

PLS2 回归计算顾客满意度指数

李勇¹, 赵艳桃²

(1 中国矿业大学 理学院, 北京 100083; 2 中华女子学院 基础部, 北京 100101)

摘要: 介绍了 PLS2 回归计算顾客满意度指数的建模方法. 对于算法中涉及的求高阶实对称阵的最大特征值及其特征向量, 采用幂法加以实现. 给出了一个具体例子, 计算出了顾客满意度、忠诚度等二级指标的值, 并对此做了简要分析. PLS2 回归较好地克服了各指标间的多重共线性问题, 通过此方法求得的顾客满意度指数更准确、合理.

关键词: 顾客满意度指数; PLS 回归; 幂法; 多重共线性

中图分类号: O212.7 文献标识码: A 文章编号: 1007-855X(2006)01-0115-04

Computing Customer Satisfaction Indices by PLS2 Regression

LI Yong¹, ZHAO Yan-tao²

(1 Maths Department, China University of Mining and Technology Beijing 100083, China)

2 Basic Department, China Women's University Beijing 100101, China)

Abstract The method of model building for the computation of customer satisfaction indices by PLS2 regression is introduced. As for computing the largest eigenvalue and its eigenvector of a higher-order real symmetric matrix in the algorithm, the power method is used. And an example is given. Then the values of level two indices customer satisfaction, customer loyalty are computed and briefly analyzed. PLS2 regression successfully solve the multicollinearity problems among the indices, which indicates that the customer satisfaction indices computed are better and more reasonable.

Key words customer satisfaction indices; PLS regression; power method; multicollinearity

0 引言

20 世纪 70 年代, 一些发达国家开始着手研究顾客满意度, 逐步开展了顾客满意度指数测评. 顾客满意度是顾客满意程度的简称, 是顾客对某种产品或服务满意程度的心理感受. 顾客满意度指数是通过特定的因果关系模型对顾客满意程度心理感受的测评结果. 瑞典设计了第一个顾客满意度指数测评标准——SCSB 标准, 美国在此基础上建立了自己的顾客满意度指数模型——ACSI 模型 (图 1), 此模型被很多国家采用.

中国也逐步开展了顾客满意度指数测评. 1999 年, 国务院颁布了《关于进一步加强产品质量工作若干问题的决定》, 提出“要研究和探索产品质量用户满意度指数评价方法, 向消费者提供真实可靠的产品质量信息”. 中国的顾客满意度指数模型是在 ACSI 模型的基础上建立的, 它可

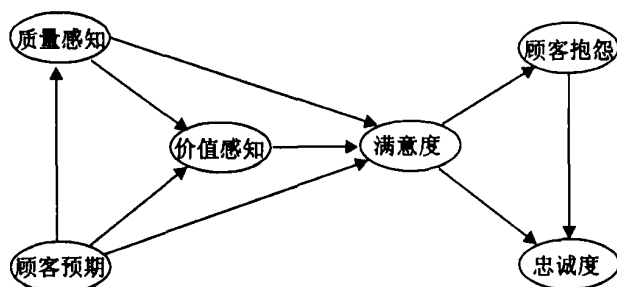


图1 ACSI模型

Fig. 1 ACSI model

收稿日期: 2005-03-15 基金项目: 国家自然科学基金资助项目 (项目编号: 10372112).

第一作者简介: 李勇 (1979.7~), 男, 在读硕士研究生. 主要研究方向: Bootstrap 方法及其应用. E-mail: ly9802@eyou.com

能与ACSI模型中的二级指标个数不同.二级指标较抽象,由一些三级指标来测定(图中没画出),被调查顾客对它们打分(五分制、七分制等),根据这些值计算出顾客满意度指数.对不同的产品或服务,三级指标的个数和内容可能不同.

PLS回归(Partial Least-squares Regression 偏部分最小二乘回归)于1983年由瑞典的S.Wold和C.Abbano等人首次提出,它最初产生于化学领域,是一种新型的多元数据统计分析方法.近年来,它在理论、方法和应用方面都得到了广泛地发展,在工农业生产、社会生活等方面均有所应用.

本文介绍了PLS2回归计算顾客满意度指数的建模方法,并对算法的实现做了介绍.通过一个实例,计算出了顾客满意度指数.

1 PLS回归

PLS回归的主要应用是多因变量对多自变量的回归建模,特别是在变量多重相关系统中建模,在样本量个数小于自变量个数的设计中建模.它将回归建模、主成分分析、典型相关分析融合在一起,在一次计算之后,可同时实现预测建模、两组变量间的相关分析以及对多变量系统的综合简化.ACSI的创立者——米西根大学的Fornell教授称PLS回归为第二代回归分析方法.

