第 6 卷第 6 期　　管 理 科 学 学 报　　Vol. 6 No. 6
2003 年 12 月　　JOURNAL OF MANAGEMENT SCIENCES IN CHINA　　Dec. , 2003

# 基于 GRPLS 回归的中国经济增长影响因素分析①

许和连[1] , 赖明勇[1] , 许青松[2]
(1. 湖南大学经济与贸易学院 , 长沙 410082 ; 2. 湖南大学数学与计量经济学院 ,长沙 410082)

摘要 : 构造一种新的方法 ——岭 - 偏最小二乘回归方法(它既有效消除了因素变量之间的多重共线性 ,又克服了传统方法的不足 ,且使模型更加稳健 ,具有更强的预测和分析能力) ;并运用广义岭 - 偏最小二乘回归方法分析了我国经济增长的影响因素 ,为我国制订持续、快速增长的经济政策提供了有益的参考.

关键词 : 经济增长 ; 广义岭 - 偏最小二乘回归 ; 影响因素
中图分类号 : F12　　文献标识码 : A　　文章编号 : 1007 - 9807(2003) 06 - 0008 - 07

## 0　引　言

经济学家从来没有停止对经济增长是由什么因素决定问题的探讨. 早期的观点认为一国财富的增长是由自然资源、劳动力和物质资本的投入引起的 ,其中物质资本是经济增长的关键因素. 20 世纪 60 年代末和 70 年代初 ,舒尔茨的人力资本理论受到广泛重视 ,人力资本被认为是比物质资本更为重要的促进经济增长的因素 ;20 世纪 70 年代以来 ,科学技术在经济增长中的重要作用受到了人们的重视 ,更多的经济学家把经济增长的关键原因归于科学技术的进步. 关于经济增长影响因素的定量分析主要有 :以丹尼森 (Denison) 为代表的增长核算分析 ,以钱纳里 (Chenery) 和巴罗 (Barro) 为代表的经济计量分析 ,以及乔根森 (Jorgenson) 的生产率度量方法. 当然 ,不同国家与地区以及不同的阶段 ,经济增长的影响因素是有差异的 ,实证研究的结论体现出这一点. 在实证研究过程中 ,研究者一般采用跨国 (地区) 的截面数据或单个国家 (地区) 的时间序列数据 ,方法则一般采用最小二乘法、多元统计分析中的主成分方法.

中国经济增长因素问题 ,已有的一些研究工作主要从实证的角度进行了探讨 ,由于数据资料收集的局限以及模型方法采用的不同得到了不同的结论. 如文献[1]对生产率与中国经济增长的研究 ,文献[2]对中国国有企业效率决定因素的研究 ,文献[3]对中国经济增长因素的实证分析等. 实证分析基本上是在生产函数的基础上运用普通最小二乘回归方法进行分析 ,考虑的因素主要涉及资本、劳动力、产出和技术等因素.

实证研究中 ,如果采用截面数据 ,运用普通最小二乘法的多元线性回归 ,要求所选取的样本点 (不同的国家或地区) 具有相同的经济结构和生产技术 ,这在现实经济中是无法满足的[4]. 影响一国经济增长的因素有很多 ,而且不同的因素变量之间都不同程度地存在多重共线性或近似多重共线性关系. 对于存在多重共线性关系的变量运用普通最小二乘法 ,将使模型极其不稳定 ,甚至得出与现实相反的结论 ,因而不能解释所要说明的问题. 采用主成分分析方法能有效地消除所选取自变量之间的多重共线性. 但主成分方法在分析中

① 收稿日期 : 2002 - 03 - 06 ; 修订日期 : 2003 - 06 - 19.
基金项目 : 高校优秀青年教师教学科研奖励计划资助项目 ; 国家自然科学基金资助项目 (70273010) ; 国家社会科学基金资助项目 (01CJY023) .
作者简介 : 许和连(1971 —) ,男 ,博士.

只考虑了自变量(解释变量)所包含的信息,没有涉及因变量(被解释变量)的信息.

本文运用一种新的方法——广义岭-偏最小二乘回归(generalized ridge partial least-squares regression,GRPLS)方法,从更加广阔的范围分析影响我国经济增长的因素,不仅可有效消除自变量之间的多重共线性,同时最大限度地考虑了因变量与自变量的信息,而且使模型更加稳健,预测与分析能力更强.

