

对灰色预测模型残差问题的探讨

刘 树, 王 燕, 胡凤阁

(河北大学 经济学院, 河北 保定 071000)

摘 要: 对利用原始经济序列 x_0 建立的灰色预测模型检验不合格或精度不理想时, 要对建立的模型进行残差修正(建立修正模型), 以提高模型的预测精度。在对以往的残差模型进行残差检验时常用

$(1)=x_1-\hat{x}_1$ 衡量, 笔者认为利用灰色模型实际预测的是 \hat{x}_0 的大小, 因此对模型进行检验时需用 $(0)=x_0-\hat{x}_0$ 衡量。本文以灰色预测模型中的 GM(1,1) 模型为例, 对两种残差检验的衡量方法进行了比较分析, 并提出了改进灰色预测模型的方法与建议。

关键词: 残差检验; 误差分析; 残差修正

中图分类号: C81

文献标识码: A

文章编号: 1002-6487(2008)01-0009-03

灰色预测法是一种对含有不确定因素的系统进行预测的方法, 它用生成模块建立微分方程模型, 可从少量的、离散的、杂乱无章的数据中找出规律性, 并且具有良好的时效性。灰色预测模型还能够根据现有的少量信息进行计算和推测, 先根据自身数据建立动态微分方程, 再预测自身的发展, 特别适合“少数据建模”。

经过近几十年的应用, 普遍反映其预测结果比较接近实际, 是一种可靠的预测方法, 目前已广泛应用于农业、社会和经济领域, 成功地解决了生产、生活和科学研究中的大量实际问题, 并已取得显著成就。灰色理论在管理学、决策学、战略学、预测学、未来学、生命科学等领域也展示了极为广阔的应用前景。本文对“以往研究时对灰色预测残差模型(经过残差修正的原始序列预测模型)的残差检验用 $(1)=x_1-\hat{x}_1$ 表示”提出了质疑, 并以灰色预测模型中常用的 GM(1,1) 模型为例, 计算了残差模型的 x_0 的误差, 即 $(0)=x_0-\hat{x}_0$, 对 (0) 进行了比较分析, 认为无论何种灰色预测模型, 对其进行残差检验时或反映模型的精度时, 应该依据 x_0 与其预测值的误差作为判断标准, 最后提出提高 GM(1,1) 模型预测精度的方法, 以丰富和完善 GM(1,1) 模型。

1 灰色预测 GM(1,1) 模型及 GM(1,1) 残差模型

1.1 GM(1,1) 模型

GM(1,1) 模型由一个单变量的一阶微分方程构成, 设原始数据列 x_0 有 n 个观察值, $x_0=\{x_0(1), x_0(2), \dots, x_0(n)\}$, 则通过累加生成一次累加序列 $x_1=\{x_1(1), x_1(2), \dots, x_1(n)\}$, 其中, $k=1, 2, \dots, n$ 。则 GM(1,1) 相应的微分方程为:

$$\frac{dx_1}{dt} + x_1 = \mu$$

式中, μ 为发展灰数; μ 为内生控制灰数。

设 $\hat{\mu}$ 为待估参数向量, $\hat{\mu} = \begin{pmatrix} \mu \\ \mu \end{pmatrix}$, 利用最小二乘法可得:

$$\hat{\mu} = (B^T B)^{-1} B^T Y_n$$

其中:

$$B = \begin{bmatrix} -\frac{1}{2}[x_1(1)+x_1(2)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x_1(n-1)+x_1(n)] & 1 \end{bmatrix}$$

$$Y_n = \begin{bmatrix} x_0(2) \\ 1 \\ \vdots \\ x_0(n) \end{bmatrix}$$

求解微分方程, 可得预测模型, 即 GM(1,1) 模型:

$$\hat{x}_1(k+1) = [x_0(1) - \frac{\mu}{2}]e^{-k\frac{\mu}{2}} + \frac{\mu}{2} \quad k=0, 1, 2, \dots, n \quad (1)$$

