文章编号:1006-544X(2004)01-0107-04

单因变量的偏最小二乘法在双曲递减中的应用

郭大浩,邓英尔,管英柱

(成都理工大学"油气藏地质及开发工程"国家重点实验室,四川 成都 610059)

摘 要:二元回归法是一种重要的求解油气产量双曲递减参数的方法,但在二元回归方程中,2个自变量的相关系数较大,存在严重的多重相关性,这样会使最小二乘法失效,得到的回归模型的拟合效果不好.为了较好地解决多重相关性问题,基于陈元千教授提出的二元回归方程,提出了用偏最小二乘法来求解递减参数.介绍了偏最小二乘法的基本原理、建模基本思想和交叉有效性判别法并讨论了该方法在双曲递减方程中的应用.实例表明偏最小二乘法解决了二元回归方程中的多重相关性问题,得到了较好的拟合效果.

关键词:偏最小二乘回归方法;双曲递减;多重相关性

中图分类号:TE319;O212.1 **文献标识码**:A①

自从J. J. ARPS于1945年提出双曲递减模型以来,国内外众多的油气藏科技人员对双曲递减方程的建立和求解进行了深入细致的研究,先后提出了试差法[1,2]、图解法[1,2]、曲线位移法[1,3]、陈氏二元回归法[4]、Towler & Bansal线性回归法[5]、非线性曲线拟合法[6]、遗传算法[7]等.这些方法各有特点,其中试差法、图解法、曲线位移法由于受人为因素的影响,会造成多解;Towler & Bansal线性回归法、非线性曲线拟合法、遗传算法弥补了多解的不足,得到了最优解,但算法较复杂;二元回归法是一种重要的求解递减参数的方法,但在回归方程中2个自变量Q和Qt的相关系数较大,存在严重的多重相关性,这样会使最小二乘法失效,破坏参数估计,对回归系数的统计检验造成了一定的困难,得到的回归模型也达不到较好的拟合效果.

偏最小二乘回归方法是近年来兴起的一种多元统计分析方法,它将主成分分析、典型相关分析和多元线性回归分析三者综合在一起,能够在自变量存在严重的多重相关性的条件下进行回归建模[8].本文针对二元回归法中的多重相关性问题讨论了单因变量的偏最小二乘法在双曲递减方程中的应用.

1 双曲递减的二元回归方程[4]

在双曲递减方程中,有

$$Q = Q_i/(1 + nD_it)^{1/n}. \tag{1}$$

由(1)式积分可得到相应时刻的累积产量方程

$$N_p = \frac{EQ_i^n}{D_i}\left(\frac{1}{1-n}\right)(Q_i^{1-n} - Q^{1-n}), \tag{2}$$

(2)式又可化为

$$N_p = \frac{EQ_i}{D_i}\left(\frac{1}{1-n}\right) - \frac{E}{D_i}\left(\frac{1}{1-n}\right)Q - \left(\frac{En}{1-n}\right)Qt.$$

若设

$$a_0 = \frac{EQ_i}{D_i}\left(\frac{1}{1-n}\right),\quad a_1 = \frac{E}{D_i}\left(\frac{1}{1-n}\right),$$

$$a_2 = En/1-n,$$

则得 $$N_p = a_0 + a_1Q + a_2Qt. \tag{3}$$

如果再设$x_1 = Q, x_2 = Qt, y = N_p$,则由(3)式可得二元回归分析的一般形式

$$y = a_0 + a_1x_1 + a_2x_2. \tag{4}$$

① **收稿日期**:2003-04-23

基金项目:国家973项目(2002CB211700)

作者简介:郭大浩(1979-),男,硕士研究生,油气田开发专业.

