

二次移动平均法中两次移动 平均步长关系研究^①

席建国

(闽南师范大学商学院)

研究目标: 获得两次移动平均步长不等的二次移动平均模型。**研究方法:** 数理推导两次移动平均步长不等的二次移动平均模型; 在样本量、截距和斜率变化时, 通过蒙特卡洛方法探讨两次移动平均步长的关系; 并实例考察模型的实用性。**研究发现:** 样本量较小时, 两次移动平均的最优步长存在等值倾向。样本量较大时, 两次移动平均的最优步长应取此大彼小或此小彼大, 倾向第一次移动平均步长大于第二次移动平均步长; 且随截距和斜率的增加, 这种趋势更加显著。**研究创新:** 构建的两次移动平均步长不等的二次移动平均模型更具一般性; 其预测精度高于或等于现有二次移动平均模型。**研究价值:** 本文构建的理论模型可以为经济、管理等领域提供更一般的基础性预测方法, 并为二次指数平滑法、二次曲线指数平滑法等方法的完善提供借鉴。

关键词 二次移动平均 移动平均步长 独立性 等值性 斜率

中图分类号 F224.0 **文献标识码** A

引言

二次移动平均法 (Double Moving Average, DMA) 是在数据呈线性递增或递减趋势时, 运用一次移动平均法所得预测值存在明显的系统性偏差, 为修正此偏差而提出的一种移动平均方法。具体而言, 当数据有线性递增趋势时, 一次移动平均法的预测值会对实际值产生系统性的低估, 进而采用第二次移动平均对预测值向上修正; 当数据有线性递减趋势时, 一次移动平均法的预测值会对实际值产生系统性的高估, 进而采用第二次移动平均对预测值向下修正。鉴于二次移动平均法原理简洁、操作简单、计算量小、结果易于解释等优点, 其被广泛应用于宏观经济、投资决策、竞技体育、生物工程、自然灾害等领域的趋势预测。梳理已有文献, 发现国内外学者对该方法进行了严谨的理论推导和大量的实际运用。特别是国内学者对二次移动平均法的研究虽然起步较晚; 但无论理论推导, 抑或实际运用, 近年来成果均较国外学者丰富。总体而言, 国内外学者关于二次移动平均法的研究主要集中于二次移动平均法的理论探讨, 宏观的经济预测和微观、行业的市场预测三大领域等。

关于二次移动平均法的理论探讨, 美国商务部下属的国家标准和技术协会 (NIST/SE-

^① 本文获得 2018 年度福建省高校杰出青年科研人才培育计划、福建省社会科学规划项目 (FJ2018X009)、福建省教育厅“商务大数据分析与应用福建省高校重点实验室”的资助。

MATECH, 2013) 在其制定的工程统计手册中, 认为二次移动平均法中的第一次移动平均步长和第二次移动平均步长应相等。Dodd (1941) 探讨了对周期数据采用二次移动平均法进行平滑时, 可能产生新的周期数据, 并推导了第一次移动平均步长和第二次移动平均步长的取值范围; 但没有指出两次移动平均的步长需要相等。田名誉和曹进 (1994) 在两次移动平均步长相等的条件下, 率先推导了二次移动平均法的计算公式, 并通过案例分析演示了该方法在投资领域的使用。华伯泉 (1995) 虽然在建模时没能摆脱两次移动平均步长相等的传统约束, 但其借助概率论的相关方法, 对二次移动平均法的计算公式进行了重新推导; 相较于田名誉和曹进 (1994) 的推导, 此方法具有明显的概率色彩; 并借助推导的公式对我国人均国民收入和我国财政收入占国民收入的比例进行了预测。Khoo 和 Wong (2008) 在 Montgomery (2005) 一次移动平均法的基础上, 提出借助二次移动平均法来刻画数据中的线性趋势。具体而言, 所构建的二次移动平均法是在第一次移动平均之后, 对一次移动平均数再进行一次与第一次移动平均等步长的移动平均。并通过数值模拟说明原始数据有中等或较大斜率时, 二次移动平均法的预测效果优于一次移动平均法。在两次移动平均步长相等的情景下, Areepong 和 Sukparungsee (2013) 探讨了零膨胀二次移动平均控制图的平均游程长度, Sukparungsee (2015) 推导了零膨胀二项式过程的二次移动平均步长的显式解。Alevizakou 和 Pantazis (2017) 比较了两次移动平均步长相等的二次移动平均法和其他预测方法的预测效果; 并通过实际数据说明, 在某些数据类型下, 二次移动平均法的预测效果较好。Phantu 等 (2018) 提出一种计算序列相关的整数型自回归泊松过程中二次移动平均控制图的显式解。从以上二次移动平均法的理论发展历程来看, 文献中基本均假定第一次移动平均步长需和第二次移动平均步长相等, 至于为何需施加这一限制条件, 则较少从理论上进行论述。

经济预测领域对二次移动平均法的使用方面, 赵益新 (1993) 在两次移动平均步长相等的情境下, 以最小均方误差准则确定两次移动平均的最优步长, 进而, 分别对全国国民生产总值、四川省国民生产总值、四川省第三产业生产总值和四川省年末总人口进行了预测。杨立勋 (1997) 认为在经济预测中, 需要正确认识移动时间项数的多少与长期趋势的变化并非成正比关系。此外, 暴奉贤和陈宏立 (2001)、白雪梅和赵松山 (2001)、张桂喜 (2003)、朱建平 (2011)、洪兴建和惠琦娜 (2011)、徐国祥 (2016) 等均指出, 当数据呈线性增加或减少时, 应采用两次移动平均步长相等的二次移动平均法来修正一次移动平均法对实际观测值预测的系统性偏差。张保法 (2004) 也认为当时间序列有线性趋势时, 宜采用二次移动平均法来修正预测偏差, 但没有认为两次移动平均的步长应相等; 而在运用二次移动平均法进行预测时, 却使用了两次移动平均步长相等的计算式。冯文权和傅征 (2018) 也说明在样本序列具有线性趋势时, 应采用二次滑动平均值预测法^①对观测值做出预测, 但所采用的仍然是第一次滑动平均步长和第二次滑动平均步长相等的方法。以上文献说明, 在经济预测方面, 对二次移动平均法的使用同样被限制在两次移动平均步长相等的条件之下。

市场预测方面有关二次移动平均法的应用, 徐阳 (2001) 认为, 对具有线性趋势的市场调查数据进行预测时, 一次移动平均数明显地存在系统性滞后 (当数据具有线性递增趋势

^① 冯文权和傅征 (2018) 所表征的“二次滑动平均值预测法”, 从预测值的计算过程考量, 就是本文表述的“二次移动平均法”。同样, 他们表述的“一次滑动平均数”, 即为本文表述的“一次移动平均数”; 他们表述的“二次滑动平均数”, 也即本文表述的“二次移动平均数”。

