

基于虚拟变量控制的GM(1,N)模型构建及其应用

丁松^{1†}, 党耀国¹, 徐宁²

(1. 南京航空航天大学 经济与管理学院, 南京 211106; 2. 南京审计大学 管理科学与工程学院, 南京 211815)

摘要: 现实中系统行为特征序列常受到虚拟变量的影响, 而此时传统 GM(1,N) 模型不能准确地描述系统特征的变化规律. 将虚拟变量引入传统 GM(1,N) 模型的灰作用量, 构建虚拟变量控制的 GM(1,N) 模型, 讨论新模型的参数求解方法; 鉴于背景值对模型精度有着重要影响, 利用粒子群优化算法对含有插值系数的背景值进行优化求解; 从两个角度提出虚拟变量有效性检验方法. 最后, 通过河南省农民人均收入预测案例表明, 新模型能够准确描述虚拟变量影响下系统特征序列的未来变化趋势.

关键词: 灰色系统; GM(1,N); 虚拟变量; 预测

中图分类号: N941.5

文献标志码: A

Construction and application of GM(1,N) based on control of dummy variables

DING Song^{1†}, DANG Yao-guo¹, XU Ning²

(1. College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; 2. College of Management Science and Engineering, Nanjing Audit University, Nanjing 211815, China)

Abstract: To solve the problems that the traditional GM(1,N) model can not describe precisely the changing patterns of the system behavior characteristic variables influenced by the dummy variables in the real life, the dummy variables are introduced into the grey acting term of the conventional GM(1,N) model, and the novel GM(1,N) model is built with discussing the estimating methods of the parameters. Due to the great effect of the background value on the precision of modeling, the interpolated coefficient of the background value is optimized by using the PSO algorithm. The test methods of validity are proposed from two perspectives. Finally, the per capital income of farmers in Henan province is simulated and predicted, which shows that the proposed model can effectively describes the future trend of system change influenced by dummy variables.

Keywords: grey system; GM(1,N); dummy variables; forecasting

0 引言

作为灰色预测理论的重要组成部分, 灰色多变量预测模型自提出以来, 受到了众多学者的关注^[1]. GM(1,N) 模型本质上是因子分析模型, 能够对含有多因子的系统进行全面、动态地分析, 鉴于其能够相对准确地描述系统与因素之间的关系, 许多学者对优化模型参数、改善模型精度、拓展模型的应用范围付出了努力. 如文献[2]在文献[1]的基础上, 提出了将驱动项视为灰常量, 解决了原始 GM(1,N) 模型的求解预测问题; 文献[3]通过背景值优化和残差修正构

建了具有全适性的 GM(1,N) 模型, 拓宽了模型的应用范围; 文献[4]在传统模型的基础上, 引入控制参数, 利用非线性优化方法对参数进行优化, 提升了模型的适应性; 文献[5]将驱动因素的变化趋势融入到系统行为特征序列建模过程中, 提出了 TGM(1,N) 模型, 实现了对系统趋势的直接预测; 文献[6]针对输入和输出存在时滞特征的系统, 提出了 GM(1,N| τ , r) 模型, 并成功应用到武汉科技投入与产出的预测当中; 文献[7]引入分数阶累加算子, 将时滞模型扩展为分数阶累加 GM(1,N, τ) 模型; 文献[8]针对传统模型未

收稿日期: 2016-12-20; 修回日期: 2017-03-07.

基金项目: 国家自然科学基金项目(71371098, 71701024); 中央高校基本科研业务费专项资金项目(2017301); 江苏省普通高校研究生科研创新计划项目(KYZZ16_0153); 南京航空航天大学博士学位论文创新与创优基金项目(BCXJ16-09); 江苏省高校自然科学研究项目(16KJD120001); 江苏省社科基金重点研究项目(16GLA001).

作者简介: 丁松(1992—), 男, 博士生, 从事灰色系统理论及其应用的研究; 党耀国(1964—), 男, 教授, 博士生导师, 从事灰色系统理论、数量经济、决策支持系统、产业经济等研究.

†通讯作者. E-mail: dingsong1129@163.com

考虑驱动因素变化趋势的问题,提出了新型多变量预测模型,并对其时间响应函数进行了优化,较准确地预测了我国未来CO₂的排放趋势;文献[9]通过引入时滞控制因子构建了多变量离散模型,并结合灰色扩维识别方法探索时滞参数的测算,在农村水环境与区域发展滞后效应测算中取得了较好的应用效果;文献[10]通过将灰色控制参数引入到传统模型的构建,并利用卷积积分技术求解出模型的精确解,在此基础上提出了一系列改进模型^[11-12],增强了新模型对多种特征序列的建模应用效果;文献[13]将单方程的GM(1, N)模型拓展到多方程联立的系统动态方程组,提出了MGM(1, N)模型,并将其拓展到非等间距序列建模,进一步拓展了GM(1, N)模型的应用领域.通过考察现有的GM(1, N)模型及其拓展模型的结构形式可以发现,相关变量的一阶累加生成序列以线性组合形式被引入灰作用量,这类相关变量均具有定量的含义,实例中变量的大小都传递了有用信息,而在灰色模型中缺少考虑定性指标变量对系统特征序列的影响.在实际建模中,一些定性变量具有不可忽视的重要影响,如工资水平预测受到文化程度的影响,农业产量和农民收入受到农业政策的影响等,这些影响因素的共同特征是定性描述的.