根据递减阶段实际的生产数据,由(4)式进行二元回归分析后,可得常系数 a_0,a_1,a_2 的数值,然后确定 Q_i,D_i,n 的数值

$$Q_i = -a_0/a_1; \tag{5}$$

$$D_i = -E/a_1(1-n); \tag{6}$$

$$n = a_2/(a_2 - E). \tag{7}$$

2 偏最小二乘法的原理[8]

2.1 基本思想

设有单因变量 y 和 p 个自变量 $\{x_1, x_2, \cdots, x_p\}$,观测 n 个样本点,构成自变量和因变量的数据表 $\boldsymbol{X} = [x_1, x_2, \cdots, x_p]_{n\times p}$ 和 $\boldsymbol{Y} = [y]_{n\times p}$.

在一般的多元线性回归模型中,当数据总体能够满足高斯 - 马尔科夫假设条件时,根据最小二乘法,有

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{8}$$

$\hat{\boldsymbol{Y}}$ 是 $\boldsymbol{Y}$ 的一个很好的估计量. 从式(8)容易看出,$(\boldsymbol{X}'\boldsymbol{X})$ 必须是可逆矩阵,但当 $\boldsymbol{X}$ 中的变量存在严重的多重相关性时,这个最小二乘估计量就会失效,并引发一系列应用方面的问题,而偏最小二乘回归分析可以避免 $(\boldsymbol{X}'\boldsymbol{X})$ 的不可逆性.

偏最小二乘回归的基本思路是:分别在 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 中提取成分 $\boldsymbol{t}_1$ 和 $\boldsymbol{u}_1$($\boldsymbol{t}_1$ 是 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p$ 的线性组合,$\boldsymbol{u}_1$ 是 $\boldsymbol{Y}$ 的线性组合,即 y 本身). 在提取这 2 个成分时,为了回归分析的需要,有下列要求:

(1) $\boldsymbol{t}_1$ 和 $\boldsymbol{u}_1$ 应尽可能多地携带其各自数据表中的变异信息;

(2) $\boldsymbol{t}_1$ 和 $\boldsymbol{u}_1$ 的相关程度能够达最大.

这 2 个要求表明,$\boldsymbol{t}_1$ 和 $\boldsymbol{u}_1$ 应尽可能好地代表数据表 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$,同时自变量的成分 $\boldsymbol{t}_1$ 对因变量的成分 $\boldsymbol{u}_2$ 有最强的解释能力.

在第 1 个成分 $\boldsymbol{t}_1$ 和 $\boldsymbol{u}_1$ 被提取后,偏最小二乘回归分别实施 $\boldsymbol{X}$ 对 $\boldsymbol{t}_1$ 的回归以及 $\boldsymbol{Y}$ 对 $\boldsymbol{t}_1$ 的回归. 如果回归方程已经达到满意的精度,则算法终止;否则,将利用 $\boldsymbol{X}$ 被 $\boldsymbol{t}_1$ 解释后的残余信息以及 $\boldsymbol{Y}$ 被 $\boldsymbol{t}_1$ 解释后的残余信息进行第 2 轮的成分提取. 如此往复,直到能达到一个较满意的精度为止. 若最终对 $\boldsymbol{X}$ 共提取了 m 个成分 $\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_m$,偏最小二乘回归将进行 y 对 $\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_m$ 的回归,然后再表达成 y 关于原变量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p$ 的回归方程.

2.2 算法推导

记 $\boldsymbol{F}_0$ 是因变量 $\boldsymbol{Y}$ 的标准化矩阵,有

$$\boldsymbol{F}_{0i} = (Y_i - \overline{Y})/s_y,\ i = 1, 2, \cdots, n.$$

式中:$\overline{Y}$,s_y 分别为 $\boldsymbol{Y}$ 的均值、标准差.

记 $\boldsymbol{E}_0$ 为自变量集合 $\boldsymbol{X}$ 的标准化矩阵,有

$$\boldsymbol{E}_{ij} = (X_{ij} - \overline{X}_j)/s_j,\ i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, p.$$

式中,$\overline{X}_j$,s_j 分别是第 j 个自变量的均值、标准差.

