第34卷 第4期 电子科技大学学报 Vol.34 No.4
2005年8月 Journal of UEST of China Aug. 2005

GMDH物价指数预测模型研究及实证

田益祥

(华中科技大学经济学院 武汉 430074)

【摘要】利用物价指数的月均值与年均值数据同时建模的数据处理群组方法的两水平算法，扩大了月均值数据的可预测范围，引入以三角函数求和形式的调和自组织的数据处理群组方法解决月均值物价指数的波动影响，提高了月均值数据预测的准确性。实证分析表明，数据处理群组方法的调和两水平方法预测月均值物价指数是有效的。

关 键 词 数据处理群组方法； 多层迭代； 调和算法； 物价指数

中图分类号 F201; N945.24; O241.5 **文献标识码** A

Study Choosing Data of Two-Level Algorithm of GMDH and Applications

TIAN Yi-xiang

(School of economy, Huazhong University of Science and Technology Wuhan 430074)

Abstract The price idex is very inportant in the economic system ,and it is difficult to predict accurately it is difficult to predict accurately because of the reason and the long intervals. We solve this problems with harmonic algorithm and two-leve algorithm of GMDH .The result that we researched the problems of prediction with the theory and methods in example show thatthis method is very efficient in example.

Key words group method of data handling; multilayer algorithm; harmonic algorithm; price idex

模型自组织理论的数据处理群组方法(Group Method of Data Handling，GMDH)是以采用多层迭代方法[1]，借助计算机客观地选择输入输出变量一种建模理论，主要用于解决模式识别，复杂系统建模及预测。但在实际应用中，由于事物的发展不仅受长期趋势支配，还受到周期波动影响，从而呈现出以趋势线为中心的变动轨迹[2]。作为随机方程的回归预测的模型自组织建模，如果仅考虑到长期趋势，没有考虑周期波动，就会导致效果不理想。文献[3]中的GMDH调和算法预测模型很好地解决了这类波动事物的预测。文献[4, 5]中的两水平算法是利用月均值数据和年均值数据同时建模，把第一水平详细预测值的可预测范围，提高到第二水平模糊预测值的可预测范围，在一定程度上扩大了第一水平详细预测值的可预测范围。

1 调和两水平算法

调和两水平算法是利用统计数据的特点建模来拓广模型的可预测范围[6]。月均值数据采用以三角函数求

收稿日期：2003－04－23

作者简介：田益祥(1963－)，男，在职博士生，教授，现在电子科技大学管理学院工作，主要从事系统工程和金融工程方面的研究.

和形式的调和算法处理后与年均值数据同时建模，利用月均值数据建模时，预测值的详细性和年均值数据建模作预测相对长的范围，在一定程度上扩大了月均值数据的可预测范围。

1.1 计算方法

1.1.1 多层迭代算法

1) 建立原始年数据表，设X_1，X_2，…，X_n，为内生变量，U_1，U_2，…，U_m为外生变量。

2) 用年均值数据，采用多层迭代算法进行系统分析，探测变量之间的关系，决定模型结构及输出变量。

第一层，对n个变量中的每一个变量用多项式给出确定性差分形式的模型[7]：

$$Q_i^t = a_0 + \sum_{k=1}^{m}\sum_{j=0}^{\alpha} a_{ik}U_k^{t-j} + \sum_{j=1}^{\alpha} b_i Q_i^{t-j} \qquad i=1，2，…，n \tag{1}$$

式中 Q_i 是年均值数据变量，a 是应考虑的滞后数，a_{ik} 和b_i为系数。

第二层，增加一个内生变量穷举它们之间的关系得：

$$Q_{ik}^t = a_0 + \sum_{k=1}^{m}\sum_{j=0}^{\alpha} a_{ik}U_k^{t-j} + \sum_{j=1}^{\alpha} b_i Q_i^{t-j} + \sum_{j=1}^{\alpha} c_k Q_k^{t-j} \quad I，k=1，2，…，n \quad i \neq k \tag{2}$$

式中 Q_{ik}^t 是变量Q_i^t在第二层的表示，C_k为系数，a为季节滞后数。

如此下去，第n层增加n个内生变量，穷举它们之间的关系得：

$$Q_i^t = a_0 + \sum_{k=1}^{m}\sum_{j=0}^{\alpha} a_{ik}U_k^{t-j} + \sum_{j=1}^{\alpha} b_i Q_i^{t-j} + \sum_{k=1}^{n}\sum_{j=0}^{\alpha} b_{kj} Q_k^{t-j} \quad i=1，2，…，n，\quad i \neq k \tag{3}$$

式中 a_{jk}，b_j，b_{kj}为系数，a为季节滞后数。

在上面的每一层中穷举的所有方程，用系统最小偏差准则$\eta_{bs} = \sum_{P\in\omega}(Y-\hat{Y})_P^2$ 选出最优模型。在选择过程中，ε可调节，在$\eta_{bs} < \varepsilon$ 时选择最优模型。设输出变量为X_1，X_2，…，X_p，输入变量为X_{p+1}，…，X_{p+q}及U_1，U_2，…，U_h，可得到F_1个方程。

