

需水量预测的线性测量误差模型

贾仁甫^{1 2} 陈守伦¹ 丁 鑫²

(1.河海大学水利水电工程学院 ,江苏 南京 210098 ;2.扬州大学建筑科学与工程学院 ,江苏 扬州 225009)

摘要 :为了克服现有需水量预测方法未考虑原始数据测量误差的缺陷 ,以扬州市 1999 ~ 2006 年的 GDP 及用水量作为原始数据 ,建立了需水量预测的线性测量误差模型 ,并对扬州市 2010 年、2015 年和 2020 年的需水量进行了预测 .与普通线性回归方法的对比结果表明 ,线性测量误差模型预测精度较高 ,其预测结果可以为节约用水规划及水资源规划提供依据 .

关键词 :需水量预测 ;线性模型 ;测量误差 ;校正似然

中图分类号 :TU991.31 文献标识码 :A 文章编号 :1000-1980(2008)04-0488-03

目前 ,常用的需水量预测方法有 3 种 :时间序列法^[1]、结构分析法^[2-3]和系统法^[4-7] .由于这些方法未考虑原始数据的测量误差 ,预测精度会受到一定的影响 .

测量误差模型是一种可以考虑原始数据测量误差的模型 .该模型虽在许多领域得到应用^[8-19] ,但在需水量预测方面的应用还未见报道 .本文以扬州市 1999 ~ 2006 年的 GDP 及用水量作为原始数据 ,建立了需水量预测的线性测量误差模型 ,并对扬州市 2010 年、2015 年和 2020 年的需水量进行了预测 .

1 模型的建立与分析

设 Y 为用水量 , X 为 GDP , X 的测量有一定的误差 ,其期望为 Z 且未知 ,则可得到如下线性测量误差模型 :

$$\begin{cases} y_i = f(x_i, \beta) + \varepsilon_i & \varepsilon_i \sim N(0, \sigma^2) \\ X = Z + \xi & \xi \sim N(0, I_n \otimes \Lambda) \end{cases} \tag{1}$$

式中 : f ——已知函数 ; y_i ——观测响应变量 ; Z —— $n \times p$ 阶第 i 行为 z_i^T 的矩阵 ; I_n —— n 阶单位矩阵 ; β —— p 维未知参数 ; X —— $n \times p$ 阶第 i 行为 x_i^T 的矩阵 ; σ^2 ——未知参数 ; Λ ——已知非负定矩阵 ; ε_i ——测量值的随机误差 ; ξ —— $n \times p$ 阶矩阵 .

因为 X 和 Z 是 2 种不同性质的协变量 , X 是测量值 ,而 Z 未知 ,所以可以用 Nakamura 的方法 ,即校正似然法来处理此类测量误差问题 .

设 $U(\beta, Z, Y)$ 、 $U(\beta, Z, Y)$ 、 $J(\beta, Z, Y)$ 和 $K(\beta, Z, Y)$ 分别为模型的似然函数、得分函数、观测信息及 Fisher 信息 ,且记 E^+ 为 Y 的数学期望 ,则

$$E^+ U(\beta, Z, Y) = 0 \qquad E^+ J(\beta, Z, Y) = K(\beta, Z, Y) \tag{2}$$

校正似然函数 $l^*(\beta, X, Y)$ 满足条件

$$E^* l^*(\beta, X, Y) = U(\beta, Z, Y) \tag{3}$$

其中 E^* 表示 Y 给定时的 X 的条件期望 .设 $U^*(\beta, X, Y) = \partial l^*(\beta, X, Y) / \partial \beta$ 、 $J^*(\beta, X, Y) = -\partial U^*(\beta, X, Y) / \partial \beta$ 、 $J^*(\beta, X, Y)$ 分别表示校正得分函数、观测信息和 Fisher 信息 .如果 E^* 与 $\partial \beta$ 可以交换 ,则

$$E^* U^*(\beta, X, Y) = U(\beta, Z, Y), E^* J^*(\beta, X, Y) = K(\beta, Z, Y) \tag{4}$$

