

15. 1stOpt 分段函数拟合

分段拟合是拟合中常遇到的问题。如何让程序自动选取并计算出分段点值？如果需要，又如何保证分段点处的连续？以两个实际案例演示 1stOpt 轻松并完美处理该类问题。

15.1 案例一

分段拟合公式、示意图及数据分别见公式 15-1、图 15.1 和表 15.1。图形两段部分用线性公式、中间部分用二次多项式，两个分段点为 k1 和 k2，根据图形趋势，k1 的范围可设定为[1000,1500]，k2 的范围则设定为[1700,2200]。

$$y = \begin{cases} a_0 + a_1x, & \text{if } x \leq k_1 \\ b_0 + b_1x + b_2x^2, & \text{if } k_1 \leq x \leq k_2 \\ c_0 + c_1x, & \text{if } x \geq k_2 \end{cases} \quad (15-1)$$

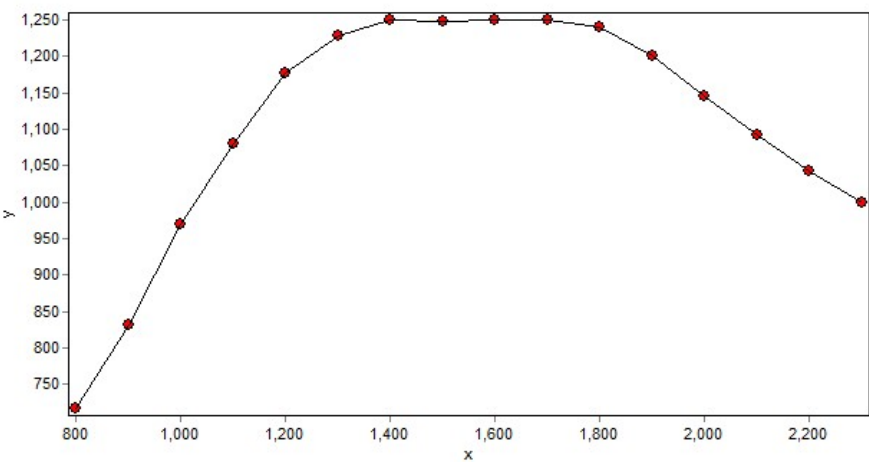


图 15.1 案例一分段拟合数据示意图

表 15.1 数据：

x	800,900,1000,1100,1200,1300,1400,1500,1600,1700,1800,1900,2000,2100,2200,2300
y	717.4,831,969.3,1079.9,1177.6,1229.1,1249.9,1249,1250.1,1250.1,1239.4,1199.9,1145.3,1093,1043.8,999.2

求解代码如下。注意拟合公式中判断条件 if 的写法，尤其是多个 if 判断条件的写法，可以无限嵌套。

代码 15-1

```
Parameter k1=[1100,1500],k2=[1700,2200];
```

Function $y = \text{if}(x < k_1, a_0 + a_1 * x, \text{if}(x > k_2, c_0 + c_1 * x, b_0 + b_1 * x + b_2 * x^2));$
 Data;
 $x = 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300;$
 $y = 717.4, 831, 969.3, 1079.9, 1177.6, 1229.1, 1249.9, 1249, 1250.1, 1250.1, 1239.4, 1199.9, 1145.3, 1093, 1043.8, 999.2;$

结果:

Root of Mean Square Error (RMSE): 5.50711106855735
 Sum of Squared Residual: 485.25235714283
 Correlation Coef. (R): 0.999381686676783
 R-Square: 0.998763755664932
 Adjusted R-Square: 0.998573564228768
 Determination Coef. (DC): 0.998763755664932
 Chi-Square: 0.218238554449713
 F-Statistic: 706.91388627685

Parameter Best Estimate

Parameter	Best Estimate
k1	1133.85615842417
k2	1721.38615294825
a0	-265.109995341884
a1	1.22579999462332
c0	2128.4657142946
c1	-0.491885714277365
b0	-186.716439486274
b1	1.86286787211923
b2	-0.000600535719245784

