文章编号：1000－7695（2004）06－0117－04

金融时间序列去噪的小波变换方法

兰秋军，马超群，文凤华
（湖南大学工商管理学院，湖南长沙 410082）

摘要：比较分析了传统滤波方法对金融数据去噪的缺陷，提出采用小波分析对金融时间序列进行去噪。根据 Donoho 提出的小波去噪中的非线性阈值理论，结合金融时间序列的特点，分析了相应去噪参数的选取问题。以深圳成份指数数据为例进行实验，结果表明了该方法的有效性。

关键词：小波；去噪；时间序列；金融数据
中图分类号：F224 **文献标识码**：A

1 引言

由于金融市场中各种偶然因素的影响，使得金融数据，特别是金融时间序列数据中存在许多噪声。这些噪声严重影响了进一步地分析和处理，因此必须预先去噪。但是金融时间序列本身具有非平稳、非线性和信噪比高的特点，采用传统的去噪处理方法往往存在诸多缺陷。而小波理论是根据时－频局部化的要求而发展起来的，具有自适应和数学显微镜性质，特别适合非平稳、非线性信号的处理。自从 1984 年法国地球物理学家 J.Morlet 在分析地震数据时真正提出小波概念后，小波变换作为一种数学理论和方法在科学技术界引起了越来越多的关注和重视。1988 年，Mallat 提出多分辨分析的概念，使小波具有带通滤波的特性，因此可利用小波分解和重构的方法滤波降噪；1992 年 Mallat 又提出奇异性检测的理论，证明了信号和噪声在不同尺度上具有不同的统计特征，从而可利用其实现信噪分离。后来，Donoho 等提出非线性小波变换阈值法去噪，并得到广泛的应用。由于对金融时间序列去噪的好坏往往关系到后续进一步处理和分析的成败，因而引入小波变换对其去噪处理是很有意义的。

2 传统去噪方法的缺陷分析

对金融时间序列去噪的传统方法，主要有移动平均法、传统滤波方法、卡尔曼滤波方法和维纳滤波方法：

（1）移动平均法就是将序列从第一项开始，逐项移动，重叠求出每移动一次的序时平均数，从而构成新的时间序列。在新序列中短期的偶然因素引起的变动被削弱，从而达到去噪目的。作为一种简单的数据平滑技术，该方法非常粗略，在去噪声的同时，把许多有用信息也一并去掉了，因而它只适用于对数据的简单处理，不适合对数据的深层分析。

（2）传统滤波方法的一个典型例子是采用傅立叶变换，将时域信号变换到频域，一般地认为低频信号是有用信号，而高频信号一般看作噪声。将高于某个阈值频率的傅立叶系数全部设为 0 后，再通过傅立叶逆变换恢复到时域，从而实现去噪。它要求有用信号和噪声的频谱相互分开。但对金融时间序列来说，比如股价时间序列和收益率序列其波动性都比较大，频谱比较宽，因而有用信号和噪声谱重叠比较严重，采用传统方法难以实现信噪的有效分离。

（3）卡尔曼滤波是以最小均方差为估计的最佳准则来寻求一套递推估计的算法，它利用前一时刻的状态估计值和当前时刻的观测值来共同确定当前状态的估计值。它需要知道系统的运动规律以建立准确的状态方程。但金融时间序列是一个非平稳、非线性的时间序列，很难用一个确定的方程来描述其状态和行为，因此采用这种方法来对金融时间序列去噪也存在固有的难度。

（4）维纳滤波法是现代滤波理论中的典型代表，其基本思想是：寻找线性滤波器的最佳冲击响应或传递函数，使得滤波器的输出波形作为输入波形的最佳估计。但其只适应平稳过程，不适于非平稳过程，并且维纳滤波需要噪声和有用信号的先验知识如它们的自相关函数、功率谱密度等。由于实际中这些先验知识很难得到或者过于简化，因而往往使理论上最优的维纳滤波达不到要求。

3 小波去噪的基本原理

小波（wavelet），即小区域的波，是一种特殊的、长度有限、平均值为 0 的波形。它有两个显著特点：一是在时域都具有紧支集或近似紧支集；二是正负交替的波动性。小波分析是将信号分解成一系列小波函数的叠加，而这些小波函数都是由一个母小波通过平移和尺度伸缩得来的。若待分析信号 $f(t)$ 为能量有限的一维函数，即 $f(t) \in L^2(R)$，则其连续小波变换定义为：

$$W_f(a,b) = \langle f, \Psi_{a,b} \rangle = |a|^{-1/2} \int_R f(t)\, \overline{\Psi\left(\frac{t-b}{a}\right)} \mathrm{d}t. \tag{1}$$