事实上,经济社会系统中广泛存在一类定性指标变量,如性别、学历高低、政策有无等,这类变量能够反映不同属性类型的驱动因素对主系统的行为特征变量的影响,对于主系统的未来趋势发展作用不容忽视.通常这类变量均是以二值信息出现,例如有无、男女、高低等,这类变量被称为虚拟变量,用0-1变量来刻画.因此,在构建多变量预测模型时,不仅需要考虑定量指标变量的单方面影响,而且也应该考虑虚拟变量对系统发展趋势的影响.本文拟将虚拟变量作用项引入传统GM(1, N)模型,建立基于虚拟变量控制的GM(1, N)模型,从而展示虚拟变量对系统特征变量的影响,讨论虚拟变量控制GM(1, N)模型的求解和优化方法,并通过实际案例验证模型的有效性和适用性.

1 基于虚拟变量控制的GM(1, N)模型构建及其求解

1.1 GM(1, N)模型定义缺陷分析

定义1 设原始序列 $X_i^{(0)}$,其一次累加生成序列为 $X_i^{(1)}$, $X_1^{(0)}$ 为系统行为序列, $X_i^{(0)}$ ($i = 2, 3, \dots, N$)为驱动因素序列,称

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(1)}(k) \quad (1)$$

为GM(1, N)模型.其中: a 为系统发展系数, $b_i x_i^{(1)}(k)$ 为驱动项, b_i 为驱动系数.

定义2 设 $\hat{a} = [a, b_2, \dots, b_N]$ 为模型的参数列,则称

$$\frac{dx_1^{(1)}}{dt} + ax_1^{(1)} = \sum_{i=2}^N b_i x_i^{(1)} \quad (2)$$

为GM(1, N)模型的白化方程或影子方程.

GM(1, N)模型能够依靠自身动态特性反映相关因素序列对系统行为变量的同步作用效果,在已知影响因子变量变化趋势信息的条件下,能够对系统行为变量进行预测.但是,容易发现经典模型在参数列求出后,再进行模拟和预测是十分困难的,至今未取得统一可信的解决方法.目前,多采用文献[2]方法,假设驱动变量变化幅度较小,将 $\sum_{i=2}^N b_i x_i^{(1)}$ 作为一个灰常量来处理,采用类似于GM(1, 1)模型解决方法进行模拟和预测.这一求解方法在多种实例中得到了验证,取得了较好的建模效果,并推广到众多应用领域.然而,经典模型考虑的驱动因素变量是可量化的变量,而忽略了政策等虚拟变量对于主系统的影响.现实生活中,确实存在一类实例,在可量化的变量对主系统发挥作用的同时,虚拟变量也存在较大的促进作用,不分析和控制这类驱动变量的作用必然会导致较大的建模误差.增加虚拟变量可以较真实地反映多种不同属性的驱动因素对于系统的影响,同时也扩大了样本容量,增加了误差的自由度,有利于降低误差,进一步提升模型的精度.因此,本文在传统GM(1, N)的基础上,构建基于虚拟变量控制的新型多变量预测模型.

1.2 DVCGM(1, N)模型定义及其参数估计

定义3 设原始序列 $X_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n))$,虚拟变量序列 $D_j^{(0)} = (d_j^{(0)}(1), d_j^{(0)}(2), \dots, d_j^{(0)}(n))$, $d_j^{(0)}(k) = 0$ or 1 ,其一次累加生成序列(1-AGO)分别为 $X_i^{(1)}$ 和 $D_j^{(1)}$, $X_1^{(0)}$ 为系统行为序列, $X_i^{(1)}$ ($i = 2, 3, \dots, M$)和 $D_j^{(1)}$ ($j = M + 1, M + 2, \dots, N$)为驱动因素序列,称下式为基于虚拟变量控制的GM(1, N)模型,记为DVCGM(1, N):

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^M b_i x_i^{(1)}(k) + \sum_{j=M+1}^N b_j d_j^{(1)}(k). \quad (3)$$

其中: $\sum_{i=2}^M b_i x_i^{(1)}(k)$ 为量化变量驱动项,背景值 $z_1^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k-1)$, $\sum_{j=M+1}^N b_j d_j^{(1)}(k)$ 为虚拟变量驱动项.如果不考虑虚拟变量或 $b_j = 0$ ($j = M$

+ 1, M + 2, \dots, N), 则该模型退化为传统GM(1, N)模型. 本文模型是对传统模型的传承, 是在虚拟变量控制下对传统多变量模型的进一步拓展和丰富. 与单变量灰色预测模型相比, 多变量模型在参数的矩阵运算过程中较容易出现矩阵漂移现象, 进而影响模型精度, 因此, 在建模之前, 对于数量级较大的原始数据最好做预处理, 可选初值化或均值化等数据变化方法.

