

基于改进模拟退火算法的模糊回归参数估计

彭宇文¹, 郭莉莎¹, 毛超²

(1.湖南师范大学 商学院,长沙 410000;2.湖南大学 工商管理学院,长沙 410082)

摘要:文章讨论了自变量、因变量以及回归系数均为模糊数的线性和非线性回归模型,提出了一种与模糊隶属度相关的回归效果评价方法,并设计了模拟退火算法进行优化求解。将隶属度为1时各模糊数取值以及各模糊数的边界进行回归产生的系数作为初始解,采用多层邻域操作产生新解,通过多准则终止运算,并且利用增加记忆装置记录了每次运算的最优解。算例和对比分析的结果表明,该方法收敛速度快、回归效果较好,适合处理较复杂的模糊回归模型。

关键词:模糊回归分析;模糊数;模拟退火算法;参数估计

中图分类号: O212.4 **文献标识码:** A **文章编号:** 1002-6487(2014)01-0017-03

0 引言

在现实生活中,统计对象的某些特性使得一些统计数据往往难以准确的记录或收集。例如,在测量河流水位高低时,会由于波浪使水位在一定范围内波动,而不是停留在某个具体的数值上,样本数据具有一定的模糊性。在研究对象的观测值具有模糊性时,模糊集理论成为重要的方法。Zadeh(1965)首次提出了模糊集理论^[1],Tanaka等(1982)在此基础上,提出了模糊数据回归问题。模糊回归在许多领域得到了广泛的应用^[2]。

国内外对模糊回归的研究包括对线性和非线性回归模型的研究,总体上可分为三类:第一类是变量之间关系是具有模糊性,即需求解的回归系数是模糊的;第二类是因变量、自变量中的部分或者全部具有模糊性,即变量本身具有模糊性;第三类是变量与系数均具有模糊性。由于

第三类较复杂,大多数研究集中在第一类和第二类。

第一类和第二类模糊回归问题是第三类模糊回归问题的特殊情形,因此,探索更具一般性的模糊回归方法具有重要意义。本文讨论第三类模糊回归问题的参数估计,其变量与系数均具有模糊性的回归模型。

1 模糊回归模型

设因变量 $\tilde{Y}$ 为模糊变量,是某一组模糊自变量 $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ 的函数,其中: $f: \mathbb{R}^n \rightarrow E$, $\tilde{Y}_i = f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in})$, i 为观测值的编号, n 为因变量个数。对于模糊线性回归,其目标是对下述模糊线性回归模型的模糊参数进行估计,以使得回归得出的因变量值与样本观测值的“最匹配”:

$$\tilde{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_{i1} + \hat{\beta}_2 \tilde{x}_{i2} + \dots + \hat{\beta}_n \tilde{x}_{in} + u_i \quad (1)$$

非线性回归的目标是对下述模糊非线性模型的参数

作者简介: 彭宇文(1976-),男,湖南双峰人,博士,副教授,研究方向:数理统计,农业经济。

表5 全社会固定资产估计值及真实值

	均值估计量	均值方差	均值区间估计	总体估计量	总体方差	总体区间估计
真实值	8753.7	—	—	271365	—	—
PPS估计结果	8753.64	0.02387	8753.64 ± 0.309	271362.72	22.94	271362.72 ± 9.58
简单随机估计结果	9521	959559	9521 ± 2 × 1960	295151	922136199	295151 ± 60734

抽样总量估计值与真实值之间相差只有2.28,抽样方差只有22.94,而简单随机抽样估计量与真实值相差到达23786,方差达到922136199。因此,无论是抽样估计结果效率、抽样误差还是95%置信度条件下的区间估计,PPS估计的效率均高于简单随机抽样。在实际调研分析过程中,可以根据研究需要,充分既定利用条件,选用信度更高的方法。

参考文献:

[1]蔡泽林.PPS分层抽样时样本单位数的设计[J].湖北师范学院学报

(自然科学版),1996,(6).

[2]马慧敏,方国斌.改进PPS抽样设计效应的Monte Carlo模拟[J].统计与决策,2009,(11).

[3]张荷观.整群比估计与PPS抽样的比较[J].无锡轻工大学学报,2001,(7).