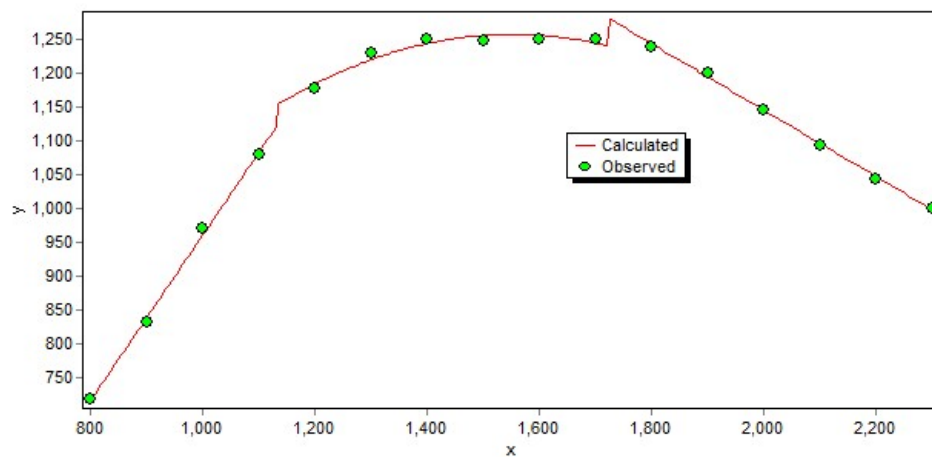


图 15.2 案例一无约束分段拟合结果示意图

从上述拟合图形可看出，分段点处是不连续的，存在“错位”现象。如果要求分段点处连续，该如何处理呢？此时仅需要加上分段点处连续的约束条件即可，即：

$$\begin{cases} a_0 + a_1 k_1 = b_0 + b_1 k_1 + b_2 k_1^2 \\ b_0 + b_1 k_2 + b_2 k_2^2 = c_0 + c_1 k_2 \end{cases} \quad (15-2)$$

该约束的添加有两种模式，一种是在拟合函数之后直接添加，两种模式效果一样。

代码 15-2

模式一	Parameter k1=[1100,1500],k2=[1700,2200]; Function y=if(x<k1,a0+a1*x,if(x>k2,c0+c1*x,b0+b1*x+b2*x^2)); a0+a1*k1=b0+b1*k1+b2*k1^2; b0+b1*k2+b2*k2^2=c0+c1*k2; Data; x=800,900,1000,1100,1200,1300,1400,1500,1600,1700,1800,1900,2000,2100,2200,2300; y=717.4,831,969.3,1079.9,1177.6,1229.1,1249.9,1249,1250.1,1250.1,1239.4,1199.9,1145.3,1093,1043.8,999.2;
模式二	Parameter k1=[1100,1500],k2=[1700,2200]; SubjectTo a0+a1*k1=b0+b1*k1+b2*k1^2, b0+b1*k2+b2*k2^2=c0+c1*k2; Function y=if(x<k1,a0+a1*x,if(x>k2,c0+c1*x,b0+b1*x+b2*x^2)); Data; x=800,900,1000,1100,1200,1300,1400,1500,1600,1700,1800,1900,2000,2100,2200,2300; y=717.4,831,969.3,1079.9,1177.6,1229.1,1249.9,1249,1250.1,1250.1,1239.4,1199.9,1145.3,1093,1043.8,999.2;

结果:

Root of Mean Square Error (RMSE): 5.8100997389641
Sum of Squared Residual: 540.116143627371
Correlation Coef. (R): 0.999311754628135
R-Square: 0.998623982937961
Adjusted R-Square: 0.99841228800534
Determination Coef. (DC): 0.998623982937931
Chi-Square: 0.241754142992597
F-Statistic: 635.01805515698
Constrained Functions: $a_0 + a_1 * k_1 - (b_0 + b_1 * k_1 + b_2 * k_1^2) = 0$
 $b_0 + b_1 * k_2 + b_2 * k_2^2 - (c_0 + c_1 * k_2) = -2.27373675443232E-13$