时)或领先(当数据具有线性递减趋势时)于实际观测值。此时,宜采用两次移动平均步长相等的二次移动平均法对预测值进行修正。同样,叶明海等(2007)在两次移动平均步长相等的情境下,借助具有线性递增趋势的数据说明,此时一次移动平均数总是滞后于实际观测值,而二次移动平均数又总是滞后于一次移动平均数,且这两个滞后量基本相等,故可以将二次移动平均数滞后于一次移动平均数的差值加到一次移动平均数上,再考虑趋势变动量,可得到比较接近实际的预测值。Areepong 和 Sukparungsee (2011)运用两次移动平均步长相等的二次移动平均法探讨了统计过程控制问题,并说明在中等趋势的线性数据下,二次移动平均法的预测效果优于二次指数平滑法和一次移动平均法。吴杨和陈兆荣(2014)也指出,在原始数据具有线性增长趋势时,需用两次移动平均步长相等的二次移动平均法来修正预测值。庄贵军(2014)同样使用第一次移动平均步长和第二次移动平均步长相等的二次移动平均法进行预测。Zakamulin (2018)以一次移动平均数和二次移动平均数的关系为基础,构建了量化交易的移动平均策略,并发现最好的情况下,移动平均策略只比买入并持有策略稍好;但从统计学的角度来看,则无法区分移动平均策略和买入并持有策略的优劣。与上述文献的表述大同小异,刘秋华(2004)、陈殿阁(2004)、周思勤和刘红霞(2005)、刘利兰(2012)、刘红霞(2007)、屈援(2007)、廖进球和李志强(2009)、丁洪福和战颂(2014)、简明(2015)、罗洪群和王青华(2016)等均采用了第一次移动平均步长和第二次移动平均步长相等的二次移动平均法。这说明,两次移动平均步长相等的二次移动平均法被广泛地应用于市场预测领域,至于为何要施加第一次移动平均步长和第二次移动平均步长相等的条件,则鲜有学者提及。

除此之外,也有学者将两次移动平均步长相等的二次移动平均法应用于其他领域。如Du等(2013)采用二次移动平均法对三峡库区的山体崩塌滑坡位移进行了预测。唐小峰(2013)运用二次移动平均法对劳动力资源的趋势进行了预测。Lian等(2014)采用二次移动平均法对山体滑坡位移进行了预测。柳瑞芝等(2017)、黄文文和肖毅(2018)等在竞技体育领域运用两次步长相等的二次移动平均法进行预测。

综上所述,国内外学者或研究机构基本认为,当数据有线性趋势时,一次移动平均法的预测值会对实际值产生系统性的偏差,此时,宜采取二次移动平均法进行预测。具体在介绍二次移动平均法的使用中两次移动平均的步长关系时,多数学者或研究机构认为应取相等的移动平均步长,仅有张保法(2004)对此可能存在异议,但其给出的计算式却是相同步长的。那么,二次移动平均法中两次移动平均的步长是否应该相等?如果不等,两者存在何种关系?进一步,观测数据的样本量、截距和斜率大小是否会对两次移动平均的步长选择产生影响?针对这些问题的研究,截至目前,尚无文献可查。为回答这些问题,本文共分为六部分:第一部分对二次移动平均法的发展历程和实际应用进行梳理;第二部分将所提问题简化为原假设及扩展假设;第三部分在两次移动平均步长不等情况下,重新推导二次移动平均法的计算公式;第四部分依据第三部分的理论推导结论,通过数值模拟对第二部分所提假设进行检验;第五部分通过一个实际案例,验证结论的合理性;第六部分进行总结。

一、二次移动平均法中两次移动平均步长的关系假设

已有文献中,一般认为二次移动平均法中的两次移动平均步长相等。这一认识有两层含义:第一次移动平均步长 N_1 和第二次移动平均步长 N_2 不独立;第一次移动平均步长 N_1 和第二次移动平均步长 N_2 完全正相关。那么,这两层含义是否具有统计上的显著性?答案

尚且未知，故作以下假设进行检验。

假设 1 二次移动平均法的两次移动平均步长 N_1 和 N_2 独立。

在假设 1 的基础上，若二次移动平均法的两次移动平均步长 N_1 和 N_2 不独立。那么，它们是否有较大概率相等？这种关系在统计上显著吗？为回答此问题，作出假设 2。

假设 2 二次移动平均法中两次移动平均步长 N_1 等于 N_2 的概率大于平均值。

倘若二次移动平均法的两次移动平均步长 N_1 和 N_2 不独立，且相等的概率不大。那么，两者的关系是 N_1 较大时， N_2 取值应较小？还是 N_1 较小时， N_2 取值应较大？^① 为检验此问题，作假设 3。

假设 3 线性数据的样本量不同时，二次移动平均法中两次移动平均步长 N_1 和 N_2 的一般选择应是： N_1 较大时， N_2 取值应较小； N_1 较小时， N_2 取值应较大。

在假设 1～假设 3 的基础上，线性数据的截距大小是否影响两次移动平均的步长选取？同样，线性数据的斜率大小是否影响两次移动平均的步长选择？为回答这两个问题，分别作假设 4 和假设 5。

假设 4 线性数据的截距越大，二次移动平均法中两次最优移动平均取值向列联表的第一行和第一列集中的趋势更明显。

假设 5 线性数据的斜率越大，二次移动平均法中两次最优移动平均取值向列联表的第一行和第一列集中的趋势更明显。

二、二次移动平均法的一般预测公式推导

设时间集为 $\{t \mid 1 \leq t \leq N, t \text{ 为正整数}\}$ ，需要预测的随机序列为 $\{Y_t, 1 \leq t \leq N\}$ ， N_1 ($2 \leq N_1 \leq N-2$) 为二次移动平均法中的第一次移动平均步长， N_2 ($2 \leq N_2 \leq N-N_1+1$) 为二次移动平均法中的第二次移动平均步长^②。则经过第一次和第二次移动平均后的序列可分别表示为：

$$M_t^{(1)} = \frac{Y_{t-N_1+1} + Y_{t-N_1+2} + \dots + Y_t}{N_1} \quad (1)$$

$$M_t^{(2)} = \frac{M_{t-N_2+1}^{(1)} + M_{t-N_2+2}^{(1)} + \dots + M_t^{(1)}}{N_2} \quad (2)$$

通过式（1）和式（2）可以计算得到二次移动平均的两次平均值序列，见表 1。在表 1 中，对随机序列 $\{Y_t, 1 \leq t \leq N\}$ 而言， Y_t 为其在 t 时刻的随机变量， $\{Y_t, 1 \leq t \leq N\}$ 的期望值序列呈线性递增或递减态势，则从时间节点 t 来看， $t+T$ 时刻的随机变量 Y_{t+T} 的期望可表示为：

$$E(Y_{t+T}) = a_t + b_t \times T \quad (3)$$

其中， a_t 为线性模型的截距项， b_t 为线性模型的斜率， $E(Y_{t+T})$ 为基于 t 时刻的信息集下 Y_{t+T} 的期望值。则 $t+T$ 时刻的随机变量 Y_{t+T} 可表示为：

$$Y_{t+T} = a_t + b_t \times T + \varepsilon_{t+T} \quad (4)$$

^① 此处，没有讨论 N_1 和 N_2 正相关的情景，是依据暴春贤和陈宏立（2001）所述，当原数据波动较大时，宜采用较大的移动平均步长。那么，在消除原数据的波动中，两步移动平均步长取值就存在你大我小或你小我大的情况。

^② N_1 和 N_2 这样限定是保证至少进行两次移动平均，使得后文中线性模型的截距及斜率可以计算。

表 1

二次移动平均法中两次平均值的计算

t	Y_t	$M_t^{(1)}$	$M_t^{(2)}$
1	Y_1	NA	NA
2	Y_2	NA	NA
3	Y_3	NA	NA
\vdots	\vdots	\vdots	\vdots
$N_1 - 1$	$Y_{N_1 - 1}$	NA	NA
N_1	Y_{N_1}	$M_{N_1}^{(1)}$	NA
$N_1 + 1$	$Y_{N_1 + 1}$	$M_{N_1 + 1}^{(1)}$	NA
\vdots	\vdots	\vdots	\vdots
$N_1 + N_2 - 2$	$Y_{N_1 + N_2 - 2}$	$M_{N_1 + N_2 - 2}^{(1)}$	NA
$N_1 + N_2 - 1$	$Y_{N_1 + N_2 - 1}$	$M_{N_1 + N_2 - 1}^{(1)}$	$M_{N_1 + N_2 - 1}^{(2)}$
$N_1 + N_2$	$Y_{N_1 + N_2}$	$M_{N_1 + N_2}^{(1)}$	$M_{N_1 + N_2}^{(2)}$
\vdots	\vdots	\vdots	\vdots
N	Y_N	$M_N^{(1)}$	$M_N^{(2)}$

