

行政院國家科學委員會專題研究計畫 成果報告

以無母數管制圖監控兩階段製程品質變數之探討 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 99-2118-M-004-007-
執行期間：99年08月01日至100年07月31日
執行單位：國立政治大學統計學系

計畫主持人：蔡紋琦
共同主持人：楊素芬
計畫參與人員：碩士班研究生-兼任助理人員：蔡瑋倫

處理方式：本計畫可公開查詢

中華民國 100 年 10 月 25 日

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

以無母數管制圖監控兩階段製程品質變數之探討

Using non-parametric cause-selecting charts to monitor dependent process steps

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 99-2118-M-004-007-

執行期間：2010 年 08 月 01 日至 2011 年 07 月 31 日

執行機構及系所：國立政治大學 統計學系

計畫主持人：蔡紋琦

共同主持人：楊素芬

計畫參與人員：蔡瑋倫

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：

赴國外出差或研習心得報告

赴大陸地區出差或研習心得報告

出席國際學術會議心得報告

國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

中 華 民 國 99 年 10 月 24 日

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

This project was planning to simulate data sets of different features of function f , describing the relationship between the in-coming and outgoing characteristics, to observe the behavior of the average run length of the proposed non-parametric cause-selecting charts under different approaches in estimating f . However, because of the time shortness, this part is not complete. Besides, the main nonparametric methods we adopted to estimate the function f for real data set were changed to B-spline and kernel regression, not including wavelet.

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

Most part of an article based on this project is complete. We plan to submit this article to Journal of Process Control.

壹、中文摘要

由於現今產品的複雜性，在產品的製造過程中，大部分都是經過多階段的製程才能完成。且一般而言，每一製程階段的品質特性通常亦會影響到下一或其後階段製程的品質特性，在這樣的情形下，並不適合只針對每一製程階段各自去畫自己的管制圖。Zhang在1984年提出了用選控圖來處理各階段品質特性間有相關性的製程。考慮最簡單的二階段製程，選控圖是針對移除掉第一階段品質特性影響的第二階段品質特性來畫一般的舒華特管制圖。而一般皆會假設第一階段的品質特性和第二階段的品質特性之間的關係為參數線性模型再加上常態分配的誤差項，然而實際上，這樣的假設有可能並不適宜。因此我們使用B樣條以及核迴歸來配適兩製程間之關係，拿掉了製程間須有參數或線性相關的假設，並使用區塊拔靴法及曲線深度來建構產品未失控的信賴帶，處理了資料為一時間序列及不放誤差項分配假設的狀況。最後再將此方法應用到無線感應資料的實例上，雖然實驗樣本數較小，但仍顯示出此方法有相當強的檢驗出製程出錯之能力。

貳、Abstract

Nowadays, most of the products are produced from several process stages and usually the quality characteristic at a current stage affects the quality characteristics at the next one or some subsequent stages. Therefore, it is not suitable to plot control charts in each stage respectively. To deal with this situation, Zhang proposed the so called cause-selecting control chart in 1984. Consider the simplest case of a two-stage process. The cause-selecting chart is based on the outgoing quality in the second stage that has been adjusted for the in-coming quality in the first stage. A linear or generalized linear model with normally distributed error terms is assumed to relate the two quality characteristics. In practice, however, it is often difficult to obtain substantial knowledge about the relationship function or the error terms. In this project, we use two non-parametric smoothing techniques including B-spline and kernel regression to describe the in-coming and outgoing quality characteristics in a two-dependent-stage process. Then block bootstrap and curve depth are used to construct simultaneous confidence band, especially for the data set which is time-related with error terms that are distribution free. Finally, a real data regarding wireless sensor is used to evaluate the efficiency of the proposed method. Although the number of experiments is not particularly large, the result strongly suggests that our proposed guide is powerful in detecting the out of control profiles.