Parameter Best Estimate

k1 1176.56821287057
k2 1845.2941122842
a0 -265.106349171658
a1 1.22579708366613
c0 2152.3287978364
c1 -0.502898900109216
b0 31.7984821765496
b1 1.549116468456
b2 -0.000489276817995342

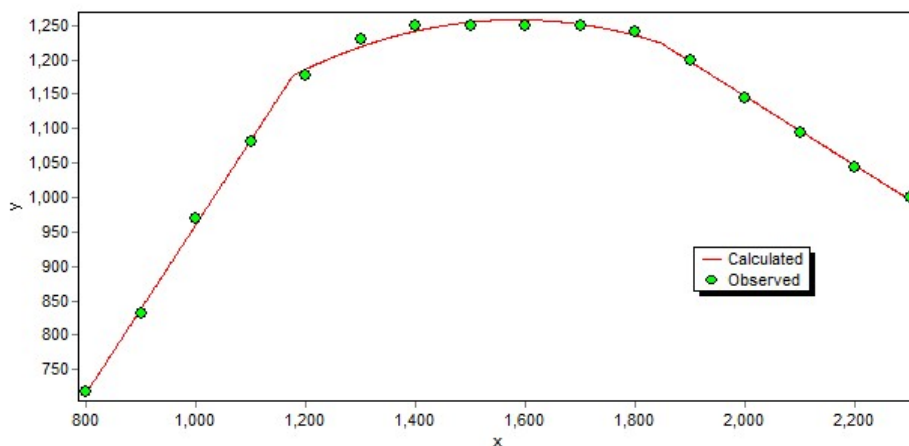


图 15.3 案例一约束分段拟合结果示意图

可以看出，增加分段点连续的约束条件后，分段点处的“错位”现象消失了，约束条件均得到满足：

$$a_0 + a_1 k_1 - (b_0 + b_1 k_1 + b_2 k_1^2) = 0$$

$$b_0 + b_1 k_2 + b_2 k_2^2 - (c_0 + c_1 k_2) = -2.27373675443232E-13$$

但另一方面，拟合度略有下降：无约束时的残差平方和为 485.252357，增加约束后则变为 540.116143，这也可视作保证分段点处连续所付出的“代价”。

15.2 案例二

分段拟合公式、示意图及数据分别见公式 15-3、图 15.4 和表 15.2。三个分段点 k_1 、 k_2 和 k_3 将数据分为四段，每段均为线性公式。根据图形趋势， k_1 、 k_2 和 k_3 的范围可分别设定为 [1,25]、[25,70] 和 [90,120]。

$$y = \begin{cases} b_1 + b_2 x, & \text{if } x \leq k_1 \\ b_3 + b_4 x, & \text{if } k_1 < x \leq k_2 \\ b_5 + b_6 x, & \text{if } k_2 < x \leq k_3 \\ b_7 + b_8 x, & \text{if } x > k_3 \end{cases} \quad (15-3)$$

