

文章编号:1000-8055(2012)11-2529-05

确定时间序列均值和方差函数的 移动多点平均方法

王治华, 傅惠民, 张勇波

(北京航空航天大学 小样本技术研究中心 北京 100191)

摘 要: 为了提高有限样本或小样本情况下时间序列均值和方差函数的确定精度, 以保证时序建模、分析和预测精度, 提出一种确定序列趋势项的移动多点平均方法. 该方法能够得到时间序列均值中非周期部分, 结合样本周期图法得到的周期项, 可综合得到其均值函数, 并可进一步得到时序的标准差函数. Monte Carlo 模拟对比结果表明: 该方法不仅能够保证时间序列段内分析精度, 而且能够有效提高时间序列的预报精度.

关 键 词: 时间序列; 均值函数; 方差函数; 移动多点平均; 预测

中图分类号: V448; O231

文献标志码: A

Moving multi-point average method for determining functions of time series mean and variance

WANG Zhi-hua, FU Hui-min, ZHANG Yong-bo

(Research Center of Small Sample Technology,

Beijing University of Aeronautics and Astronautics,

Beijing 100191, China)

Abstract: A moving multiple-point average (MMPA) method was proposed to enhance the determining precision for time series mean and variance functions, and consequently to improve the accuracy for time series modeling, analyzing and forecasting. Non-periodic part of trend item which can be obtained by the MMPA approach, and the periodic part of trend item which can be calculated by sample periodogram method, comprise the whole trend function. We can further determine the series standard deviation function. Monte Carlo simulation study shows that not only the analysis precision but also the prediction accuracy can be ensured by the established methodology.

Key words: time series; mean function; variance function;
moving multi-point average; prediction

时序分析方法已成为众多领域处理工程实际问题的有效手段. 传统平稳时序理论在协方差平稳性假设这一前提条件下, 系统建立了时序分析和预测方法. 然而, 具有时变均值和方差的非平稳

时间序列在故障诊断、自动控制、信号处理、性能测试, 以及经济等领域非常常见, 其分析与处理长期以来一直受到国内外的广泛重视, 也是当前学术界研究的热点问题之一^[1].

收稿日期: 2012-08-28

网络出版地址: <http://www.cnki.net/kcms/detail/11.2297.V.20121115.1028.019.html>

基金项目: 国家自然科学基金(11202011); 国家重点基础研究发展计划(2012CB720000); 凡舟青年科学基金(20100511)

作者简介: 王治华(1981—), 女, 辽宁辽阳人, 讲师、硕士生导师, 博士, 主要从事可靠性工程、小样本技术研究.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

平稳性假设是建立线性模型的基础,但当时间序列具有时变方差时,其处理结果将使模型参数失去最小方差特性,且方差估计不再具有无偏性^[2]. Engle 于 1982 年提出了条件异方差的概念,其方法在金融风险评估中得到广泛应用,解决了无条件方差为常数的时间序列的条件异方差估计问题^[3]. 尽管 Engle 的成果于 2003 年获得诺贝尔经济学奖,但其方法仍无法解决方差本身随时间变化的时间序列的分析与处理问题^[4-7].

在大量随机过程(序列、信号)实测数据的研究基础上,文献[8]提出相关系数平稳随机过程的概念,文献[9-11]进一步建立相关系数平稳序列模型,文献[12-15]提出广义时变 AR(auto regression), MA(moving average) 和 ARMA(auto regressive moving average) 模型及其参数估计方法,并建立广义时变序列预测方法和回归与时变自回归模型. 上述非平稳时序处理方法需要已知均值和方差的函数形式,在一些问题中他们可根据理论分析、物理性质和实际经验获得,但大多数情况事先无法知道均值和方差的函数形式. 为此,文献[16]提出一种确定时间序列均值和方差函数形式的方法. 为了提高时间序列均值和方差函数的确定精度,特别是当时间序列样本量有限或较小时有效提高精度,本文将进一步建立一种移动多点平均方法.