## 1 理论与方法

### 1.1 偏最小二乘回归

偏最小二乘(PLS)回归是一种新型的多元统计数据分析方法,H. Wold 和 C. Albano 等提出以后,PLS 回归方法得到了广泛的应用,尤其是在化学和化工领域. PLS 回归方法是一种消除自变量多重共线性的有效方法,从某种意义说,PLS 回归方法是改进了的主成分回归(principal components regression,PCR)分析方法,但又不同于主成分分析方法,PLS 方法在成分提取的过程中不仅考虑自变量(解释变量)的信息,同时考虑了因变量(被解释变量)的信息. PLS 回归方法有单因变量的 PLS 回归与多因变量的 PLS 回归,本文只涉及单因变量.

#### 1.1.1 单因变量 PLS 回归方法建模思路

设因变量为 $Y$ 和 $p$ 个自变量 $\{x_1, \cdots, x_p\}$,为了研究因变量与自变量间的统计关系,观测了 $n$ 个样本点,构成了因变量 $Y$ 的 $n$ 次观测值构成的 $n$ 维向量和由自变量 $x_1, \cdots, x_p$ 的观测值构成的 $n \times p$ 的观测矩阵 $X = [x_1, \cdots, x_p]_{n\times p}$. PLS 回归方法是首先在矩阵 X 中提取成分 $t_1$($t_1$ 为 $x_1, \cdots, x_p$ 的线性组合),要求 $t_1$ 尽可能大地携带 X 中的变异信息,且与 $Y$ 的相关程度最大,这样,$t_1$ 尽可能好地综合了 X 的信息,同时对 $Y$ 又有最强的解释能力. 在第一个成分 $t_1$ 被提取后,PLS 回归分别实施 X 对 $t_1$ 的回归及 $Y$ 对 $t_1$ 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 X 被 $t_1$ 解释后的残余信息进行第二轮的成分提取. 如此反复计算,直到达到满意的精度为止. 若最终对 X 共提取了 $k$ 个成分 $t_1, \cdots, t_k$,PLS 回归将通过实施 $Y$ 对 $t_1, \cdots, t_k$ 的回归,表达成 $Y$ 关于原变量 $x_1, \cdots, x_p$ 的线性方程[5].

#### 1.1.2 模型表述

考虑回归模型

$$Y = X\beta + \varepsilon,\ E(\varepsilon) = 0,\ Cov(\varepsilon) = \delta^2 I \quad (1)$$

其中:$\beta$ 为 $p \times 1$ 的回归系数向量;$X = [x_1, \cdots, x_p]_{n\times p}$,为样本观测矩阵;$\varepsilon$ 为 $n \times 1$ 的随机误差向量;Y 为 $n \times 1$ 的观测向量;I 为 $n \times n$ 单位矩阵.

设矩阵 X 的秩 $R(X) = q$,将 X 分解,分解后的矩阵记为

$$X = \sum_{i=1}^{q} t_i p_i^T = TP^T \quad (2)$$

这里:$T = [t_1, t_2, \cdots, t_q]$;$P = [p_1, p_2, \cdots, p_q]$;$t_i$ 为 $n \times 1$ 的向量;$p_i$ 为 $p \times 1$ 的向量;T 为 $n \times q$ 矩阵;P 为 $p \times q$ 矩阵;且向量 $t_i$、$p_i$ 和矩阵 $X_i$($X_i$ 为 $n \times p$ 的矩阵,$i = 1, 2, \cdots, q$)满足

$$t_i = X_i w_i,\quad X_{i+1} = X_i - t_i p_i^T$$

$$p_i = X_i^T t_i / \|t_i\|^2,\ \text{向量}\ w_i = X_i^T Y / \|X_i^T Y\|$$

其中:$\|\cdot\|$ 表示向量的模;$w_i$ 为 $p \times 1$ 的向量;$X_i^T$ 表示矩阵 $X_i$ 的转置;其它变量相同符号表示的含义相同.

再考察向量 $t_i$ 与矩阵 $X_i$ 之间的关系

$$t_1 = X_1 w_1 = Xh_1$$

$$\begin{aligned} t_i &= X_i w_i = (X_{i-1} - t_{i-1} p_{i-1}^T) w_i = \\ &X_{i-1}(I - w_{i-1} p_{i-1}^T) w_i = \cdots = \\ &X_1 (I - w_1 p_1^T)(I - w_2 p_2^T) \cdots \\ &(I - w_{i-1} p_{i-1}^T) w_i = X_1 h_i = Xh_i \end{aligned}$$

这里:$X_1 = X$;$h_1 = w_1$;$h_i = (I - w_1 p_1^T)(I - w_2 p_2^T) \cdots (I - w_{i-1} p_{i-1}^T) w_i$,$i \geq 2$;I 为 $p \times p$ 的单位矩阵. 令 $H = (h_1, h_2, \cdots, h_q)$,H 为 $p \times q$ 矩阵,则有

$$T = XH$$

将等式(2)代入式(1),有

$$Y = TP^T\beta + \varepsilon = T\alpha + \varepsilon \quad (3)$$

其中,$\alpha = P^T\beta$.