注：NA 表示没有相关数据。

其中， ε_{t+T} 满足 $E(\varepsilon_{t+T}) = 0$ ， $Var(\varepsilon_{t+T}) = \sigma_\varepsilon^2$ 且 $cov(\varepsilon_m, \varepsilon_s) = 0$ ($m \neq s$ 时)。为对 a_t 和 b_t 做出估计，首先需计算 $M_t^{(1)}$ 和 $M_t^{(2)}$ ，再计算 $M_t^{(1)}$ 和 $M_t^{(2)}$ 的期望，最后用 $M_t^{(1)}$ 和 $M_t^{(2)}$ 的估计值 $\hat{M}_t^{(1)}$ 和 $\hat{M}_t^{(2)}$ 代替各自的期望求解 a_t 和 b_t 的估计值 \hat{a}_t 和 \hat{b}_t 。分别计算如下：

$$\begin{aligned} M_t^{(1)} &= \frac{Y_{t-N_1+1} + Y_{t-N_1+2} + \dots + Y_t}{N_1} \\ M_t^{(1)} &= \frac{a_t + b_t \times (-N_1 + 1) + \varepsilon_{t-N_1+1} + a_t + b_t \times (-N_1 + 2) + \varepsilon_{t-N_1+2} + \dots + a_t + b_t \times (-N_1 + N_1) + \varepsilon_t}{N_1} \\ M_t^{(1)} &= a_t - \frac{N_1 - 1}{2} \times b_t + \frac{\varepsilon_{t-N_1+1} + \varepsilon_{t-N_1+2} + \dots + \varepsilon_t}{N_1} \end{aligned}$$

对上式中 $M_t^{(1)}$ 两边求期望得：

$$E(M_t^{(1)}) = a_t - \frac{N_1 - 1}{2} \times b_t \quad (5)$$

$$\begin{aligned} M_t^{(2)} &= \frac{M_{t-N_2+1}^{(1)} + M_{t-N_2+2}^{(1)} + \dots + M_t^{(1)}}{N_2} = \frac{1}{N_2} \left[\left(a_t - \frac{N_1 - 1}{2} \times b_t + \frac{\varepsilon_{t-N_1+1} + \varepsilon_{t-N_1+2} + \dots + \varepsilon_t}{N_1} \right) \right. \\ &\quad \left. + \left(a_t - \frac{N_1 - 1}{2} \times b_t + \frac{\varepsilon_{t-N_1+1} + \varepsilon_{t-N_1+2} + \dots + \varepsilon_t}{N_1} \right) \right. \\ &\quad \left. + \dots + \left(a_t - \frac{N_1 - 1}{2} \times b_t + \frac{\varepsilon_{t-N_1+1} + \varepsilon_{t-N_1+2} + \dots + \varepsilon_t}{N_1} \right) \right] \end{aligned}$$

对上式中 $M_t^{(2)}$ 两边求期望得：

$$E(M_t^{(2)}) = a_t - \frac{N_1 + N_2 - 2}{2} \times b_t \quad (6)$$

假定从随机序列 $\{Y_t, 1 \leq t \leq N\}$ 抽取了一个样本量为 n 的连续等间隔观测值序列 y_t ($1 \leq t \leq n \leq N \cup N_1 + N_2 - 1 \leq n$)，那么对 $E(M_t^{(1)})$ 和 $E(M_t^{(2)})$ 的估计值可分别表示为：

$$\hat{M}_t^{(1)} = \hat{a}_t - \frac{N_1 - 1}{2} \times \hat{b}_t \quad (7)$$

$$\hat{M}_t^{(2)} = \hat{a}_t - \frac{N_1 + N_2 - 2}{2} \times \hat{b}_t \quad (8)$$

$$\text{其中, } \hat{M}_t^{(1)} = \frac{y_{t-N_1+1} + y_{t-N_1+2} + \dots + y_t}{N_1}, \quad \hat{M}_t^{(2)} = \frac{\hat{M}_{t-N_2+1}^{(1)} + \hat{M}_{t-N_2+2}^{(1)} + \dots + \hat{M}_t^{(1)}}{N_2}.$$

\hat{a}_t 表示线性模型截距的样本估计值， \hat{b}_t 表示线性模型斜率的样本估计值。联立式 (7) 和式 (8)，可得：

$$\hat{a}_t = \hat{M}_t^{(1)} + \frac{N_1 - 1}{N_2 - 1} (\hat{M}_t^{(1)} - \hat{M}_t^{(2)}) \quad (9)$$

$$\hat{b}_t = \frac{2}{N_2 - 1} (\hat{M}_t^{(1)} - \hat{M}_t^{(2)}) \quad (10)$$

式 (9) 和式 (10) 是更一般的二次移动平均法的截距和斜率计算式。当 $N_1 = N_2$ 时，可简化为其特殊形式：

$$\hat{a}_t^* = 2\hat{M}_t^{(1)} - \hat{M}_t^{(2)} \quad (11)$$

$$\hat{b}_t^* = \frac{2}{N_1 - 1} (\hat{M}_t^{(1)} - \hat{M}_t^{(2)}) \quad (12)$$

进一步，在式 (4) 的基础上，结合式 (9) 和式 (10)， Y_{t-T} 均值的预测值可表示为：

$$\hat{y}_{t+T} = \hat{a}_t + \hat{b}_t \times T \quad (13)$$

当 $N_1 = N_2$ 时，结合式 (11) 和式 (12)， Y_{t+T} 均值的预测值可表示为：

$$\hat{y}_{t+T} = \hat{a}_t^* + \hat{b}_t^* \times T \quad (14)$$

为了及时捕捉当期信息下截距和斜率对未来预测值的影响，一般取 $T=1$ 。此情境下，二次移动平均法预测的均方误差的计算公式为：

$$MSE(N_1, N_2) = \frac{1}{n - N_1 - N_2 + 1} \sum_{t=N_1+1}^n (y_t - \hat{y}_t)^2 \quad (15)$$

通过最小化式 (15) 的取值，即可获得二次移动平均法下的一组 (N_1, N_2) ，该组两次移动平均步长即为所求二次移动平均的最优移动步长。

三、数值模拟情景下的假设检验

根据以上理论推导，为检验前文所提五个假设，首先需要模拟产生不同线性数据下两次移动平均步长的最优值，进而通过多次模拟产生的最优移动平均步长，构成一个列联表。具体步骤如下^①：

第一步：一般情况下，两次移动平均的步长 N_1 和 N_2 不会太大，故两者最大步长均取 $n/3$

① 如需模拟程序，可向作者索取。

四舍五入后的整数^①和10中的较小数，并产生一个行列均为 $\min([n/3] - 1, 9)$ 的零矩阵；

第二步：随机产生一个呈线性变化的观测值序列 y_t , $1 \leq t \leq n$;