1.1.2 调和算法

由于月均值物价指数由于季节的影响往往具有波动性，对于这类具有周期趋势的预测对象，为提高预测的准确性，引入三角函数求和形式的调和算法。

1) 求出调和算法的频率w_k，

$$q_i(x) = \sum_{k=1}^{m}[A_k \sin(w_k x) + B_k \cos(w_k x)] \qquad I=1，2，…，p+q \tag{4}$$

式中 A_k，B_k为系数；$q_i(x)$为月均值变量；$w_i \neq w_j \quad i \neq j$，$0 < w_i < \pi$；$i$=1，2，…，$p+q$；m为一自然数，根据实际需要而选定(如在后面的例子中取m =3)，一般$m_{max} < N - N/3$，N为所选数据的月数。为求出w_k，记$\Phi(w_k x) = A_k \sin(w_k x) + B_k \cos(w_k x)$，由$q(x+p) + q(x-p) + 2\sum_{k=1}^{m}\cos(mw_k)\Phi(w_k x)$得：

$$\sum_{k=1}^{m} u_p[q(x+p)+q(x-p)] = 2\sum_{k=1}^{m}\Phi(w_k x)[u_0 + \sum_{p=1}^{m-1}u_p\cos(pw_k)] =$$

$$2\sum_{k=1}^{m}\cos(mw_k)\Phi(w_k x) = q(x+m) + q(x-m) \tag{5}$$

式中 $u_0 + \sum_{p=1}^{m-1}u_p\cos(pw_k) = \cos(mw_k)$，所以有：

$$q(x+m) + q(x-m) = \sum_{k=1}^{m-1}u_p[q(x+p)+q(x-p)] \tag{6}$$

由式(6)给定的m和月均数据，用最小二乘法求出u_0，u_1，…，u_{m-1}估计值，并代入$\hat{u}_0 + \sum_{p=1}^{m-1}\hat{u}_p\cos(pw_k) = \cos(mw_k)$，用牛顿插值法求出频率$w_k$。

2) 用输出变量X_1，X_2，…，X_p建立月均值系统方程，用二维时间读数年均值和月均值同时建模。

第一层，穷举方程$q_i(t,T) = \sum_{j=1}^{\alpha}a_i Q_i(T-j) + \sum_{k=1}^{n}\sum_{k=1}^{m}[A_{jk}\sin(w_k q_i(t-j)) + B_{jk}\cos(w_k q_i(t-j))]$，式中$a_i$，$A_{jk}$，$B_{jk}$为系数；$i=1，2，…，p$；$a$是考虑年的滞后数，一般取$a$ =5；n是考虑的月滞后数，一般取n =12；m =1，2，3，…，由实际需要而定。

第二层，增加一个内生变量穷举它们之间的关系：

$$q_{ik}(t,T)=\sum_{j=1}^{\alpha}a_iQ_i(T-j)+\sum_{k=1}^{n}\sum_{k=1}^{m}[A_{jk}\sin(w_kq_i(t-j)+B_{jk}\cos(w_kq_i(t-j))]+$$

$$\sum_{j=1}^{n}\sum_{k=1}^{m}[A_{jk}\sin(w_kq_h(t-j)+B_{jk}\cos(w_kq_h(t-j))]\qquad i\neq h;\ \mathrm{i},\ h=1,\ 2,\ \cdots,\ p \tag{7}$$

如此进行下去，直到第 $p+q$ 层为止。

对上面每一层穷举的方程，用η_{bs} 和$\Delta(c)$选择输出变量得到F_2个方程[8]。

3) 对穷举方程，用预测平衡准则 $\beta_L^2=\sum b_i^2\to\min$ 选出X_1，X_2，…，X_p在整体范围内的最优方程。预测平衡准则 $\beta_L^2=\sum b_i^2\to\min$中，$b_i^2=[Q_i-(q_{1i}+q_{2i}+,\ \cdots,\ +q_{Li})/L]^2$，$\beta_L^2$ 是区间N_1到N_2的总平衡，L表示一年的季节数。

4) min β_L^2 指出了年、季节的协调预测，即第二水平最优预测Q^*和第一水平最优复杂预测模型(具有最优预测q^*_{iT})。

2　用模型自组织GMDH调和两水平算法预测物价指数

为了论证GMDH调和两水平算法预测月均值物价指数的有效性，用月均值数据与年均值数据同时建模，对具有明显周期波动的月均值，引入具有三角函数求和形式的调和算法。为简单起见，直接确定物价的月均值为输出变量，它的滞后变量为输入变量。对1981～1997年的月统计数据，选取1981～1995年的数据作GMDH训练数据，对1996～1997年24个月的数据作为预测效果进行检验其有效性。