相应的重构公式（小波逆变换）为：

收稿日期：2004－10－03
基金项目：国家自然科学基金项目（70371028）；教育部优秀青年教师资助计划（教人司［2003］355 号）

$$f(t) = \frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{a^2}W_f(a,b)\Psi(\frac{t-b}{a})\mathrm{d}a\mathrm{d}b. \quad (2)$$

式中，称 a 为尺度因子，b 为平移因子，$\Psi(t)$ 为母小波。尺度因子 a 越高对应的信息频率越低。目前最通行的方法是尺度按幂级数取值，即尺度 $a=2^0,2^1,\Lambda,2^J,J\in Z$。尺度越大，意味着小波函数在时间上越长，亦即被分析的信号区间也越长，主要获取的是信号的低频部分；反之尺度越小，意味着只与信号的非常小的局部进行比较，获取的是信号的高频部分。小波分析理论的一个重要特色是可以进行多分辨率分析。如图(1)所示，信号 S 可通过多层分解为反映高频信息的细节部分 cDj 和反映低频信息的概貌部分 cAj（j 为分解所在层次）。

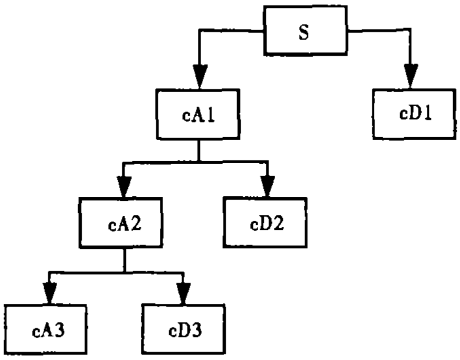

图 1 小波的多分辨率分解

通过这种多分辨率分解，信号和噪声通常会有不同的表现，从而可达到信噪分离的目的。1992 年 Mallat 提出了奇异性检测理论，指出小波变换与刻画信号奇异性的 Lipschitz 指数之间存在密切关系，利用奇异信号和随机噪声在小波变换各尺度空间中模极大值的不同传播特性，提出了一种基于模极大值的去噪方法，但用模极大值进行重构时采用的是交替投影法，为保证重构精度，通常要进行多次迭代，计算速度非常慢。另外，搜寻传播点也是一个问题，由于受到各种因素的干扰，跟踪非常困难。因此，本文对金融时间序列去噪处理，采用另外一中应用更广泛的方法——非线性阈值处理方法。

非线性阈值处理方法又称“小波收缩”法（wavelet shrinkage)，由美国 Standford 大学 Donoho 等人提出。该方法的基本原理是基于了小波变换的“集中”能力。即通过小波变换后有用信号的能量集中于少数小波系数上；而白噪声在小波变换域上仍然分散在大量小波系数之上。因而相对来说，有用信号的小波系数值必然大于那些能量分散，且幅值较小的噪声的小波系数值。因此，从谱的幅度上(不是谱的位置)，有用信号和噪声可以实现分离。该方法可分为以下三个步骤：

(1) 选择合适的正交小波基和分解层数 J，对含噪信号进行小波变换分解到 J 层；

(2) 对分解得到的小波系数进行阈值处理，可以使用两种处理方法：硬阈值和软阈值法。硬阈值法保留较大的小波系数并将较小的小波系数置零。即，

$$\hat{W}_f(j,k)=\begin{cases}W_f(j,k), & W_f(j,k)\geq Tr\\ 0, & W_f(j,k)<Tr\end{cases}. \quad (3)$$

软阈值法将较小的小波系数置零，而对较大的小波系数向零进行了收缩：

$$\hat{W}_f(j,k)=\begin{cases}\mathrm{sgn}(W_f(j,k))(W_f(j,k)-Tr), & W_f(j,k)\geq Tr\\ 0, & W_f(j,k)<Tr\end{cases}. \quad (4)$$

Donoho 在文献［4］中证明了用软阈值法能使估计信号实现最大均方误差最小化，即去噪后的估计信号是原始信号的近似最优估计。该方法具有广泛的适应性，是应用最为广泛的一种小波去噪方法，其计算速度也很快。

