

时间序列预测的研究与实现

陈 崢 杉

(闽发证券有限责任公司 福建 福州 350003)

【摘 要】 目前人们对经济上的时间序列预测越来越感兴趣。例如在股票交易中,人们希望从历史的交易数据中预测出未来一段时间里股票的走势,更好地指导投资者进行投资。本文在分析时间序列预测相关概念的基础上,研究了两种预测算法即移动平均法和指数平滑法的基本思想及实现方法,并结合实例阐述了它们的具体应用。

【关键词】 时间序列预测 移动平均法 指数平滑法

1 引言

时间序列是指同一种现象在不同时间上的相继观察值排列而成的一组数字序列。时间序列预测方法的基本思想是:预测一个现象的未来变化时,用该现象的过去行为来预测未来。即通过时间序列的历史数据揭示现象随时间变化的规律,将这种规律延伸到未来,从而对该现象的未来做出预测。

现实中的时间序列的变化受许多因素的影响,有些起着长期的、决定性的作用,使时间序列的变化呈现出某种趋势和一定的规律性;有些则起着短期的、非决定性的作用,使时间序列的变化呈现出某种不规则性。时间序列的变化大体可分解为以下四种:

(1)长期或趋势变化(long-term or trend movement),指现象随时间变化朝着一定方向呈现出持续稳定地上升、下降或平稳的趋势。

(2)循环变动或循环变化(cyclic movement or cyclic variation),指现象按不固定的周期呈现出的波动变化。

(3)季节性变动或季节性变化(seasonal movement or seasonal variation),指现象受季节性影响,按一固定周期呈现出的周期波动变化。

(4)非规则或随机变化(irregular or random movement),指现象受偶然因素的影响而呈现出的不规则波动。

时间序列一般是以上几种变化形式的叠加或组合。时间序列预测方法分为两大类:一类是确定型的时间序列模型方法;另一类是随机型的时间序列分析方法。确定型时间序列预测方法的基本思想是用一个确定的时间函数 $y = f(t)$ 来拟合时间序列,不同的变化采取不同的函数形式来描述,不同变化的叠加采用不同的函数叠加来描述。具体可分为趋势预测法、平滑预测法、分解分析法等。随机型时间序列分析方法的基本思想是通过分析不同时刻变量的相关关系,揭示其相关结构,利用这种相关结构来对时间序列进行预测。本文讨论的时间序列预测法指的是确定型时间序列模型方法。

2 时间序列预测的相关概念

时间序列就是某一系统在不同时间(地点、条件等)的响应。这个定义从系统运行的观点出发,不仅指出时间序列是按一定顺序排列而成的;这里的“一定顺序”既可以是时间顺序,也可以是具有各种不同意义的物理量,如代表温度、速度或其他单调递增的取值的物理量。

在市场预测中,经常遇到按时间排列的统计数据,如按月份、季度和年度统计的数据,称为时间序列。时间序列预测就是通过对预测目标本身时间序列的处理,研究预测目标的变化趋势。

用时间序列预测法进行预测必须具有以下条件:

一是预测变量的过去、现在和将来的客观条件基本保持不变,历史数据解释的规律可以延续到未来。

二是预测变量的发展过程是渐变的,而不是跳跃式的或大起大落的。

股票价格的变动可以作为时间序列。金融时间序列特指各种不同金融产品的时间序列,比如汇率、基金、股票价格等。对于期货和股票市场,常用的趋势预测技术是移动平均法和指数平滑法,或由此衍生的其他常规技术指标分析方法。移动平均法是在算术平均数的基础上发展起来的预测方法,指数平滑法则以加权平均为基础,是移动平均的一种改进,两者均属时间序列下的预测方法,假设预测对象的变化仅与时间有关。根据它的变化特征,以惯性原理推测其未来状态。

表 1 是中国证券 A 股市场上证指数及深证成指近年来按交易时间排好序的月线收盘价,取自 2003 年 1 月 31 日到 2004 年 4 月 30 日,共 16 条记录。

表 1 A 股市场上证指数及深证成指月线收盘价

交易时间	上证指数 1A0001	深证成指 399001
2003.01	1499.81	3051.22
2003.02	1511.93	3057.24
2003.03	1510.58	3101.86
2003.04	1521.44	3260.75
2003.05	1576.26	3503.69
2003.06	1486.02	3224.81
2003.07	1476.74	3318.42
2003.08	1421.98	3192.37
2003.09	1367.16	3054.89
2003.10	1348.3	3167.34
2003.11	1397.22	3267.94
2003.12	1497.04	3479.8
2004.01	1590.73	3716.87
2004.02	1675.07	3919.92
2004.03	1741.62	4060.14
2004.04	1595.59	3613.17