1 确定时序方差函数的定理

定理 设 $\{x_t\}$ 为相关系数平稳序列, $\mu(t) = E(x_t)$ 和 $\sigma^2(t) = \text{Var}(x_t)$ 分别为其均值和方差函数,若 $\frac{\Delta\sigma(t)}{\sigma(t)}$ 为常数或 $\left[\frac{\Delta\sigma(t)}{\sigma(t)}\right]_{\max}$ 与 1 相比是可以忽略不计的高阶小量,即 $\left[\frac{\Delta\sigma(t)}{\sigma(t)}\right]_{\max} \ll 1$, 其中 $\left[\frac{\Delta\sigma(t)}{\sigma(t)}\right]_{\max}$ 为 $\frac{\Delta\sigma(t)}{\sigma(t)}$ 的最大值, $\Delta\sigma(t) = \sigma(t+1) - \sigma(t)$, $t = 1, 2, \dots, n-1$. 那么,序列 $\{x_t\}$ 的标准差函数 $\sigma(t)$ 与其差分绝对值序列 $\{|\nabla x_t|\}$ 的均值(趋势项)函数仅相差一个常数倍,其中

$$\nabla x_t = x_{t+1} - \mu(t+1) - x_t + \mu(t)$$

即

$$\sigma(t) = cE|x_{t+1} - \mu(t+1) - x_t + \mu(t)| \quad (1)$$

其中 c 为一个正实数.

证明 根据相关系数平稳序列的定义, x_t 可

表示为

$$x_t = \mu(t) + \sigma(t)\varepsilon_t \quad \varepsilon_t \sim N(0, 1)$$

则

$$\begin{aligned} E|x_{t+1} - \mu(t+1) - x_t + \mu(t)| &= \\ E|\sigma(t+1)\varepsilon_{t+1} - \sigma(t)\varepsilon_t| &= \\ \sigma(t)E\left|\frac{\sigma(t) + \Delta\sigma(t)}{\sigma(t)}\varepsilon_{t+1} - \varepsilon_t\right| &= \\ \sigma(t)E\left|1 + \frac{\Delta\sigma(t)}{\sigma(t)}\varepsilon_{t+1} - \varepsilon_t\right| \end{aligned}$$

令随机变量

$$\xi_t = \left[1 + \frac{\Delta\sigma(t)}{\sigma(t)}\varepsilon_{t+1} - \varepsilon_t\right]$$

则 ξ_t 仍服从正态分布,其均值为

$$E(\xi_t) = E\left[1 + \frac{\Delta\sigma(t)}{\sigma(t)}\varepsilon_{t+1} - \varepsilon_t\right] = 0$$

方差为

$$\begin{aligned} \text{Var}(\xi_t) &= E\left\{\left[1 + \frac{\Delta\sigma(t)}{\sigma(t)}\varepsilon_{t+1} - \varepsilon_t\right]^2\right\} = \\ &\left(1 + \frac{\Delta\sigma(t)}{\sigma(t)}\right)^2 + 1 - 2\rho_1\left(1 + \frac{\Delta\sigma(t)}{\sigma(t)}\right) \end{aligned}$$

因此,当 $\frac{\Delta\sigma(t)}{\sigma(t)}$ 为常数或 $\left[\frac{\Delta\sigma(t)}{\sigma(t)}\right]_{\max}$ 与 1 相比是可以忽略不计的高阶小量时, $\text{Var}(\xi_t)$ 为常数,且 $E|\xi_t| = E\left|1 + \frac{\Delta\sigma(t)}{\sigma(t)}\varepsilon_{t+1} - \varepsilon_t\right| = \sqrt{\frac{2}{\pi}\text{Var}(\xi_t)}$ 为常数,令 $c = 1/\sqrt{\frac{2}{\pi}\text{Var}(\xi_t)}$, 可知式(1)成立.

当上述定理条件不能满足时,可采用下式确定序列 $\{x_t\}$ 的方差函数

$$\sigma(t) = c_1E|x_t - \mu(t)| \quad (2)$$

式(2)中 c_1 为一正实数. 式(2)的证明与式(1)相同,这里不再赘述.

2 确定时序趋势项的方法

相关系数平稳序列 $\{x_t\}$ 的均值函数 $\mu(t)$ 是其本身的趋势项,标准差函数 $\sigma(t)$ 可根据差分绝对值序列 $\{|\nabla x_t|\}$ 的趋势项求得, $t = 1, 2, \dots, n$.

因此,可首先根据序列 $\{x_t\}$ 的一个样本 $x_1, x_2, \dots, x_n$, 确定序列 $\{x_t\}$ 的趋势项估计值 $\hat{\mu}(t)$; 然后,据此趋势项估计结果得到差分绝对值序列 $\{|\nabla x_t|\}$ 的一个样本 $|\nabla x_t| = |x_{t+1} - x_t - \hat{\mu}(t+1) + \hat{\mu}(t)|$, $t = 1, 2, \dots, n-1$, 在此基础上根据定理结论确定序列 $\{|\nabla x_t|\}$ 的趋势项,从而得到标准差函数 $\sigma(t)$.