4 金融时间序列小波去噪方法分析

尽管非线性阈值法相对简单，然而还是存在几个重要参数的确定问题，即：小波函数的选取、阈值和分解层次的确定。下面针对金融时间序列的特性分析如何确定相应参数。

4.1 *金融时间序列的特点与一般去噪要求*。股票市场数据是金融数据的典型代表，常见的有股票价格数据和由股价计算出的收益率数据。如图（2a）、图（3a）所示。这两类序列都表现为很不平稳的特性。数据的起伏波动比较大，信号中的奇异点比较多。其中收益率数据波动性更大，奇异点数更加密集，信号中的高频成分较多，很难将其与噪声区别开来。除了大幅波动之外，数据中还有很频繁的小幅波动，随机性很强，几乎贯穿整个时间。一般的，信号的大幅波动蕴含着比较重要的信息，因而我们一般将其作为有用信号，在消噪时要保留。而大量的小幅波动一般将其作为噪声，因为即使在金融市场非常平稳，没有什么重大新闻、政策出台或暗箱操作，它也会由于股票市场的流动性而表现这种小幅波动性，不具有分析和预测价值，反而具有干扰作用。由于去噪只是进行进一步数据分析的预处理。一般地，我们只要求去掉这种小的波动性，而尽可能的保留有用信息，防止信号的失真。

4.2 *小波函数的选取*。目前有几十种小波函数，它们性质各异，适合不同的应用场合。本文的目的是金融时间序列的去噪，不同的小波函数有不同的去噪效果，因而选取合适的小波函数是很重要的。分析以下一些与去噪关系紧密的小波函数特性：

(1) 正交性。具有正交性的小波函数一方面消除了冗余，保持小波系数间的不相关性，因而可提高除噪性能；另一面正交性或双正交性是可实现快速离散小波变换的条件。由于除噪过程需将原信号在多个层次分解，显然这个特性是必须首先考虑的；

(2) 紧支撑性。如果描述尺度函数的低通滤波器组 $h(n)$ 可表征为 FIR 滤波器，那么尺度函数和小波函数只在有限区间非零，此时称小波函数具有紧支撑性。支撑宽度越小，小波的局部分辨能力越好，除噪更精细。

(3) 消失矩。若 $\int\Psi(t)t^m\mathrm{d}t=0,(m=0,1,\Lambda,M-1)$ 则称小波具有 M 阶消失矩。消失矩特性使小波展开时消去

信号的高阶平滑部分，也即函数展开为多项式时的前 $M-1$ 项对应函数的光滑部分，小波系数将非常小，因而小波变换仅仅反映函数的高阶变换部分，从而反映信号奇异性的能力强。针对金融时间序列具有突变性的特点，一定的消失矩是需要的。但需注意太高的消失矩，若信号中奇异点比较多时，在对小波系数进行阈值处理后，重构失真度可能增大，因此象收益率这样的序列，消失矩不能太高，否则会丢掉很多信息，而股价数据可适当高一点。

(4) 对称性。越对称的小波，在经过小波变换后，其偏差越可能小。因而有利于除噪后信号的恢复和重建。但该特性相对上述其它特性来说对除噪效果影响要小。

因为特性之间往往存在相互牵制，找到一个上述各方面特性都最优的小波是不可能的。我们针对一些常用的小波函数进行考察，如表 (1) 列示了它们的一些特性。

表 1　一些常用小波基的特性

	紧支撑	支撑宽度	对称性	正交性	消失矩
morlet	no	inf	yes	no	
haar	yes	1	yes	yes	1
mexican	no	inf	yes	no	1
meyer	no	inf	yes	yes	inf
dbN	yes	2N－1	far	yes	N
symN	yes	2N－1	near	yes	N
coifN	yes	6N－1	near	yes	2N
sinc	no	inf	yes	yes	no
biorNr.Nd	yes	2Nr+1, 2Nd+1	yes	no	Nr
gaussian	no	inf	yes	no	

根据上述分析，从综合情况来看 dbN (即 Daubechies 系列小波)、symN (Symlets 系列小波)、coifN (Coiflet 系列小波) 都比较合适，并且相对的，就小波系列而言 symN 是最佳选择。并且对于股价序列等相对比较平缓的序列可选择消失矩阶数稍高一点，即对应小波序列 N 取 4～8 都是可以的。但对收益率数据，因其奇异点密度非常大，消失矩不能太高，建议不要超过 4，也即 db2～db4，sym2～sym4 比较恰当。

4.3　*阈值的确定*。由前述阈值法消噪的原理可知阈值是区分信号和噪声的分水岭。显然它对除噪性能有至关重要的影响。阈值太高，会引起信号失真，太低则又去噪不完全。一般地，阈值的确定主要基于这样几项准则：

(1) 无偏风险估计准则 (rigrsure) 即一种基于 Stein 的无偏似然估计原理的自适应阈值选择方法。对每个阈值，求出对应的风险值，风险最小的即为所选。其具体算法为：