第 1 步,从 $\boldsymbol{F}_0$ 中提取一个成分 $\boldsymbol{u}_1$,$\boldsymbol{u}_1 = \boldsymbol{F}_0\boldsymbol{c}_1$,$\boldsymbol{c}_1$ 是 $\boldsymbol{F}_0$ 的第一主轴,并且 $\|\boldsymbol{c}_0\| = 1$;从 $\boldsymbol{F}_0$ 中提取一个成分 $\boldsymbol{t}_1$,$\boldsymbol{t}_1 = \boldsymbol{F}_0\boldsymbol{w}_1$,$\boldsymbol{w}_1$ 是 $\boldsymbol{F}_0$ 的第一主轴,并且 $\|\boldsymbol{w}_1\| = 1$,如果要求 $\boldsymbol{t}_0$,$\boldsymbol{u}_1$ 能分别很好地代表 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 中的数据变异信息,根据提取成分的要求有:$\boldsymbol{t}_1$ 与 $\boldsymbol{u}_1$ 的标准差 $\mathrm{Var}(\boldsymbol{t}_1)$,$\mathrm{Var}(\boldsymbol{u}_1)$ 趋于最大,$\boldsymbol{t}_1$ 与 $\boldsymbol{u}_1$ 的相关系数 $\mathrm{r}(\boldsymbol{t}_1, \boldsymbol{u}_1)$ 趋于最大,即要求 $\boldsymbol{t}_1$ 与 $\boldsymbol{u}_1$ 的协方差达到最大,即

$$\mathrm{Cov}(\boldsymbol{t}_1, \boldsymbol{u}_1) = \sqrt{\mathrm{Var}(\boldsymbol{t}_1)\mathrm{Var}(\boldsymbol{u}_1)}\,\mathrm{r}(\boldsymbol{t}_1, \boldsymbol{u}_1) \to \max.$$

也就是说,在 $\|\boldsymbol{w}_0\| = 1$ 和 $\|\boldsymbol{c}_0\| = 1$ 的约束条件下去求 $\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{c}_1$ 的最大值,采用拉格朗日算法,记对

$$s = \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{c}_1 - \lambda_1(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 - 1) - \lambda_2(\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{c}_1 - 1) \to \max.$$

对 s 分别求关于 $\boldsymbol{w}_1$,$\boldsymbol{c}_1$,λ_1,λ_2 的偏导,并令之为零,有

$$\partial s/\partial \boldsymbol{w}_1 = \boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{c}_1 - 2\lambda_1\boldsymbol{w}_1 = 0,$$

$$\partial s/\partial \boldsymbol{c}_1 = \boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{w}_1 - 2\lambda_2\boldsymbol{c}_1 = 0,$$

$$\partial s/\partial \lambda_1 = -(\boldsymbol{w}_0^{\mathrm{T}}\boldsymbol{w}_1 - 1) = 0,$$

$$\partial s/\partial \lambda_2 = -(\boldsymbol{c}_0^{\mathrm{T}}\boldsymbol{c}_1 - 1) = 0.$$

通过推导可得

$$\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{w}_1 = \theta_1^2\boldsymbol{w}_1;\ \boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{c}_1 = \theta_1^2\boldsymbol{c}_1.$$

式中:$\theta_1 = 2\lambda_1 = 2\lambda_2 = \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{c}_1$.

可见,$\boldsymbol{w}_1$ 是矩阵 $\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0$ 的特征向量,对应的特征值为 θ_1^2. 由上可知 θ_1 为优化问题的目标函数值,它要求取最大值,所以,$\boldsymbol{w}_1$ 是对应于 $\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0$ 矩阵特征值的单位特征向量. 另一方面,$\boldsymbol{c}_1$ 是对应于 $\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{E}_0\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{F}_0$ 矩阵最大特征值 θ_1^2 的单位特征向量.