定理1 若 $X_i^{(0)}, X_i^{(1)}, D_j^{(0)}(k), D_j^{(1)}(k)$ 如定义3所示, $\hat{b} = (a, b_2, \dots, b_N)^T$ 为参数列, $Y = [x_1^{(0)}(2), x_1^{(0)}(3), \dots, x_1^{(0)}(n)]^T$, 且有

$$B = \begin{bmatrix} -z_1^{(1)}(2)x_2^{(1)}(2) & \dots & x_M^{(1)}(2) & d_{M+1}^{(1)}(2) & \dots & d_N^{(1)}(2) \\ -z_1^{(1)}(3)x_2^{(1)}(3) & \dots & x_M^{(1)}(3) & d_{M+1}^{(1)}(3) & \dots & d_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(n)x_2^{(1)}(n) & \dots & x_M^{(1)}(n) & d_{M+1}^{(1)}(n) & \dots & d_N^{(1)}(n) \end{bmatrix},$$

则参数列的最小二乘法估计满足: 1) 当 $n = N + 1$ 时, 有 $\hat{b} = B^{-1}Y, |B| \neq 0$; 2) 当 $n > N + 1$ 时, 有 $\hat{b} = (B^T B)^{-1} B^T Y, |B^T B| \neq 0$; 3) 当 $n < N + 1$ 时, 有 $\hat{b} = B^T (B^T B)^{-1} Y, |B^T B| \neq 0$.

证明 将 $k = 2, 3, \dots, n$ 代入DVCGM(1, N)模型(3), 可得

$$\begin{aligned} x_1^{(1)}(2) &= -ax_1^{(1)}(2) + \sum_{i=2}^M b_i x_i^{(1)}(2) + \sum_{j=M+1}^N b_j d_j^{(1)}(2), \\ x_1^{(1)}(3) &= -ax_1^{(1)}(3) + \sum_{i=2}^M b_i x_i^{(1)}(3) + \sum_{j=M+1}^N b_j d_j^{(1)}(3), \\ &\vdots \\ x_1^{(1)}(n) &= -ax_1^{(1)}(n) + \sum_{i=2}^M b_i x_i^{(1)}(n) + \sum_{j=M+1}^N b_j d_j^{(1)}(n), \end{aligned}$$

即由最小二乘法可得 $Y = B\hat{b}$.

1) 当 $n = N + 1$ 且 $|B| \neq 0$ 时, B 存在逆矩阵, 方程组有唯一解, 即可得 $\hat{b} = B^{-1}Y$;

2) 当 $n > N + 1$ 且 B 为列满秩时, 有 B 的满秩分解为 $B = DC$, 进而可以得到 B 的广义逆矩阵 B^+ 为

$$\begin{aligned} B^+ &= C^T(CC^T)^{-1}(D^T D)^{-1}D^T, \\ \hat{b} &= C^T(CC^T)^{-1}(D^T D)^{-1}D^T Y. \end{aligned}$$

由于 B 为满秩矩阵, C 可以取为单位矩阵, $B = DI_N, B = D$, 即

$$\begin{aligned} \hat{b} &= C^T(CC^T)^{-1}(D^T D)^{-1}D^T Y = \\ &(D^T D)^{-1}D^T Y = (B^T B)^{-1}B^T Y. \end{aligned}$$

3) 当 $n < N + 1$ 且 B 为行满秩矩阵时, D 可取为单位矩阵, $B = I_{n-1}C, B = C$, 则

$$\hat{b} = C^T(CC^T)^{-1}(D^T D)^{-1}D^T Y =$$

$$C^T(CC^T)^{-1}Y = B^T(BB^T)^{-1}Y. \quad \square$$

定义4 设 $\hat{b} = (a, b_2, \dots, b_N)$, 则称

$$\frac{dx_1^{(1)}(t)}{dt} + ax_1^{(1)}(t) = \sum_{i=2}^M b_i x_i^{(1)}(t) + \sum_{j=M+1}^N b_j d_j^{(1)}(t) \quad (4)$$

为DVCGM(1, N)模型的白化方程.