参数 β 的估计满足条件 $U^*(\hat{\beta}, X, Y) = 0$ 称 $\hat{\beta}$ 为校正最大似然估计 (CMLE) .记 $E = E^+ E^*$,显然有

$$EU^*(\beta, X, Y) = E^+ E^* U^*(\beta, X, Y) = E^+ U(\beta, Z, Y) = 0 \tag{5}$$

收稿日期 :2008-02-28

基金项目 :国家自然科学基金(50609023) 扬州大学科技创新培育基金(2007CXJ007)

作者简介 :贾仁甫(1965 —) 男 ,江苏泰州人 ,副教授 ,博士研究生 ,主要从事水利水电系统分析与技术经济研究 .

这表明校正得分函数是无偏的.对于上述正态测量误差模型(1),Nakamura 证明了参数估计的渐近正态性和相合性,并且有

$$LD(\boldsymbol{\beta})=2\{l^*(\hat{\boldsymbol{\beta}},\boldsymbol{X},\boldsymbol{Y})-l^*(\boldsymbol{\beta},\boldsymbol{X},\boldsymbol{Y})\}\rightarrow\chi^2(p)$$

(6)

将上述结果应用于模型(1)则似然函数为

$$\ell(\boldsymbol{\beta},\boldsymbol{Z},\boldsymbol{Y})=-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-f(z_i,\boldsymbol{\beta}))^2$$

(7)

校正似然函数可表示为

$$l^*(\boldsymbol{\beta},\boldsymbol{X},\boldsymbol{Y})=-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n(y_i-f(x_i,\boldsymbol{\beta}))^2-\boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta}\right]$$

(8)

通过计算可知 $E^*l^*(\boldsymbol{\beta},\boldsymbol{X},\boldsymbol{Y})=\ell(\boldsymbol{\beta},\boldsymbol{Z},\boldsymbol{Y})$.由于校正似然函数性质良好,根据它可很容易求出 $\boldsymbol{\beta}$ 的 CMLE.

特别地,如果式(1)中的 $f(x_i,\boldsymbol{\beta})=x_i^T\boldsymbol{\beta}$,那么将式(8)对 $\boldsymbol{\beta}$ 求导,得

$$U^*(\boldsymbol{\beta},\boldsymbol{X},\boldsymbol{Y})=\frac{\partial l^*}{\partial \boldsymbol{\beta}}=\sigma^{-2}(X^T\boldsymbol{Y}-X^T\boldsymbol{X}\boldsymbol{\beta}+n\boldsymbol{\Lambda}\boldsymbol{\beta})$$

(9)

则 $\boldsymbol{\beta}$ 的参数估计 $\hat{\boldsymbol{\beta}}$ 可由方程 $U^*(\boldsymbol{\beta},\boldsymbol{X},\boldsymbol{Y})=0$ 求得:

$$\hat{\boldsymbol{\beta}}=(X^T\boldsymbol{X}-n\boldsymbol{\Lambda})^{-1}X^T\boldsymbol{Y}$$

(10)

则 $\boldsymbol{Y}$ 的拟合值为 $\hat{\boldsymbol{Y}}=\boldsymbol{H}\boldsymbol{Y}$.其中

$$\boldsymbol{H}=X(X^T\boldsymbol{X}-n\boldsymbol{\Lambda})^{-1}X^T$$

(11)

其元素为 h_{ij} , $\hat{\boldsymbol{e}}=\boldsymbol{Y}-\hat{\boldsymbol{Y}}$ 称为残差向量. σ^2 的估计为 $\hat{\sigma}^2=\|\hat{\boldsymbol{e}}\|^2/(n-p)$.由式(8)可知,校正 Fisher 信息阵为

$$I^*(\boldsymbol{\beta},\boldsymbol{X},\boldsymbol{Y})=\sigma^{-2}(X^T\boldsymbol{X}-n\boldsymbol{\Lambda})$$

(12)

从上述讨论可知,校正似然法也适用于一些广义线性模型以及非线性模型.