则等式(3)的最小二乘解为

$$\hat{\alpha} = (T^T T)^{-1} T^T Y$$

得向量 Y 的拟合向量为

$$\hat{Y} = T\hat{\alpha} = XH(T^T T)^{-1} T^T Y$$

因此,若将所有的成分放入模型,则 $\beta$ 的 PLS 估计为

$$\hat{\beta}_p = H(T^T T)^{-1} T^T Y$$

这时,$\hat{\beta}_p$ 与式(1)中的 LS 估计 $\hat{\beta}_L$ 相同.

由于 PLS 估计 $\hat{\beta}_p$ 是一个非线性估计,其均方

误差 MSE(mean square error) 很难有明确的表达式. 但是,当随机误差不大时,可近似将矩阵 H 和 T 看成非随机的[6]. 为了研究的需要,本文将矩阵 H 和 T 均近似视为确定性的,由此可得

$$
\begin{aligned}
MSE(\hat{\beta}_p) &= E(\hat{\beta}_p - \beta)^T(\hat{\beta}_p - \beta) = \\
&E(\mathrm{tr}((\hat{\beta}_p - E(\hat{\beta}_p))^T(\hat{\beta}_p - E(\hat{\beta}_p)))) + \\
&(E(\hat{\beta}_p) - \beta)^T(E(\hat{\beta}_p) - \beta)) \approx \\
&\sigma^2 \sum_{i=1}^{q} h_i^T h_i / t_i^T t_i + \left\| E(\hat{\beta}_p) - \beta \right\|^2
\end{aligned}
\tag{4}
$$

这里,tr(·) 表示矩阵的迹. 式(4) 第 1 项是估计 $\hat{\beta}_p$ 的方差,第 2 项是估计 $\hat{\beta}_p$ 的偏差.

在建模的过程中,并非所有的成分都是模型必须的,当变量之间存在多重共线性时,有一些 PLS 成分的方差非常小($t_i^T t_i \to 0$). 由式(4) 可知,方差很小的 PLS 成分对 $MSE(\hat{\beta}_p)$ 影响非常大,甚至可使估计 $\hat{\beta}_p$ 失效,使模型失去解释能力. 为了解决这个问题,通常的做法是将方差很小的 PLS 成分从模型中删除,降低模型的维数,即提取最能代表系统的信息和对系统具有最强解释能力的成分.

因此将矩阵 T, P 和 H 做如下划分:

$$
\left.
\begin{aligned}
T &= [t_1, t_2, \cdots, t_k, \cdots t_{k+1}, \cdots, t_q] = \\
&\quad [T_1 \cdots T_2] \\
P &= [p_1, p_2, \cdots, p_k, \cdots p_{k+1}, \cdots, p_q] = \\
&\quad [P_1 \cdots P_2] \\
H &= [h_1, h_2, \cdots, h_k, \cdots h_{k+1}, \cdots, h_q] = \\
&\quad [H_1 \cdots H_2]
\end{aligned}
\right\}
\tag{5}
$$

将式(5) 代入式(3) ,有

$$Y = T_1 P_1^T \beta + T_2 P_2^T \beta + \varepsilon \tag{6}$$

令 $\alpha^T = [\alpha_1^T, \alpha_2^T] = [(P_1^T \beta)^T, (P_2^T \beta^T)]$

则式(6) 可写为

$$Y = T_1 \alpha_1 + T_2 \alpha_2 + \varepsilon \tag{7}$$

假设矩阵 X 最终共提取了 $k$ 个成分,意味着前面的 $k$ 个综合变量已经最大限度地携带了 X 的信息,且对 Y 又有最强的解释能力. 则将后面的 $q - k$ 个成分从模型中删去,模型(7) 转化为如下降维模型

$$Y = T_1 \alpha_1 + \varepsilon \tag{8}$$

模型(8) 的最小二乘估计为

$$\hat{\alpha}_1 = (T_1^T T_1)^{-1} T_1^T Y = (T_1^T T_1)^{-1} H_1^T X^T Y$$

从而得到含有前面 $k$ 个 PLS 成分的参数 $\beta$ 的 PLS 估计为

$$\hat{\beta}_p(k) = H_1 (T_1^T T_1)^{-1} H_1^T X^T Y \tag{9}$$

而且 $\hat{\beta}_p$ 的模满足

$$
\begin{aligned}
\left\| \hat{\beta}_p(k) \right\| &= \left\| H_1 (T_1^T T_1)^{-1} H_1^T X^T X (X^T X)^{-1} X^T Y \right\| \leq \\
&\left\| H_1 (T_1^T T_1)^{-1} H_1^T X^T X \right\| \left\| \hat{\beta}_L \right\| = \left\| \hat{\beta}_L \right\|
\end{aligned}
\tag{10}
$$

其中,$\hat{\beta}_L$ 为模型(1) 的 LS 估计.