第三步：在二次移动平均法的第一次移动平均步长 $N_1 = 2$ 时，计算第一次移动平均值序列 $\hat{M}_t^{(1)}$ ；

第四步：在二次移动平均的第二次移动平均步长 $N_2 = 2$ 时，计算第二次移动平均值序列 $\hat{M}_t^{(2)}$ ；

第五步：根据第三步和第四步计算的 $\hat{M}_t^{(1)}$ 和 $\hat{M}_t^{(2)}$ ，按照式（9）和式（10）分别计算截距 \hat{a}_t 和斜率 \hat{b}_t ；

第六步：取 $T=1$ ，计算式（13）和式（15）的取值；

第七步：依次选择第二次移动平均步长 $N_2 = 3, 4, \dots, \min([n/3], 10)$ ，重复第四步~第六步，计算 N_2 在不同取值下的 $MSE(2, N_2)$ ；

第八步：依次选择第一次移动平均步长 $N_1 = 3, 4, \dots, \min([n/3], 10)$ ，重复第三步~第七步，并选择使 $MSE(N_1, N_2)$ 最小的那对 (N_1, N_2) 作为本次模拟的最佳移动平均步长。在第一步产生的矩阵的第 $N_1 - 1$ 行第 $N_2 - 1$ 列元素上加1，其他元素不变，并替换第一步产生的矩阵；

第九步：重复第二步~第八步 10000 次，形成一个 $\min([n/3] - 1, 9) \times \min([n/3] - 1, 9)$ 的列联表。此列联表中最后一行和最后一列所对应的 (N_1, N_2) 不一定使均方误差最小，故将这个列联表的最后一行和最后一列予以剔除，最终形成一个 $\min([n/3] - 2, 8) \times \min([n/3] - 2, 8)$ 的新列联表，这个新列联表为后续检验的基础。

1. N_1 和 N_2 的独立性和等值性检验

在假设1和假设2的基础上，不失一般性，分别模拟在样本量为15、20、25、30、35、40、45和50的情况下列联表的计算结果，并将统计量和伴随概率绘制如图1和图2。通过图1，可以发现：在样本量为15时，二次移动平均法的最优两次移动平均步长取值在统计上独立；但随着样本量的增加，最优两次移动平均步长取值快速落入统计上显著的不独立区域，且显著性水平一般在1%以下。

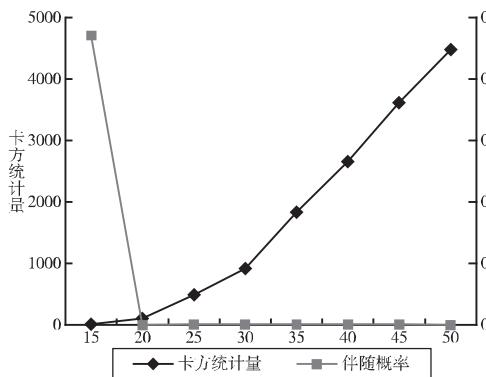


图1 两次移动平均步长取值的独立性检验

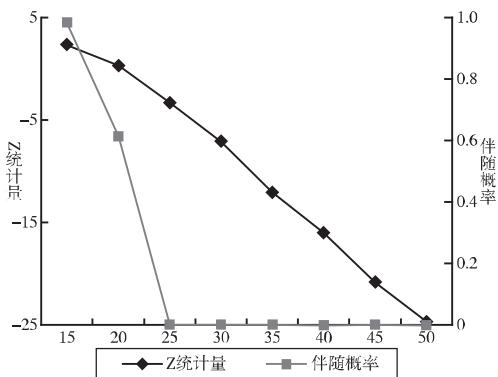


图2 两次移动平均步长取值的等值性检验

① 后文用取整函数表示，即 $[n/3]$ 。

图 2 刻画了最优两次移动平均步长等值的检验结果。可知：当样本量较小（小于等于 20）时，有较大的伴随概率不拒绝两次移动平均的步长取值相等的假设；随着样本量的增加，均拒绝最优两次移动平均步长相等，且拒绝效率不断提高，并在样本量为 25 时，拒绝水平提高到 1% 以上。实际应用中，意味着样本量较小（小于等于 20）时，可设定两次移动平均的最优步长相等；样本量较大（大于 20）时，则应分别探寻两次移动平均的最优步长。

值得注意的是，图 1 和图 2 中的结论看似矛盾，实则相成。因为图 1 仅检验两次移动平均的最优步长取值是否独立问题，而图 2 则检验的是两次移动平均的最优步长是否相等问题。即便两次移动平均的最优步长取值相互独立，也不能否认两者可能存在取相同值的倾向；同样即便两次移动平均步长取值有较大概率相等，也同样存在两者独立的可能性。

2. 不同样本量下 N_1 和 N_2 取值关系的进一步检验

初步模拟发现二次移动平均法的第一次移动平均最优步长 N_1 和第二次移动平均最优步长 N_2 存在这样的关系： N_1 和 N_2 呈反方向变动；两者中存在一次的移动平均最优步长取最小值（一般为 2），另一次的最优移动平均步长取较大值倾向。具体而言，当 N_1 较大时， N_2 取值会向 2 靠拢； N_1 取 2 时， N_2 倾向于取较大步长。为检验这种直观认识在统计上是否显著，模拟统计量和概率见表 2。

表 2 不同样本量下两次移动平均步长取值关系检验

样本量	(N_1, N_2) 向列联表的第一行、第一列集中的检验		$N_1=2$ 时， N_2 倾向取较大值；或 $N_2=2$ 时， N_1 倾向取较大值的检验	
	原假设	Z 统计量	原假设	模拟概率
15	$\sum_{i=1}^3 p_{1,i} \geq 0.333$	10.093	$p_{1,j+1} > p_{1,j}, j=1, 2 (N_1=2)$	0.62
	$\sum_{i=1}^3 p_{i,1} \geq 0.333$	20.755	$p_{j+1,1} > p_{j,1}, j=1, 2 (N_2=2)$	0.93
20	$\sum_{i=1}^5 p_{1,i} \geq 0.2$	6.499	$p_{1,j-1} > p_{1,j}, j=1, 2, 3, 4 (N_1=2)$	0.67
	$\sum_{i=1}^5 p_{i,1} \geq 0.2$	26.773	$p_{j-1,1} > p_{j,1}, j=1, 2, 3, 4 (N_2=2)$	0.89
25	$\sum_{i=1}^6 p_{1,i} \geq 0.167$	4.821	$p_{1,j+1} > p_{1,j}, j=1, 2, 3, 4, 5, 6 (N_1=2)$	0.97
	$\sum_{i=1}^6 p_{i,1} \geq 0.167$	39.635	$p_{j+1,1} > p_{j,1}, j=1, 2, 3, 4, 5, 6 (N_2=2)$	0.89
30	$\sum_{i=1}^8 p_{1,i} \geq 0.125$	10.513	$p_{1,j+1} > p_{1,j}, j=1, 2, 3, 4, 5, 6, 7 (N_1=2)$	0.83
	$\sum_{i=1}^8 p_{i,1} \geq 0.125$	57.143	$p_{j+1,1} > p_{j,1}, j=1, 2, 3, 4, 5, 6, 7 (N_2=2)$	0.88
35	$\sum_{i=1}^8 p_{1,i} \geq 0.125$	6.846	$p_{1,j+1} > p_{1,j}, j=1, 2, 3, 4, 5, 6, 7 (N_1=2)$	0.99
	$\sum_{i=1}^8 p_{i,1} \geq 0.125$	76.696	$p_{j+1,1} > p_{j,1}, j=1, 2, 3, 4, 5, 6, 7 (N_2=2)$	0.98
40	$\sum_{i=1}^8 p_{1,i} \geq 0.125$	11.600	$p_{1,j+1} > p_{1,j}, j=1, 2, 3, 4, 5, 6, 7 (N_1=2)$	1.00
	$\sum_{i=1}^8 p_{i,1} \geq 0.125$	88.968	$p_{j+1,1} > p_{j,1}, j=1, 2, 3, 4, 5, 6, 7 (N_2=2)$	0.99