在顾客满意度指数测评中,解释变量(三级指标)较多,变量间往往存在多重相关性,这就很难利用传统的最小二乘回归方法.为解决此问题,通常仅选取少数几个解释变量,但这必然造成信息的损失和难以估计预测精度的问题.主成分分析和岭回归分析是两个常用方法.但当解释变量很多时,这两种方法的计算量很大;主成分分析还有如何决定哪些主成分要删去的问题,岭回归分析还有估计岭参数的问题.此外,如图1所示,顾客满意度指数测评中被解释变量顾客满意度、忠诚度等二级指标也存在多重相关性的问题.PLS回归较好地克服了多重共线性问题.它不仅吸取了主成分分析中从解释变量提取信息的思路,又注意了主成分分析中所忽略的自变量对因变量的解释问题.PLS回归可在比主成分分析少用因子的情况下达到较小的均方误差,其计算量比主成分分析和岭回归分析小,计算结果更可靠.

按因变量的个数,PLS回归模型可分为PLS1、PLS2回归模型,它们分别为单因变量、多因变量的偏最小二乘回归模型.

2 PLS2回归计算顾客满意度指数

设顾客满意度指数测评中有 p 个自变量 X_1, X_2, \dots, X_p , q 个因变量 Y_1, Y_2, \dots, Y_q .不妨设自变量组和因变量组的 n 次标准化观测数据阵分别为 $X_0 = (x_{ki}^0)_{n \times q} = (x_{11}, x_{12}, \dots, x_{1p})_{n \times p}$, $Y_0 = (y_{kj}^0)_{n \times q} = (y_{11}, y_{12}, \dots, y_{1q})_{n \times q}$ 其中 $x_i = (x_{1i}^0, x_{2i}^0, \dots, x_{ni}^0)^T$, $y_j = (y_{1j}^0, y_{2j}^0, \dots, y_{nj}^0)^T$, $i = 1, 2, \dots, p$; $j = 1, 2, \dots, q$

分别从 X_0, Y_0 中提取第一成分 t_1, u_1 ,使它们尽可能多地包含 X_0, Y_0 中的变异信息,且使它们相关性最大.如达到精度要求,就不再抽取成分,求出 X_0, Y_0 关于 t_1 的回归方程;若否,再分别从残差阵 X_1, Y_1 中提取成分 t_2, u_2 ,使它们尽可能多地包含 X_1, Y_1 中的变异信息,且使它们相关性最大.再次提取成分直到满足精度要求为止.设 X_0 的秩为 m .

算法如下:

1) 求矩阵 $(Y_{s-1}^T X_{s-1})^T Y_{s-1}^T X_{s-1}$ 的最大特征值 λ_s 及其特征向量 ω_s ,将 ω_s 单位化为 ω_s^0 .求矩阵 $(X_{s-1}^T Y_{s-1})^T X_{s-1}^T Y_{s-1}$ 的最大特征值 μ_s 及其特征向量 c_s ,将 c_s 单位化为 c_s^0 .再求

$$t_s = (t_{1s}, t_{2s}, \dots, t_{ms})^T = X_{s-1} \omega_s^0, p_s = \frac{X_{s-1}^T t_s}{\|t_s\|^2}, X_s = X_{s-1} - t_s p_s^T; u_s = (u_{1s}, u_{2s}, \dots, u_{qs})^T = Y_{s-1} c_s^0, r_s = \frac{Y_{s-1}^T t_s}{\|t_s\|^2},$$

$$Y_s = Y_{s-1} - t_s r_s^T, s = 1, 2, \dots, m;$$

2) 分别求 X_0, Y_0 关于 t_1, t_2, \dots, t_m 的回归方程

$$X_0 = (x_1, x_2, \dots, x_p)_{n \times p}$$

$$= t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T + X_m$$

$$\approx X_0 \{ \omega_1^0 p_1^T + (E_p - \omega_1^0 p_1^T) \omega_2^0 p_2^T + \dots + (E_p - \omega_1^0 p_1^T) (E_p - \omega_2^0 p_2^T) \dots (E_p - \omega_{m-1}^0 p_{m-1}^T) \omega_m^0 p_m^T \}$$

$$= (x_1, x_2, \dots, x_p)_{n \times p} (a_{ij})_{p \times q}$$

$$Y_0 = (y_1, y_2, \dots, y_q)_{n \times q}$$

$$= t_1 r_1^T + t_2 r_2^T + \dots + t_m r_m^T + Y_m$$

$$= X_0 \{ \omega_1^0 r_1^T + (E_p - \omega_1^0 p_1^T) \omega_2^0 r_2^T + \dots + (E_p - \omega_1^0 p_1^T) (E_p - \omega_2^0 p_2^T) \dots (E_p - \omega_{m-1}^0 p_{m-1}^T) \omega_m^0 r_m^T \}$$

$$= (x_1, x_2, \dots, x_p)_{n \times p} (b_{ij})_{p \times q}$$

其中 $E_p = \text{diag}(1, 1, \dots, 1)$ 为 p 阶单位阵;