选取的社会消费品指数1981～1997年样本为数据表1所示。

表1　1981～1997 年社会消费品指数

年	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1981	1.144	1.002	1.149	1.123	1.120	1.102	1.092	1.081	1.090	1.129	1.131	1.145
1982	1.118	0.983	1.138	1.102	1.094	1.103	1.091	1.078	1.067	1.062	1.049	1.103
1983	1.022	1.272	1.097	1.096	1.933	1.086	1.083	1.107	1.098	1.082	1.138	1.135
1984	1.113	1.017	1.098	1.118	1.141	1.169	1.205	1.232	1.267	1.298	1.277	1.318
1985	1.378	1.585	1.543	1.456	1.421	1.418	1.408	1.375	1.364	1.326	1.282	1.250
1986	1.128	1.084	1.091	1.141	1.146	1.157	1.163	1.179	1.194	1.201	1.189	1.171
1987	1.167	1.137	1.168	1.166	1.171	1.178	1.200	1.186	1.161	1.179	1.162	1.167
1988	1.186	1.326	1.246	1.253	1.255	1.300	1.319	1.410	1.368	1.289	1.243	1.182
1989	1.209	1.178	1.249	1.208	1.177	1.088	1.059	1.98	0.976	1.984	1.990	1.003
1990	1.025	0.934	0.943	0.963	0.992	1.007	1.011	1.026	1.039	1.078	1.103	1.104
1991	1.085	1.227	1.121	1.162	1.114	1.125	1.126	1.129	1.135	1.142	1.143	1.125
1992	1.143	1.151	1.199	1.105	1.146	1.153	1.160	1.167	1.173	1.181	1.179	1.234
1993	1.205	1.100	1.221	1.309	1.318	1.327	1.301	1.273	1.233	1.236	1.248	1.354
1994	1.220	1.303	1.239	1.243	1.261	1.274	1.299	1.340	1.364	1.374	1.410	1.365
1995	1.344	1.283	1.313	1.323	1.306	1.280	1.297	1.273	1.258	1.257	1.245	1.237
1996	1.192	1.281	1.213	1.198	1.197	1.199	1.163	1.171	1.187	1.182	1.176	1.201
1997	1.200	1.124	1.143	1.136	1.130	1.130	1.130	1.118	1.099	1.116	1.108	1.089

样本为1982～1995年年均值数据建模得到GMDH的一水平最优方程为：

$$Q_t=2.583-0.294\,2\,Q_{t-2}-0.414\,3Q_{t-4}-0.486\,9\,Q_{t-5} \tag{8}$$

这里选取ε=0.027 1(调整ε 的值直到选出最优方程)，样本为1982～1995年月均值与年均值数据，同时建模得到GMDH调和两水平最优方程为：

$$q_t = -0.058 + 0.574\sin(1.226\, q_{t-2}) - 0.483\cos(1.226\, q_{t-2}) + 0.206\cos(0.948\, q_{t-3}) + 0.049\cos(1.874\, q_{t-5}) + 0.601 Q_T - 0.031 Q_{T-1} \tag{9}$$

这里选取 m=3，B=0.2，ε=10 (调整ε 的值直到选出最优方程)，预测效果如表2所示。

表2　1996～1997 年预测结果与实际值

		1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1996年	实际值	1.192	1.281	1.213	1.198	1.197	1.199	1.163	1.173	1.187	1.182	1.176	1.201
	预测值	1.168	1.212	1.201	1.203	1.198	1.202	1.198	1.202	1.199	1.202	1.200	1.202
	误差	0.024	0.069	0.012	0.005	0.001	0.003	0.035	0.029	0.012	0.020	0.024	0.001
1997年	实际值	1.198	1.124	1.143	1.136	1.130	1.130	1.130	1.118	1.099	1.116	1.108	1.089
	预测值	1.161	1.162	1.135	1.143	1.126	1.135	1.125	1.135	1.127	1.135	1.129	1.134
	误差	0.037	0.038	0.008	0.007	0.004	0.005	0.005	0.017	0.028	0.019	0.021	0.045

虽然直接确定物价的月均值为输出变量，它的滞后数据为输入变量。但从表中可以看出，预测了两年24个月的值，效果令人满意，说明这种方法对月均值的物价指数的预测有一定的有效性。这一方法可广泛的用于预测其他指标的月均值。

参 考 文 献

[1] 田益祥. 中长期预测模型的GMDH不同水平算法优势比较及实证分析[J]. 预测. 1999, 18(6): 73-75

[2] 刘光中. 王　绥. 自组织方法中准则的抗干扰性[J]. 系统工程理论与实践. 1995,15(11): 67-72

[3] 田益祥. 中长期预测模型的GMDH两水平算法改进及实证分析[J]. 系统工程理论与实践, 2002, 22(8): 73-78

[4] Rorbert S P, Daniel L R. econometric models and econometric Forecasts[M]. The McGraw-Hill Companies. Inc., 1998

[5] William H G. Econometric Analysis[M]. New Jersey: Prentice-Hall.Inc., 1997

[6] 田益祥. 中长期预测模型的GMDH两水平算法及收敛性[J]. 系统工程理论方法应用, 1997, 6(4): 72-76

[7] 田益祥, 章军祥, 李红松. GMDH调和算法预测模型及其应用[J]. 数量经济技术经济研究, 2000, 17(8): 66-68

[8] Madala H R,Ivakhnenko A G. Inductive learning Athgorithm of complex systems modeling[M]. CRC: Press. Inc., 1994

编　辑　熊思亮