



# HOKKAIDO UNIVERSITY

Title	PLRのある種の改良
Author(s)	杉村, 徹; Sugimura, Tooru; 佐藤, 義治 他
Citation	北海道大學工學部研究報告, 81, 93-98
Issue Date	1976-08-30
Doc URL	<a href="https://hdl.handle.net/2115/41380">https://hdl.handle.net/2115/41380</a>
Type	departmental bulletin paper
File Information	81_93-98.pdf



## PLR のある種の改良

杉 村 徹\* 佐 藤 義 治\* 河 口 至 商\*  
(昭和 51 年 3 月 31 日受理)

### An Improvement of PLR

— A Piecewise Linear Regression Method —

Tooru SUGIMURA, Yoshiharu SATO and Michiaki KAWAGUCHI  
(Received March 31 1976)

#### Abstract

McGee-Carleton's PLR is suitable for estimating plural switching points. In this article, the authors have attempted to improve the PLR method, taking advantage of the chi-square distribution of goodness-of-fit: Potential clusters are fixed when they are smaller than or equal to the specified percentage point. When the newest fixed clusters intersect with each other, these will be combined if the probability as a result of F-test is small enough. The improved PLR is examined with a few data.

#### 1. は じ め に

時系列データにおいて、その時系列のいくつかの時点において変動が生じているために、その時系列データの回帰分析を適当ないくつかの区間に区切ってその区間ごとに行う場合がある。

このことに関して次の 2 つの立場がある。その 1 つは、区間と区間の境界となる時点 (switching point) を観測者から与えられたものとして扱う立場である。残りの 1 つは、switching point を与えられたものとしてではなく、時系列のデータから推定する立場である。後者の立場は Quandt<sup>3)</sup> (1958) に始まるが、彼の問題にした複数個の switching points を取り扱うことに関して、McGee-Carleton<sup>1)</sup> (1970) は階層的クラスター分析の手法を導入することで switching points の個数を決定することに 1 つの解決法を与えた。それは **Piecewise Linear Regression** (略記 PLR) と言われる。(区分的線形回帰分析と訳す)

PLR の手法を

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (t=1, 2, \dots, N)$$

但し、 $\varepsilon_t \sim N(0, \sigma^2)$ 、互に独立の場合について説明する。その手続は次の通りである。

i) 3 個の時点からなる (1, 2, 3), (2, 3, 4), ..., (N-2, N-1, N) の N-2 個の **type-0** の **potential cluster** (PC) を構成する。(これを **minimal cluster** (略記 MC) と呼ぶ)

ii) 各 PC ごとに  $\beta_0, \beta_1$  を推定し、それによる残差の平方和を自由度で除した値  $\phi$  を求める。 $\phi$  をその PC の goodness-of-fit という。

iii)  $\phi$  の最小な値をもつ PC を決定する (fix する)。この PC を **newest fixed cluster** (略記 NFC) という。

\* 情報工学専攻 情報数理第一講座

iv) NFC と交わりのある PC を消去し、次のように新たな PC を構成する。

NFC を  $(t_1, \dots, t_2)$  とすると、時点  $t_1-1$  が fixed cluster に含まれていないとき、その点と NFC から **type-1** の PC を構成する。他の場合は、 $t_1-1$  を含む fixed cluster と NFC から **type-3** の PC を構成する。時点  $t_2+1$  についても、時点  $t_1-1$  と同様に、**type-2** 又は **type-3** の PC が構成される。

v) NFC が  $N$  個の時点から構成されているならばここで手続きを終る (最終 level)。その他の場合は ii) へもどって手続きを続行する。

以上の i) から v) の手続きの中で、NFC を1つ決定するごとに level が1つ更新される。但し、i) での level は0である。また、fixed された PC を **fixed cluster** (略記 FC) といい、FC に含まれない時点を **isolated point** (略記 IP) という。

i) から v) の手続き中の各 level で、時系列は FC によって何個かの区間 (regime) に分割される。どの分割の仕方、即ちどの level を採用するかは、最終的には、観測者の決定に任されることにはなるが、しかしその際の判断基準 (decision rule) として McGee-Carleton は次の2つをあげている。