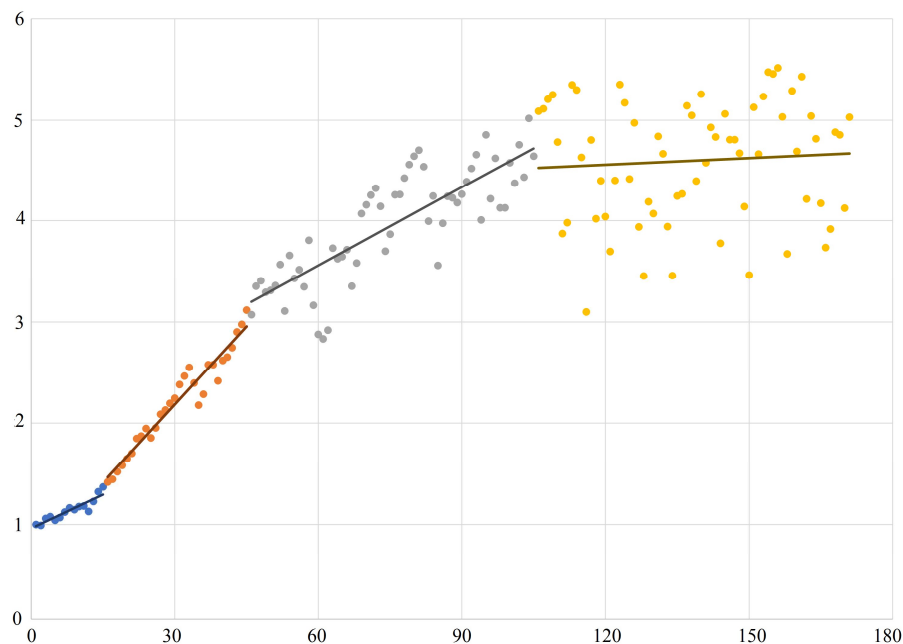


图 15.4 案例二分段拟合数据示意图

表 15.2 数据：

x	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172
y	1.00,0.99,1.06,1.08,1.04,1.07,1.12,1.16,1.15,1.18,1.18,1.13,1.23,1.32,1.37,1.42,1.44,1.52,1.58,1.65,1.70,1.85,1.87,1.95,1.86,1.95,2.09,2.13,2.19,2.25,2.38,2.46,2.55,2.40,2.18,2.28,2.58,2.58,2.42,2.62,2.65,2.74,2.90,2.98,3.12,3.07,3.35,3.41,3.29,3.31,3.36,3.57,3.11,3.66,3.43,3.52,3.34,3.80,3.16,2.88,2.83,2.92,3.73,3.62,3.64,3.71,3.35,3.58,4.07,4.16,4.25,4.32,4.14,3.70,3.87,4.26,4.26,4.42,4.56,4.64,4.70,4.54,3.99,4.24,3.56,3.97,4.24,4.22,4.18,4.26,4.39,4.52,4.66,4.01,4.85,4.22,4.62,4.13,4.12,4.57,4.37,4.75,4.43,5.02,4.64,5.09,5.11,5.21,5.25,4.78,3.87,3.98,5.35,5.30,4.63,3.10,4.80,4.02,4.39,4.04,3.70,4.40,5.36,5.17,4.41,4.97,3.94,3.45,4.19,4.07,4.84,4.67,3.94,3.46,4.24,4.26,5.14,5.05,4.39,5.26,4.58,4.93,4.83,3.78,5.06,4.81,4.80,4.67,4.14,3.46,5.13,4.66,5.24,5.48,5.46,5.52,5.03,3.67,5.29,4.69,5.43,4.21,5.04,4.81,4.17,3.73,3.92,4.88,4.85,4.12,5.03,3.75

