研究計畫成果報告

計畫名稱：廣義排序容忍度的選取在迴歸分析的應用  
(Application of Tolerance Selection of the Extended Ranks in Regression Analysis)

計畫編號：NSC 87-2118-M-032-001  
執行期限：86年8月1日至87年7月31日  
主持人：陳主智（淡江大學數學系）

## 一、中文摘要

本研究利用排序轉換將迴歸分析中自變數  $X$  與因變數  $Y$  以其排序值代入，用以預估因變數的值，並計算樣本變異數以判斷估計的良窳。在排序轉換的過程，由於自變數可能受到一些與因變數無關的變數的干擾，統稱之為測量誤差。這些測量誤差會使得預估值值的變異增加，而影響準確度。因應於此，本研究採取以自變數的廣義排序轉換，而非一般排序的方式。此二者的差別在於廣義排序可允許容忍度的選取，以反應測量誤差的變異程度。

在容忍度  $e$  的選取方面，利用區域相關性的觀念，可以找到在一定範圍內  $X$  與  $Y$  會缺乏相關性，因而在此範圍內可以容許  $X$  值的變動，而不致影響  $Y$  值。在模式估計上，一部份的模式誤差也可預期被吸收至此容忍度  $e$  的變異，因而增加預測值的準確性。

理論部份的結果則是可以將誤差平方和分解成自變數  $X$  可解釋的部份，容忍度  $e$  所解釋的部份，與模式誤差。而各項之間可預期近似獨立。因而減少模式誤差，提高準確度。進一步的結果有待探討。

關鍵詞：排序轉換、測量誤差、區域相關性、模式誤差

## Abstract

This research applies the rank transformations of the dependent variable  $Y$  and the independent variable  $X$  in regression analysis to reduce the variability in model fitting caused by nonlinearity and outlier problems. In finding the relationship between  $Y$  and  $X$ , there may exist some factors such as measurement error in  $X$  which is uncorrelated with  $Y$ . As a result, the coefficient parameter is under-estimated and the precision is reduced.

In stead of replacing the ranks of  $X$  by their ordinary ranks, we employ extended

ranks (Chen , 1993) of X in model fitting. A scale called tolerance is plugged in the ranking of X, which is dependent on the correlation between Y and X, as well as the strength of other factors such as measurement error in X which is un-correlated with Y. By the concept of local correlation, the scale of tolerance is selected.

The sum of squared error is then decomposed into three parts: the variability explained by X, the variability explained by tolerance, and unexplained variability of model error. Each part of it is expected to be independent, and the precision in model fitting is expected to be enhanced by decomposing the model error using ordinary ranks of X further into two components using extended ranks of X: variability by tolerance and residual error.

Keywords: Rank Transformation, Measurement Error, Local Correlation, Model Error

## 二、緣由與目的

在迴歸分析與共變異數分析中，利用排序轉換是極為簡便且有效的方式 (Iman and Conover, 1979; Conover and Iman, 1981, 1982)。其最主要的優點為對於單調非線性函數的估計與檢定有很高的效度 (power) 且可避免非正常值 (outlier) 的影響。但是由於排序轉換並不具有相加性 (additive)，因此，學界在變異數分析上的應用較有所保留。基本上，在漸進相對效率及效度上，排序轉換的應用仍具有一定的優勢 (Akritas, 1990; Hora and Conover, 1984, 1988)。至於排序轉換在當自變數 X 有測量誤差時應用，文獻上提及極為有限 (Chen, 1997)。

在 Chen (1997) 一文中所探討的 X (有測量誤差) 為一般的排序轉換，相較於使用真實數據的最小平方法，排序估計的準確性在某些情況下會更高。但是實際上，與 Y 無關的測量誤差在模式係數估計時並沒有排除開。因此，若能事先有效排除來自此誤差的變異，則估計的準確性應可相對的提高。這是本研究採取廣義排序 (Chen, 1993) 利用容忍度的篩選，期望在估計時降低若干來自測量誤差的變異的原因。

在進一步的理論探討部份，類似於變異數分析的平方誤差和的分解，我們也預期排序轉換有近似的性質，並進而將模式誤差分解為容忍度可解釋的誤差與無法解釋的項，以更進一步縮減模式誤差，而提高估計準確性。

### 三、結果與討論

在結果方面，由模擬結果，確實顯示存在若干證據會在適當的容忍度之下，誤差平方和會比一般的排序轉換為小。但是尚無法找到其規則。利用區域相關性找適當容忍度也許可行，但仍須更多的模擬或理論結果支持。

至於排序轉換的誤差平方和的分解，相關的理論架構也尚待釐清，仍有極大的理論空間需要探討，期望在不久的將來能有豐碩的成果。

### 四、計畫成果自評

本計畫的成果正在整理之中，暫時編寫成技術報告。俟有更完整的理論結果，則將送審學術刊物發表。

大致而言，本研究結果已達成預期目標。但是能仍缺乏較完善的理論結果，以致無法立即送審發表，是為美中不足，並需要改進之處。

### 五、參考文獻

Akritis, M. G. (1990) "The rank transform method in some two-factor designs", *Journal of the American Statistical Association* 85, 73-78.

Chen, C. C. (1993) *Extended Rank and Matched-pair Analysis of Covariance*, Doctoral dissertation, Department of Biostatistics, University of North Carolina-Chapel Hill.

Chen, C. C. (1997) "Rank transformations when covariables are measured with error or mismodelled", *Communications in Statistics—Theory and Methods*, 26, 2967-2982.

Conover, W. J., and Iman, R. L. (1981) "Rank transformations as a bridge between parametric and nonparametric statistics", *The American Statistician*, 35, 124-129.

Conover, W. J., and Iman, R. L. (1982) "Analysis of covariance using the rank transformation", *Biometrics*, 38, 715-724.

Hora, S. C., and Conover, W. J. (1984) "The F statistic in the two-way layout with rank-score transformed data", *Journal of the American Statistical Association* , 79, 668-673.

Hora, S. C., and Conover, W. J. (1988) "Asymptotic relative efficiencies of the rank-transformation procedure in randomized complete block design", *Journal of the American Statistical Association* , 83, 462-470.