- (a) 採用しようとする level での IP の総数が与えられた一定の数以下であること
- (b) PC を fix したときの回帰係数の共通性についての検定結果の考慮

PLR の概念はおおよそ以上に述べたごとくであるが、改良にあたり着目された点及び改良点<sup>4)</sup>、goodness-of-fit  $\phi$  が  $\chi^2$  分布するのであるから、 $\phi$  が最小な PC のみを fix するだけでなく、何らかの基準以下の  $\phi$  をもつ PC も同時に fix することにした点である (これを PLR 1 と呼ぶ)。さらに PLR 1 を検討し、次の8つに問題点を整理した<sup>4)</sup>。

- (1) 多くのデータを使つての検討の必要。Monte Carlo test も必要。
- (2)  $\mu$  の変化、MC の size の変化を加えた影響を調べること。
- (3) decision rule を加味して、decision rule と  $\lambda$  との関係の追求。
- (4) PLR 1 の手続きを他の方法に変更してその比較検討を行う。
- (5) PLR と PLR 1 とのさらに詳細な比較検討。
- (6) PLR を含めての理論的な側面の追求の必要。
- (7) 多くの実的なデータを利用するの検討、及び実用性の検討。
- (8) 他の統計的手法との比較検討の必要。

本報告は上の問題点のうち、(1) と (4) にしほり、PLR 1 に変更を加えたものを PLR 2 と呼び、これを用いて今後の指針をさぐってみた。

## 2. PLR 1 と PLR 2

PLR の手続き ii) で計算される goodness-of-fit は  $\chi^2$  分布をする：

$$\phi \sim \chi^2(n-2)\sigma^2/(n-2)$$

但し、 $n$  は PC に含まれる時点の総数。

そこで PLR 1 では  $\lambda$  という値を与えておき、そのパーセント点以下の  $\phi$  をもつ PC を同時に fix する。従つて  $\lambda$  を小さく与えると PC の fix される確率が増加し、最終 level までの level の数が減少する。 $\lambda$  を大きく与えるとその逆の場合が生じる。このことにより  $\lambda$  による調節が可能となる。

しかし PLR 1 では複数個の PC を同時に fix するために、NFC 同士の交わりが生ずること



促進を control できる。しかし、PLR 1 では  $\lambda$  のパーセント点以下の  $\phi$  をもつ PC が存在しなかった level から後の level では通常の PLR の手続きを取るようにしてあるので、高い level では  $\lambda, \mu$  による効果が得られなくなることが多いと考えられる。実際に PLR 1 のデータを用いての検討の場合、そのことが生じている。そこで今回はその点に着目し、 $\lambda$  のパーセント点以下の  $\phi$  をもつ PC が存在しない level のみに通常の PLR の手続きを実行させるよう、手続きを変更した。即ち、その level だけ最小の  $\phi$  をもつ PC を 1 個 fix させる。この手続きを変更した PLR 1 を、区別するために **PLR 2** と言うことにする。

PLR 2 の General flow chart は Fig. 2 に示した。この program の実行は北大大型計算機センターの FACOM 230-75 を利用した。注意として、PLR 2 は、PLR 及び PLR 1 としても命令を変更するだけで使える。