(续)

样本量	(N_1, N_2) 向列联表的第一行、第一列集中的检验		$N_1=2$ 时, N_2 倾向取较大值; 或 $N_2=2$ 时, N_1 倾向取较大值的检验	
	原假设	Z统计量	原假设	模拟概率
45	$\sum_{i=1}^8 p_{1,i} \geq 0.125$	10.209	$p_{1,j+1} > p_{1,j}, j = 1, 2, 3, 4, 5, 6, 7$ ($N_1=2$)	1.00
	$\sum_{i=1}^8 p_{i,1} \geq 0.125$	106.808	$p_{j+1,1} > p_{j,1}, j = 1, 2, 3, 4, 5, 6, 7$ ($N_2=2$)	1.00
50	$\sum_{i=1}^8 p_{1,i} \geq 0.125$	11.940	$p_{1,j+1} > p_{1,j}, j = 1, 2, 3, 4, 5, 6, 7$ ($N_1=2$)	1.00
	$\sum_{i=1}^8 p_{i,1} \geq 0.125$	122.657	$p_{j+1,1} > p_{j,1}, j = 1, 2, 3, 4, 5, 6, 7$ ($N_2=2$)	1.00

注: $p_{i,j}$ 为列联表中第 i 行和第 j 列元素占总频数的比例。

表 2 中的第二和第三列, 首先检验了 N_1 和 N_2 取值是否有向列联表的第一行或第一列集中的趋势。发现不同样本量下, 在均值意义上, 两次移动平均的最优步长均有向列联表的第一行或第一列集中的倾向, 且均在 10% 的显著水平上无法拒绝原假设。不同样本量下, N_1 和 N_2 在第一行上集中程度的检验统计量, 大致随样本量的增加而增加, 但增加幅度较小; N_1 和 N_2 在第一列上集中程度的检验统计量, 随样本量的增加而增加, 且增加幅度较大。说明虽然 N_1 和 N_2 的取值均有向列联表的第一行或第一列集中的倾向, 但更偏重于向第一列集中。

表 2 中的第四和第五列, 是在前两列检验的基础上, 进一步对 N_1 和 N_2 的最优取值在行或列上集中程度的度量。检验发现, 在列联表的第一行上, N_2 的最优取值向较大值靠拢的速度更快; 在列联表的第一列上, N_1 向较大值靠拢的速度较慢。结合前两列的结果, 发现 N_1 和 N_2 的最优取值集中在第一列上的趋势更明显, 但在第一行上向 N_2 的最大值靠拢的速度更快。这验证了前文所提的假设 3。实际运用中, 样本量有限时, 更倾向于直接取 N_1 等于 2, N_2 较大的二次移动平均; 也可取 N_2 等于 2, N_1 较大的二次移动平均。

3. 截距对 N_1 和 N_2 取值的影响

为验证假设 4, 不失一般性, 分别在样本量为 15、20、25、30、35、40、45 和 50 的情景下, 截距为 10、100、200、500 和 1000 时, 构建 Z 统计量企图从统计上进行研究。表 3 主要回答了不同样本量下, N_1 和 N_2 的最优取值向列联表的第一行或第一列的集中程度是否具有统计上的显著差异。

表 3 不同样本量下截距大小对两次移动平均步长取值的影响

样本量	原假设	不同截距下的 Z 统计量				
		10	100	200	500	1000
15	$\sum_{i=1}^3 P_{1,i} + \sum_{i=2}^3 P_{i,1} \geq \sum_{i=1}^3 p_{1,i} + \sum_{i=2}^3 p_{i,1}$	NA	0.314	0.541	0.590	-0.072
20	$\sum_{i=1}^5 P_{1,i} + \sum_{i=2}^5 P_{i,1} \geq \sum_{i=1}^5 p_{1,i} + \sum_{i=2}^5 p_{i,1}$	NA	1.039	0.203	-0.152	-0.478
25	$\sum_{i=1}^6 P_{1,i} + \sum_{i=2}^6 P_{i,1} \geq \sum_{i=1}^6 p_{1,i} + \sum_{i=2}^6 p_{i,1}$	NA	-0.333	0.307	-0.329	-0.246
30	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	0.793	0.077	-0.375	0.595

(续)

样本量	原假设	不同截距下的 Z 统计量				
		10	100	200	500	1000
35	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-1.495*	1.954	0.495	0.290
40	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-0.102	-1.519*	-0.496	0.087
45	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-0.154	1.285	0.729	0.681
50	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-0.154	0.829	-0.893	1.461

注: $P_{i,j}$ 、 $p_{i,j}$ 分别表示较大、较小截距下各自列联表中第 i 行和第 j 列元素占各自总频数的比例; *、**、*** 分别表示在 10%、5%、1% 的水平上显著。同表 1。

表 3 说明, 随着截距增加, 检验统计量取值的伴随概率整体增加, 两次移动平均的最优步长向列联表的第一行和第一列进一步靠拢, 仅在截距为 10~100 和 100~200 时各有一次在 10% 的显著水平上拒绝原假设, 其他则均不能拒绝。列维度结果显示, 不同样本量、同一截距下, 拒绝原假设的检验效力变化不大。总体而言, 随数据截距增加, 两次移动平均最优步长的频数有向列联表的第一行和第一列进一步集中趋势。这说明, 当采样数据的截距越大, 越有理由选择 N_2 等于 2, N_1 较大的二次移动平均。

4. 斜率对 N_1 和 N_2 取值的影响

为验证不同斜率的原始数据是否会影响两次移动平均的最优步长取值, 在上文模拟程序基础上, 不失一般性, 取原始数据的斜率分别为 -10、-5、1、5 和 10, 得到表 4 的检验结果。

表 4 不同样本量下斜率大小对两次移动平均步长取值的影响

样本量	原假设	不同斜率下的 Z 统计量				
		-10	-5	1	5	10
15	$\sum_{i=1}^3 P_{1,i} + \sum_{i=2}^3 P_{i,1} \geq \sum_{i=1}^3 p_{1,i} + \sum_{i=2}^3 p_{i,1}$	NA	4.823	4.867	-4.948***	-3.133***
20	$\sum_{i=1}^5 P_{1,i} + \sum_{i=2}^5 P_{i,1} \geq \sum_{i=1}^5 p_{1,i} + \sum_{i=2}^5 p_{i,1}$	NA	-1.782**	1.570	0.929	1.817
25	$\sum_{i=1}^6 P_{1,i} + \sum_{i=2}^6 P_{i,1} \geq \sum_{i=1}^6 p_{1,i} + \sum_{i=2}^6 p_{i,1}$	NA	-4.253***	-1.576*	2.374	4.581
30	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-6.106***	-0.911	0.877	3.680
35	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-3.666***	-1.324*	1.818	4.867
40	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-3.437***	-0.555	0.748	2.937
45	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-2.250**	-0.737	-0.812	1.790
50	$\sum_{i=1}^8 P_{1,i} + \sum_{i=2}^8 P_{i,1} \geq \sum_{i=1}^8 p_{1,i} + \sum_{i=2}^8 p_{i,1}$	NA	-2.689***	0.443	0.075	1.195