① 把用来估计阈值的小波系数向量取绝对值(设其长度为 n)，由小到大排序，然后将各元素平方，得到新的待估计向量 NV；

② 对 NV 的每个元素下标 k，按下式计算风险向量：

$$Risk(k)=\frac{n-2k+\sum_{j=1}^{k}NV(j)+(n-k)\cdot NV(k)}{n}. \tag{5}$$

③ 求出风险向量 $Risk$ 的最小点所对应的下标 k 值，从而得到阈值 Tr 为：

$$Tr=\sqrt{NV(k)}. \tag{6}$$

(2)固定阈值准则(*sqtwolog*) 设 n 为小波系数向量长度，则对应的阈值为：

$$Tr=\sqrt{2\log n}. \tag{7}$$

(3)混合准则(heursure) 它是 rigrsure 和 sqtwolog 准则的混合，当信噪比很低时，rigrsure 准则估计有很大噪声，这时采用固定阈值。其阈值计算方法为：首先判断两个变量 Eta 和 $Crit$ 的大小，它们的表达式分别为

$$\begin{cases} Eta=\dfrac{\sum_{i=1}^{n}|w_i|^2-n}{n} \\ Crit=\sqrt{\dfrac{1}{n}\left[\dfrac{\log n}{\log 2}\right]^3} \end{cases}. \tag{8}$$

其中 n 是待估计小波系数向量的长度，若 $Eta<Crit$，则选取固定阈值，否则选取 rigrsure 准则和 sqtwolog 准则的较小者作为本准则阈值。

(4) 极小极大准则(minimaxi) 也是一种固定阈值选择形式。该原理是在统计学中为设计估计量而采用的，由于去噪信号可假设为未知回归函数的估计量，则极小极大估计量是实现在最坏条件下最小均方误差最小的量。其阈值计算公式为：

$$Tr=\begin{cases}0, & n\le 32 \\ 0.3936+0.1829\cdot\dfrac{\log n}{\log 2}, & n>32\end{cases}. \tag{9}$$

以上阈值是针对标准差(小波域)为 1 的高斯白噪声而言，因此实际阈值应取为 $Tr\cdot\sigma$，其中 σ 为噪声的标准差，一般认为最小尺度上的小波系数大部分由噪声引起，因而以其估计 σ 的值。估计方法为，若 M_x 为含噪信号最小尺度上的小波系数绝对值向量的中位数，则：

$$\sigma=\frac{M_x}{0.6745}. \tag{10}$$

上述准则中 rigrsure 准则和 minimaxi 准则相对比较保守(仅将部分系数置零)，因此在信号的高频信息中有很少一部分在噪声范围内时，这两种阈值比较适合，而另外两种阈值选取规则，特别是固定阈值方法，能消除更多的噪声，但也有可能将有用信号的高频部分去掉。考虑到本文针对金融数据去噪的目的只是为方便后续分析，并不要求完全去掉噪声。对股票价格等数据而言，其信号频率较少地与噪声重叠，因此我们可以选用 sqtwolog 和 heursure 准则，使去噪效果明显些。但对收益率这样的高频数据，尽量采用保守的 rigrsure 或 minimaxi 准则来确定阈值，以保留较多的信号。

4.4　*分解层次确定*。根据多分辨率分析理论，高层分解的小波系数对应的是低频部分，而低频部分主要是由信号构成。因此分解层次越高，去掉的低频成分越多，去噪效果

明显，但失真度也相应增大。由于本文的目的旨在部分去除白噪声的干扰，因此为保守起见，分解层次不宜太高，最大不超过 5 层。对于那些本身波动性强的序列，比如收益率序列，由于其信号本身高频成分较多，更是不能取太高的层次，一般不超过 3 层。

5 实验结果及结论

选取深圳成指 2003 年 1 月 2 日～2003 年 12 月 31 日交易日收盘价数据，进行去噪实验。根据上述分析，选用 sym6 小波，采用 sqtwolog 阈值估计准则，将其进行多分辨率分解到第 4 层，结果如图 2 所示。很明显大部分小幅波动已经被去除，而信号的主要波动特征得到保留，因此，去噪效果是令人满意的。同样选取深圳成指 2003 年 1 月 2 日～2003 年 12 月 31 日按收盘价计算的日收益率数据。根据前面的分析，我们选用 sym2 小波、minimaxi 阈值估计准则、分解到第 2 层，结果如图 3 所示，噪声明显被消除，结果也是令人满意的。