### 3. データによる PLR 2 の検討

PLR 1 の場合は Fig. 3 の #1 についてのみ検討を行ったが、今回の PLR 2 は Fig. 3 及び

No.	McGee-Carleton's Data				Quandt's Data	
	X	Trial #1 Y	X	Trial #2 Y	X	Y
1	17	22.181	9	13.909	4	3.473
2	17	24.311	10	14.539	13	11.555
3	22	26.812	12	15.919	5	5.714
4	13	19.115	20	26.294	2	5.710
5	17	23.172	15	22.954	6	6.046
6	24	27.378	8	11.835	8	7.650
7	21	25.428	14	17.878	1	3.140
8	20	18.121	21	20.878	12	10.312
9	13	16.258	18	16.035	17	13.353
10	21	17.806	19	18.326	20	17.197
11	16	17.678	13	14.396	15	13.036
12	5	11.469	24	19.965	11	8.264
13	19	18.950	24	22.047	3	7.612
14	14	18.179	7	8.914	14	11.802
15	7	9.815	23	24.331	16	12.551
16	11	13.058	24	25.127	10	10.296
17	7	7.721	5	6.923	7	10.014
18	14	16.506	9	12.333	19	15.472
19	19	21.998	16	16.908	18	15.650
20	5	6.795	7	8.380	9	9.871

Fig. 3

データ	$\lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
TRIAL #1		2	4	4	4	3 7	3 7	4 9	2 13	2 13
TRIAL #2	なし	なし	4 8	5 9	5 9	5 11	5 13	2 13	2 13	2 13
Quandt's Data		2	2	2	なし	3	3 5	3 7	3 5	3 7 8

Fig. 5  $\lambda$  のパーセント点以下の  $\phi$  をもつ PC が存在しなかった level

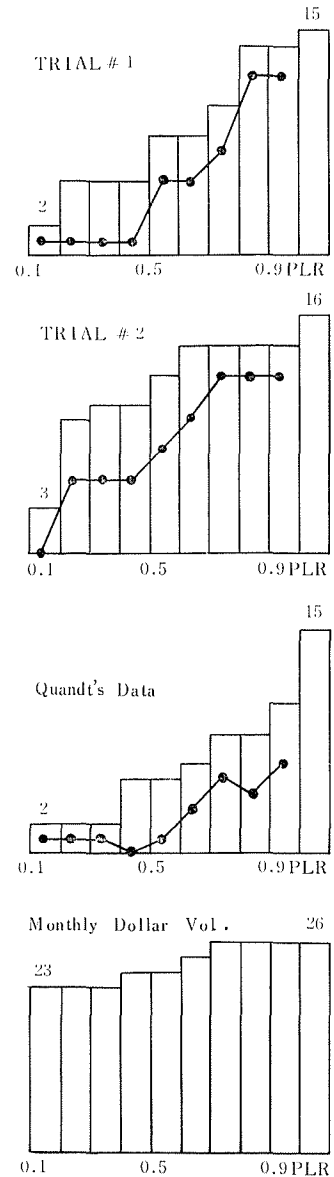
Fig. 4 にある 4 つのデータを試みることにする。全て McGee-Carleton が用いたものであるが、とくに Quandt's data は Quandt が最初に用いたものである<sup>3)</sup>。

$\lambda$  は 0.1 から 0.9 まで 0.1 間隔で変化させて調べた。Fig. 5 は各  $\lambda$  についての  $\lambda$  のパーセント点以下の  $\phi$  をもつ PC が存在しなかった level である。この場合 PLR 2 では  $\phi$  が最小な PC を fix する。Fig. 6 は各  $\lambda$  についての level の数と  $\lambda$  パーセント点以下の  $\phi$  をもつ PC

No.	月 (t)	New York and American stock exchanges (x)	Boston stock exchange (y)
1	'67 1月	10581.6	78.8
2	2月	10234.3	69.1
3	3月	13299.5	87.6
4	4月	10746.5	72.8
5	5月	13310.7	79.4
6	6月	12835.5	85.6
7	7月	12194.2	75.0
8	8月	12860.4	85.3
9	9月	11955.6	86.9
10	10月	13351.5	107.8
11	11月	13285.9	128.7
12	12月	13784.4	134.5
13	'68 1月	16336.7	148.7
14	2月	11040.5	94.2
15	3月	11525.3	128.1
16	4月	16056.4	154.1
17	5月	18464.3	191.3
18	6月	17092.2	191.9
19	7月	15178.8	159.6
20	8月	12774.8	185.5
21	9月	12377.8	178.0
22	10月	16856.3	271.8
23	11月	14635.3	212.3
24	12月	17436.9	139.4
25	'69 1月	16482.2	106.0
26	2月	13905.4	112.1
27	3月	11973.7	103.5
28	4月	12573.6	92.5
29	5月	16566.8	116.9
30	6月	13558.7	78.9
31	7月	11530.9	57.4
32	8月	11278.0	75.9
33	9月	11263.7	109.8
34	10月	15649.5	129.2
35	11月	12197.1	115.1