注: 同表 3。

行维度的结果显示，在同一样本量下，除样本量为15时，检验统计值均表现出增加态势；样本量为15时，则表现为减少。说明在样本量较小（小于等于15）时，随数据斜率的增加，两次移动平均最优步长的频数没有向列联表的第一行和第一列集中；样本量较大（大于等于20）时，两次移动平均最优步长的频数有进一步向列联表的第一行和第一列集中趋势。列维度的结果显示，不同样本量下，数据的斜率从-10到-5时，基本可以在1%的显著水平上拒绝原假设，说明两次移动平均最优步长的取值向列联表的第一行和第一列集中效应不显著。当数据的斜率从-5到1、1到5、5到10时，多数在10%的显著水平上不能拒绝原假设，但检验统计量取值有下降趋势，说明随样本量增加，两次移动平均最优步长的取值频数依然会向列联表的第一行和第二行集中，但集中速度变慢。这一结果验证了前文所提出的假设5。

针对具体数据而言，当原始数据的斜率较大且样本量较大（大于等于20）时，可以更有信心地采用第一次移动平均步长 N_1 较大，第二次移动平均步长 N_2 等于2；或采用第一次移动平均步长 N_1 等于2，第二次移动平均步长 N_2 较大进行预测。当原始数据的斜率较大且样本量较小（小于等于15）或原始数据的斜率较小时，第一次移动平均步长 N_1 和第二次移动平均步长 N_2 取值关系不确定。

四、应用举例

为验证所得结论的合理性，采用张保法（2004）书中第103页的线性数据。针对该案例，张保法（2004）在两次移动平均步长相等的条件下，将第一次移动平均步长和第二次移动平均步长均设置为3。至于为何需要将第一次移动平均的步长和第二次移动平均的步长先验地取相同的值，则没有给出充分的理由。考虑到这一原因，本文将第一次移动平均的步长和第二次移动平均的步长不施加任何先验条件，而是在尽可能多的取值组合下，通过最小化均方误差，选择第一次移动平均步长和第二次移动平均步长的最优组合。

1. 两次移动平均最优步长确定

针对张保法（2004）书中第103页的线性数据而言，该案例的样本量为12，故分别计算第一次移动平均步长和第二次移动平均步长在2~4^①的范围内变化时，每种不同步长组合下的均方误差，并最终根据均方误差最小原则选取最优步长组合，结果见表5。结果显示，当第一次移动平均步长为4，第二次移动平均步长均为2时，所得均方误差为最小的21.15。该结果说明张保法（2004）将第一次移动平均的步长和第二次移动平均的步长均设置为3，没能使得均方误差达到最小值，存在提高样本内预测精度的其他步长组合。

表5 两次移动平均步长不同组合下的均方误差

N_1	N_2	2	3	4
2		25.57	28.86	27.10
3		28.09	30.89	34.57
4		21.15	26.34	24.38

注：表中右下角模块为第一次移动步长和第二次移动步长的不同组合下均方误差的计算结果。

^① 因该案例的样本量为12，所以文中最大移动平均步长取4；这与前文的表述一致。

同时，表 5 的结论与图 2 的模拟结果也是相符的。图 2 的结论说明样本量较小时，最优步长的组合中，第一次移动平均步长和第二次移动平均步长有取相等值的倾向；但在具体的实际案例中，也会以一定的概率（图 2 中这个概率大约为 0.2）存在第一次移动平均步长和第二次移动平均步长不等的情况。该案例正好是样本量小于 20 时，最优的移动平均步长组合下，第一次移动平均步长和第二次移动平均步长不相等的情况。

2. 新旧方法的预测效果对比分析

为比较本文构建的一般性二次移动平均法和以往的二次移动平均法在预测效果上的差异，首先，采用两次移动平均步长相等的二次移动平均法对张保法（2004）书中第 103 页的线性数据进行预测，结果列于表 6。表 6 中的第 3 列和第 4 列是在第一次移动平均步长和第二次移动平均步长均为 3 时^①，计算的一次移动平均数和二次移动平均数。因为两次移动平均步长相等，所以采用式（11）和式（12）分别计算 \hat{a}_t^* 和 \hat{b}_t^* ，结果分列于第 5 列和第 6 列。最后，在向前一步预测期数 $T=1$ 时，根据式（14）得到各期的预测值，列于表 6 中的第 7 列。

表 6 两次移动平均步长相等的向前一期预测值^②

年份	利润（万元）	$N_1 = N_2 = 3$		\hat{a}_t^*	\hat{b}_t^*	预测值
		$\hat{M}_t^{(1)}$	$\hat{M}_t^{(2)}$			
1989	120.87	NA	NA	NA	NA	NA
1990	125.58	NA	NA	NA	NA	NA
1991	131.66	126.04	NA	NA	NA	NA
1992	130.42	129.22	NA	NA	NA	NA
1993	130.38	130.82	128.69	132.95	2.13	NA
1994	135.54	132.11	130.72	133.51	1.40	135.08
1995	144.25	136.72	133.22	140.23	3.50	134.91
1996	147.82	142.54	137.12	147.95	5.41	143.73
1997	148.57	146.88	142.05	151.71	4.83	153.36
1998	148.61	148.33	145.92	150.75	2.42	156.54
1999	149.73	148.97	148.06	149.88	0.91	153.17
2000	154.56	150.97	149.42	152.51	1.54	150.79

注：同表 1。

表 7 为本文构建的二次移动平均法下各项的计算结果。其中，第 3 列为第一次移动平均步长取 4 时，计算的一次移动平均数，第 4 列为第二次移动平均步长取 2 时，计算的二次移动平均数；第 5 列和第 6 列为采用式（9）和式（10），分别计算的各期 \hat{a}_t 和 \hat{b}_t ；最后一列为向前一步预测期数（即 $T=1$ ）时，根据式（13）计算的各期预测值。

① 这里沿用了张保法（2004）书中第 103 页对两次移动平均步长的设置。

② 需要说明，在张保法（2004）书中第 103 页，其第一次移动平均和第二次移动平均都是采用加权移动平均，而不是简单移动平均，反映了该书中第 102 页关于“二次移动平均数法”的理论表述与其实际案例计算之间的矛盾。本文则采用简单移动平均法来计算一次移动平均数和二次移动平均数。使得本文的计算结果和该书中的结果存在差异。

表 7 两次移动平均步长不等的向前一期预测值

年份	利润(万元)	$N_1=4, N_2=2$		\hat{a}_t	\hat{b}_t	预测值
		$\hat{M}_t^{(1)}$	$\hat{M}_t^{(2)}$			
1989	120.87	NA	NA	NA	NA	NA
1990	125.58	NA	NA	NA	NA	NA
1991	131.66	NA	NA	NA	NA	NA
1992	130.42	127.13	NA	NA	NA	NA
1993	130.38	129.51	128.32	133.08	2.38	NA
1994	135.54	132.00	130.76	135.74	2.49	135.46
1995	144.25	135.15	133.57	139.87	3.15	138.23
1996	147.82	139.50	137.32	146.02	4.35	143.02
1997	148.57	144.05	141.77	150.87	4.55	150.37
1998	148.61	147.31	145.68	152.21	3.27	155.42
1999	149.73	148.68	148.00	150.74	1.37	155.48
2000	154.56	150.37	149.53	152.90	1.69	152.11