无断点连续要求的求解代码：

代码 15-3

```

Variable x,y;
Parameter k1=[1,25], k2=[25,70], k3=[90,120], b(8);
Function y=if(x<=k1, b1+b2*x,if((x>k1)and(x<=k2),b3+b4*x, if(x>k3, b7+b8*x, b5+b6*x)));
Data;
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172;
1.00,0.99,1.06,1.08,1.04,1.07,1.12,1.16,1.15,1.18,1.18,1.13,1.23,1.32,1.37,1.42,1.44,1.52,1.58,1.65,1.70,1.85,1.87,1.95,1.86,1.95,2.09,2.13,2.19,2.25,2.38,2.46,2.55,2.40,2.18,2.28,2.58,2.58,2.42,2.62,2.65,2.74,2.90,2.98,3.12,3.07,3.35,3.41,3.29,3.31,3.36,3.57,3.11,3.66,3.43,3.52,3.34,3.80,3.16,2.88,2.83,2.92,3.73,3.62,3.64,3.71,3.35,3.58,4.07,4.16,4.25,4.32,4.14,3.70,3.87,4.26,4.26,4.42,4.56,4.64,4.70,4.54,3.99,4.24,3.56,3.97,4.24,4.22,4.18,4.26,4.39,4.52,4.66,4.01,4.85,4.22,4.62,4.13,4.12,4.57,4.37,4.75,4.43,5.02,4.64,5.09,5.11,5.21,5.25,4.78,3.87,3.98,5.35,5.30,4.63,3.10,4.80,4.02,4.39,4.04,3.70,4.40,5.36,5.17,4.41,4.97,3.94,3.45,4.19,4.07,4.84,4.67,3.94,3.46,4.24,4.26,5.14,5.05,4.39,5.26,4.58,4.93,4.83,3.78,5.06,4.81,4.80,4.67,4.14,3.46,5.13,4.66,5.24,5.48,5.46,5.52,5.03,3.67,5.29,4.69,5.43,4.21,5.04,4.81,4.17,3.73,3.92,4.88,4.85,4.12,5.03,3.75

```

87,1.95,1.86,1.95,2.09,2.13,2.19,2.25,2.38,2.46,2.55,2.40,2.18,2.28,2.58,2.58,2.42,2.62,2.65,2.74,2.90,2.98,3.1
 2,3.07,3.35,3.41,3.29,3.31,3.36,3.57,3.11,3.66,3.43,3.52,3.34,3.80,3.16,2.88,2.83,2.92,3.73,3.62,3.64,3.71,3.35,
 3.58,4.07,4.16,4.25,4.32,4.14,3.70,3.87,4.26,4.26,4.42,4.56,4.64,4.70,4.54,3.99,4.24,3.56,3.97,4.24,4.22,4.18,4.
 26,4.39,4.52,4.66,4.01,4.85,4.22,4.62,4.13,4.12,4.57,4.37,4.75,4.43,5.02,4.64,5.09,5.11,5.21,5.25,4.78,3.87,3.9
 8,5.35,5.30,4.63,3.10,4.80,4.02,4.39,4.04,3.70,4.40,5.36,5.17,4.41,4.97,3.94,3.45,4.19,4.07,4.84,4.67,3.94,3.46,
 4.24,4.26,5.14,5.05,4.39,5.26,4.58,4.93,4.83,3.78,5.06,4.81,4.80,4.67,4.14,3.46,5.13,4.66,5.24,5.48,5.46,5.52,5.
 03,3.67,5.29,4.69,5.43,4.21,5.04,4.81,4.17,3.73,3.92,4.88,4.85,4.12,5.03,3.75;

结果:

Stop Reason of Computation: Convergence tolerance reached

Root of Mean Square Error (RMSE): 0.410725033837938

Sum of Squared Residual: 29.0155491884421

Correlation Coef. (R): 0.944631369767115

R-Square: 0.892328424748097

Adjusted R-Square: 0.891054204922631

Determination Coef. (DC): 0.892328424748097

Chi-Square: 3.41337014730407

F-Statistic: 133.428787032104

Parameter Best Estimate

Parameter	Best Estimate
k1	13.9987052329631
k2	58.9971732032522
k3	110.908420311061
b1	0.984999992171523
b2	0.0174175830188084
b3	0.5887255172086
b4	0.0535230565478818
b5	1.64668572911672
b6	0.0302872876947118
b7	3.75673516061524
b8	0.00554455946331597