1.2 GM(1,1) 残差模型

若用原始序列建立的 GM(1,1) 模型检验不合格时, 要对建立的 GM(1,1) 模型进行残差修正。设原始经济时间序列建立的模型为:

$$\hat{x}_1(i+1) = [x_0(1) - \frac{\mu}{2}]e^{-i\frac{\mu}{2}} + \frac{\mu}{2}$$

可获得生成序列 x_1 的预测值 \hat{x}_1 , 即对于 $x_1=\{x_1(1), x_1(2), \dots, x_1(n)\}$, 有预测序列 $\hat{x}_1=\{\hat{x}_1(1), \hat{x}_1(2), \dots, \hat{x}_1(n)\}$, 定义残差为: $e_i(j)=x_1(j)-\hat{x}_1(j)$ 。若取 $j=i, i+1, \dots, n$, 则与 x_1 及 \hat{x}_1 对应的残差序列为: $e_i=\{e_i(i), e_i(i+1), \dots, e_i(n)\}$, 其累加生成序列为: $e_i=\{e_i(i), e_i(i+1), \dots, e_i(n)\}$, 利用 e_i 建立相应的 GM(1,1) 模型:

$$\hat{e}_i(k+1) = [e_i(1) - \frac{\mu_i}{2}]e^{-k\frac{\mu_i}{2}} + \frac{\mu_i}{2}$$

$$\hat{e}_i(k+1) \text{ 的导数为: } (-\frac{\mu_i}{2})[e_i(1) - \frac{\mu_i}{2}]e^{-k\frac{\mu_i}{2}} \text{ 加上 } \hat{e}_i(k+1) \text{ 修正 } \hat{x}_1$$

(k+1), 得修正模型:

$$\hat{x}_1(k+1) = [x_0(1) - \frac{\mu}{2}]e^{-k\frac{\mu}{2}} + \frac{\mu}{2} + \partial(k-1)(-\frac{\mu_i}{2})[e_i(1) - \frac{\mu_i}{2}]e^{-k\frac{\mu_i}{2}}$$

其中 $\partial(k-1) = \begin{cases} 1, & k=2 \\ 0, & k < 2 \end{cases}$ 为修正系数。

最后给出经过残差修正的原始序列预测模型:

$$\hat{x}_0(k+1) = \hat{x}_1(k+1) - \hat{x}_1(k) \quad (k=1, 2, \dots, n) \quad (2)$$

2 GM(1,1)模型与 GM(1,1)残差模型的残差检验

2.1 模型残差检验的判断标准及方法

模型的检验是建立模型后进行的必不可少的工作,对灰色预测模型的检验一般有残差检验、关联度检验和后验差检验,在此重点介绍下模型的残差检验。按预测模型计算 $\hat{x}_i(i)$,并将 $\hat{x}_i(i)$ 累减生成 $\hat{x}_0(i)$,然后计算原始序列 $x_0(i)$ 与 $\hat{x}_0(i)$ 的绝对误差序列及相对误差序列。

$$e_0(i) = |x_0(i) - \hat{x}_0(i)| \quad i=1,2, \dots, n \quad (3)$$

$$r_0(i) = \frac{e_0(i)}{x_0(i)} \times 100\% \quad i=1,2, \dots, n \quad (4)$$

模型的残差检验的判断标准有两种:“给定 α ,当 $\alpha < 0.01$ —0.05”1;成立时,认为模型为残差合格模型(一般取0.01—0.05)1;

如果模型的相对误差平均值在1%及以下,说明模型的精度等级为一级;如果其大于1%且小于等于5%,模型的精度等级为二级;如果其大于5%且小于等于10%,则为三级;当其大于10%且小于等于20%,则模型的精度等级为四级。

对原始时间序列建立GM(1,1)模型的实质是为了预测该时间序列的发展趋势,预测时间序列未来的发展值,因此分析GM(1,1)模型的残差检验或模型的预测精确度应该分析序列 x_0 的相对误差大小,如式(3)所示。当取 α 值为0.05时,如果序列 x_0 的相对误差平均值的小于等于0.05则认为该模型通过检验。

对于GM(1,1)残差模型的残差检验问题,笔者观察到,在以往的研究中是用累加序列 x_1 的相对误差大小来衡量的。例如,徐国祥教授的《统计预测和决策》一书中“GM(1,1)残差模型及GM(n,h)模型”在原模型的残差检验用 x_0 的误差衡量没有通过时,对模型进行了修正,而对经过残差修正的模型的残差检验时用累加序列 x_1 的相对误差进行衡量,最后得出了通过检验。笔者对此提出了质疑,认为这背离了灰色预测的实质。虽然用累加序列 x_1 的绝对误差与相对误差来衡量能够通过残差检验,但 x_0 的误差情况并没有改善,模型预测的精确度就没有得到提高,只是对原模型的残差起到一定的修匀作用,残差模型的建立并没有达到真正的修正意义。因此,GM(1,1)残差模型的残差检验或精确度仍需要用 x_0 的相对误差大小来衡量。