注：同表 1。

在表 6 和表 7 的基础上，将新旧方法计算的各时期实际观测值与预测值的相对误差列于表 8。其中，第 4 列为第一次移动平均步长和第二次移动平均步长均为 3 时，实际观测值与预测值的相对误差；第 6 列为第一次移动平均步长为 4，第二次移动平均步长为 2 时，实际观测值与预测值的相对误差。最后两行为新旧方法分别计算的衡量所有观测值与预测值差异程度的均方误差 (MSE)，以及衡量所有观测值与预测值平均绝对偏差程度的平均相对误差绝对值 (MAPE)^①。

针对该案例，倘若采用两次移动平均步长相等的二次移动平均法，实际观测值与预测值之间的均方误差为 30.89，两者之间的平均相对误差绝对值为 3.27%；而使用两次移动平均步长不等的二次移动平均法，所得均方误差仅为 21.15，平均相对误差绝对值也仅为 2.67%。由此可见，两次移动平均步长不等的均方误差比两次移动平均步长相等时，绝对值减少了 9.74（或降低了 31.53%）；两次移动平均步长不等的平均相对误差绝对值比两次移动平均步长相等时，绝对值下降了 0.6%（或降低了 18.26%）。这说明采用两次移动平均步长不等的二次移动平均法，在总体预测精度上具有明显的优势。

表 8 新旧方法计算的向前一期预测值评价

年份	利润(万元)	两次移动平均步长相等的预测值	相对误差	两次移动平均步长不等的预测值	相对误差
1989	120.87	NA	NA	NA	NA
1990	125.58	NA	NA	NA	NA
1991	131.66	NA	NA	NA	NA

① 平均相对误差绝对值指标采用的是徐国祥（2016）的定义，计算公式为 $MAPE = (\frac{1}{T} \sum_{t=1}^T |\frac{y_t - \hat{y}_t}{y_t}|) \times 100\%$ 。这和易丹辉（2014）在《统计预测：方法与应用》书中第 134 页定义的平均绝对百分误差内涵一致；同时，与冯文权和傅征（2018）在《经济预测与决策技术》书中第 286 页定义的相对平均绝对误差（ ΔARE ）内涵也一致。

(续)

年 份	利 润(万元)	两次移动平均步长相等的预测值	相 对 错 差	两次移动平均步长不等的预测值	相 对 错 差
1992	130.42	NA	NA	NA	NA
1993	130.38	NA	NA	NA	NA
1994	135.54	135.08	0.34%	135.46	0.06%
1995	144.25	134.91	6.48%	138.23	4.18%
1996	147.82	143.73	2.77%	143.02	3.25%
1997	148.57	153.36	-3.22%	150.37	-1.21%
1998	148.61	156.54	-5.34%	155.42	-4.58%
1999	149.73	153.17	-2.30%	155.48	-3.84%
2000	154.56	150.79	2.44%	152.11	1.59%
均 方 误 差(MSE)		30.89	NA	21.15	NA
平均相对误差绝对值(MAPE)	NA	3.27%	NA	2.67%	

注：同表 1。

具体到对每个观测值的预测情况来看，两次移动平均步长相等时，相对误差在 1995 年和 1998 年均出现了预测偏差大于 5%，占总预测样本量的 28.57%，说明该方法的预测效果较差。相反，当两次移动平均步长不等时，针对观测值的预测，则没有一个预测值与实际观测值的相对误差偏差大于 5%，说明此方法的预测效果相较以往有明显改善。同时，两种方法在针对每个观测值的预测上，当两次移动平均步长不等时，相对误差的绝对值仅在 1996 年和 1999 年高于两次移动平均步长相等的，其他年份均对两次移动平均步长相等时的情况有所改善。说明对每个观测值的预测效果上，两次移动平均步长不等时的预测效果也好于两次移动平均步长相等的情景。

3. 针对该案例数据的检验

此案例的数据仅为一次抽样所得，倘若将其视为来自某一随机过程的一次实现，则可使用该案例的数据来估计此随机过程均值序列的截距和斜率。那么，从该随机过程获取的不同样本的第一次移动平均步长是否和第二次移动平均步长独立？第一次移动平均步长是否和第二次移动平均步长有取相等值的倾向？进一步，当第一次移动平均步长取较大值时，第二次移动平均步长是否有取较小值的倾向？抑或当第一次移动平均步长取较小值时，第二次移动平均步长是否有取较大值的倾向？为解决这些问题，首先通过样本估计该随机过程的均值序列形式如下。

拟合均值序列的表达式为：

$$\hat{y}_t = 119.491 + 3.001 \times t \quad (t = 1, 2, \dots, 12; \hat{\sigma}_e = 2.846)$$

则该随机过程在 t 时刻的随机变量 y_t 的分布为：

$$y_t \sim N(119.491 + 3.001 \times t, 8.1) \quad (t = 1, 2, 3, \dots)$$

其次，从所求随机过程中，每次以时间间隔 1 连续抽样 12 个^①，连续模拟 10000 次，并计算检验统计量和该检验统计量的伴随概率，检验结果见表 9。

① 因为是对连续等间隔的时间序列抽取样本，所以抽得的样本也是连续等间隔的，并且间隔和原随机序列一致。

表 9

拟合随机过程再抽样的假设检验

原假设	统计量	自由度	伴随概率	判断结果
两次移动平均的步长 N_1 和 N_2 独立	22.20	1	0.00	拒绝
两次移动平均的步长 $N_1=N_2$	-4.94		0.00	拒绝
(N_1, N_2) 向列联表第一行和第一列集中	21.26		1.00	不拒绝
$N_1=2$ 时, N_2 倾向取较大值			0.98	不拒绝
$N_2=2$ 时, N_1 倾向取较大值			0.99	不拒绝

以上结果说明, 针对该随机过程而言, 经过多次抽样, 其两次移动平均步长不独立; 且拒绝第一次移动平均步长和第二次移动平均步长相等的假设, 这和前文的模拟结果是相符的。另外, 针对该随机过程, 在特定的样本量、截距和斜率下, 两次移动平均的最优步长有向列联表的第一行和第一列集中的趋势; 且当第一次移动平均步长取最小值 2 时(或第二次移动平均步长取最小值 2 时), 第二次移动平均步长有取较大值的倾向(或第一次移动平均步长有取较大值的倾向), 这和表 2 与表 5 中的结论也具有一致性。总体而言, 通过张保法(2004)一书中的案例, 从应用视角进一步佐证了本文所得结论的合理性。

五、总 结

为解决有线性递增或递减趋势的时间序列预测问题, 以往的学者提出可以在一次移动平均法的基础上, 采用两次移动平均步长相等的二次移动平均法来对一次移动平均法的预测偏差进行修正。至于为何选择两次移动平均步长相等的二次移动平均法, 以往文献鲜有论述。

鉴于此, 本文从理论上推导了两次移动平均步长不等情况下的二次移动平均法的预测式, 并通过数值模拟回答了所提出的几个假设。结论如下: 第一, 样本量小于等于 15 时, 两次移动平均的最优步长取值在统计上独立; 而样本量大于等于 20 时, 两次移动平均的最优步长取值在统计上不独立。第二, 样本量小于等于 20 时, 两次移动平均的最优步长存在取等值倾向; 样本量大于等于 25 时, 两次移动平均的最优步长取等值的倾向不复存在。第三, 两次移动平均的最优步长取值有此大彼小或此小彼大的倾向, 且更偏向于第一次移动平均取较大值, 第二次移动平均步长取较小值集中。第四, 整体而言, 随线性数据截距或斜率的增加, 两次移动平均最优步长取值中此大彼小或此小彼大的倾向更加突出。本文还通过一个实例, 佐证了结论的合理性。

虽然本文致力于二次移动平均法中两次移动平均的最优步长取值研究, 并从理论、数值模拟和应用三个方面对该问题进行了论证, 形成了一些研究结论; 但因线性数据样本量、截距和斜率的多变性, 文中无法详尽各种情景设定^①。

参 考 文 献

- [1] Alvezakou E. G., Pantazis G., 2017, *A Comparative Evaluation of Various Models for Prediction of Displacements* [J], *Applied Geomatics*, 9 (2), 93~103.
- [2] Areepong Y., Sukparungsee S., 2011, *An Analytical ARL of Binomial Double Moving Average Chart* [J], *International Journal of Pure & Applied Mathematics*, 73 (4), 477~488.