[4]李战江,闫在在. PPS估计与比估计的比较[J].内蒙古农业大学学报(自然科学版),2005,(4).

[5]俞纯权. PPS抽样下子总体参数的估计[J]. 统计与决策,2006,(18).

[6]卢静莉,闫在在,丁昌江. 多元简单随机抽样与二相抽样的最优分配[J]. 统计与决策,2010,(19).

[7]倪加勤,张勇.调查内比估计效果的数据模拟分析——关于简单抽样设计与PPS系统抽样设计的比较[J].统计研究,2005,(11).

[8]Nibia,Aires. Comparisons between Conditional Poisson Sampling and Pareto π ps Sampling Designs[J]. Journal of Statistical Planning and Inference, 2000,(88).

(责任编辑/亦民)

进行估计:

$$\hat{Y}_i = \hat{\beta}_0 e^{\hat{\beta}_1 \hat{x}_{i1} + \hat{\beta}_2 \hat{x}_{i2} + \dots + \hat{\beta}_n \hat{x}_{in}} + u_i \quad (2)$$

模糊非线性回归模型可通过取对数转化为模糊线性规划模型如下:

$$\ln \hat{Y}_i = \ln(\hat{\beta}_0) + \hat{\beta}_1 \hat{x}_{i1} + \hat{\beta}_2 \hat{x}_{i2} + \dots + \hat{\beta}_n \hat{x}_{in} + u_i \quad (3)$$

模型(1)和(2)中,待估计的参数 $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n$ 均为模糊数。所谓“最匹配”实际上是对拟合效果的评价,目前尚无统一的评价标准。

2 基于模拟退火算法的模糊回归参数估计

2.1 拟合效果的评价

拟合效果的评价与对模糊数距离的定义相关,文献^[3]提出了两个模糊数的贴近度的概念,并以平均贴近度最小为目标进行参数估计。文献^[4]将每一组的模糊输出与观测值的误差定义为: $e_i = \frac{1}{2} \sum_a (\hat{Y}_{ia}^L - \tilde{Y}_{ia}^L)^2 + \frac{1}{2} \sum_a (\hat{Y}_{ia}^U - \tilde{Y}_{ia}^U)^2$, 以总误差(所有误差之和)最小为目标进行参数估计。本文认为误差也具有模糊性,单组误差产生的可能性将影响到拟合的总误差,因此将与隶属度关联的总误差最小做为目标来评价拟合效果。

设 $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n$ 为参数的一组估计值,隶属度为 α 时,其对应的拟合效果为:

$$\begin{aligned} E(\hat{\beta}_1, \dots, \hat{\beta}_n) &= [E(\hat{\beta}_1, \dots, \hat{\beta}_n)]_\alpha^L + [E(\hat{\beta}_1, \dots, \hat{\beta}_n)]_\alpha^U \\ &= \frac{1}{2} \sum_{i=1}^m (\hat{Y}_{ia}^L - \tilde{Y}_{ia}^L)^2 + \frac{1}{2} \sum_{i=1}^m (\hat{Y}_{ia}^U - \tilde{Y}_{ia}^U)^2 \end{aligned} \quad (4)$$

与隶属度关联的总误差可以表示为:

$$e = \int_0^1 \alpha [E(\hat{\beta}_1, \dots, \hat{\beta}_n)]_\alpha d\alpha = \frac{1}{2} \int_0^1 \alpha \sum_{i=1}^m [(\hat{Y}_{ia}^L - \tilde{Y}_{ia}^L)^2 + (\hat{Y}_{ia}^U - \tilde{Y}_{ia}^U)^2] d\alpha \quad (5)$$

为简化运算,将(8)式改写为离散形式并将样本值代入,可得:

$$\begin{aligned} e &= \frac{1}{2} \sum_a \left(\alpha \sum_{i=1}^m [(\hat{Y}_{ia}^L - \tilde{Y}_{ia}^L)^2 + (\hat{Y}_{ia}^U - \tilde{Y}_{ia}^U)^2] \right) \\ &= \frac{1}{2} \sum_a \left(\alpha \sum_{i=1}^m \left[\left(\sum_{j=0}^n (\tilde{x}_{ij} \times \tilde{\beta}_j)_a^L - \tilde{Y}_{ia}^L \right)^2 + \left(\sum_{j=0}^n (\tilde{x}_{ij} \times \tilde{\beta}_j)_a^U - \tilde{Y}_{ia}^U \right)^2 \right] \right) \end{aligned} \quad (6)$$