定理2 设 $X_i^{(0)}, X_i^{(1)}, D_j^{(0)}(k), D_j^{(1)}(k), B, Y$ 如定义1和定义3所述, 则有:

1) DVCGM(1, N)模型的白化微分方程的解为

$$\begin{aligned} x_1^{(1)}(t) &= e^{-at} \left[x_1^{(1)}(0) - t \left(\sum_{i=2}^M b_i x_i^{(1)}(0) + \sum_{j=M+1}^N b_j d_j^{(1)}(0) \right) + \right. \\ &\left. \sum_{i=2}^M \int \left(\sum_{i=2}^M b_i x_i^{(1)}(t) + \sum_{j=M+1}^N b_j d_j^{(1)}(t) \right) e^{at} dt \right]. \quad (5) \end{aligned}$$

2) 当驱动因素序列 $X_i^{(1)}(i = 2, 3, \dots, N)$ 变化幅度较小时, 模型驱动项 $\sum_{i=2}^M b_i x_i^{(1)}(t)$ 和 $\sum_{j=M+1}^N b_j d_j^{(1)}(t)$ 可以被视为灰常量, 则DVCGM(1, N)模型灰色微分方程的近似时间相应函数数列为

$$\begin{aligned} \hat{x}_1^{(1)}(k) &= \frac{1}{a} \left[\sum_{i=2}^M b_i x_i^{(1)}(k) + \sum_{j=M+1}^N b_j d_j^{(1)}(k) \right] - \\ &\frac{e^{-a(k-1)}}{a} \left[\sum_{i=2}^M b_i x_i^{(1)}(k) + \sum_{j=M+1}^N b_j d_j^{(1)}(k) \right] + \\ &e^{-a(k-1)} x_1^{(1)}(1). \quad (6) \end{aligned}$$

3) 一阶累减还原式为

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k). \quad (7)$$

4) DVCGM(1, N)差分模拟式为

$$\hat{x}_1^{(0)}(k) = -az_1^{(1)}(k) + \sum_{i=2}^M b_i \hat{x}_i^{(1)}(k) + \sum_{j=M+1}^N b_j \hat{d}_j^{(1)}(k). \quad (8)$$

证明 1) 由白化方程(4)可得通解公式为

$$\begin{aligned} x_1^{(1)}(t) &= e^{-at} \sum_{i=2}^M \int \left[\sum_{i=2}^M b_i x_i^{(1)}(t) + \right. \\ &\left. \sum_{j=M+1}^N b_j d_j^{(1)}(t) \right] e^{at} dt + ce^{-at}, \quad (9) \end{aligned}$$

其中 c 为待定常数. 将初始条件 $x_i^{(1)}(0)$ 代入式(4)可得

$$c = - \sum_{i=2}^N \int \left[\sum_{i=2}^M b_i x_i^{(1)}(t) + \sum_{j=M+1}^N b_j d_j^{(1)}(t) \right] e^{at} dt + x_1^{(1)}(0). \quad (10)$$

2)~4)将驱动项看作灰常量,与GM(1,1)模型的证明过程类似,由此得证. \square

定理1和定理2分别给出了DVCGM(1, N)模型的参数估计和模拟值计算公式,当主系统行为和驱动因素作用机制已知时,便可通过上述定理进行计算.

1.3 DVCGM(1, N)模型的参数优化

对于灰色预测模型,背景值是影响模型建模精度的关键因素之一.目前,优化方法主要分为3种:第1种是基于纯指数率的背景值改进^[3];第2种是基于背景值的几何意义,利用求和函数近似求解^[4];第3种是引入插值系数^[15],结合优化算法进行求解.前两种方法对于实际数据的适应能力有限,而插值系数能够根据目标函数进行不断优化和调整,故适应性较强.本文拟通过插值系数最大限度优化DVCGM(1, N)背景值,其表达式为 $z_1^{(1)}(k) = \lambda x_1^{(1)}(k) + (1 - \lambda)x_1^{(1)}(k - 1)$, $\lambda \in [0, 1]$,因此,上述模型的背景值优化模型为:以式(6)作为时间响应式,在还原数据平均相对误差最小的目标下,求解优化的背景值插值系数

$$\begin{aligned} \text{Min}_{\lambda} \text{avg}(e(k)) &= \frac{1}{n-1} \sum_{k=2}^n \left| \frac{\hat{x}_1^{(0)}(k) - x_1^{(0)}(k)}{x_1^{(0)}(k)} \right|. \\ \text{s.t.} \quad &\begin{cases} \hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(0)}(k); \\ \hat{b} = [a, b_2, \dots, b_N]^T, k = 2, 3, \dots, n; \\ \lambda \in [0, 1], i = 2, 3, \dots, N. \end{cases} \end{aligned} \quad (11)$$

鉴于本文数据计算的复杂性以及粒子群算法具有计算简单、收敛速度快,及较强的适应性等优点,拟采用Matlab中粒子群算法工具箱,对本文背景值插值系数进行优化.粒子群(PSO)算法用于求解 λ 的最优取值.就算法原理而言,粒子群算法建立在多个个体构成的群落基础上,根据一定随机方式寻找适应度最好的位置,进而获得最优的参数估计.算法本身对最优值搜索速度较快,且在可行域内不易落入局部最优值的情况.就算法构建角度而言,令个体数量设置为正整数 M , M 根据问题的计算规模设置足够大,同时设置算法循环的终止条件,即 I 为最大迭代次数,同时以模型计算精度为基础计算当前最优位置的适应度函数,通过分析迭代计算的最优适应度取值及其对应位置来确定最终最优取值.