^① 研究者可以借助作者后期发布的 R 软件包对数据进行分析。

- [3] Areepong Y., Sukparungsee S., 2015, *Explicit Expression for the Average Run Length of Double Moving Average Scheme for Zero-Inflated Binomial Process* [J], International Journal of Applied Mathematics & Statistics, 53 (3), 33~43.
- [4] Dodd E. L., 1941, *The Problem of Assigning a Length to the Cycle to Be Found in a Simple Moving Average and in a Double Moving Average of Chance Data* [J], Econometrica, 9 (1), 25~37.
- [5] Du J., Yin K., Lacasse S., 2013, *Displacement Prediction in Colluvial Landslides, Three Gorges Reservoir, China* [J], Landslides, 10 (2), 203~218.
- [6] Khoo M. C., Wong V. H., 2008, *A Double Moving Average Control Chart* [J], Communications in Statistics-Simulation and Computation, 37 (8), 1696~1708.
- [7] Lian C., Zeng Z., Yao W., Tang H., 2014, *Ensemble of Extreme Learning Machine for Landslide Displacement Prediction Based on Time Series Analysis* [J], Neural Computing and Applications, 24 (1), 99~107.
- [8] Montgomery D. C., 2005, *Introduction to Statistical Quality Control* [M], Hoboken, N.J.: John Wiley.
- [9] NIST SEMATECH, 2013, *E-Handbook of Statistical Methods* [DB/OL], <http://www.itl.nist.gov/div898/handbook/prmc/section4/prmc422.htm>.
- [10] Phantu S., Sukparungsee S., Areepong Y., 2018, *DMA Chart Monitoring of the First Integer-Valued Autoregressive Processes of Poisson Counts* [J], Advances and Applications in Statistics, 52 (2), 97~119.
- [11] Sukparungsee S., 2013, *Average Run Length of Double Moving Average Control Chart for Zero-Inflated Count Processes* [J], Far East Journal of Mathematical Sciences, 1 (1), 85~103.
- [12] Zakamulin V., 2018, *Revisiting the Profitability of Market Timing with Moving Averages* [J], International Review of Finance, 18 (2), 317~327.
- [13] 白雪梅、赵松山:《移动平均法应用之我见》[J],《统计教育》2001年第1期。
- [14] 暴奉贤、陈宏立:《经济预测与决策方法》[M],暨南大学出版社,2001年。
- [15] 陈殿阁:《市场调查与预测》[M],清华大学出版社、北京交通大学出版社,2004年。
- [16] 丁洪福、战颂:《市场调查与预测》[M],东北财经大学出版社,2014年。
- [17] 冯文权、傅征:《经济预测与决策技术(第六版)》[M],武汉大学出版社,2018年。
- [18] 洪兴建、惠琦娜:《统计预测》[M],浙江工商大学出版社,2011年。
- [19] 华伯泉:《统计预测中的二次移动平均法》[J],《统计研究》1995年第2期。
- [20] 黄文文、肖毅:《基于二次移动平均法的乒乓球运动员张继科比赛态势分析》[J],《中国体育科技》2018年第1期。
- [21] 简明:《市场研究定量分析(第二版)》[M],中国人民大学出版社,2015年。
- [22] 廖进球、李志强:《市场调查与预测》[M],湖南大学出版社,2009年。
- [23] 刘红霞:《市场调查与预测》[M],科学出版社,2007年。
- [24] 刘利兰:《市场调查与预测(第三版)》[M],经济科学出版社,2012年。
- [25] 刘秋华:《市场调查与预测》[M],中国社会科学出版社,2004年。
- [26] 柳瑞芝、施之皓、任杰:《基于二次移动平均法的乒乓球比赛势态分析模型》[J],《上海体育学院学报》2017年第3期。
- [27] 罗洪群、王青华:《市场调查与预测(第二版)》[M],清华大学出版社,2016年。
- [28] 屈援:《市场预测与决策》[M],经济科学出版社,2007年。
- [29] 唐小峰:《基于二次移动平均法的阜阳市劳动力资源预测分析》[J],《中国西部科技》2013年第5期。
- [30] 田名管、曹进:《时间序列二次移动平均预测法——公式的推导及应用》[J],《基建管理优化》1994年第4期。
- [31] 吴杨、陈兆荣:《市场调查与预测(第二版)》[M],中国科学技术大学出版社,2014年。
- [32] 徐国祥:《统计预测和决策(第五版)》[M],上海财经大学出版社,2016年。

-
- [33] 徐阳:《市场调查与市场预测》[M],高等教育出版社,2001年。
 - [34] 杨立勋:《关于合理运用移动平均法的三点建议》[J],《江苏统计》1997年第7期。
 - [35] 叶明海、于磊、胡志莹:《市场调查与预测》[M],同济大学出版社,2007年。
 - [36] 易丹辉:《统计预测:方法与应用(第二版)》[M],中国人民大学出版社,2014年。
 - [37] 张保法:《经济预测与经济决策》[M],经济科学出版社,2004年。
 - [38] 张桂喜:《经济预测、决策与对策》[M],首都经济贸易大学出版社,2003年。
 - [39] 赵益新:《二次移动平均模型在经济预测中的应用》[J],《西南民族学院学报(自然科学版)》1993年第4期。
 - [40] 周思勤、刘红霞:《市场调查与预测》[M],科学出版社,2005年。
 - [41] 朱建平:《经济预测与决策》[M],厦门大学出版社,2011年。
 - [42] 庄贵军:《市场调查与预测(第二版)》[M],北京大学出版社,2014年。

Research on the Relationship of Twice Moving Average's Steps in Double Moving Average

Xi Jianguo

(School of Business, Minnan Normal University)

Research Objectives: Obtaining a double moving average's model with different step sizes. **Research Methods:** Deriving double moving average's models with different step sizes, and discussing the relationship between twice moving average's steps in different sample size, intercept and slope by Monte Carlo method. The practicability of the model is also investigated by an example. **Research Findings:** It is found that when the sample size is small, the steps of the twice moving average have a tendency to equal. When the sample size is large, the steps of the twice moving average should be one big and other small, or on the contrary. The first moving average's step tends to be larger than the second. With the increase of intercept and slope, this trend is more significant. **Research Innovations:** The double moving average model with different step sizes is more general, and its accuracy of prediction is higher than or equal to the existing double moving average's model. **Research Value:** The theoretical model established in this paper can provide more general basic method of prediction in economy, management and other fields, and give reference for the improvement of double exponential smoothing method and double curve exponential smoothing method.

Key Words: Double Moving Average; Moving Average's Step; Independence; Equivalence; Slope

JEL Classification: C22; C32; M21

(责任编辑:焦云霞)