求得主轴 $\boldsymbol{w}_1$ 和 $\boldsymbol{c}_1$ 后,即可得到成分

$$\boldsymbol{t}_1 = \boldsymbol{E}_0\boldsymbol{w}_1;\quad \boldsymbol{u}_1 = \boldsymbol{F}_1\boldsymbol{c}_1.$$

然后,分别求 $\boldsymbol{E}_0$ 和 $\boldsymbol{F}_0$ 对 t_1 的 2 个回归方程

$$\boldsymbol{E}_0 = \boldsymbol{t}_1\boldsymbol{P}_1^{\mathrm{T}} + \boldsymbol{E}_1;\quad \boldsymbol{F}_0 = \boldsymbol{t}_1\boldsymbol{r}_1^{\mathrm{T}} + \boldsymbol{F}_1.$$

式中，回归系数向量为

$$\boldsymbol{P}_1 = \boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{t}_1 / \|\boldsymbol{t}_1\|^2;\ \boldsymbol{r}_1 = \boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{t}_1 / \|\boldsymbol{t}_1\|^2.$$

而 $\boldsymbol{E}_1$，$\boldsymbol{F}_1$ 分别是两个回归方程的残差矩阵.

第 2 步，用残差矩阵 $\boldsymbol{E}_1$ 和 $\boldsymbol{F}_1$ 取代 $\boldsymbol{E}_0$ 和 $\boldsymbol{F}_0$，用同样方法求第 2 个轴 $\boldsymbol{w}_2$ 和 $\boldsymbol{c}_2$ 以及第 2 个主成分 $\boldsymbol{u}_2$ 和 $\boldsymbol{t}_2$，如此计算下去，如果进行了 m 次运算，则会有

$$\boldsymbol{E}_0 = \boldsymbol{t}_1\boldsymbol{P}_1^{\mathrm{T}} + \boldsymbol{t}_2\boldsymbol{P}_2^{\mathrm{T}} + \cdots + \boldsymbol{t}_{\mathrm{m}}\boldsymbol{P}_{\mathrm{m}}^{\mathrm{T}} + \boldsymbol{E}_{\mathrm{m}},$$

$$\boldsymbol{F}_0 = \boldsymbol{t}_1\boldsymbol{r}_1^{\mathrm{T}} + \boldsymbol{t}_2\boldsymbol{r}_2^{\mathrm{T}} + \cdots + \boldsymbol{t}_{\mathrm{m}}\boldsymbol{r}_{\mathrm{m}}^{\mathrm{T}} + \boldsymbol{F}_{\mathrm{m}}.$$

2.3　交叉有效性判别

在偏最小二乘回归建模中，究竟应该选取多少个成分为宜，这可通过考察增加一个新的成分以后，能否对模型的预测功能有明显的改进来考虑. 用交叉有效性进行判别：除去某个样本点 i 的所有样本集合作为一个并使用 h 个成分拟合一个回归方程；然后把排除的样本点 i 代入前面拟合的回归方程，得到 y_i 在样本点 i 上的拟合值 $\hat{y}_{h(-i)}$，对每一个样本点重复上述计算，定义 y_i 的预测误差平方和为 PRESS_h，有

$$\mathrm{PRESS}_h = \sum_{i=1}^{n} (y_i - \hat{y}_{h(-i)})^2.$$

另外，再采用所有的样本点，拟合含 h 个成分的回归方程. 这时，记第 i 个样本点的预测值为 $\hat{h}i$，则可以定义 y_1 的误差平方和为 SS_h，有

$$\mathrm{SS}_h = \sum_{i=1}^{n} (y_i - \hat{y}_{hi})^2.$$

一般说来，总是有 $\mathrm{PRESS}_h > \mathrm{SS}_h$，而 $\mathrm{SS}_h < \mathrm{SS}_{h-1}$. 下面比较 SS_{h-1} 和 PRESS_h，SS_{h-1} 是用全部样本点拟合的具有 $(h-1)$ 个成分的方程的拟合误差. PRESS_h 增加了一个成分 t_h，但却含有样本点的扰动误差，如果 PRESS_h 在一定程度上小于 SS_{h-1}，则认为增加一个成分 t_h，会使预测的精度明显提高. 因此希望 $(\mathrm{PRESS}_h/\mathrm{SS}_{h-1})$ 的比值越小越好. 一般认为：当 $\mathrm{PRESS}_h/\mathrm{SS}_{h-1} \leqslant 0.95^2$ 时，增加成分是有益的；否则，就认为增加新的成分 t_h 对减少方程的预测误差无明显的改善作用.