式(10) 表明 PLS 估计 $\hat{\beta}_p(k)$ 为 LS 估计的压缩估计,且有

$$MSE\left(\hat{\beta}_p(k) \approx \sigma^2 \sum_{i=1}^{k} h_i^T h_i / t_i^T t_i + \left\| E(\hat{\beta}_p) - \beta \right\|^2\right)$$

### 1.2 岭 - 偏最小二乘回归

从式(10) 可看出,总体上 PLS 估计压缩了 LS 估计,但并不是所有的变量都得到压缩,相反 PLS 估计 $\hat{\beta}_p$ 往往膨胀了 LS 估计 $\hat{\beta}_L$ 的前面若干分量[7],在回归分析中,这是首先要预防的. 借助岭估计与主成分方法相结合的思想,把岭估计与 PLS 估计有机地结合起来,得到普通的岭 - 偏最小二乘 (RPLS) 回归与广义岭 - 偏最小二乘 (GRPLS) 回归方法.

#### 1.2.1 普通岭 - 偏最小二乘(ridge partial least-squares, RPLS) 回归

在模型(1) 中,系数向量 $\beta$ 的普通岭估计为

$$\hat{\beta}_R = (X^T X + dI)^{-1} X^T Y \tag{11}$$

其中,$d > 0$,为岭参数. 从式(11) 可以看出,岭估计是把 $X^T X$ 用 $X^T X + dI$ 来代替,这样做的理由很明显,因为当自变量之间存在多重共线性时,$X^T X$ 的特征根至少有一个非常接近于 0,而 $X^T X + dI$ 的特征根是在原来 $X^T X$ 每个特征根上加了一个 $d$,从而改善了特征根接近于 0 的可能性.

这里对式(9) 的 $\hat{\beta}_p$ 作如下变化,得到 RPLS 估计 $\hat{\beta}_{RPLS}$,令

$$\hat{\beta}_{RPLS}(k) = H_1 (T_1^T T_1 + dI)^{-1} H_1^T X^T Y \tag{12}$$

同理可以得到

$$
\begin{aligned}
MSE\Big(\hat{\beta}_{RPLS}(k) &\approx \sigma^2 \sum_{i=1}^{k} h_i^T h_i / (t_i^T t_i + d) + \\
&\left\| E(\hat{\beta}_{RPLS}(k)) - \beta \right\|^2\Big)
\end{aligned}
\tag{13}
$$

对于矩阵 $H_1$,其秩 $R(H_1) = k$,则总存在 $P \times P$ 初等方阵 F,使得 $FH_1 = \begin{bmatrix} I_{k \times k} \\ 0 \end{bmatrix}$,有

$$\|\beta_{RPLS}(k)\| = \|H_1(T_1^T T_1 + dI)^{-1} T_1^T T_1 (FH_1)^T F H_1 (T_1^T T_1)^{-1} H_1^T X^T Y\| =$$
$$\|H_1(T_1^T T_1 + dI)^{-1} T_1^T T_1 (FH_1)^T F \beta_p(k)\| \le \|H_1(T_1^T T_1 + dI)^{-1} T_1^T T_1 (FH_1)^T F\| \, \|\beta_p(k)\| =$$
$$\max_{1 \le i \le k} \{ t_i^T t_i / (t_i^T t_i + d) \}^{1/2} \|\beta_p(k)\| \le \|\beta_p(k)\|$$

从上式可知, $\beta_{RPLS}(k)$ 是 $\beta_p(k)$ 的压缩估计, RPLS 估计比 PLS 估计有更多的优点, 它可用比 PLS 更多的成分, 避免信息的损失, 而且由于岭参数的加入, 可减少估计的均方误差, 提高模型的稳定性与预测能力, 显然 $\beta_p(k)$ 是 $\beta_{RPLS}(k)$ 的一种特殊情形, 如果岭参数 $d$ 选择适当, $\beta_{RPLS}(k)$ 有望比 $\beta_p(k)$ 更小的均方误差.