Monthly dollar volume of sales (in millions) on Boston stock exchange and combined New York and American stock exchanges

Fig. 4



level の数  
 $\lambda$  のパーセント点以下の  $\phi$  をもつ PC の存在しない level の数

Fig. 6

が存在しなかった level の数をグラフにまとめたものである。

Fig. 6 において Fig. 4 のデータは他の3つのデータと比べて、 $\lambda$  の変化によって level の数の変化をほとんど生じない。Fig. 4 のデータは他の3つのデータよりは時間の系列は長い。一般に長い列をもつデータは PC の数が短い系列よりもより多いから、 $\lambda$  が小さくなれば level の減少する数も多いと考えられる。従って Fig. 4 のデータの level の減少が生じない原因は、他の3つのデータとの質的な差異が考えられる。即ち、Fig. 3 の3つのデータは Fig. 4 のデータとは異なり、乱数を基にして作られた人工的なデータである。人工的なデータがモデルに含まれている仮定を充分満足するように作られているのに対して、現実で得られるデータは必ずしもモデルに含まれている仮定を充分満足するとは限らない。実際、Fig. 4 のデータの  $\phi$  の値は期待されるそれよりも 10 倍又は  $10^2$  倍の値が生じた。他の3つのデータについてはそのような場合はなかった。従って現実で得られるデータに対しては“重み”を考慮したり、その他の影響を除去する方法を取る必要がある。

高い level での **type-4** の cluster (NFC 同士の交わりによって生じた FC) の発生は、4つのデータ全てについて得られなかった。このことは、level が高くなるにつれて clustering が進み、各 level での PC の数が減少してくるために、 $\lambda$  のパーセント点以下の  $\phi$  をもつ PC が存在する率が小さくなることが原因の1つとなっている。従って、高い level でも PC の数が少なくならないように、長い時系列のデータを用いた実験が必要である。また Fig. 5 から、 $\lambda$  のパーセント点以下の  $\phi$  をもつ PC が存在しない level の数が最終 level までの level の数の 50%以上を占めている場合が非常に多い。このことも、もう1つの原因と考えられる。これに対しては、 $\lambda$  の値が高い level では小さくなるようにする方法を取るなどが必要となる。このことは  $\lambda$  の効果を高めることにもなる。

#### 4. あ と が き

今回の報告では PLR 1 を若干修正した PLR 2 についてデータを用いた検討を行ったが、通して言えることは、低い level での **type-4** cluster の発生の結果、level の数を減少させることが出来るが、高い level での発生がないために  $\lambda$  及び  $\mu$  の効果が十分得られない。従って実験の対象とするデータの数を増すばかりでなく、高い level での **type-4** cluster の発生する条件を考察することがまず必要である。このことは理論的な側面の追求と合せて行われるべきであると考え。今後の研究の方針をこれに向けて進めて行きたいと思う。

#### 文 献

- 1) McGee Victor, E. and Carleton Willard. T.: "Piecewise Regression," Journal of American Statistical Association, Vol. 65 (1970), pp. 1109-1124.
- 2) 奥野忠一, 久米 均, 芳賀敏郎, 吉沢 正: 多変量解析法, (日科技連, (1971)), pp. 391-411.
- 3) Quandt, R. E.: "The Estimation of the Parameters of a Linear Regression System Odeying Two Separate Regimes," Journal of the American Statistical Association, 53 (1958), pp. 873-880.
- 4) 杉村 徹: "PLR の処理過程に関するある種の改良", 北海道大学工学部修士論文, (1976).