3) 分别求 $x_i^* = x_b$, $y_j^* = y_p$ 关于 $x_i = x_b$, $i = 1, 2, \dots, p$ 的回归方程

$$x_i \approx \sum_{k=1}^p a_{ki} x_k, y_j \approx \sum_{k=1}^p a_{kj} x_k, x_i^* = x_b, y_j^* = y_p \text{ 则 } x_i^* \approx \sum_{k=1}^p a_{ki} x_k^*, y_j^* \approx \sum_{k=1}^p b_{kj} x_k^*, i = 1, 2, \dots, p; j = 1, 2, \dots, q$$

4) 由以上几步, 就可求出各变量观测值的对应预测值. 将各三级指标的预测值进行主成分分析, 加权平均求出二级指标的值.

注: ① 相关著作介绍了 PLS 回归的理论依据、简化算法等^[4-6]. ② 算法中涉及的求高阶实对称阵的最大特征值及其对应的特征向量, 可采用幂法加以实现^[7]. 求 n 阶实对称阵 A 的最大特征值 λ_1 及对应的特征向量 x_1 的步骤: 取 $u_0 = v_0 = (1, 1, \dots, 1)^T$ (n 维向量); $v_m = A u_{m-1}$, $\mu_m = \max(|v_m|) / \max(|u_{m-1}|)$ (为向量 v_m 的绝对值最大的分量), $u_m = v_m / \mu_m$; 取 μ_m 作为 λ_1 的近似, u_m 作为 x_1 的近似. ③ 通常, PLS 回归并不需要选用 m (X_0 的秩) 个成分来建立回归方程, 只选用前几个成分就可得到预测效果较好的回归模型. 对于抽取成分数的确定, 可采用 QR 方法分别求出 X_0, Y_0 的所有特征值^[7], X_0, Y_0 的最大特征值分别除以它们的所有特征值的和为累计贡献率, 由此可粗略判断出抽取成分数. 如某残差阵的各元素近似为 0 也可认为达到了精度要求. SAS 8.0 中的 PLS 过程也可实现 PLS 回归. 该软件采用了舍一交叉验证、分批交叉验证、分裂样本交叉验证方法确定抽取成分数. 较准确地确定抽取成分数, 现在通常采用交叉有效性.

3 例子及分析

本文选用 2003 年调查获得的某品牌女鞋的顾客满意度指数测评数据 (数据由北京某数据科技有限公司提供), 用 PLS2 回归计算其满意度、忠诚度等二级指标的值. 有效调查问卷共 151 份. 自变量 16 个, 分别为: X_{11} —质量期望, X_{12} —时尚性, X_{13} —档次, X_{21} —总体质量, X_{22} —皮革质量, X_{23} —做工精细程度, X_{24} —款式多样性, X_{25} —款式漂亮程度, X_{26} —穿着舒适度, X_{27} —耐穿性, X_{28} —服务态度, X_{29} —服务及时性, X_{210} —对促销方式的满意度, X_{211} : 对生产厂商售后服务的满意度, X_{31} —相对该产品品质, 其价格的合理程度, X_{32} —与价格相比, 对产品品质的满意度; 因变量 6 个, 分别为: Y_{11} —总体满意度, Y_{21} —抱怨与否, Y_{22} —抱怨次数, Y_{31} —再次购买的可能性, Y_{32} —推荐给亲友的可能性, Y_{33} —如其它同档次品牌以 9 折促销, 仍购买该品牌的可能性. 三级指标 $X_{11}, X_{12}, X_{13}, X_{21}, X_{22}, X_{23}, X_{24}, X_{25}, X_{26}, X_{27}, X_{28}, X_{29}, X_{210}, X_{211}, X_{31}, X_{32}, Y_{11}, Y_{21}, Y_{22}, Y_{31}, Y_{32}, Y_{33}$ 对应的二级指标分别为: X_1 —顾客预期; X_2 —质量感知; X_3 —价值感知; Y_1 —满意度; Y_2 —顾客抱怨; Y_3 —顾客忠诚.