3 时间序列预测的移动平均法算法描述及应用

3.1 移动平均法算法的描述

移动平均法是以过去某一段时期的数据平均值作为将来某时期预测值的一种方法。该方法按对过去若干历史数据求算术平均数,并把该数据作为以后时期的预测值。

(1)预测模型:

$$x_t = T + R_t \quad \text{或} \quad x_t = TR_t$$

假定 R_t 存在正自相关

对于时间数列,项之间距离越近,相关性越强,所以在预测中应更加重视近期数据,根据这种思想我们采用移动平均法:首先选定一个自然数 N ,预测某一期的值时,只利用最近 N 期的数据取平均,而把更糟的数据弃之不用, N 称为其步长。

设给定时间序列观察值 $x_1, x_2, \dots, x_t (t \geq N)$, 移动平均法的基本公式:

$$M_t = \frac{x_{t-N+1} + x_{t-N+2} + \dots + x_t}{N}$$

预测方程: $\hat{x}_{t+1} = M_t$

M_t 称为第 t 期的步长为 N 的移动平均值, \hat{x}_{t+1} 为第 $t+1$ 期的预测值。移动平均法就是以第 t 期的移动平均值作为第 $t+1$ 的预测值。

(2) N 的选择:

采用移动平均法进行预测,用来求平均数的时期数 N 的选择非常重要。这也是移动平均的难点。事实上,不同 N 的选择对所计算的平均数是有较大影响的。 N 值越小,表明对近期观测值预测的作用越重视,预测值对数据变化的反应速度也越快,但预测的修匀程度较低,估计值的精度也可能降低。反之, N 值越大,预测值的修匀程度越高,但对数据变化的反映程度较慢。因此, N 值的选择无法二者兼顾,应视具体情况而定。

不存在一个确定时期 N 值的规则。 N 一般在 $3 \sim 200$ 之间,视序列长度和预测目标情况而定。一般对水平型数据, N 值的选取较为随意;一般情况下,如果考虑到历史上序列中含有大量随机成分,或者序列的基本发展趋势变化不大,则 N 应取大一点。对于具有趋势性或阶跃型特点的数据,为提高预测值对数据变化的反应速度,减少预测误差, N 值取较小一些,以使移动平均值更能反映目前的发展变化趋势。

3.2 移动平均法算法的应用

取 $N=5$,用程序分别对表 1 数据计算出其移动平均法预测值,结果如表 2 所示:

表 2 上证指数及深证成指月线收盘价移动平均法预测值

交易时间	上证指数	预测值	深证成指	预测值
2003.01	1499.81		3051.22	
2003.02	1511.93		3057.24	
2003.03	1510.58		3101.86	
2003.04	1521.44		3260.75	
2003.05	1576.26		3503.69	
2003.06	1486.02	1524.006	3224.81	3194.953
2003.07	1476.74	1521.248	3318.42	3229.67
2003.08	1421.98	1514.21	3192.37	3281.906
2003.09	1367.16	1496.491	3054.89	3300.007
2003.10	1348.3	1465.634	3167.34	3258.835
2003.11	1397.22	1420.043	3267.94	3191.565
2003.12	1497.04	1402.283	3479.8	3200.191
2004.01	1590.73	1406.343	3716.87	3232.467
2004.02	1675.07	1440.092	3919.92	3337.367
2004.03	1741.62	1501.674	4060.14	3510.372
2004.04	1595.59	1580.338	3613.17	3688.933

3.3 移动平均法算法的分析

移动平均法有着简单易行、容易掌握的主要优点,但它存在下列不足:

(1) 只是在处理水平型历史数据时才相对有效,每计算一次移动平均需要最近的 n 个观测值。而在现实经济生活中,历史数据的类型远比水平型复杂,使移动平均法的应用范围受到较大限制。

(2) 在大多数情况下只用于以月度或周为单位的近期预测,因此比较适用于短期预测。

4 时间序列预测的指数平滑法算法描述及应用

4.1 指数平滑法算法的描述

指数平滑法认为,数据的重要程度按时间上的近远成非线性递减。指数平滑法又称指数加权平均法,实际是加权的移动

平均法,它是选取各时期权重数值为递减指数数列的均值方法。指数平滑法解决了移动平均法需要几个观测值和不考虑 $t-n$ 前时期数据的缺点,通过某种平均方式,消除历史统计序列中的随机波动,找出其中主要的发展趋势。

(1) 常数模型指数平滑法的一般公式:

$$s_t = \alpha x_t + (1-\alpha)s_{t-1} \quad (4-1)$$

$$\hat{x}_t(l) = s_t$$

其中, s_t -----第 t 期平滑值;