其中,

$$\begin{aligned} (\tilde{x}_{ij} \times \tilde{\beta}_j)_a^L &= \min([\tilde{x}_{ij}]_a^L [\tilde{\beta}_j]_a^L, [\tilde{x}_{ij}]_a^L [\tilde{\beta}_j]_a^U, [\tilde{x}_{ij}]_a^U [\tilde{\beta}_j]_a^L, [\tilde{x}_{ij}]_a^U [\tilde{\beta}_j]_a^U) \\ (\tilde{x}_{ij} \times \tilde{\beta}_j)_a^U &= \max([\tilde{x}_{ij}]_a^L [\tilde{\beta}_j]_a^L, [\tilde{x}_{ij}]_a^L [\tilde{\beta}_j]_a^U, [\tilde{x}_{ij}]_a^U [\tilde{\beta}_j]_a^L, [\tilde{x}_{ij}]_a^U [\tilde{\beta}_j]_a^U) \end{aligned}$$

2.2 算法设计

式(6)的最小值是一个无约束二次函数最值问题,但由于模糊运算的复杂性,无法直接求出其最优解。采用启发式算法进行求解成为重要思路,模拟退火算法具有对初始

可行解要求低,通过 Metropolis 接受准则避免陷入局部最优等优点^[5],本文对传统模拟退火算法进行改进以获得最佳的参数估计值。

(1) 算法流程。参数估计的模拟退火算法具体步骤为:

① 获取一个初始可行解 $x_0 = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$, 设定初始温度 t_0 , 令当前解 $x_i = x_0$, 当前迭代步数 $k=0$, 当前温度 $t_k = t_0$ 。

② 若在该温度达到内循环停止条件,则转第③步;否则,从邻域 $N(x_i)$ 中随机选择一个邻解 x_j , 并计算两个解对应的拟合效果之差 $\Delta E_{ij} = E(x_j) - E(x_i)$ 。若 $\Delta E_{ij} \leq 0$, 则 $x_i = x_j$; 否则,若 $\exp(-\Delta E_{ij}/t) > \text{rand}(0, 1)$ ($\text{rand}(0, 1)$ 表示一个 0 到 1 之间的均匀随机数), 则 $x_i = x_j$, 重复第②步。

③ $k = k + 1$, $t_{k+1} = y(t_k)$ (温度控制函数), 若满足终止条件,转第④步;否则,转第(2)步。

④ 输出计算结果,算法停止。

内循环为第②步,它表示在同一个温度下进行随机搜索。外循环主要包括第③步的温度下降变化,迭代步数的增加和停止准则。

(2) 初始解的构造。为获得初始解,本文用所有模糊样本的左边界、右边界、以及隶属度为 1 时的样本取值分别进行回归,将各参数的回归系数中的最小值作为模糊系数的左边界,最大值作为模糊系数的右边界,中间值作为模糊系数隶属度为 1 时的值。

(3) 邻域操作。为获得更优的解,设计了多层次邻域操作方法。以三角模糊数为例,各参数的形式为 $\tilde{\beta}_i = (\tilde{\beta}_{iL}, \tilde{\beta}_{iM}, \tilde{\beta}_{iR})$ 。首先,随机选取 $N = \left\lceil \frac{n+1}{2} \right\rceil$ 个系数作为需要调整的对象;其次, N 个对象分别随机选定 $\tilde{\beta}_{iL}$, $\tilde{\beta}_{iM}$, $\tilde{\beta}_{iR}$ 中的一个作为要调整的部分;最后,确定需要调整的幅度:若调整 $\tilde{\beta}_{iL}$, 则调整的幅度为 $\frac{\tilde{\beta}_{iM} - \tilde{\beta}_{iL}}{K}$, 否则,