1.2.2　广义岭 - 偏最小二乘 (generalized ridge partial least-squares, GRPLS) 回归

在 RPLS 回归中, 对不同的 PLS 成分都使用了同一个参数 $d$, 即对矩阵 $T^T T$ 的每一个特征根上加了一个常数 $d$. 这里结合广义岭回归的思想, 将式 (12) 改写为如下的形式, 得到广义岭 - 偏最小二乘回归 $\beta_{GRPLS}$

$$\beta_{GRPLS}(k) = H_1(T_1^T T_1 + D)^{-1} H_1^T X^T Y \qquad (14)$$

其中, $D = \mathrm{diag}[d_1, d_2, \cdots, d_k]$, $d_i \ge 0$ $(i = 1,2, \cdots, k)$, 为岭参数.

显然, 在 $\beta_{GRPLS}$ 中, 不同的 PLS 的成分使用了不同的参数, 且 $\beta_{GRPLS}(k)$ 的均方误差可近似写为

$$MSE(\beta_{GRPLS}(k)) \approx \sigma^2 \left( \sum_{i=1}^{k} h_i^T h_i / (t_i^T t_i + d) \right) + \|E(\beta_{GRPLS}(k) - \beta)\|^2$$

由于

$$\|\beta_{GRPLS}(k)\| = \|H_1(T_1^T T_1 + D)^{-1}(T_1^T T_1 + dI)(FH_1)^T F H_1 (T_1^T T_1 + dI)^{-1} H_1^T X^T Y\| \le$$
$$\|H_1(T_1^T T_1 + D)^{-1}(T_1^T T_1 + dI)(FH_1)^T F\| \, \|\beta_{RPLS}(k)\| =$$
$$\max_{1 \le i \le k} \left\{ (t_i^T t_i + d) / (t_i^T t_i + d_i) \right\}^{1/2} \|\beta_{RPLS}(k)\|$$

当选择适当的岭参数 $d_i$ $(i = 1,2, \cdots, k)$ 时, 可使

$$\max_{1 \le i \le k} \left\{ (t_i^T t_i + d) / (t_i^T t_i + d_i) \right\}^{1/2} \le 1$$

由此可得

$$\|\beta_{GRPLS}(k)\| \le \|\beta_{RPLS}(k)\|$$

因此, 当岭参数合适时, $\beta_{GRPLS}$ 是 $\beta_{RPLS}$ 的压缩估计. 引入更多的岭参数的目的, 一是为了降低估计的均方误差, 二是为了使不同方差 PLS 的成分得到较公平的对待, 比RPLS 更合理. 所以, 当岭参数合适时, 通过 GRPLS 得到的模型更加稳定, 有更强的解释和预测能力.

### 1.3　PLS 成分数和岭参数的确定

岭 - 偏最小二乘回归方法有两种参数要确定: 一是 PLS 成分数的确定, 二是岭参数的确定.

PLS 成分数采用国外广泛应用的交互检验 (cross validation, CV) 方法确定, 先构造统计量预测误差平方和 (prediction residual error sum of squares, PRESS), 然后求使其达到最小的成分数 $k^*$, 即为所求. 关于 PRESS 的构造: 把所有 $n$ 个样本点分成两部分, 第 1 部分是除去某个样本点 $i$ 的所有样本点集合, 用这部分样本点并用 $k$ 个 PLS 成分拟合一个回归方程, 第 2 部分是把被排除的样本点 $i$ 代入前面拟合的回归方程, 得到 $Y$ 在样本点 $i$ 上的拟合值 $\hat{y}_{(-i)}$, 对于每一个 $i = 1,2, \cdots, n$, 重复上述测试, 则可得到 $Y$ 的 $PRESS(k)$.

$$PRESS(k) = \sum_{i=1}^{n} (y - \hat{y}_{(-i)})^2$$

由此可得

$$PRESS(k^*) = \min_{1 \le k \le q} \left[ \sum_{i=1}^{n} (y - \hat{y}_{(-i)})^2 \right] \qquad (15)$$

其中, $k^*$ 即为所求的 PLS 成分数.

岭参数的确定也采用交互检验方法. 对于 RPLS (见式 (12)), 先对每一给定的成分数 $k$, 计算最优的 $PRESS$ 值及所对应的 $d$, 然后从 $q$ 个结果 $(k = 1,2, \cdots, q)$ 中选出最好的, 这样, 同时确定了 $k$ 与 $d$. 对于 GRPLS (见式 (14)), 计算要复杂一些, 本文给出一个类似向前法的算法来确定成分数 $k$ 和岭参数 $d_i$ $(i = 1,2, \cdots, q)$.

算法步骤:

0° 取 $k = 1$

1° 利用等式(12) 和(15) ,求出使 *PRESS* 最小的 $d$,记为 $d^*$.

令 $D_0 = \mathrm{diag}[d^*, d^*, \cdots, d^*]$.

2° 令 $D_1 = \mathrm{diag}[d, d^*, \cdots, d^*]$, 求出使 *PRESS* 最小的 $d$,记为 $d_1^*$.