2.2 GM(1,1)残差模型(修正模型)与 GM(1,1)模型(原模型)的误差计算

以徐国祥教授的《统计预测和决策》一书中“GM(1,1)残差模型及GM(n,h)模型”的数据为例,笔者计算了残差模型的 x_0 的误差和GM(1,1)模型的 x_1 的误差。

设原始序列 x_0 为1991—1996年国内生产总值构成中的总消费,其累加序列为 x_1 , \hat{x}_0 (原)、 \hat{x}_1 (原)分别为用GM(1,1)模型预测的国内生产总值构成中的总消费序列和其累加序列, \hat{x}_0 (修正)、 \hat{x}_1 (修正)分别为GM(1,1)残差模型预测的国内生产总值总消费序列和其累加序列,如表1所示。其中 e_0 (原)、 r_0 (原)、 e_1 (原)、 r_1 (原)分别表示GM(1,1)模型的 x_0 的绝对误差、 x_0 的相对误差、 x_1 的绝对误差、 x_1 的相对误差, e_0 (修正)、 r_0 (修正)、 e_1 (修正)、 r_1 (修正)分别表示GM(1,1)残差模型的 x_0 的绝对误差、 x_0 的相对误差、 x_1 的绝对误差、 x_1 的

表1 预测值表 (亿元)

年份	x_0	x_1	\hat{x}_0 (原)	\hat{x}_1 (原)	\hat{x}_0 (修正)	\hat{x}_1 (修正)
1991	386.06	386.06	386.06	386.06	386.06	386.06
1992	476.57	862.63	530.43	916.49	513.95	900.01
1993	679.35	1541.98	653.07	1569.57	722.55	1622.56
1994	873.89	2415.80	804.06	2373.63	804.07	2426.63
1995	1085.33	3501.20	989.96	3363.60	992.57	3419.20
1996	1252.33	4753.53	1218.84	4582.44	1202.86	4622.06

表2 误差表

	1992	1993	1994	1995	1996	总误差	平均	最大值	最小值
e_0 (原)	53.87	26.27	69.82	95.37	33.49	278.82	46.47	95.37	26.27
r_0 (原)%	11.30	3.87	7.99	8.79	2.67	34.62	5.77	11.30	2.67
e_1 (原)	53.87	27.59	42.23	137.60	171.08	432.38	72.06	171.08	27.59
r_1 (原)%	6.24	1.79	1.75	3.93	3.60	17.31	2.89	6.24	1.75
e_0 (修正)	37.38	43.20	69.82	92.76	49.47	292.63	48.77	92.76	37.38
r_0 (修正)%	7.84	6.36	7.99	8.55	3.95	34.69	5.78	8.55	3.95
e_1 (修正)	37.38	80.58	10.76	82.00	131.47	342.19	57.03	131.47	10.76
r_1 (修正)%	4.33	5.23	0.45	2.34	2.77	15.11	2.52	5.23	0.45

注:由于原始序列与累加序列的第一项相等,因此第一期的误差为0,在本表中未列出。

相对误差,计算结果如表2所示。

2.3 GM(1,1)残差模型与 GM(1,1)模型的误差比较分析

笔者利用计算的误差对残差模型和原模型进行了几个方面的比较分析。

2.3.1 残差模型与原模型的 x_0 的误差比较

由表2可以看出,无论从绝对误差还是从相对误差来说,残差模型的 x_0 的总误差和平均误差都大于原模型的 x_0 的总误差和平均误差,并且残差模型的 x_0 的误差最小值大于原模型的 x_0 的误差最小值,即经过残差修正的模型其预测情况并不及原模型。依照模型的残差检验标准,在给定 α 取值0.05时残差模型与原模型都没有通过检验,残差模型与原模型的预测精度等级为三级。由此,可以说经过残差修正后的模型其精确度并没有提高,原模型与残差模型都未通过残差检验,可见残差模型并未达到修正目的。

2.3.2 残差模型的 x_0 的误差与 x_1 的误差比较

无论是相对误差的总值还是平均值,残差模型的 x_1 的误差都低于 x_0 的误差,并且 x_1 的相对误差最大值小于 x_0 的相对误差最大值, x_1 的相对误差最小值小于 x_0 的相对误差最小值。如果残差模型以 x_1 的相对误差来进行残差检验,因为其值为2.52%小于0.05,则模型通过检验,且模型的预测精度为二级。如果残差模型以 x_0 的相对误差来进行残差检验,因为其值为5.78%>0.05,则模型通不过检验,模型的预测精度为三级。