令 $Q(i, j) = \lambda(i, j)$,即第 i 次迭代中粒子 j 对应的位置.其中: $0 \leq i \leq I, 0 \leq j \leq M$.首先初始化粒子的位置,即 $i = 0$ 时 $Q(0) = \lambda(0, j) (j = 1, 2, \dots, M)$ 为可行域内的随机初始化值.基于拟合误差的累积绝对误差最小设置适应度函数

$$\text{fitness}[\lambda(i, j)] = \frac{1}{n-1} \sum_{k=2}^n \left| \frac{\hat{x}_1^{(0)}(k) - x_1^{(0)}(k)}{x_1^{(0)}(k)} \right|. \quad (12)$$

单个粒子依靠自身经验搜索到的最优解称为个体极值,第 i 次迭代极值点位置记为 $\text{pbest}(i)$;所有粒子搜索到的最优解称为全局极值,全局极值点记为 gbest .粒子更新速度公式设置为

$$\begin{aligned} v(i+1, j) &= \\ wv(i, j) + c_1[\text{pbest}(i) - Q(i, j)] + c_2[\text{gbest} - Q(i, j)]. \end{aligned} \quad (13)$$

其中: $v(i, j)$ 为粒子 $Q(i, j)$ 的速度向量; w 为惯性权重,惯性权重在算法中用于保持全局搜索和局部搜索的平衡; c_1, c_2 为粒子加速系数,取值为随机数,以此保障算法能够在全局搜索最优位置.利用速度公式计算粒子新的位置,即

$$Q(i+1, j) = Q(i, j) + v(i+1, j). \quad (14)$$

其中: $v(i+1, j)$ 为更新后的粒子速度, $Q(i+1, j)$ 为更新后的粒子位置向量.当迭代次数达到最大值 M 时停止迭代,分析变量 gbest 取值,并获取待定系数 $\hat{\lambda}$ 的最优取值.将 $\hat{\lambda}$ 最优值代入定理1,求出待估参数 \hat{a}, \hat{b} ,进一步利用定理2求解模型的最优还原响应式并进行模拟和预测.

1.4 DVCGM(1, N)模型中虚拟变量的有效性检验

对于虚拟变量的有效性识别,除了虚拟变量本身要具有一定的物理意义和实际意义以外,还将从下面两个角度进行考量.

1)改善建模精度检验.考察在引入虚拟变量前后是否对提升建模精度有帮助.若引入虚拟变量建模后,模型精度显著提高,则说明虚拟变量引入是正确的;反之,则虚拟变量引入无意义.本文采用相对误差和平均相对误差两个指标作为衡量模型精度的手段.将 k 时刻的相对误差记为 $\text{RPE}(k)$,其公式为

$$\text{RPE}(k) = \left| \frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\%; \quad (15)$$

各点的平均相对误差记为 $\text{ARPE}(k)$,其公式为

$$\text{ARPE}(k) = \frac{1}{n-1} \sum_{k=2}^n \text{RPE}(k). \quad (16)$$

i) 如果 $k \leq n$,则称 $\text{RPE}(k)$ 为 k 点模拟相对误差,

ARPE为平均模拟相对误差;若 $k > n$,则称RPE(k)为k点预测相对误差,称ARPE为平均预测相对误差.

ii) 给定 α ,当 $ARPE < \alpha$ 且 $RPE(n) < \alpha$ 时,称模型为残差合格模型. 为凸显虚拟变量对系统行为的影响程度,本文参考文献[16]中模型精度检验标准,设定 $\alpha = 10\%$,以高标准要求检验虚拟变量对系统的作用,进而保证本文模型能够高精度模拟和预测.

2) 序列相似性检验. 对于虚拟变量是否引入的考察标准除了常用的相对误差和平均相对误差作为模型建模效果检验标准以外,本文将通过关联模型从拟合序列相似程度角度做检验,对比拟合序列与原始序列的形状相似程度,以此作为是否需要考虑虚拟变量以及进行外推预测的依据,从而保证建模得到的拟合序列与原始序列保持趋势的一致性. 如果引入虚拟变量建模后,模拟序列的灰色关联度有所改善,则虚拟变量引入通过检验;反之,则属于无效虚拟变量. 采用绝对关联度作为拟合序列与原始序列形状相似性的判断依据,计算公式如下:

$$|s_0| = \sum_{k=2}^{n-1} |x_0^0(k)| + \frac{1}{2}|x_0^0(n)|,$$

$$|s_j| = \sum_{k=2}^{n-1} |x_j^0(k)| + \frac{1}{2}|x_j^0(n)|,$$

$$|s_j - s_0| = \sum_{k=2}^{n-1} |x_j^0(k) - x_0^0(k)| + \frac{1}{2}|x_j^0(n) - x_0^0(n)|.$$

其中: X_0^0 和 X_j^0 分别为序列 X_0 和 X_j 的始点零化像,则关联度计算公式为

$$\gamma_{0j} = \frac{1 + |s_0| + |s_j|}{1 + |s_0| + |s_j| + |s_j - s_0|}. \quad (17)$$

1.5 DVCGM(1,N)模型建模步骤

Step 1: 模型变量的选取. 根据实际应用案例,初步分析系统变量及包含定量和虚拟变量的驱动变量.