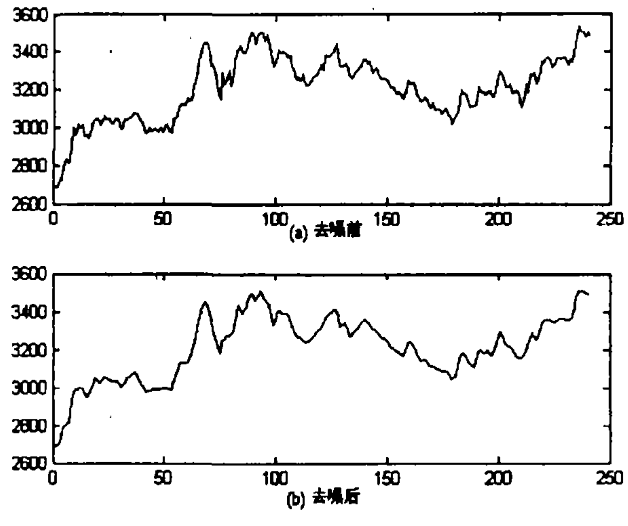

图 2　深圳成指小波去噪结果

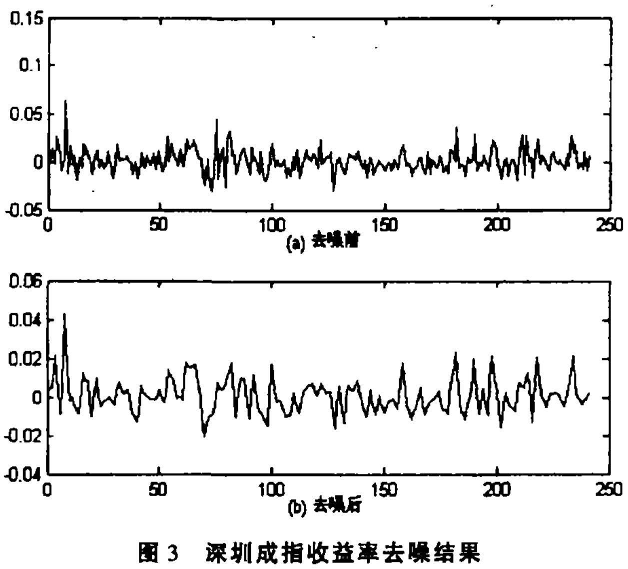

图 3　深圳成指收益率去噪结果

由此可见，小波方法可有效消除金融时间序列中的噪声，并能充分保留原信号的特征，为进一步的分析创造条件。但是，在具体应用时，相关的一些去噪参数的选取是非常重要的，必须根据序列的一些特征灵活选取，本文通过对金融时间序列特征分析，给出的参数选取方法是有效的。

参考文献：

[1] 彭玉华. 小波变换与工程应用 [M]. 北京：科学出版社，1999.

[2] Mallat S. Theory for multi - resolution signal decomposition: The wavelet representation [J]. IEEE transactions on pattern analysis and machine intelligence, 1989, 11 (7): 674～693.

[3] Mallat S, Hwang W L. Singularity detection and processing with wavelets [J]. IEEE Transactions on information theory, 1992, 38 (2): 617～643.

[4] Donoho D L, Johnstone I M. Ideal spatial adaptation by wavelet shrinkage [J]. Biometrika, 1994, 81 (3): 425～455.

[5] Donoho D L, Johnstone I M. Wavelet shrinkage asymptopia [J]. Journal of royal stastical society, 1995, 57 (2): 301～369.

[6] Donoho D L. Denoising by soft - thresholding [J]. IEEE transaction on Information, 1995, (3): 613～627.

[7] 詹姆斯 B 汉密尔顿. 时间序列分析 [M]. 北京：中国社会科学出版社，1999.

[8] 胡鹏飞. 基于小波变换的非线性去噪方法的研究 [C]. 西安：长安大学，2002.

[9] 郑君里，应启珩，杨为理. 信号与系统（第二版）[M] 北京：高等教育出版社，2000.

[10] Mathworks. Matlab wavelet toolbox user's guide (V6.5.1) [M]. Mathworks Inc, 2003.

[11] 胡昌华，张军波，等. 基于 MATLAB 的系统分析与设计—小波分析 [M]. 西安：西安电子科技大学出版社，1994.

[12] 郭代飞，高振朋，张坚强. 利用小波门限法进行信号去噪 [J]. 山东大学学报（自科版），2001, 36 (3): 306～311.

作者简介：兰秋军（1972－），男，湖南冷水江人，博士生，主要研究领域为金融信息处理、数据挖掘、管理决策支持系统；马超群（1963－），男，湖南岳阳人，教授，博士生导师，主要研究领域为金融工程、投资决策与风险管理；文凤华（1974－），男，博士生，湖南益阳人，主要研究领域为金融风险管理。

（本文责编：廖政权）