2.3.3 残差模型与原始模型的 x_1 的误差比较

由表2可以看出,经过残差修正后的模型的误差总值、误差平均值都低于原始模型的误差总值、误差平均值。但是,当以 x_1 的相对误差来进行残差检验时,因为其值小于0.05,所以无论是残差模型还是原始模型都能通过检验,并且两个模型的预测精度都为二级。可见,若以此为检验标准则残差模型就没有了建立的必要。

2.3.4 原始模型的 x_1 的误差与 x_0 的误差比较

对于原始模型而言,无论是绝对误差还是相对误差, x_1 的误差总值、平均值都低于 x_0 的误差总值、平均值。以 x_1 的相对误差平均值来衡量原模型的精确度,依据残差检验标准,则原模型的预测精确度为二级,模型能通过检验。然而,当用 x_0 的相对误差平均值来衡量原模型的精确度时,依据残

差检验标准,原模型的预测精确度为三级,模型通不过检验。

由以上的分析可以看出:无论是原模型还是残差模型进行残差检验时,如果以 x_1 的相对误差进行衡量,则模型都能通过检验,并且模型的精确度为二级;如果原模型和残差模型进行残差检验时,以 x_0 的相对误差进行衡量,则两个模型都通不过检验,模型的预测精确度为三级。由此,可以得出:徐国祥教授之所以认为残差模型能通过检验,是因为其用 x_1 的相对误差进行衡量的结果,而这背离了灰色预测的实质,使预测失去了真实性;而实际上,经过残差修正的模型与原模型相比,其依旧是通不过残差检验,并且没有提高模型的预测精确度;对模型进行检验的目的是“为了预测而检验”而不是“为了检验而检验”,不能为了使模型通过检验而用 x_1 的相对误差衡量,否则就背离模型预测的实质,使预测失去意义。

2.4 x_1 的误差与 x_0 的误差关系分析

在以上的分析中,观察到 GM(1,1)残差模型与 GM(1,1)模型进行残差检验时 x_0 的相对误差总值、平均值都大于 x_1 的相对误差总值、平均值,并且如果用 x_0 的误差进行衡量则都通不过检验,而如果用 x_1 的误差进行衡量则能通过检验。笔者分析了其中的原因,认为对于任意 GM(1,1)模型或其残差模型而言, x_0 的相对误差平均值肯定都大于或至少等于 x_1 的相对误差平均值。

在表 2 中计算相对误差时各项数据都取了绝对值,笔者计算了不取绝对值的 x_0 的预测值与 x_0 的差和 x_1 的预测值与 x_1 的差,如表 3 所示。其中 ∂_0 原、 ∂_1 原分别为原模型的预测值减去的 x_0 值、 x_1 预测值减去 x_1 的值, ∂_0 修正、 ∂_1 修正分别为残差模型的 x_0 预测值减去 x_0 的值、 x_1 预测值减去 x_0 的值。在表 3 中可以看到: ∂_1 原、 ∂_1 修正分别为 ∂_0 原的累加序列、 ∂_0 修正的累加序列,这是由于 x_1 为 x_0 的累加序列的缘故。以原模型为例进行分析,第一期 ∂_0 原为 0, ∂_1 原也为 0;在第 2 期时 ∂_0 原为负值,因此对前 2 期进行累加得到 ∂_1 原的第 2 期值时,由于负值的作用使得 ∂_1 原的第 2 期值小于第 1 期值;在第 3 期时 ∂_0 原为正值,因此对前 3 期进行累加得到 ∂_1 原的第 3 期值时,由于正值的作用使得 ∂_1 原的第 3 期值大于第 2 期值;同理,对于残差修正模型也有此规律。由此可以得出: ∂_0 的值为正数时, ∂_1 的同期值会增加,一旦 ∂_0 的值出现负数时,就会降低同期 ∂_1 的值;由于累加序列求相对误差时基数比原序列的大,再加上负值出现的抵消作用,使得 ∂_1 值减小,因此使得累加序列的相对误差要小于原序列的相对误差。所以,当用 x_1 的相对误差进行残差检验时模型预测精度高于用的相对误差进行检验的模型预测精度。

表 3 误差比较分析表

年份	1991	1992	1993	1994	1995	1996
∂_0 原	0	-53.8687	26.2747	69.8247	95.3661	33.488
∂_1 原	0	-53.8687	-27.594	42.2307	137.597	171.085
∂_1 (原)	0	53.8687	27.594	42.2307	137.5968	171.0848
∂_0 修正	0	-37.38	-43.2	69.82	92.76	49.47
∂_1 修正	0	-37.38	-80.58	-10.76	82	131.47
∂_1 (修正)	0	37.38	80.58	10.76	82	131.47