Step 2: 对DVCGM(1,N)进行建模. 根据定理1建立DVCGM(1,N)模型,并对模型参数 a, b 进行估计;根据定理2建立DVCGM(1,N)的预测模型,进行模拟和预测.

Step 3: 由式(11)建立背景值优化的DVCGM(1,N)模型,利用粒子群算法进行背景值插值系数优化,然后替换定理1中的背景值,依据Step 2的流程进行建模和预测.

Step 4: 利用两种虚拟变量的有效性检验方法对引入虚拟变量后的DVCGM(1,N)及其优化模型的建模效果进行检验. 如果通过检验,则利用该模型进行预测;如果没有通过检验,则返回Step 1,筛选驱动变量进行重新建模,直到获得满意的建模效果为止.

2 实例分析

为验证本文模型在实际应用中的有效性,选取河南省农民人均纯收入作为研究对象,分别建立传统GM(1,N)模型, DVCGM(1,N)模型和DVCGM(1,N)优化模型,比较他们的建模精度. 河南省地处我国中西部地区,人口较多,是我国重要的农业大省,被誉为“中国粮仓”. 近几年,河南省农民纯收入在稳定增长,生活水平有了较大的改善,尤其是自2005年以来,河南省农民收入持续较快增长,这与中央政府取消农业税,对农民种地实行补助等一系列政策措施有着密切的联系. 根据河南省统计年鉴,2013年农民全年人均总收入为11345元,其中家庭经营收入为6804元,占总收入的59.97%,而第一产业收入为5304元,占家庭经营收入的77.95%. 由此可见,河南省农民收入的主要来源是农业生产,农业税的取消对于提升农民积极性和提高农民收入有着重要意义. 因此,根据Step 1,在对农民收入影响因素指标的选择方面,本文选取粮食产量、农业财政支出和农业税政策3个指标作为投入变量(相关因素变量),农民人均收入作为产出变量(系统行为特征变量). 另外,鉴于数据统计指标和口径的变化,本文选取2000~2011年数据进行建模,从而验证本文模型的有效性,数据见表1.

表1 2000~2011年河南省农民人均收入及影响因素

年份	农民人均收入/元	粮食产量/万吨	农业财政支出/亿元	取消农业税政策/是1否0
2000	2726	4101.5	34.19	0
2001	2916	4119.88	36.94	0
2002	3060	4209.98	44.77	0
2003	3036	3569.47	47.92	0
2004	3536	4260	65.99	0
2005	3946	4582	82.98	1
2006	4459	5112.3	111.34	1
2007	5197	5245.22	152.51	1
2008	5994	5365.48	209.59	1
2009	6414	5389	361.6	1
2010	7293	5437	399.19	1
2011	8725	5542.5	480.48	1

由表1可见,4个变量数据的量纲不一致,为了避免多变量数据灰色建模中出现矩阵漂移现象,需要对各变量进行初值化处理,得到预测序列之后,进行初值化的逆变换,还原系统行为特征序列的量纲和数量级. 根据Step 2和Step 3,通过对初始化数据进行建模可得传统GM(1,N)模型, DVCGM(1,N)模型和DVCGM(1,N)优化模型的计算公式如下所示:

传统GM(1,N)模型为

$$\hat{x}_1^{(1)}(k) = (1.1223x_2^{(1)}(k) + 0.1159x_3^{(1)}(k)) \times (1 - e^{-0.9742(k-1)}) + e^{-0.9742(k-1)}.$$

DVCGM(1, N) 模型为

$$\hat{x}_1^{(1)}(k) = (1.0843x_2^{(1)}(k) + 0.0996x_3^{(1)}(k) + 0.1087d_4^{(1)}(k)) \times (1 - e^{-1.1637(k-1)})d + e^{-1.1637(k-1)}$$

DVCGM(1, N) 优化模型为

$$\hat{x}_1^{(1)}(k) = (1.0988x_2^{(1)}(k) + 0.0999x_3^{(1)}(k) + 0.0974d_4^{(1)}(k)) \times (1 - e^{-1.4204(k-1)}) + e^{-1.4204(k-1)}$$

根据 Step 4 从两个角度进行虚拟变量的有效性

检验. 对于 DVCGM(1, N) 优化模型, 其插值系数可通过粒子群算法求得, $\lambda = 0.7063$, 并且粒子群的收敛情况如图 1 所示.