參、關鍵詞

B 樣條、區塊拔靴法、選控圖、信賴帶、曲線深度、無母數剖面監控、核迴歸

肆、Keywords

B-spline, block bootstrap, cause-selecting control chart, confidence band, curve depth, nonparametric profile monitoring, kernel regression

伍、報告內容

Traditional profile monitoring methods may have some unrealistic assumptions. They include (i) linear parametric relationship or generalized linear relationship is assumed between the response variable and the explanatory variable(s); (ii) the error terms are normally distributed or follow some specific distribution; (iii) the within-profile data are independent. This project tries to propose a more flexible and computationally efficient method so that these unrealistic assumptions are not required. Based on the observed in-control profiles we wish to establish an adequate confidence band for the underlying functional relationship. Then this confidence band can serve as a control chart for phase II process monitoring. The proposed guide is mainly divided into six steps. Step 1: the two-sided median method, an automated approach, is used to clean each profile data set; step 2: an adequate B-spline or kernel regression is fitted to each profile data; step 3: the moving block bootstrap method is applied to generate several correlated samples for each profile data; step 4: again the B-spline or kernel regression is fitted to each of the bootstrap samples; step 5: the corresponding curve depths of fitted curves in step 4 are calculated and those fitted bootstrap curves with smaller curve depths are removed for each profile data respectively; step 6: The resulting confidence bands of all profiles are pooled so to obtain a simultaneous confidence band for the underlying functional relationship.

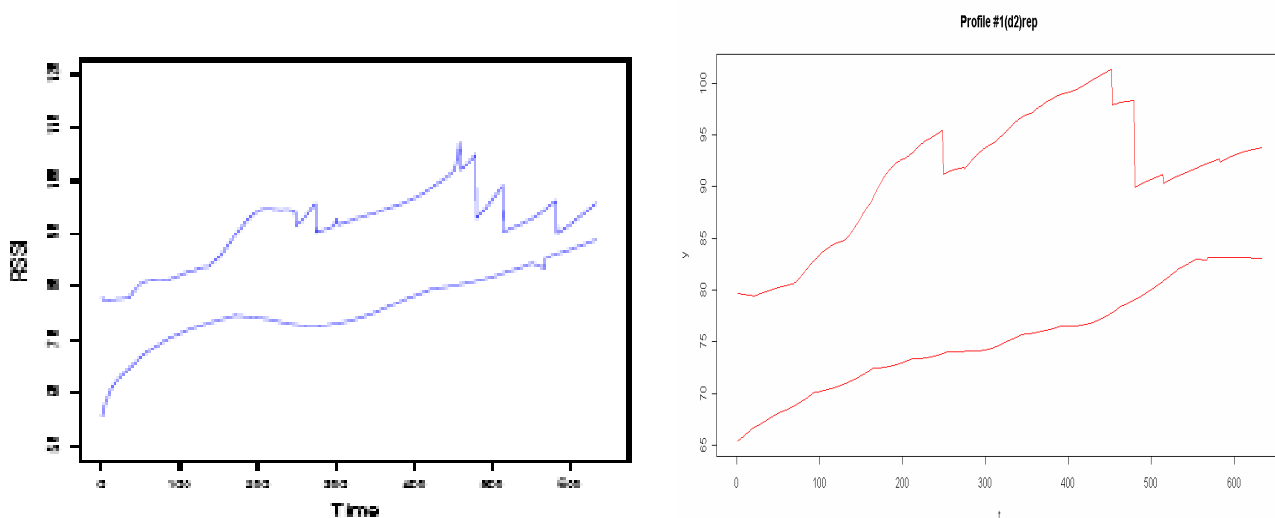
To simplify the formulation, we discuss the simple case of a two-stage process with only one covariate in the first stage. Suppose there are M independent profiles from a typical design of an in-control process and the i th profile has $n_{\{i\}}$ observations which are in a time-order. Let $y_{\{ij\}}$ be the measurement of j th observation of the i th profile in the second stage and $x_{\{ij\}}$ be the corresponding explanatory variable in the first stage. More precisely, $y_{\{ij\}}=f(x_{\{ij\}})+\epsilon_{\{ij\}}$ for $i=1,\dots,M$ and $j=1,\dots,n_{\{i\}}$, where f is a smooth function and $\epsilon_{\{ij\}}$'s, the associated error terms, can be dependent within profiles from some unknown distribution. Since unusual observations (outliers) for time series data can lead to intervention of analysis for the underlying process, we first take off those observations $y_{\{ij\}}$ which are far from the median of its neighborhood. This automated data cleaning approach, called two-sided median method, is proposed by Basu and Meckesheimer (2007).