2 应用实例

以扬州市 1999~2006 年的 GDP 和总用水量(表 1)作为原始数据,建立需水量预测的线性测量误差模型.

根据模型式(1),用水量 $y_i=\beta_0+\beta_1x_i+\varepsilon_i(x_i\sim N(z_i,\lambda))$.由上述讨论可知,参数可通过极大似然估计求得.通过计算,得到 β_0,β_1,σ^2 和 λ 的参数估计为 $\hat{\beta}_0=38.0506,\hat{\beta}_1=0.0017,\hat{\sigma}^2=4.6959,\hat{\lambda}=0.0004$.由此可得扬州市需水量预测的线性测量误差模型,即

$$y_i=38.0506+0.0017x_i+\varepsilon_i$$

(13)

式中: y_i ——扬州市需水量,亿 m^3 ; x_i ——扬州市的 GDP,亿元.

从线性测量误差模型与普通线性回归模型的拟合结果(表 2)可得,普通线性回归模型的 $\hat{\sigma}^2=6.2609$,而线性测量误差模型的 $\hat{\sigma}^2=4.6959$.可见,本文所建立的线性测量误差模型的精度比普通线性回归模型高.

表 1 扬州市 1999~2006 年用水量及 GDP
Table 1 Water consumption of Yangzhou City from 1999 to 2006

年份	GDP/ 亿元	用水量/ 亿 m^3
1999	427	36.73
2000	472	36.86
2001	505	42.57
2002	544	41.58
2003	632	39.28
2004	760	40.35
2005	922	39.41
2006	1 100	39.66

表 2 2 种模型拟合结果的对比
Table 2 Comparison of fitting results between the presented model and the ordinary linear regression method

年份	普通线性回归模型				线性测量误差模型			
	真实值/ 亿 m^3	拟合值/ 亿 m^3	绝对误 差/亿 m^3	相对误 差/%	真实值/ 亿 m^3	拟合值/ 亿 m^3	绝对误 差/亿 m^3	相对误 差/%
1999	36.73	38.77	2.04	5.55	36.73	38.84	2.11	5.74
2000	36.86	38.85	1.99	5.40	36.86	38.90	2.04	5.53
2001	42.57	38.90	-3.67	-8.62	42.57	38.95	-3.62	-8.50
2002	41.58	38.97	-2.61	-6.28	41.58	39.01	-2.57	-6.18
2003	39.28	39.12	-0.16	-0.41	39.28	39.14	-0.14	-0.36
2004	40.35	39.33	-1.02	-2.53	40.35	39.33	-1.02	-2.53
2005	39.41	39.60	0.19	0.48	39.41	39.57	0.16	0.41
2006	39.66	39.90	0.24	0.61	39.66	39.83	0.17	0.43

根据扬州市社会经济发展规划,扬州市 2010 年、2015 年和 2020 年的 GDP 将分别达到 1 924 亿元、3 545 亿元和 5 709 亿元.运用式(13)对 2010 年、2015 年及 2020 年扬州市需水量进行预测,同时根据“三 σ 准则”确定预测区间,结果见表 3.

3 结 语

为了克服现有需水量预测方法未考虑原始数据测量误差的缺陷,本文以扬州市 1999~2006 年的 GDP 及用水量作为原始数据,建立了需水量预测的线性测量误差模型,并对扬州市 2010 年、2015 年和 2020 年的需水量进行了预测.拟合对比结果表明,线性测量误差模型的预测精度比普通线性回归模型高.由于本文所采用的原始数据量偏少,线性测量误差模型的作用未得到明显的发挥.但是,如果原始数据量足够的大,其作用就比较明显了.

参考文献:

[1] WANG Q ,HELLER M. Hybrid box-jenkins and neural network forecasting of potable water demand[C]//Proceedings of the 1996 Artificial Neural Networks in Engineering ANNIE 96. Missouri : ASME ,1996 801-807.