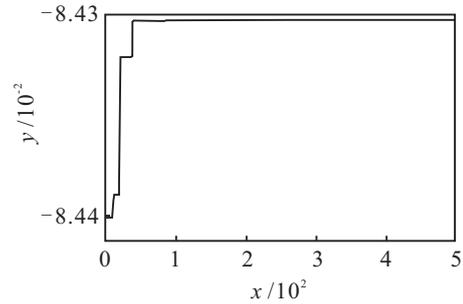


图 1 最优适应度踪迹

表 2 河南省农村居民人均收入的模拟和预测结果

年份	原始数据 / 元	GM(1, N) 模型		DVCGM(1, N) 模型		DVCGM(1, N) 优化模型	
		模拟值	误差 / %	模拟值	误差 / %	模拟值	误差 / %
2000	2726	2726.00	0	2726.00	0	2726.00	0
2001	2916	2529.73	13.25	2588.06	11.25	2916.00	0
2002	3060	4002.31	30.80	3867.10	26.37	3935.53	28.62
2003	3036	3613.86	19.04	3342.62	10.10	3268.75	7.67
2004	3536	4069.37	15.09	3771.42	6.66	3734.62	5.61
2005	3946	4335.70	9.88	4334.07	9.83	4305.44	9.11
2006	4459	4917.16	10.28	4896.44	9.81	4896.71	9.82
2007	5197	5358.23	3.10	5299.89	1.98	5314.88	2.27
2008	5994	5955.76	0.64	5831.73	2.71	5855.45	2.31
2009	6414	7368.38	14.88	7052.43	9.95	7083.51	10.44
模拟值平均相对误差 / %		12.99		9.85		9.48	
关联度		0.8114		0.8417		0.9069	
		预测值	误差 / %	预测值	误差 / %	预测值	误差 / %
2010	7293	8287.29	13.63	7384.19	1.25	7417.99	1.71
2011	8725	9598.91	10.02	8105.22	7.10	8142.83	6.67
预测值平均相对误差 / %		11.82		4.18		4.19	

从表 2 可以看出, 经典 GM(1, N) 模型的模拟和预测的平均相对误差分别为 12.99% 和 11.82%, 模拟和预测的平均误差均大于 10%, 主要是由于没有考虑政策等定性因素对于提升农民种地积极性、改善农民种地条件的影响, 因此经典 GM(1, N) 模型属于残差不合格模型; 而基于虚拟变量控制的 DVCGM(1, N) 模型模拟平均相对误差为 9.85%, 预测误差缩小至 4.18%, 充分说明, 考虑了取消农业税政策对于提升农民积极性, 提高农民收入具有较大的促进作用, 在建模时应予以充分考虑才能实现准确模拟和预测. 对于 DVCGM(1, N) 优化模型, 模拟误差实现理论上最小, 达到 9.48%, 显著优于经典模型, 其预测误差与 DVCGM(1, N) 相近, 也能充分展示良好的模拟和预测效果. 同时也发现, 增加背景值插值系数的优化可以降低样本内的平均模拟误差, 但不能确保提高样本外推误差最小. 从误差检验角度看, 基于虚拟变量控制的 DVCGM(1, N) 及其优化模型, 均属于残差合格模型, 从图 2 也能得出类似结论.

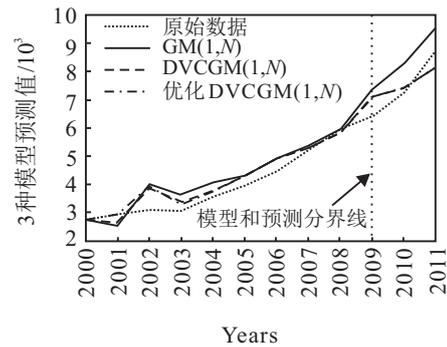


图 2 3种模型的模拟和预测

就取消农业税政策实施前后建模效果而言, 考虑了取消农业税政策后, 2006~2009年 DVCGM(1, N) 及其优化模型的模拟精度分别为 6.11% 和 6.21%, 小于没考虑取消农业税政策的经典 GM(1, N), 其值为 7.22%. 就两步外推预测效果而言, 2010~2011年 DVCGM(1, N) 及其优化模型的模拟精度分别为 4.18% 和 4.19%, 远小于经典模型的 11.82%, 进而说明取消农业税的政策在实施后几年对农民收入影响

较大,不能忽视取消农业税政策的作用。

另外,就灰色关联方法的有效性检验角度而言,本文DVCGM(1,N)模型及其优化模型与原始序列的绝对关联度分别为0.8417和0.9069,而传统灰色GM(1,N)模型得到的拟合序列与原始序列之间的绝对关联度为0.8114。由此可知,本文模型相对于传统模型对序列特征的描述更为精准,说明虚拟变量的引入具有实际意义,基于虚拟变量控制的多变量模型及其优化模型用于河南省农民收入预测的可行性更高。因此,综合采用两种检验方法对比3种模型的建模效果可以发现,引入政策等虚拟控制变量到灰作用量后,虚拟变量控制的DVCGM(1,N)模型相对于传统GM(1,N)模型能够更准确地描述和预测河南省农民收入增长的变化规律。