In order to simplify the notation, we now denote the cleaned data of a particular profile by $\{x_{\{j\}},y_{\{j\}}\}$, $j=1,\dots,n$. and $x_1<\dots<x_n$. We then fit the data a smooth function \underline{f} by B-spline or kernel regression. The B-spline involves the choices of number and position for the knots. Since B-spline at any given point depends only on the observations falling in a fixed length of window. If we wish to re-compute the entire spline curve after one control point is changed, then only those terms whose window contains that point need to be recomputed. This important feature allows us employ the leave-one-out cross-validation strategy for choosing the optimal number of knots, if we consider equal knot spacing. The leave-one-out cross-validation score is defined as the mean square error of prediction (MSEP) as $MSPE(k)= [1/(n+1)] [(y_1-f_{(-1)}(x_1))^2+\dots+(y_n-f_{(-n)}(x_n))^2]$, where $f_{(-j)}$ is the obtained B-spline with k knots by removing the j th observation. Then the optimal number of knots is the k minimizes $MSPE(k)$. Similarly, the kernel regress involves the choice of optimal bandwidth h , which is a tradeoff between model fidelity and roughness. We again try to minimize the mean squared error of prediction to obtain the optimal h .

After obtain the fitted curve \hat{f} , either by B-spline or kernel regression, the residuals can be computed by $e_{\{j\}}=y_{\{j\}}-\hat{f}(x_{\{j\}})$. Since we allow a dependent structure between residuals and do not put any assumption on their distribution, we apply the moving block bootstrap (MBB), for which overlapping blocks of the same length are draw randomly with replacement. For the MBB block length, we use the diagnostic plot of the sample autocorrelation function (ACF) to decide this length, after which the values of ACF has a sharp decay.

To construct the simultaneous confidence band for the underlying function f , we first establish the bootstrap sampling percentile confidence band for each profile and then glue all the bands together. Suppose for i th profile, we using MBB generate B bootstrap samples. Then we fit the bootstrap samples by B-spline or kernel regression as described previously and obtain B fitted curves $\{f_{\{i1\}}, \dots, f_{\{iB\}}\}$. The confidence band for each profile can be constructed based on the curve depth, which was first proposed by Yeh (1996). The smaller the curve depth is, the further the curve is located from the benchmark curve. Therefore, we can accordingly exclude $100\alpha\%$ bootstrap curves with the lowest depths to obtain the $100(1-\alpha)\%$ bootstrap confidence band. Finally collecting all the confidence bands for all profiles together, we obtain the simultaneous confidence band for f over the entire data space.

We finally evaluate the proposed method by a real data set from wireless sensors. The babyfinder is a wireless sensor designed to monitor the physical or environmental conditions. When there is “distance” between the transceiver and the receiver, a wireless signal is generated and its strength measured (in decibels, dBs) by the Received Signal Strength Indicator (RSSI). With this simple design, one can monitor the occurrence of some “unexpected event” by observing the change of RSSI values over time. For example, suppose the babyfinder is designed to monitor the event that a bicycle is stolen. It is natural to assume there exists a functional relationship between the observed RSSI value and the corresponding point in time under the usual circumstance that the bicycle is not stolen. Once the bicycle is stolen, the observed RSSI values should reveal inconsistency with the original functional relationship. We collect 17 in-control data sets (the bicycle is not stolen) and 18 out-of-control data sets (the bicycle is stolen). Then we apply the proposed method to the 17 in-control data sets and obtain the following two 99% simultaneous confidence bands for f ; the left one is using B-spline and the right panel is using kernel regression. As one can see, both of the simultaneous



confidence bands are similar and not smooth. This is likely due to that each profile has different observed length. To evaluate the power of our proposed framework, we plot the fitted curves for in-control and out-of-control data sets together with the simultaneous confidence bands. If the fitted line falls outside the band at some point for an experiment, we issue an alarm then. The result is organized in the following tables. Table above is for B-spline and table below is for kernel regression. There is no false alarm. And out of the 18

In-control	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
false alarm																		
Out-of-control	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
No true alarm									X									

In-control	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
false alarm																		
Out-of-control	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
No true alarm	X										X	X						

out-of-control experiments, about 83% can be identified when kernel regression is used. When the B-spline is adopted, the power is even satisfactorily 94%. As to the reason why the two fitted approaches do not give true alarm to different experiments, it is still not clear yet.

We proposed a practical guide for monitoring nonparametric profiles with mild model assumptions. Furthermore, the numerical results from a real application show our proposed method is effective in detecting out-of-control profiles. However, there are still some potential problems. There is no criterion that can reasonably compare the confidence bands obtained from different modeling techniques. Also, the control chart we built is not in real time. These are obviously more challenging tasks.

陸、参考文献

[1]. Asadzadeh, S.; Aghaie, A.; and Shahriari, H. (2009). "Monitoring dependent process steps using robust cause-selecting control charts". *Quality and Reliability Engineering International*, vol. 25, pp. 851-874.