α -----平滑系数(取值范围 $0 < \alpha < 1$);

x_t -----第 t 期实际观察值。

将

$$s_{t-1} = \alpha x_{t-1} + (1-\alpha)s_{t-2}$$

$$s_{t-2} = \alpha x_{t-2} + (1-\alpha)s_{t-3}$$

代入公式(4-1)可得:

$$S_t = \sum_{i=0}^{t-1} \alpha(1-\alpha)^i x_{t-i} \quad (4-2)$$

公式(4-2)中各项系数和为:

$$\alpha + \alpha(1-\alpha) + \dots + \alpha(1-\alpha)^{t-1} + (1-\alpha)^t = \alpha \left[\frac{1-(1-\alpha)^t}{1-(1-\alpha)} \right] + (1-\alpha)^t$$

当 $t \rightarrow \infty$ 时, $(1-\alpha)^t \rightarrow 0$, 系数和 $\rightarrow 1$ 。

所以,可以说 s_t 是 t 期以及以前各期观察值的指数加权平均值,观察值的权数按递推周期以几何级数递减,各期的数据离第 t 期越远,它的系数愈小,因此它对预测值的影响也越小。

公式(4-1)稍作变换可得:

$$s_t = s_{t-1} + \alpha(x_t - s_{t-1}) \quad (4-3)$$

如果用平滑值 s_t 作为下一期的预测值,则:

$$\hat{x}_t(l) = \hat{x}_{t-1}(l) + \alpha[x_{t-1} - \hat{x}_{t-1}(l)] = \hat{x}_{t-1}(l) + \alpha e_t$$

其中, e_t 为一期预测误差。因此,简单指数平滑法用于预测实际上是根据本期预测误差对本期预测值作出一定的调整后得到的下一个预测值,即:

新的预测值 = 老的预测值 + $\alpha \times$ 老预测值的误差

对老预测值所作的调整的幅度视 α 的大小而定, α 越大,所作的调整也越大, α 越小,所作的调整也越小。

(2) 平滑系数: α 的取值对平滑效果影响很大, α 越小平滑效果越显著。 α 取值的大小决定了在平滑值中起作用的观察值的项数的多少,当 α 取值较大时,各观察值权数的递减速度很快,因此在平滑值中起作用的观察值的项数就较少;而当 α 取值较小时,各观察值权数的递减速度很慢,因此在平滑值中起作用的观察值的项数就较多。

如果用移动平均数与指数平滑法相比,要使两者具有相同的灵敏程度,移动平均数 N 的取值与指数平滑法中 α 的取值有如下关系:

$$\frac{N-1}{2} = \frac{1-\alpha}{\alpha}$$

当 α 取值 $0.05 \sim 0.3$ 之间时,如果要使移动平均具有相应的灵敏程度,则 N 的取值为:

α	0.05	0.1	0.2	0.3
N	39	19	9	5.66 \approx 6

可见 α 取值大相当于移动平均法中 N 的取值小,反之则为 N 取值大。因此,当 α 取值较小时,指数平滑法的平滑能力较强,而 α 取值较大时,模型对现象变化的反应速度较快。一般来说 α 取值的大小应当视所预测对象的特点及预测期的长短而定。一般情况下,观测值呈较稳定的水平发展, α 值取 $0.1 \sim 0.3$ 之间;观测值波动较大时, α 值取 $0.3 \sim 0.5$ 之间;观测值呈波动很大时, α 值取 $0.5 \sim 0.8$ 之间。

4.2 指数平滑法算法的应用

取 $\alpha=0.2$,用程序分别对表 1 数据计算出其指数平滑法预测值,结果如表 3 所示:

表 3 上证指数及深证成指月线收盘价指数平滑法预测值

交易时间	上证指数	预测值	深证成指	预测值
2003.01	1499.81		3051.22	
2003.02	1511.93	1499.81	3057.24	3051.22
2003.03	1510.58	1502.23	3101.86	3052.42
2003.04	1521.44	1503.9	3260.75	3062.31
2003.05	1576.26	1507.41	3503.69	3102
2003.06	1486.02	1521.18	3224.81	3182.34
2003.07	1476.74	1514.15	3318.42	3190.83
2003.08	1421.98	1506.67	3192.37	3216.35
2003.09	1367.16	1489.73	3054.89	3211.55
2003.10	1348.3	1465.22	3167.34	3180.22
2003.11	1397.22	1441.84	3267.94	3177.64
2003.12	1497.04	1432.92	3479.8	3195.7
2004.01	1590.73	1445.74	3716.87	3252.52
2004.02	1675.07	1474.74	3919.92	3345.39
2004.03	1741.62	1514.81	4060.14	3460.3
2004.04	1595.59	1560.17	3613.17	3580.27