对于趋势项为周期函数的情况,可根据文献[16]中周期趋势项的样本周期图方法,可得到时

间序列 $\{x_t\}$ 趋势项中的周期部分

$$x_t = \bar{x} + \sum_{j=1}^M \{ \alpha_j \cos[\omega_j(t-1)] + \beta_j \sin[\omega_j(t-1)] \} \quad t = 1, 2, \dots, n \tag{3}$$

其中参数 $\bar{x}$, M , α_j , β_j 和 ω_j 可根据文献[16]中的方法得到.

对于趋势项为非周期函数的情况,可根据序列 $\{x_t\}$ 样本 $x_1, x_2, \dots, x_n$ 的波动情况,设置各次多点平均的点数 m 及移动间隔 Δ ,进行 m 点移动平均

$$\bar{t}_i = \frac{1}{m} \sum_{t=\Delta i-1}^{\Delta i+m-2} t \quad i = 1, 2, \dots, l \tag{4}$$

$$\bar{x}_i = \frac{1}{m} \sum_{t=\Delta i-1}^{\Delta i+m-2} x_t \quad i = 1, 2, \dots, l \tag{5}$$

得到 1 组平均值 $(\bar{t}_i, \bar{x}_i)$, $i = 1, 2, \dots, l$. 其中 $l = \lceil \frac{n-m}{\Delta} \rceil + 1$, “ $\lceil \cdot \rceil$ ”为取整函数,通常选取的 m 可为数据总量 n 的 $1/50 \sim 1/5$, Δ 为多点平均点数 m 的 $1/10 \sim 1/2$.

然后,对该组平均值 $(\bar{t}_i, \bar{x}_i)$, $i = 1, 2, \dots, l$, 进行拟合,得到拟合函数 $f(t)$, 即为序列 $\{x_t\}$ 的趋势项.

通常,趋势项中既有非周期函数,又有周期函数,此时可按下面步骤确定序列 $\{x_t\}$ 的趋势项.

1) 采用样本周期图方法对样本 $x_1, x_2, \dots, x_n$ 进行分析,得到序列 $\{x_t\}$ 趋势项中的周期函数

$$\sum_{j=1}^M \{ \alpha_j \cos[\omega_j(t-1)] + \beta_j \sin[\omega_j(t-1)] \}$$

2) 对样本

$$x_t - \sum_{j=1}^M \{ \alpha_j \cos[\omega_j(t-1)] + \beta_j \sin[\omega_j(t-1)] \} \quad t = 1, 2, \dots, n$$

采用移动多点平均方法求得序列 $\{x_t\}$ 趋势项中的非周期函数 $f(t)$. 此时, $m = \lceil \frac{2\pi}{\omega_{\min}} \rceil$, 其中 $\omega_{\min}$ 为诸频率 ω_j 中的最小频率.

3) 对样本 $x_t - f(t)$, $t = 1, 2, \dots, n$, 采用样本周期图方法进行分析,求得序列 $\{x_t\}$ 趋势项中的周期函数

$$\sum_{j=1}^M \{ \alpha_j \cos[\omega_j(t-1)] + \beta_j \sin[\omega_j(t-1)] \}$$

4) 重复步骤 2) 和步骤 3), 直到非周期函数和周期函数中的各参数计算值稳定或满足一定的精度要求为止,最终得到序列 $\{x_t\}$ 的趋势项

$$\hat{\mu}(t) = f(t) +$$

$$\sum_{j=1}^M \{ \alpha_j \cos[\omega_j(t-1)] + \beta_j \sin[\omega_j(t-1)] \}$$

由上述方法可得到序列 $\{x_t\}$ 的趋势项估计值 $\hat{\mu}(t)$; 然后,可以得到差分绝对值序列 $\{|\nabla x_t|\}$ 的一个样本

$$|\nabla x_t| = |x_{t+1} - x_t - \hat{\mu}(t+1) + \hat{\mu}(t)| \quad t = 1, 2, \dots, n-1$$

将上面步骤 1)~步骤 4) 中的样本 x_t , $t = 1, 2, \dots, n-1$, 换成差分绝对值序列样本

$$|\nabla x_t| = |x_{t+1} - x_t - \hat{\mu}(t+1) + \hat{\mu}(t)| \quad t = 1, 2, \dots, n-1$$

由上述方法即可得到差分绝对值序列 $\{|\nabla x_t|\}$ 的趋势项,即序列 $\{x_t\}$ 的标准差函数.