3 对于 GM(1,1)模型的改进和建议

在本例中, GM(1,1)残差模型的建立没有提高原始模型的精度,因此 GM(1,1)残差模型的建立并未真正起到误差修正的作用,在此运用另一种提高灰色模型预测精度的模型——带残差修正的 GM(1,1)模型对 GM(1,1)模型的进行改进。

带残差修正的 GM(1,1)模型,把残差值作为补充量加在预测值上,从而使原来的残差值大为减小,从而使预测精度得以显著地提高。建立带残差修正的 GM(1,1)模型的步骤为:

运用 GM(1,1)模型求得 $\hat{x}_1, \hat{x}_0, e_0$ 。

对残差序列进行处理,要求有两点:非负 e_0 ; 单调升: $e(k+1) > e(k)$ 。

如果残差序列中有值小于 0, 则应在残差序列上加一个适当的正数,使其中最小值变为 0, 取 $e_{\min}\{e(1), \dots, e(n)\}$, 将其绝对值加在残差序列上得到非负序列 $\hat{e}, \hat{e}_i(k) = \sum_{i=1}^k e_0(i)$, 且 \hat{e}_i 必为单调升的序列。

利用 GM(1,1)模型求得 \hat{x}_1 , 将其累减生成序列 \hat{x}_0 , 再减去 e_{\min} 得到:

$$e_0 = \hat{x}_0 - e_{\min}$$

将 \hat{e}_0 加在预测序列 \hat{x}_1 上, 得到修正后的预测值 $\hat{x}_1 = \hat{x}_1 + \hat{e}_0$, 将其累减生成序列 \hat{x}_0 。

重新计算残差,即 x_0 与 \hat{x}_0 的误差,作后验差检验。

利用本例中的数据,运用带残差修正的 GM(1,1)模型,对 GM(1,1)模型进行修正,求得的结果如表 4 所示。由表 4 可以看出:经过修正后的模型 x_0 的相对误差值最大值为 0.078 小于 GM(1,1)模型的 x_0 的相对误差值最大值 0.11;经计算 x_0 的相对误差值平均值为 0.0318, 小于运用 GM(1,1)模型与 GM(1,1)残差模型计算的 x_0 的相对误差平均值,依据残差检验的标准模型能够通过检验,且由于 $0.0318 < 0.05$, 模型的预测精度达到二级,达到了修正原模型误差、提高模型预测精度的目的。

表 4 带残差修正的 GM(1,1)模型计算结果

\hat{e}_1	\hat{e}_0	\hat{e}_0	\hat{x}_1	\hat{x}_0	x_0	相对误差
53.8687	53.8687	0.0000	386.0600	386.0600	386.0600	0.0000
53.6004	-0.2683	-54.1370	862.3617	476.3017	476.5700	0.0006
132.8796	79.2792	25.4105	1594.9845	732.6229	679.3500	0.0784
250.1399	117.2603	63.3916	2437.0309	842.0464	873.8900	0.0364
423.5772	173.4373	119.5686	3483.1718	1046.1409	1085.3300	0.0361
680.1048	256.5276	202.6589	4785.1041	1301.9322	1252.3300	0.0396

利用原始序列数据建立灰色预测模型是为了更准确地预测出事物未来的发展趋势,如果原始序列 x_0 的误差较大,即便是 GM(1,1)残差模型用 x_1 的误差衡量能够通过检验,此模型的建立也是毫无意义的。检验残差模型的目的是为了“预测而检验”,而不是为了“检验而检验”。因此,对于残差模型的检验而言,如要真实反映模型预测的精确度,就需要以 x_0 的预测值与其误差大小作为判断标准,只有这样才不失建立预测模型的意义。

参考文献:

- [1]斯林.灰色预测系统在城镇居民商品房销售预测中的应用[J].统计与决策,2004,(12).
- [2]杨红平,马大川.灰色预测理论在信息分析与预测中的应用[J].情报方法,2004,(3).
- [3]樊新海,苗卿敏,王华民.灰色预测 GM(1,1)模型及其改进与应用[J].装甲兵工程学院学报,2003,(6).
- [4]张大海,江世芳,史开泉.灰色预测公式的理论缺陷及改进[J].系统工程理论与实践,2002,(8).
- [5]徐国祥.统计预测与决策[M].上海:上海财经大学出版社,1998.
- [6]董力.现代经济管理预测与决策方法[M].北京:地震出版社,1999.

(责任编辑/亦民)