交叉有效性也可定义为

$$Q_h^2 = 1 - \frac{\mathrm{PRESS}_h}{\mathrm{SS}_{h-1}}.$$

当 $Q_h^2 \geqslant (1 - 0.95^2) = 0.097\ 5$ 时，表明加入成分能改变模型质量，否则不能.

3　算例分析

以某井的实际开发数据(表 1)为例[2]，该井数据曾被多篇文章引用，比较有代表性.

表 1　某气井的实际生产数据与偏最小二乘法的预测结果比较

Table 1　Comparison of actual data and predicted results by partial least squares

t/a	$Q/(10^4\mathrm{m}^3 \cdot \mathrm{d}^{-1})$		$N_\mathrm{p}/10^4\mathrm{m}^3$	
	实际	预测	实际	预测
0	28.47	28.47	0	−5.65
0.5	23.91	23.88	4 753.77	4 687.44
1	20.27	20.33	8 767.44	8 921.2
1.5	17.53	17.52	12 240.25	12 201.1
2	15.26	15.27	15 229.15	15 222.7
2.5	13.44	13.42	17 847.99	17 767.3
3	11.90	11.89	20 153.71	20 121.4
3.5	10.59	10.62	22 146.31	22 291.2
4	9.56	9.54	24 025.05	23 956.9
相对误差平均/%	0.016		0.027	

设 $y = N_p, x_1 = Q, x_2 = Qt$，根据实例得到了 $n(n=9)$ 个样本点，由此构成了因变量数据表 $\boldsymbol{Y} = [y]_{9\times2}$ 和自变量数据表 $\boldsymbol{X} = [x_1, x_2]_{9\times2}$，由于 y，x_1，x_2 数量级相差太大，所以应作标准化处理，设

$$y_i^* = \frac{y_i - \bar{y}}{s_y},\ x_{i1}^* = \frac{x_1 - \bar{x}_1}{s_{x_1}},\ x_{i2}^* = \frac{x_2 - \bar{x}_2}{s_{x_2}}$$

$$(i = 1,2,\cdots,9).$$

式中 $\bar{y}, \bar{x}_1, \bar{x}_2$ 分别是 y, x_1, x_2 的均值，S_y, S_{x_1}, S_{x_2} 分别是 y, x_1, x_2 的标准差.

由此构成了标准化的因变量数据矩阵 $\boldsymbol{F}_0 = (y^*)_{9\times1}$ 和标准化的自变量数据矩阵 $\boldsymbol{E}_0 = (x_1^*, x_2^*)_{9\times2}$.

3.1　用最小二乘法计算

y^* 对 x_1^*，x_2^* 进行二元回归，经最小二乘计算后，结果如下：

测定系数 $R^2 = 0.594\ 058$；F 检验值 $F = 19.806\ 6$；回归方程的回归系数及 t 检验值见表 2；同时计算得到 x_1^* 与 x_2^* 的相关系数 $r = -0.992$，可见存在严重的多重相关性.

由上述计算结果可知，F 检验通过，但是所有回归系数均未通过 t 检验，且自变量 x_2^* 的回归系数显著为零，这与实际模型相违背.

表 2　最小二乘法回归系数及 t 检验值

Table 2　Regression coefficients and t testing value by least squares

变量	回归系数	t 检验值
x_1^*	−1.62	−0.782
x_2^*	0.005	0.002 4

3.2 用偏最小二乘法计算

利用前面的分析，用偏最小二乘法计算得到：

成分 $t_1 = -0.71\ x_1^* + 0.70\ x_2^*$；

$t_2 = -0.71\ x_1^* + 0.70\ x_2^*$.