3 计算机仿真验证

大量 Monte Carlo 模拟验证表明,与传统的多点平均方法相比,本文的移动多点平均方法能够有效提高预测精度. 下面以一具体算例说明本文方法的建模和预测效果. 图 1 给出 1 组 100 个相关系数 AR(1)序列模拟数据 x_t , $t = 1, 2, \dots, 100$, 其一维分布为正态分布,其中

$$\mu(t) = 1.5t^{1.5} \tag{6}$$

$$\sigma(t) = 12 + 0.05t + 0.001t^2 \tag{7}$$

相关系数 AR(1)序列模型为

$$\frac{x_t - \mu(t)}{\sigma(t)} = 0.8 \frac{x_{t-1} - \mu(t-1)}{\sigma(t-1)} + \epsilon_t \quad \epsilon_t \sim N(0, 1) \tag{8}$$

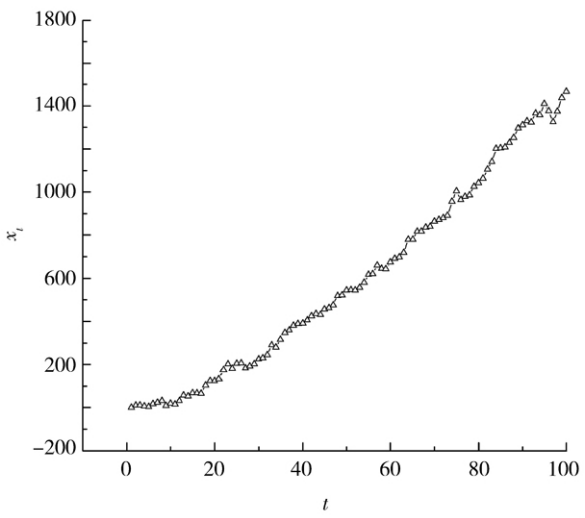


图 1 相关系数 AR(1)序列模拟数据

Fig. 1 Simulating value for correlation coefficient AR(1) series

采用本文移动多点平均方法对图 1 中数据进行处理,设多点平均的点数 $m = 20$ 及移动间隔 $\Delta = 10$,得到相关系数平稳序列 $\{x_t\}$ 和其差分绝对值序列的趋势项分别为

$$\mu(t) = 1.482\,907t^{1.501\,532} \tag{9}$$
$$\hat{\sigma}(t) = 0.002\,149t^2 - 0.090\,710t + 12.923\,966 \tag{10}$$

进一步建立相关系数 AR(1)模型

$$\frac{x_t - \mu(t)}{\sigma(t)} = 0.624\,000 \frac{x_{t-1} - \mu(t-1)}{\sigma(t-1)} + \epsilon_t$$
$$\epsilon_t \sim N(0, 1.446\,392) \tag{11}$$

采用多点平均法对图 1 中数据进行处理,设多点平均的点数 $m = 20$,得到相关系数平稳序列 $\{x_t\}$ 和其差分绝对值序列 $\{|\nabla x_t|\}$ 的趋势项分别为

$$\mu(t) = 1.466\,913t^{1.503\,825} \tag{12}$$
$$\hat{\sigma}(t) = 0.001\,498t^2 - 0.011\,197t + 11.061\,298 \tag{13}$$

进一步建立相关系数 AR(1)模型

$$\frac{x_t - \mu(t)}{\sigma(t)} = 0.720\,867 \frac{x_{t-1} - \mu(t-1)}{\sigma(t-1)} + \epsilon_t$$
$$\epsilon_t \sim N(0, 1.510\,126) \tag{14}$$

现采用两种方法得到的结果对 $x_t, t = 101, 102, \dots, 105$,进行预测,对比结果见表 1.从中可以看到,本文移动多点平均方法较传统多点平均方法相比,可有效提高建模和预测精度.

表 1 预报对比结果

Table 1 Comparison results of prediction

t	真值	本文方法	多点平均法
101	1518.13	1498.74	1470.89
102	1541.64	1527.63	1505.58
103	1604.57	1554.32	1537.12
104	1645.31	1579.66	1566.39
105	1659.11	1604.18	1594.03

4 结 论

1) 提出一种确定序列趋势项的移动多点平均方法,该方法同样可用于确定时间序列的方差函数.与传统多点平均方法相比,本文方法不仅能够保证序列段内分析精度,而且能够有效提高序列的预报精度.

2) 当序列样本容量较大且精度要求不是很

高时,可采用本文方法得到序列的趋势项和标准差,并通过 $y_t = [x_t - \hat{\mu}(t)]/\hat{\sigma}(t)$ 将相关系数平稳序列 $\{x_t\}$ 变换成相关函数平稳序列 $\{y_t\}$,然后按传统方法进行参数估计和预测,见模拟算例.