令 $D_2 = \mathrm{diag}[d_1^*, d, d^*, \cdots, d^*]$, 又求出 *PRESS* 最小的 $d$,记为 $d_2^*$.

令 $D_3 = \mathrm{diag}[d_1^*, d_2^*, d, d^*, \cdots, d^*]$, …, 如此向前,直到得到 $d_k^*$.

3° 令 $D_0 = \mathrm{diag}[d_1^*, d_2^*, \cdots, d_k^*]$,返回步骤 2°,如此循环,直到收敛.

4° 令 $k = k + 1$, 返回 1°.

最后,从 $q$ 次循环中($k = 1, 2, \cdots, q$) 取最优的 *PRESS* 值所对应的 $D$ 和 $k$,记为 $k^*$ 和 $d_i^*$($i = 1, 2, \cdots, k^*$) 即为所求.

## 2 实证分析

运用广义岭 - 偏最小二乘(GRPLS) 回归方法分析我国经济增长的影响因素.

### 2.1 变量和数据的选取

中国目前所处的发展阶段,在一个较长的时期内其经济增长的过程,也是工业化的过程,城市化的过程. 增长的过程涉及生产、投资、贸易及收入分配等. 本文在分析我国经济增长的影响因素变量时,以文献[8,9] 的研究为基础,除了以经济理论为基础和吸收已取得的研究成果之外,还考虑了变量数据的客观性和可获得性,选出 12 个变量做为影响我国经济增长(GDP) 的因素:劳动投入(Lab),人力资本投入(HC),物质资本投入(MC),技术水平(Tec),人口增长(PG),进口(IM),出口(EX),外资(FC),总消费支出(CE),城乡结构(CCS),产业结构(IS),政府消费支出(GE).

以国内生产总值(GDP,亿元) 表示我国的经济增长. 在 12 个变量指标中,劳动投入以我国历年社会劳动者人数(万人) 表示;人力资本以我国历年中等及以上学校在校学生人数(万人) 表示;物质资本存量的计算遵循文献[10] 的计算方法,对 1980—2000 年期间我国全社会固定资产投资总额进行了推算;技术水平根据"余值法"计算,取 1979 年数据为初始数据;人口增长以人口自然增长率表示;进口、出口分别以历年我国进口商品额(亿元)、出口商品额(亿元) 表示;外资(亿元) 包括外商直接投资、对外借款及外商其他投资;总消费支出以历年全社会消费支出额(亿元) 表示;城乡结构以我国城镇人口占总人口比例表示;产业结构以第二、三产业所占比例表示;政府消费支出以国家财政支出用于国防、行政管理方面的费用(亿元) 表示.

所有变量的数据均选取 1980—2000 年的年度数据(原始数据资料来自《中国统计年鉴》、《中国经济年鉴》各期). 为了研究的方便,考虑对各时序数据取对数以后并不影响变量之间的关系,对各变量数据作对数处理,处理后的时序变量分别记为: $Y = \ln(\mathrm{GDP})$, $x_1 = \ln(\mathrm{Lab})$, $x_2 = \ln(\mathrm{HC})$, $x_3 = \ln(\mathrm{MC})$, $x_4 = \ln(\mathrm{Tec})$, $x_5 = \ln(\mathrm{PG})$, $x_6 = \ln(\mathrm{IM})$, $x_7 = \ln(\mathrm{EX})$, $x_8 = \ln(\mathrm{FC})$, $x_9 = \ln(\mathrm{CE})$, $x_{10} = \ln(\mathrm{CCS})$, $x_{11} = \ln(\mathrm{IS})$, $x_{12} = \ln(\mathrm{GE})$.

### 2.2 PLS 成分数与岭参数的选取

首先分析 12 个变量间的相关性,发现它们之间存在严重的多重共线性关系,因此不能用普通最小二乘法的多元线性回归分析,因为这样可能会导致模型的无效,因而运用 GRPLS 回归方法来建模分析.

先运用交互检验(CV) 方法及前面提出的方法确定 PLS 成分和岭参数,从计算结果发现取 4 个 PLS 成分满足要求,在 RPLS 中 $d$ 取 0.48,GRPLS 中取 $D = \mathrm{diag}(0.85, 0, 2.88, 0)$ 满足算法要求,且各方法计算出的 *PRESS* 值见下图(PLS 成分数均取 4 个).

图 1 中,PLS 的 *PRESS* 值为 0.066 9,RPLS 的 *PRESS* 值为 0.065 9, GRPLS 的 *PRESS* 值仅为 0.035 6,说明了 GRPLS 回归方法建立的模型具有比其它两种方法更好的预测能力,由于样本数有限,本文没有计算各方法的均方误差(MSE).