[2]. Basu, S. and Meckesheimer, M. (2007). Automatic outlier detection for time series: an application to sensor data, *Knowl. Inf. Syst.* 11, 137-154.

[3]. Boor, C. (2001). *A Practical Guide to Splines*, Revised Edition, Springer-Verlag.

[4]. Chatterjee, S. and Qiu, P. (2009). "Distribution-free cumulative sum control charts using bootstrap-based control limits". *Annals of Applied Statistics*, vol. 3, pp. 349-369.

[5]. Genovese, C. and Wasserman, L. (2005). "Confidence sets for nonparametric wavelet regression". *Annals of Statistics*, vol. 33, pp. 698-729.

[6]. Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall.

[7]. Hall, P (1992). "On bootstrap confidence intervals in nonparametric regression". *Annals of Statistics*, vol.

- 20, pp. 695-711.
- [8]. Hall, P. ; Horowitz, J.L. ; and Jing, B-Y. (1995). On blocking rules for the bootstrap with dependent data, *Biometrika* 82, 561-574.
- [9]. Härdle, W. (1990). "Applied nonparametric regression". Cambridge University Press.
- [10]. Hawkins, M. (1991). "Multivariate quality control based on regression-adjusted variables". *Technometrics*, vol. 33, pp. 61-75.
- [11]. Künsch, H.R. (1989). The Jackknife and the Bootstrap for General Stationary Observations, *Ann. Statist.* 17, 1217-1241.
- [12]. Mandel, B. (1969). "The regression control chart". *Journal of Quality Technology*, vol. 1, pp. 1-9.
- [13]. Molinari, N. ; Durand, J.F. ; and Sabatier, R.(2004) Bounded optimal knots for regression splines, *Comput. Statist. Data Anal.*, 45159-178.
- [14]. Nadaraya, E. A. (1964). "On Estimating Regression". *Theory of Probability and its applications*, vol. 9, pp. 141-142.
- [15]. Pena, D. (2001). Outliers, influential observations, and missing data, In: A course in time series analysis, Wiley, New York, pp. 136-170.
- [16]. Shu, L. and Tsung, F. (2003). " On multistage statistical process control". *Journal of Chinese Institute of Industrial Engineers*, vol. 20, pp.1-8.
- [17]. Shu, L.; Tsung, F.; and Kapur, K. (2004). "Design of Multiple Cause-Selecting Charts for Multistage Processes with Model Uncertainty". *Quality Engineering*, vol. 16, pp. 437-450.
- [18]. Wade, M. and Woodall, W. (1993). "A review and analysis of causing-selecting control charts". *Journal of Quality Technology*, vol. 25, pp. 161-169.
- [19]. Watson, G. (1964). "Smooth regression analysis". *Sankhya Ser. A*, vol. 26, pp.101-116.
- [20]. Yang, S. (1998). "Optimal process control for multiple processes". *Quality and Reliability Engineering International*, vol.14, pp. 347-355.
- [21]. Yang, S. and Su, H. (2007). " Adaptive control schemes for two dependent process steps". *Journal of Loss Prevention in the Process Industries*, vol. 20, pp. 15-25
- [22]. Yeh, A.B. (1996). Bootstrap percentile confidence bands based on the concept of curve depth, *Comm. Statist. Simulation Comput.*, 25, 905-922.
- [23]. Zhang, G. X. (1984). "A new type of control charts - cause-selecting control charts and a diagnosis theory with control charts", *Proceedings of World Quality Congress '84*, pp.175-85.

國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/20

國科會補助計畫	計畫名稱: 以無母數管制圖監控兩階段製程品質變數之探討
	計畫主持人: 蔡紋琦
	計畫編號: 99-2118-M-004-007- 學門領域: 工業統計
無研發成果推廣資料	

99 年度專題研究計畫研究成果彙整表

計畫主持人：蔡紋琦		計畫編號：99-2118-M-004-007-					
計畫名稱：以無母數管制圖監控兩階段製程品質變數之探討							
成果項目		量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）	
		實際已達成數（被接受或已發表）	預期總達成數（含實際已達成數）	本計畫實際貢獻百分比			
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 （本國籍）	碩士生	1	1	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	1	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 （外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

<p>其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p>	<p>無</p>
----------------------------------------------------------------------------------------	----------

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

請見精簡報告(未超過 200 字, 但卻無法輸入)

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：(以 100 字為限)

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

Most part of an article based on this project is complete. We plan to submit this article to Journal of Process Control.