3) 当序列样本量较小时,可首先根据本文方法得到均值和方差函数,得到序列均值和方差的函数形式,然后根据文献[9-11]方法计算其中的待定参数.迭代时可将本文方法得到的参数值作为初始值.

参考文献:

[1] Hamilton J D. Time series analysis [M]. Princeton: Princeton University Press,1994.

[2] Gujarati D. Essentials of econometrics [M]. New York: McGraw Hill,1998.

[3] 罗开位,王晓刚. 2003 年度诺贝尔经济学奖评述[J]. 外国经济与管理,2003,25(11):34-38.
LUO Kaiwei, WANG Xiaogang. Commentary on the Nobel economics prize in 2003[J]. Foreign Economies and Management, 2003,25(11):34-38. (in Chinese)

[4] Ibrahim A. Nonstationarity characteristics of the S&P500 returns: an approach based on the evolutionary spectral density[J]. Economics Bulletin,2003,3(32):1-7.

[5] Ibrahim A. A complementary test for the KPSS test with an application to the US Dollar/Euro exchange rate[J]. Economics Bulletin,2004,3(4):1-5.

[6] Loretan M,Phillips P C B. Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial data sets[J]. Journal of Empirical Finance,1994,1(2):211-248.

[7] Mikosch T, Starica C. Change of structure in financial time series, long range dependence and the GARCH model [J]. The Reviw Economics and Statistics, 2004, 86 (1): 378-390.

[8] 傅惠民. 相关系数平稳过程[J]. 机械强度,2002,24(3): 400-404.
FU Huimin. Method for correlation coefficient stationary process[J]. Journal of Mechanical Strength,2002,24(3): 400-404. (in Chinese)

[9] 傅惠民,刘成瑞,马小兵. 相关系数平稳过程时频分析方法[J]. 航空动力学报,2002,17(5):505-512.
FU Huimin, LIU Chengrui, MA Xiaobing. Time frequency analysis of correlation coefficient stationary process[J]. Journal of Aerospace Power, 2002, 17 (5): 505-512. (in Chinese)

[10] 傅惠民,马小兵. 相关系数 MA(q) 序列分析方法[J]. 机械强度,2003,25(2):183-186.
FU Huimin, MA Xiaobing. Analysis method of correlation coefficient MA (q) series [J]. Journal of Mechanical Strength,2003,25(2):183-186. (in Chinese)

[11] 傅惠民,刘成瑞,马小兵. 相关系数 ARMA(p,q) 序列分析

方法[J]. 航空动力学报, 2003, 18(2): 161-166.
FU Huimin, LIU Chengrui, MA Xiaobing. Analysis method of correlation coefficient ARMA(p, q) series[J]. Journal of Aerospace Power, 2003, 18(2): 161-166. (in Chinese)

[12] 傅惠民, 王治华. 广义时变 ARMA 模型参数函数的确定方法[J]. 机械强度, 2004, 26(6): 636-641.
FU Huimin, WANG Zhihua. Method for determining parametric functions of generalized time-varying ARMA model [J]. Journal of Mechanical Strength, 2004, 26(6): 636-641. (in Chinese)

[13] 王治华, 傅惠民. 广义时变 ARMA 序列预测方法[J]. 航空动力学报, 2005, 20(5): 713-717.
WANG Zhihua, FU Huimin. Prediction method for generalized time-varying ARMA series[J]. Journal of Aerospace Power, 2005, 20(5): 713-717. (in Chinese)

[14] 王治华, 傅惠民. 回归与时变自回归模型[J]. 机械强度, 2005, 27(5): 612-615.
WANG Zhihua, FU Huimin. Method of regression and time-varying autoregression[J]. Journal of Mechanical Strength, 2005, 27(5): 612-615. (in Chinese)

[15] 王治华, 傅惠民. 时变序列分析方法[J]. 机械强度, 2006, 28(3): 636-641.
WANG Zhihua, FU Huimin. Method for time-varying series analysis[J]. Journal of Mechanical Strength, 2006, 28(3): 636-641. (in Chinese)

[16] 傅惠民, 刘成瑞, 马小兵. 时间序列均值和方差函数的确定方法[J]. 机械强度, 2004, 26(2): 164-169.
FU Huimin, LIU Chengrui, MA Xiaobing. Method for determining functions of time series mean and variance[J]. Journal of Mechanical Strength, 2004, 26(2): 164-169. (in Chinese)