### 2.3 GRPLS 回归方法计算的模型结果

利用 GRPLS 方法计算得到我国经济增长影响因素的回归模型,结果见表 1.

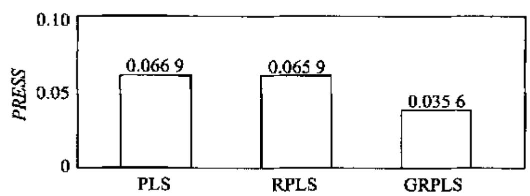


图 1 三种方法计算的 *PRESS* 值

表 1　我国经济增长影响因素的 GRPLS 回归模型结果

| 因变量 $Y$ | | | | | |
|---|---|---|---|---|---|
| 自变量 | 系数值 | 自变量 | 系数值 | 自变量 | 系数值 |
| 常数项 | 9.934 5 | $x_5$ | -0.030 1 | $x_{10}$ | 0.024 7 |
| $x_1$ | 0.138 5 | $x_6$ | 0.056 9 | $x_{11}$ | 0.059 9 |
| $x_2$ | 0.170 5 | $x_7$ | 0.134 3 | $x_{12}$ | 0.196 2 |
| $x_3$ | 0.048 6 | $x_8$ | 0.111 5 | | |
| $x_4$ | 0.078 8 | $x_9$ | 0.226 9 | | |

### 2.4　结果分析

分析表 1 的模型计算结果,可以得到以下几点:

1) 所有变量前的系数符号均符合经济含义. 只有人口增长变量前的系数为负,说明人口增长对我国经济增长起阻碍作用. 改革开放以来,我国一直把计划生育作为一项基本国策来实施,人口增长率从 1998 年开始已下降到 1 %以内,这与本文的研究结果一致.

2) 总消费支出变量前的系数为 0.226 9,说明消费支出对我国经济增长的影响显著. 为保证经济持续增长,应该积极鼓励社会消费. 政府消费对我国经济增长的影响也很明显(系数为 0.192 6),因此适当增加政府的消费,尤其是国防费用的投入,一方面可以通过国防技术水平的提高,带动整个技术的提升;另一方面,可以增加国家的安全.

3) 对经济增长影响最基本的三个要素:劳动、资本、技术中,人力资本的作用表现显著,物质资本的作用较小,劳动的投入对经济增长的影响作用也很明显,而技术水平对经济增长的作用表现相对较弱,这与文献[3]的结论相似. 说明改革开放以来,我国经济增长主要依靠资本及劳动力投入等生产要素的增加,技术水平增长的贡献相对来说有限. 而从长期的角度,经济增长主要依靠技术的进步及人力资本的积累,人力资本积累主要表现在对教育的投入上. 因此,为了实现我国经济的持续稳定增长及经济增长方式的转变,必须大力增加科技投入,以及技术的引进,不断提高科技因素在经济增长中的贡献,同时在现有人力资本积累的基础上,对专业教育和职业教育等方面进一步加大投入,不断提高国民的整体教育水平,增加人力资本积累.

4) 在对外的因素中,出口对我国经济增长的作用最为明显,其次是外资的引入,然后是进口. 有实证研究表明,改革开放以来,我国实施的是出口导向型的外贸政策[11],出口对经济增长的作用明显,进口不仅要受国内经济政策的影响,同时要受国内消费、投资和出口需求的影响,表现出相对较弱的作用. 外资对我国经济增长作用显著,主要是因为,一方面外资流入为我国经济的持续高速增长注入了大量资本,弥补了改革初期我国的资金短缺;另一方面引进外资推动了我国改革的进程,带来了新的事物和符合国际惯例的制度,通过外资进入的"干中学"的示范效应对我国技术的提升起了积极的作用.

5) 从结构指标中城乡结构与产业结构对我国经济增长的作用来看,表现不是很突出,这可能与指标的选取有关系. 产业结构对经济增长的作用要明显地大于城乡结构. 从数据可以看出,我国的第二、三产业,尤其是第三产业尽管从改革开放以来有了比较大的提高,但是与国外发达国家相比还有比较大的差距. 经济学家钱纳里、库兹涅茨等通过对 100 多个不同收入水平国家的分析得出结论:经济增长过程其核心是产业结构的转变过程. 我国的产业结构还有待于进一步的调整与提升,以促进经济的高速增长. 城乡结构代表一个国家城市化水平,2000 年我国的城镇人口占总人口比重为 36.22 %,而 1993 年全世界城市人口占总人口已达 44 %,中等收入国家的城市人口比重更达达 60 %. 这说明我国的城市化水平还很低,而城市化水平是一个国家工业化程度的标志. 因此,我国必须加大城镇化建设,加快工业化进程,以保证经济持续、快速增长.