由交叉有效性及多次实验表明, 分别抽取 3 个成分 $t_1, u_1; t_2, u_2; t_3, u_3$, 然后进行回归, 就可得到精度较高的 PLS2 回归模型. 采用 2 中的算法, 得到 $X_1, X_2, X_3, Y_1, Y_2, Y_3$ 的值分别为 86.4 78.9 75.1 77.1 0.96 77.5 (百分制).

就该品牌女鞋而言, 顾客满意度指数测评中的两重要指标——满意度和忠诚度的分值还不太高, 这就说明了该品牌女鞋在某些方面还没很好地达到顾客的要求. 进一步分析各指标间的相关性, 并与同行业其它品牌相比较, 明确该品牌的优势和劣势, 有针对性地加以改进, 工作重点放在产品质量和服务上, 努力加以改善与提高, 及时处理顾客的抱怨与投诉, 减小他们的负面影响, 努力使他们成为忠诚顾客.

4 结论

PLS2 回归较好地克服了顾客满意度指数中各指标间的多重共线性问题. 本文给出了由 PLS2 回归计

算顾客满意度指数的建模方法,通过此方法求出的顾客满意度、忠诚度等顾客满意度指数比直接由三级指标主观赋权数,加权平均求出的结果更合理、准确,且易于编程,易于上机实现。

参考文献:

- [1] 唐晓芬. 顾客满意度测评[M]. 上海:上海科学技术出版社, 2001
- [2] 国家质检总局质量管理司,清华大学中国企业研究中心. 中国顾客满意度指数指南[M]. 北京:中国标准出版社, 2003
- [3] CLAES F. A National Customer Satisfaction Barometer: The Swedish Experience[J]. Journal of Marketing, 1992, 56
- [4] TENENHAUS M, GAUCHI J P, MENARDO C. Regression PLS et application[J]. Revue de Statistiques Appliquées, 1995, 53(1): 7 ~ 63
- [5] TENENHAUS M. La régression PLS théorie et pratique[M]. Paris: Editions Technip, 1998
- [6] 王惠文. 偏最小二乘回归及其应用[M]. 北京:国防工业出版社, 1999
- [7] 李庆扬,王能超,易大义. 数值分析[M]. 北京:清华大学出版社, 2002
- [8] 高惠璇. 实用统计方法与SAS系统[M]. 北京:北京大学出版社, 2001
- [9] SAS Institute. SAS OnlineDoc(HTML Format)[OL/EB]. <http://v8doc.sas.com/sashtml>
- [10] 高惠璇,等. SAS系统 SAS/STAT软件使用手册[M]. 北京:中国统计出版社, 1997

(上接第114页)

3 结论

本文通过改变标的资产的价格行为模式的假设,即假定标的资产的价格过程为混合过程,由市场的完备性,利用鞅方法讨论了无风险利率 $r(t)$ 和资产价格波动率 $\sigma(t)$ 均为时间 t 的函数时欧式未定权益的定价公式,在特定条件下对 Black-Scholes公式作了有限的推广.全文依然假设利率是时间 t 的函数,而在实际金融市场中往往并非如此,因此可进一步探讨随机利率和随机波动率混合模型的未定权益定价问题。

参考文献:

- [1] BLACK F, SCHOLES M. The pricing of options and corporate liabilities[J]. Journal of Political Economy, 1973, 81: 133~155
- [2] MERTON R C. Theory of rational option pricing[J]. Bell J Econ and Management Science, 1973, (4): 141~183
- [3] AASE K K. Contingent claims valuation when the security price is a combination of an Ito process and a random point process[J]. Stochastic Processes and their Applications, 1988, (28): 185~220
- [4] JEAN BLANC P M, PONTIER M. Optimal portfolio for a small investor in a market model with discontinuous prices[J]. Applied Mathematics and Optimization, 1990, (22): 287~310
- [5] DRIFSCHEL M, PROTTER P. Complete markets with discontinuous security price[J]. Finance and Stochastics, 1999, 3: 203~214
- [6] HARRISON JM, KTEPS D M. Martingales and arbitrage in multiperiod securities markets[J]. J Economic Theory, 1979, (20): 381~408
- [7] HARRISON JM, PLSKA S R. Martingales and stochastic integrals in the theory of continuous trading[J]. Stochastic Processes and their Applications, 1981, (11): 215~260
- [8] DUFFIE D. Dynamic Asset Pricing Theory[M]. Ed. New Jersey: Princeton Univ. Press, 1996