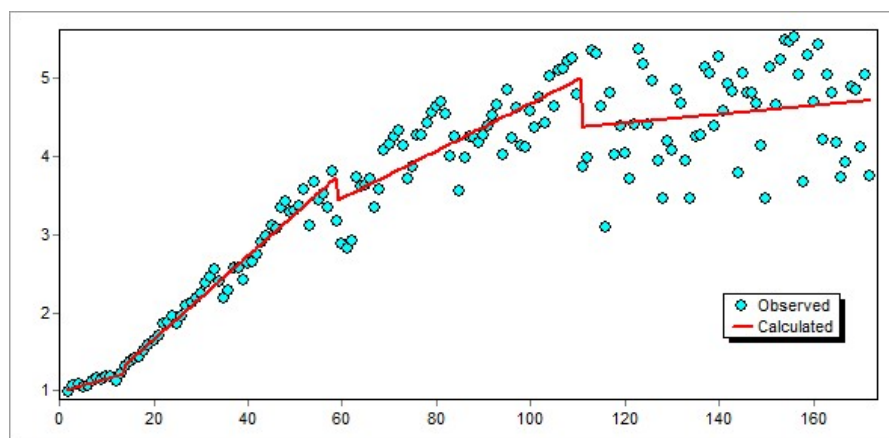


图 15.5 案例二无约束分段拟合示意图

有断点连续要求的求解代码：

代码 15-4

```
SubjectTo b1+b2*k1=b3+b4*k1,b3+b4*k2=b5+b6*k2,b5+b6*k3=b7+b8*k3;  
Variable x,y;  
Parameter k1=[1,25], k2=[25,70], k3=[90,120], b(8);  
Function y=if(x<=k1, b1+b2*x,if((x>k1)and(x<=k2),b3+b4*x, if(x>k3, b7+b8*x, b5+b6*x)));  
Data;  
....
```

结果：

Root of Mean Square Error (RMSE): 0.4224409789876
Sum of Squared Residual: 30.6944974852163
Correlation Coef. (R): 0.941327893280842
R-Square: 0.886098202668548
Adjusted R-Square: 0.884750252404271
Determination Coef. (DC): 0.886098144331686
Chi-Square: 3.61320346192678
F-Statistic: 125.314050231831
Constrained Functions: $b5+b6*k3-(b7+b8*k3) = 4.54400961302781E-11$
 $b3+b4*k2-(b5+b6*k2) = 1.53144164016794E-11$
 $b1+b2*k1-(b3+b4*k1) = 1.97815097635612E-11$

Parameter	Best Estimate
k1	12.001190214113
k2	48.0000000000012
k3	92.4793277320192
b1	0.975501032657809
b2	0.0190712439364663
b3	0.545634202091451
b4	0.0548899271568244
b5	1.75340454591759
b6	0.0297280449934614
b7	4.35308992665992
b8	0.00161705581844306

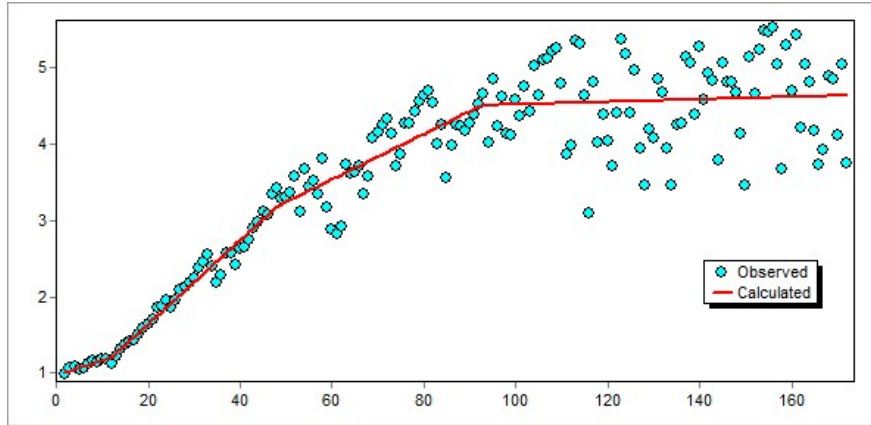


图 15.6 案例二有约束分段拟合数据示意图

与案例一类似，增加断点连续的约束后，断点处“错位”消失，均保持连续。

15.3 小结

1stOpt 可以非常容易处理任意分段函数拟合问题，自动计算出各分段点值，同时，如果有分段点连续的要求，也可以通过添加约束轻松实现。