调整幅度为 $\frac{\tilde{\beta}_{iR} - \tilde{\beta}_{iM}}{K}$ 。其中, K 的取值与 $\tilde{\beta}_{iM} - \tilde{\beta}_{iL}$ 的数量级相关, K 由小到大依次确定解的粗调整、二次调整和微调的幅度。可根据试探性运算的结果,确定 k_1 值,使得当连续 k_1 次目标函数值得不到改善时,改变 K 的取值,进入下一层邻域操作。其他类型的模糊数同样可按此思路进行多层次邻域操作。

(4) 终止准则与结果输出。采用双终止准则,第一,完成三层邻域操作后,仍然出现了连续 k_2 次目标函数值得不到改善,则终止程序,输出结果;第二,温度 t 下降到预先设定的某个值,则终止程序,输出结果。为获得该次运算的最优结果,在外循环中增加变量记录每次目标函数得到了改进的解。

3 算例及结果分析

3.1 算例分析

为了说明拟合效果评价方法与求解算法的有效性,采用文献^[6]的算例作为模糊线性回归算例,对文献^[4]的模糊非线性规划算例的自变量模糊化,作为模糊非线性回归算例。

3.1.1 线性回归算例

文献^[6]设计的算例具有一个模糊因变量和两个模糊自变量,共有15对样本。采用本文的算法求解时,算法参数的设定对结果的优劣有一定的影响,本文通过多次调整和运算来获取相对最佳参数,对模拟退火算法参数设定如下:马尔科夫链长度取200,初始温度取100,衰减因子为0.98。 K 值依次取500、1000、1500。初始解取: $x_0 = ((3.0450, 3.6524, 4.0531), (0.4937, 0.4970, 0.4986), (0.0089, 0.0092, 0.0097))$ 。求解得到的模糊线性回归方程为如下:

$$\tilde{Y}_i = (3.3511, 3.6913, 3.8762) + (0.4957, 0.4973, 0.4982) \tilde{x}_{i1} + (0.0091, 0.0092, 0.0094) \tilde{x}_{i2} \quad (7)$$

图1为退火过程,图2对拟合值与样本值进行了比较。

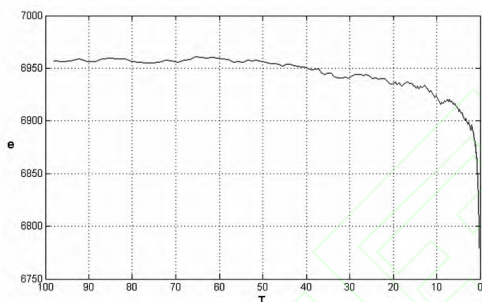


图1 算例1的退火图

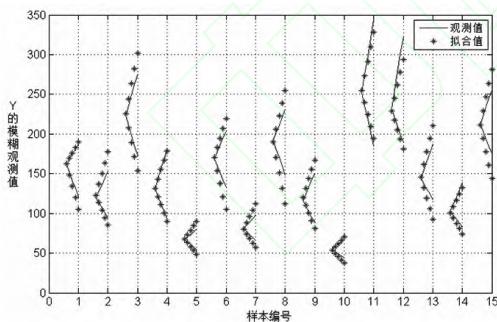


图2 算例1的样本值与拟合值比较

3.1.2 非线性回归算例

对文献^[4]的模糊非线性规划算例的自变量进行模糊化,得到模糊非线性回归样本值如表2。

表2 模糊非线性回归样本值

i	$\tilde{Y}_i$			$\tilde{x}_{i1}$		
1	128	139	158	0.4	1	1.6
2	164	190	240	1.8	2	2.3
3	300	340	358	2.6	3	3.4
4	340	376	400	3.7	4	4.4
5	400	470	585	4.5	5	5.5

调用模拟退火算法,马尔科夫链长度取200,初始温度取100,衰减因子为0.98。 K 值依次取100、500、1000。分

表3

不同模糊回归方法的比较

i	$\tilde{x}_{i1}$	$\tilde{Y}_i$	误差1(文献 ^[17] 的评价标准)				误差2(本文的评价标准)			
			Tanaka	Kao	Mosleh	SA	Tanaka	Kao	Mosleh	SA
1	(1,1,1)	(6.2,8,9.8)	62.4071	21.2671	19.9077	21.8621	26.5808	10.3361	9.91	11.5327
2	(2,2,2)	(4.2,6.4,8.6)	40.4291	42.3993	42.6939	41.1564	17.2198	21.1838	21.3448	20.0712
3	(3,3,3)	(6.9,9.5,12.1)	10.6631	4.0022	4.0018	3.9823	3.6128	1.9149	1.9145	1.8419
4	(4,4,4)	(10.9,13.5,16.1)	23.2031	32.1667	32.2204	31.4643	9.8828	15.9971	16.0952	15.5126
5	(5,5,5)	(10.6,13,15.4)	28.1421	2.9535	2.9723	3.5364	11.7583	1.4697	1.4224	1.8808
总误差			164.8446	102.7888	101.7962	102.0015	69.0546	50.9016	50.6869	50.8392