## 3　结论

1) 通过分析指出了当自变量之间存在多重共线性或近似多重共线性关系时,采用普通最小二乘法和主成分方法有许多不足之处. 把 PLS 回归与岭回归相结合,得到岭-偏最小二乘回归方法,包括 RPLS 和 GRPLS 回归,尤其是 GRPLS 回归方法,不仅能克服上述方法的不足,有效地消除自变量之间的多重共线性,而且使得模型更加稳定,有更强的解释和预测能力.

2) 运用 GRPLS 回归方法分析了改革开放以来(1980—2000 年)我国经济增长的影响因素,结果发现,在所选取的 12 个变量中,只有人口增长

对我国经济增长起着阻碍作用,其它变量指标的增长对我国经济增长均表现为促进作用.其中总消费支出对我国经济增长的影响显著;以行政管理和国防费用表示的政府消费对经济增长的作用超出了作者的预料;对经济增长影响最为基本的三个要素——劳动、资本、技术中,人力资本与劳动投入的作用显著,技术水平的作用相对比较小;对外的因素中,出口、进口、外资引入三个变量以出口与外资引入表现突出,进口的作用相对较小;在结构指标中,产业结构对经济增长的作用明显大于城乡结构的作用.

因此,为了使我国经济能保持高速、稳定的增长,可以采取以下措施:积极鼓励和刺激社会消费,适当加大政府消费,尤其是国防费用的投入;通过加大教育投入,增加人力资本积累;增加科技投入和技术引进,提高技术水平对经济增长的作用;实行积极的出口政策和外资引进政策,特别是在我国已加入 WTO 的情况下进一步加大出口,尤其是附加值高的工业制成品出口,根据我国目前的产业政策有选择地加大外资的引入,并适当增加符合我国经济发展需求的商品进口;积极调整产业结构,提高第二和第三产业的份额,提升和优化产业结构;加快城镇化速度,提高城市化水平;一如继往地实施计划生育政策,降低人口增长率.

参 考 文 献:

[1]李京文,郑友敬,等. 中国经济增长分析[J]. 中国社会科学, 1992, (1): 15—36
[2]刘小玄,郑京海. 国有企业效率的决定因素(1985—1994)[J]. 经济研究, 1998, (1): 37—46
[3]沈坤荣. 1978—1997 年中国经济增长因素的实证分析[J]. 经济科学, 1994, (4): 14—24
[4]许和连,赖明勇. 出口导向经济增长(ELG)的经验研究:综述与评论[J]. 世界经济, 2002, (2): 43—49
[5]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
[6] Xu Qing-song, Liang Yi-zeng, Shen Hai-lin. Generalized PLS regression[J]. Journal of Chemometrics, 2001, 15: 135—148
[7]Iidiko E F, Jerome H F. A statistical view of some chemometrics regression tools[J]. Technometrics, 1993, 35: 109—135
[8]Temple J. The new growth evidence[J]. Journal of Economic Literature, 1999, (March): 112—156
[9]姚愉芳,贺菊煌,等. 中国经济增长与可持续发展——理论、模型与应用[M]. 北京:社会科学文献出版社, 1998
[10]贺菊煌. 我国资产的估计[J]. 数量经济技术经济研究, 1992, (8): 24—27
[11]沈程翔. 中国出口导向型经济增长的实证分析(1977—1998)[J]. 世界经济, 1999, (12): 26—30
[12]Frankel J A, David R. Does trade cause growth[J]. American Economic Review, 1999, (7): 379—399
[13]Skuldsson H A. PLS regression methods[J]. Journal of Chemometrics, 1988, (2): 211—228
[14]Lucas R E. On the mechanics of economic development[J]. Journal of Monetary Economics, 1998, 22: 3—42

## GRPLS regression on influencing factors on China's economic growth

*XU He-lian*[1], *LAI Ming-yong*[1], *XU Qing-song*[2]
1. College of Economics and Trade, Hunan University, Changsha 410082, China;
2. College of Mathematics and Econometrics, Hunan University, Changsha 410082, China

**Abstract:** This paper constructed a new method —Gernenalized ridge partial least-squares regression (GRPLS). It can not only effectively eliminate the multi-collinearity among the factors and overcome the shortcoming of the traditional methods, but also makes the model more steady and more strong to predict and analysis. Then, the GRPLS is applied to analysis the impact factors of economic growth on China. The conclusion can offer beneficial advice to work out steady and quickly increasing economic polices in China.

**Key words:** economic growth; generalized ridge partial least-squares regression; impact factors