[2] 张雅君,刘全胜,冯萃敏.多元线性回归分析在北京城市生活需水量预测中的应用[J].给水排水,2003,29(4):26-29.

[3] 顾月红,葛朝霞,薛梅,等.北京市生活用水年预报模型[J].河海大学学报:自然科学版,2008,36(1):19-22.

[4] 周刚,王弘宇,胡雪春,等.应用灰色新陈代谢 GM(1,1)模型预测中长期城市需水量[J].中国农村水利水电,2005(8):16-18.

[5] 楼玉,张清,刘国华,等.城市用水量预测中的多变量灰色预测模型[J].水资源保护,2005,21(1):11-13.

[6] BECCALI M ,CELLURA M ,LO BRANO V , et al. Forecasting daily urban electric load profiles using artificial neural networks[J]. Energy Conversion and Management 2004 45(18):2879-2900.

[7] 刘俊良,臧景红,何延青.系统动力学模型用于城市需水量预测[J].中国给水排水,2005,21(6):31-34.

[8] 陈希孺,王松桂.近代回归分析[M].合肥:安徽教育出版社,1987.

[9] FULLER W A. Measurement error models[M]. New York : Wiley ,1987.

[10] CARROLL R J ,RUPPERT D ,STEFANSKI L A. Measurement error in nonlinear model[M]. London :Chapman and Hall ,1995.

[11] HANFELT J J ,LIANG K Y. Approximate likelihood for generalized linear errors-in-variables model[J]. J R Statist Soc B ,1997 ,59 : 627-637.

[12] NAKAMURA T. Corrected score function for errors-in-variables models : methodology and application to generalized linear models[J]. Biometrika ,1990 ,77 : 127-137.

[13] STEFANSKI L A ,CARROLL R J. Conditional scores and optimal scores for generalized linear measurement error models[J]. Biometrika ,1987 ,74 :703-716.

[14] KELLY G E. The influence function in the errors in variables problems[J]. Ann Statist ,1984 ,12 :87-100.

[15] WELLMAN J M ,GUNST R F. Influence diagnostics for linear measurement errors model[J]. Biometrika ,1991 ,78 :373-380.

[16] 宗序平.线性测量误差模型及其诊断[J].系统工程学报,1999,14(2):135-139.

[17] ZONG Xu-ping , FUNG W K. WEI bo-cheng ,et al. Influence analysis for generalized linear measurement error models[J]. Ann Inst Stat Math 2000 52(2):367-379.

[18] ZHAO Y ,Lee A H. Assessment of influence in non-linear measurement error model[J]. J Appl Statist ,1995 22 :215-225.

[19] ZONG Xu-ping , MENG Guo-ming , WANG Hai-bin , et al. Influence analysis on nonlinear measurement error model[J].应用概率统计,2003,19(2):31-39.

Linear measurement error model for water demand forecasting

JIA Ren-fu^{1,2},CHEN Shou-lun¹,DING Xin²

(1. College of Water Conservancy and Hydropower Engineering , Hohai University , Nanjing 210098 , China ;
2. College of Architectural Science and Engineering , Yangzhou University , Yangzhou 225009 , China)

Abstract : The general water demand forecasting methods don 't consider raw data measurement errors. In this paper , a linear measurement error model is proposed to improve the general methods , and the model is verified with the GDP and water consumption data from Yangzhou City from 1999 to 2006. Then , the model is applied to forecast the water demand of Yangzhou City in the years 2010 , 2015 and 2020. The results show that , compared with the ordinary linear regression method , the presented model has higher precision , and the forecast results can provide a basis for water conservation and water resources planning.

Key words : forecasting of water demand ; linear model ; measurement error ; corrected likelihood

表3 总需水量预测结果

Table 3 Forecasted water demand

年份	GDP/ 亿元	需水量/亿 m ³	
		预测值	预测区间
2010	1924	41.32	[34.82 47.82]
2015	3545	44.08	[37.58 50.58]
2020	5709	47.76	[41.25 54.26]