别对左边界、中间值和右边界进行非线性回归,得到初始解: $x_0 = ((118.6336, 124.0292, 125.8052), (0.2679, 0.2705, 0.2895))$ 。求解得到的模糊线性回归方程如下:

$$\tilde{Y}_i = (98.8741, 109.0108, 123.5765) e^{(0.3009, 0.3068, 0.3131) \tilde{x}_{i1}} \quad (8)$$

图3对拟合值与样本值进行了比较。

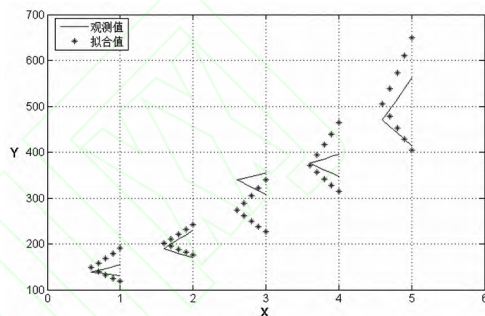


图3 算例2的样本值与拟合值比较

3.2 比较分析

本文采用 Tanaka 等(1989)、Kao 等(2003)、Mosleh 等(2010)共同用来说明各自模糊回归方法的算例进行比较分析^[4,7,8]。算例与计算结果见表3。

Tanaka等(1989)的最终回归方程为:

$$\tilde{Y}_T = (0, 3.850, 7.700) + 2.100 \tilde{x}_1 \quad (9)$$

Kao等(2003)的最终回归方程为:

$$\tilde{Y}_K = (2.606, 4.926, 7.246) + 1.718 \tilde{x}_1 \quad (10)$$

Mosleh等(2010)采用神经网络方法获得的最终回归方程为:

$$\tilde{Y}_M = (3.1100, 4.9499, 6.7898) + (1.5500, 1.7100, 1.8701) \tilde{x}_1 \quad (11)$$

本文的最终回归方程为:

$$\tilde{Y}_{SA} = (3.2176, 4.7118, 6.7813) + (1.5283; 1.7814; 1.8643) \tilde{x}_1 \quad (12)$$

各种方法的拟合效果比较见图4。

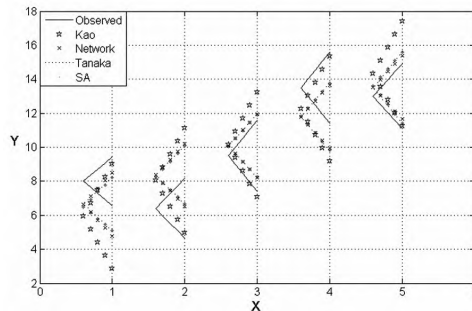


图4 Tanaka方法、Kao方法、Mosleh方法与SA方法的拟合效果

从图4可以看出,基于多层次邻域操作的模拟退火算法的模糊回归方法拟合效果要优于 Tanaka 等(1989)、Kao 等(2003)的方法,非常接近 Mosleh 等(2010)的方法。Mosleh

一类基于交叉熵的区间直觉模糊多属性群决策的新方法

李宝萍^{1,2}, 陈华友²

(1.安徽三联学院 基础部, 合肥 230601; 2.安徽大学 数学科学学院, 合肥 230601)

摘 要:针对决策信息为区间直觉模糊数且属性和专家权重完全未知的多属性群决策问题, 文章提出了一类新的基于交叉熵的群决策方法。引入两个直觉模糊集交叉熵的概念, 并以交叉熵作为决策信息的差异程度的度量, 从而提出了多属性群决策中确定属性权重和专家权重的方法。文中给出了基于交叉熵原理的区间直觉模糊多属性群决策方法的计算步骤。