在进行交叉有效性检查时，以 $Q_h^2 \geqslant 0.0975$ 作为是否显著的标志，由交叉有效性结果（表 3）可知，需采用 t_1, t_2 两个成分进行偏最小二乘回归，y^* 关于 t_1, t_2 的回归方程为

$$y^* = 0.70\ t_1 + 1.59\ t_2.$$

将 t_1, t_2 代入得到

$$y^* = -1.62\ x_1^* - 0.63\ x_2^*.$$

将标准化变量 y^*, x_1^*, x_2^* 分别还原成原始变量 y，x_1, x_2，则回归方程为

$$y = 58\,927.5 - 2\,070.01x_1 - 397x_2.$$

由上式可知回归系数分别为：$a_0 = 58\,927.5$，$a_1 = -2\,070.01$，$a_2 = -397$.

将 a_0, a_1, a_2 的值代入(5)，(6)，(7) 式分别计算递减参数，可得

$$n = 0.52, Q = 28.47, D_i = 0.368.$$

根据上面的参数，利用双曲递减方程即可进行瞬时产量与累积产量的预测（表 1），从预测结果及误差分析可知，精度令人满意.

表 3 偏最小二乘的交叉有效性判别

Table 3 Cross validation table by partial least squares

成分个数	Q_h^2	临界值
1	0.994 8	0.097 5
2	0.997 6	0.097 5

4 结 论

通过以上实例表明，偏最小二乘法能很好地应用于油气产量双曲递减的二元回归方程中，较好的处理了二元回归方程中的多重相关性问题. 同时也表明该方法具有很高的理论意义和实用价值.

符号说明：

a_0, a_1, a_2—待定系数；N_p—累积产量，$10^4 m^3$；Q—瞬时产量，$10^4 m^3/d$ 或 $10^4 m^3/a$；Q_i—递减初期的瞬时产量，$10^4 m^3/d$ 或 $10^4 m^3/a$；D_i—初始递减率，1/a 或 1/mon；n—递减指数，$0 < n < 1$；N—测试点数；t—递减期生产时间，a；E—单位换算量.

参考文献

[1] 童孝华，匡建超. 油气藏工程基础 [M]. 北京：石油工业出版社，1996. 173 - 182.

[2] 陈元千，李 璽. 现代油藏工程 [M]. 北京：石油工业出版社，2001. 127 - 140.

[3] 陈元千. 产量递减分析的一个简单方法 [J]. 天然气工业，1989，10 (2)：24 - 27.

[4] 陈元千. 确定递减类型的新方法 [J]. 石油学报，1990，11 (1)：74 - 80.

[5] Towler B，Bansal S. Hyperbolic decline curve analysis using linear regression [J]. J. of Petr. Science and Engineering，1993，(8)：257 - 268.

[6] 陈林媛，杜志敏. 双曲递减参数求解的非线性曲线拟合法 [J]. 西南石油学院学报，1999，21 (3)：1 - 4.

[7] Sinisha A Jikich，Andrei S Popa. Hyperbolic Decline Parameter Identification Using Optimization Procedures [C]. Paper SPE 65634 presented at the 2000 SPE Eastern Regional Meeting held in Morgantown，West Virginia，17 - 19 October 2000.

[8] 王惠文. 偏最小二乘回归方法及其应用 [M]. 北京：国防工业出版社，1999.

Application of simple partial least squares regression to hyperbolic decline curve

GUO Da-hao，DENG Ying-er，GUAN Ying-zhu

(*State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation*，*Chengdu University of Technology*，*Chengdu* 610059，*China*)

Abstract：Binary line regression is an important method to calculate Arps hyperbolic decline parameters，but in the regression function the relativity of two independent variables Q and Qt is high，which will bring about the problem of serious multicollinearity. Generally least-squares regression method can not deal with multicollinearity. To resolve this problem，Partial Least Squares (PLS) to obtain the decline parameters based on the binary line regression function proposed by Prof. Chen Yuanqian is adopted . The basic principle ，the basic modeling method and the cross validation of PLS regression are introduced and the application of simple PLSs regression to hyperbolic decline curve is discussd . It is shown that the PLS regression can be used to the binary line regression function of Arps hyperbolic decline to solve multicolnearity perfectly.

Key words：partial least square regression ；hyperbolic decline ；multicollinearity