参考文献

- [1] Jiawei Han and Micheline Kamber 著, 范明等译《数据挖掘概念与技术》, 机械工业出版社, 2001 年。
- [2] 宁宣熙、刘思峰著《管理预测与决策方法》, 科学出版社, 2003 年。
- [3] G. E. P. Box and G. M. Jenkins 著, 顾岚等译:《时间序列分析预测与控制》, 中国统计出版社, 1997 年。
- [4] M. P. Niemira and P. A. Klein 著, 邱东等译:《金融与经济周期预测》, 中国统计出版社, 1997 年。
- [5] 董文泉、高铁梅、姜诗章和阵磊著,《经济周期波动的分析与预测方法》, 吉林大学出版社, 1998 年。
- [6] Robert S. Pindyck and Daniel L. Rubinfeld 著,《计量经济模型与经济预测》, 机械工业出版社, 1999 年。

(上接第 21 页)

3.4 可重用性

重用性是软件产业发展的基础, 提供什么样的重用机制和在哪个层次上提供重用, 对软件的发展有不同的意义。组件化程序设计方法比传统面向对象方法提供了更高层次的重用性。

在传统的面向对象程序设计方法中, 对对象的重用表现在源代码一级的重用性上, 通过面向对象程序设计语言的类继承实现。派生类可以继承基类的非私有数据成员和函数成员, 但这使得派生类和基类具有较强的耦合关系。基类的改动要求派生类必须重新编译或修改相应的代码, 才能适应新的基类。所以在最终的可执行代码中, 基类和派生类在同一模块中, 重用性只体现在程序模块的内部; 对模块外部而言, 重用性只体现在对代码的有效管理上。在面向对象方法中的重用还要求对象程序和使用对象的客户程序使用同样的编程语言, 如某个类库是用 C++ 编写的, 那么用其他语言编写的应用程序基本上不可能重用该代码。

组件化程序设计方法中, 对组件的重用性表现在二进制代码一级的重用性上。组件软件系统由可重用的二进制形式的软件组件模块组成, 只需要相当小的改动就可以将这些来自不同软件开发商的组件模块组合在一起。特别重要的是这样的组合并不需要源代码, 也不需要重新编译, 组件之间通过基于

4.3 指数平滑法算法的分析

常数模型指数平滑法是在移动平均法的基础上发展来的。移动平均法在用于预测时有以下主要弱点: 一是计算移动平均数需要存贮最近的几个数据, 在进行大量项目预测时, 这需要占用很大的计算机存贮空间; 其次, 计算移动平均数对最近期的几个数值加以相同的权数, 然而最近期的数值应当包含了最多关于未来可能发生情况的信息, 因此应当给最近期的观察值加上最大的权数。指数平滑法能满足这种要求, 同时还能减少所需存贮空间。指数平滑法是把所有历史数据都予以保留。平均时赋予每个数据一个权重, 这些权重彼此不同, 随着时间的向前推移, 这些权重单调地减少, 逐渐趋向于零。该法合理地体现了重视近期数据的思想。

5 结束语

时间序列预测法是一种重要的预测方法, 其预测模型都比较简单, 它对资料的要求比较单一, 只需变量本身的历史数据, 因此, 在实际情况中有着广泛的适用性。时间序列预测可以对客户购买行为模式、Web 访问模式、疾病诊断、天气情况、DNA 序列等进行预测, 从而提供一定程度的决策支持, 在生产、销售、金融、保险、医学、气象分析等领域发挥巨大的经济和社会效益。

二进制的规范进行通讯。组件模块是独立于编程语言和开发环境的, 使用组件的客户程序和组件之间除了通过标准的接口进行通讯以外, 彼此不做任何限定, 从而组件拥有更高层次的重用性。

3.5 多态性

同一消息为不同的对象接受时, 可导致完全不同的行为, 这就是面向对象中的多态性, 它利用函数名重载技术实现。

在组件化程序设计中, 组件对象也具有多态性, 它的多态性通过接口来体现。组件对象的多态性使得客户程序可以用统一的方法处理不同的组件对象, 甚至是不同类型的组件对象, 只要它们实现了同样的接口。由于组件的接口遵循一定的标准, 支持相同接口的不同组件可以互换使用, 所以可以很方便的建立插件系统。

4 结束语

组件化程序设计方法继承并发展了面向对象的程序设计方法, 它将对象技术应用于系统, 对面向对象程序设计的实现过程作了进一步的抽象。组件化程序设计方法是构造应用系统体系结构层次的有效方法, 而使用面向对象程序设计方法则可以很方便地实现组件。因此我们在构造实际的软件应用系统时, 要综合运用这两种程序设计方法。

参考文献

- [1] 潘爱民. COM 原理与应用. 北京: 清华大学出版社, 1999.
- [2] [美] Don Box. COM 本质论. 潘爱民译. 北京: 中国电力出版社, 2001.
- [3] 杨晓红, 朱庆生. 组件化程序设计方法及组件标准. 重庆大学学报(自然科学版), 2001(6).