3 结论

虚拟变量普遍存在于经济社会系统中,虚拟变量和定量变量共同对系统产生影响。针对传统模型没有考虑虚拟变量作用的缺陷,本文将虚拟变量作为灰作用量引入传统模型构建中,建立了基于虚拟变量控制的DVCGM(1,N)模型,给出模型参数求解方法和优化方法,利用粒子群算法对背景值插值系数进行优化,插值系数取值的多样性赋予了优化模型可以适应多样性的数据特点,为解决系统变量受到虚拟变量影响时的小样本建模问题提供了一种定量解决方法。另外,从两个角度提出了虚拟变量有效性检验的衡量办法,确保虚拟变量的引入不仅具有本身的物理意义,还具有实际应用价值。新建模型不仅满足了灰色理论建模原理,而且是传统GM(1,N)的延伸和拓展,若不存在虚拟变量,则新模型将退化为经典模型。通过对河南省农民人均收入的模拟和预测发现,考虑取消农业税政策对农民收入影响的DVCGM(1,N)模型及其优化模型的建模效果显著优于传统模型。

参考文献(References)

- [1] Deng J L. The control problem of grey systems[J]. System Control Letter, 1982, 1(5): 288-294.
- [2] Liu S F, Dang Y G, Fang Z G. Grey system theory and its application[M]. Beijing: Science Press, 2010: 149-157.
- [3] Guo H, Xiao X P, Forrest J. A research on a comprehensive adaptive grey prediction model CAGM(1,N)[J]. Applied Mathematics and Computation, 2013, 225: 216-227.
- [4] Wang Z, Hao P. An improved grey multivariable model for predicting industrial energy consumption in China[J]. Applied Mathematical Modelling, 2016, 40(11): 5745-5758.
- [5] 丁松,党耀国,徐宁. 基于驱动变量增长趋势的TGM(1,N)预测模型[J]. 系统工程, 2015, 33(3): 149-153.
- [6] 黄继. 灰色多变量GM(1,N| τ, r)模型及其粒子群优化算法[J]. 系统工程理论与实践, 2009, 29(10): 145-151. (Huang J. Grey GM(1,N| τ, r) model and its particle swarm optimization algorithm[J]. Systems Engineering—Theory & Practice, 2009, 29(10): 145-151.)
- [7] 毛树华,高明运. 分数阶累加时滞GM(1,N,)模型及其应用[J]. 系统工程理论与实践, 2015, 35(2): 430-436. (Mao S H, Gao M Y. Fractional order accumulation time-lag GM(1,N,) model and its application[J]. Systems Engineering—Theory & Practice, 2015, 35(2): 430-436.)
- [8] Ding S, Dang Y G, Li X M, et al. Forecasting Chinese CO₂ emissions from fuel combustion using a novel grey multivariable model[J]. J of Cleaner Production, 2017, 162: 1527-1538.
- [9] 张可,曲品品. 时滞多变量离散灰色模型及其应用[J]. 系统工程理论与实践, 2015, 35(8): 2092-2103. (Zhang K, Qu P P. Delay multi-variables discrete grey model and its application[J]. Systems Engineering—Theory & Practice, 2015, 35(8): 2092-2103.)
- [10] Tien T L. The indirect measurement of tensile strength of material by the grey prediction model GMC(1,N)[J]. Measurement Science Technology, 2005, 16(6): 1322-1328.
- [11] Tien T L. Forecasting CO₂ output from gas furnace by grey prediction model IGMC(1,N)[J]. J of the Chinese Society of Mechanical Engineers, 2010, 31(1): 55-65.
- [12] Tien T L. The indirect measurement of tensile strength by the new model FGMC(1,N)[J]. Measurement, 2011, 44(10): 1884-1897.
- [13] 熊萍萍,党耀国,朱晖. 基于非等间距的多变量MGM(1,m)模型[J]. 控制与决策, 2011, 26(1): 49-53. (Xiong P P, Dang Y G, Zhu H. Research of modeling of multi-variable non-equidistant MGM(1,m) model[J]. Control and Decision, 2011, 26(1): 49-53.)
- [14] 蒋诗泉,刘思峰. 基于复化梯形公式的GM(1,1)模型背景值的优化[J]. 控制与决策, 2014, 29(12): 2221-2225. (Jiang S Q, Liu S F. Optimization of background value in GM(1,1) based on compound trapezoid formula[J]. Control and Decision, 2014, 29(12): 2221-2225.)
- [15] Hsu L C. Forecasting the output of integrated circuit industry using genetic algorithm based multivariable grey optimization models[J]. Expert Systems with Applications, 2009, 36(4): 7898-7903.
- [16] Xu N, Dang Y, Gong Y. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China[J]. Energy, 2017(118): 473-480.