关键词:群决策; 区间直觉模糊数; 交叉熵; 权重

中图分类号: C934; N945

文献标识码: A

文章编号: 1002-6487(2014)01-0020-04

0 引言

多属性群决策问题即利用多个决策者给出的决策信息对一组备选方案进行排序和择优的具体过程, 是现代决策科学的重要组成部分, 已被广泛的应用于社会学、经济学、管理学等领域。群决策过程中, 专家针对具体的属性对方案进行的评价是方案最终结果优劣的关键。由于不

同专家的知识结构、个人偏好以及对方案和属性的了解程度不一, 所做评价的结果也会存在很大差异, 因而如何选取适当的权重也成为决策的关键。基于区间直觉模糊信息的群决策问题方面, 目前研究相对较少。徐泽水等^[1]提出区间直觉模糊偏好信息的有序加权集成算子和混合集成算子, 通过引入得分矩阵和精确矩阵, 并基于集成算子给出了相应的群决策方法。卫贵武^[2]提出直觉模糊诱导几何集成算子, 并应用于 MAGDM 问题中, 同时将其方法拓

基金项目:国家自然科学基金资助项目(71071002); 安徽省高等学校自然科学基金资助项目(KJ2012A026); 安徽大学学术创新团队项目(KJTD001B, SKTD007B)

作者简介:李宝萍(1976-), 女, 安徽巢湖人, 硕士研究生, 讲师, 研究方向: 预测和决策分析。

陈华友(1969-), 男, 安徽和县人, 博士, 教授, 研究方向: 预测和决策分析。

等(2010)是通过构造模糊神经网络, 采用的梯度下降法进行求解的, 虽然其结果较好, 但其计算过程只适用于三角模糊数以及输入为精确数的模糊回归模型。而本文的方法可在获得近似最优解的前提下, 处理各种类型的模糊数据以及输入、输出和系数均为模糊数的回归模型。

4 结论

模糊回归分析是分析模糊数据之间关系的一种有效方法, 其在处理一些无法获得精确数, 只能得到可能性数据的对象时发挥了重要作用。本文考虑了模糊数据的回归中不同隶属度下误差的模糊性, 认为不同隶属度下的误差对总误差的贡献不同, 提出了与隶属度关联的拟合效果评价方法。从启发式算法入手, 采用模拟退火算法并进行相应的改进, 达到了在获得近似最优解的前提下, 克服了以往仅考虑系数和变量中的一种具有模糊性的局限性, 给出了能处理各类模糊观测值以及系数和变量均具模糊性的线性和非线性回归模型的启发式求解思路。对拟合系数显著性的检验、模糊回归与时间序列分析的融合以及模糊样条回归方法等是模糊回归的进一步研究方向。

参考文献:

- [1] Zadeh L. A. Fuzzy Sets[J]. Information and Control, 1965, 8(3).
- [2] Tanaka H, Uejima S, Asia K. Linear Regression Analysis with Fuzzy Model[J]. IEEE Trans. Sys. Man and Cyber. Smc., 1982, 12(6).
- [3] 李竹渝, 张成. 模糊数据的回归模型结构分析[J]. 统计研究, 2008, 25(8).
- [4] M. Mosleha, M. Otadi, S. Abbasbandy. Evaluation of Fuzzy Regression Models by Fuzzy Neural Network[J]. Journal of Computational and Applied Mathematics, 2010, 234(3).
- [5] 康立山, 谢云, 尤矢勇等. 非数值并行计算, 模拟退火算法(第一册)[M]. 北京: 科学出版社, 1997.
- [6] Hsien-Chung Wu. Fuzzy Estimates of Regression Parameters in linear Regression Models for Imprecise Input and Output Data[J]. Computational Statistics & Data Analysis, 2003, 42(1~2).
- [7] H. Tanaka, I. Hayashi, J. Watada. Possibilistic Linear Regression Analysis for Fuzzy Data[J]. European Journal of Operational Research, 1989, 40(3).
- [8] C. Kao, C.L. Chyu. Least-squares Estimates in Fuzzy Regression Analysis[J]. European Journal of Operational Research, 2003, 148(2).

(责任编辑/亦 民)