赫斯特指数（Hurst）指数及在 Excel 中的实现

韩海波

（兰州商学院统计学院，甘肃 兰州 730020）

摘要: Hurst 指数是描述非函数长周期的重要指标。它有别于传统单位根检验，可以发现时间序列存在的超长周期性，可以用于判断市场风险，但运算相当繁琐，单独利用 Excel 计算费时又费力，作者在充分理解 Hurst 指数内涵和应用的基础上，利用 Excel 的宏语言 VBA 编写宏程序轻松实现 Hurst 指数的计算，通过这一工作也希望能使 Hurst 指数能够得到广泛的应用。

关键词: Hurst 指数；R/ S 分析；Excel；VBA

中图分类号: F224.0　　　　**文献标识码:** A

1 Hurst 指数的基本概念

1.1 重标极差法(rescaled range analysis，R/S)

1.1.1 R/S 分析的产生和发展

标准统计分析假定系统基本是随机的，也就是说产生时间序列的过程有较大的自由度，而且序列内部的关系是非常复杂的，对它们进行确定性的说明是不可能的。只有概率论能帮助我们理解和利用这一过程。自然界和资本市场中充斥着大量的非线性随机和确定性系统。为了研究这些系统，就需要一个非参数的统计方法。

在标准高斯统计中，要求被观测的对象一定是独立同分布(IDD)的，但是在系统不是 IDD，要用什么样的方法去解决呢？1951 年英国水文学家 H·E·Hurst（1900-1978）发现了一个非常稳健的无参数统计方法。这一方法最初是用来考察尼罗河的流量变化的。但 Hurst 将他的研究扩大到许多自然系统，而且通过这个方法可以区分随机和非随机系统、趋势的持续、循环的持续。这一方法叫做“重标极差法"（rescaled range analysis），或 R/S 分析。通过这一方法可以区分具有长期非函数周期时间序列与随机序列。

爱因斯坦于 1908 年发表了关于布朗运动的论文，使得布朗运动成为随机游走的基本模型。爱因斯坦发现，分子随机运动的半径可以用涵盖时间的平方根来测度：

R=分子随机运动的半径；T=时间 $R = T^{0.5}$　　（1）

上式称作二分之一法则。在金融学中这一法则通常被用来对测量风险的标准差进行的年度化，比如用月收益的标准差乘 12 的平方根就得到年收益的标准差。Hurst 就是从这一原则出发通过新的统计量来对序列的随机性进行检验。

下面简要说明这一方法：

假设 $x=x_1$，x_2，…，x_n 为一时间序列，记做 x_i，i=1，2，…n。该时间序列的均值为 x_m，

$$\bar{x} = (x_1 + \cdots + x_n)/n \qquad (2)$$

标准差 s_n

$$s_n = \sqrt{\frac{(x_r - \bar{x})^2}{n}} \tag{3}$$

重标极差方法需要对原序列进行标准化，

$$Z_i = (x_i - \bar{x}); \qquad r = 1, \cdots, n \tag{4}$$

这样 Z 就具有零均值，$\bar{Z} = 0$，下面由 Z 产生一个累积时间序列 Y：

$$Y_1 = (Z_1 - Z_i); \qquad r = 2, \cdots, n \tag{5}$$

依据这个规则便可产生新的时间序列 Y，由定义可知，序列的最后一个 Y(Y_n)将一定为零。重标的极差 R_n，是 Y_i 的最大值减去最小值。

$$R_n = \max(Y_1, \cdots, Y_n) - \min(Y_1, \cdots, Y_n) \tag{6}$$

对于 R_n，下标 n 表明对于 x_1，x_2，…，x_n 是一个调整过的极差。因为 Z 已经被调整为零均值，Y 的最大值总是大于或等于 0，而最小值总是小于或等于 0。这样 R_n 将永远为非负。

运用重标极差 R_n，可以将等式（1）一般化，因为等式（1）仅仅适用于布朗运动，也就是随机过程。其形式如下：

$$(R/S)_n = c \cdot n^H \tag{7}$$

n 表示对应 x=x_1，x_2，…，x_n 的 R/S 值；c=a 常数

这里 R/S（rescaled range）被称为重标极差。通常情况下，R/S 随时间区间的增大而变大，H 通常称做赫斯特指数(Hurst Exponent)。由于对 R/S 进行了标准化处理，重标极差允许研究者对各种现象和时间序列进行比较。这是因为重标极差分析方法可以对没有特征规模变化的时间序列进行描述和分析。

1.1.2 赫斯特指数（Hurst Exponent）的计算与涵义

（1）赫斯特指数的计算

赫斯特指数可以由绘制的 Log(R/S)与 Log(n)图形逼近，其近似值就是其回归模型的斜率。

$$Log(R/S)_n = Log(c) + H \cdot Log(n) \tag{8}$$

假如一个系统是独立分布的，那么 H=0.5。Hurst 对尼罗河的流量研究表明，其 H=0.91，也就是说重标极差以快于时间平方根的速度在增长，这就意味着尼罗河的流量时间序列数据点的范围超过了一个随机过程可能覆盖的范围。而要覆盖更多的范围，序列间各点必定是相互影响的。尽管短期的自回归（Autoregressive, AR）会引起序列相关。但通过一阶和二阶自回归模型[AR(1)，AR(2)]对数据进行处理并不能消除这种相关性。

（2）赫斯特指数的含义

时间序列的 Hurst 指数居于 0-1 之间。以 0.5 为间隔，时间序列在不同的区间会表现不同的特性：

①H∈(0，0.5)：分形布朗运动。此时，时间序列的未来数据倾向于返回历史点,因此其发散的比标准布朗运动慢。可以证明，该序列在理论上会无数次的返回它的历史出发点。

②H=0.5：标准布朗运动。此时，序列为随机游走，表现马尔可夫链特性。

③H∈(0.5，1)：长期和无周期的循环。此时，时间序列有混沌性，其增量会表现出长期增长的特性。因此,一定范围的记录会持续相当长的时期，从而形成一个个大的循环。但是这些循环没有固定的周期,难以依靠过去的数据预测未来的变化。

④H=1：完全预测围。通常把这种过程称做均值回复过程。

1.2 R/S 分析实际操作

（1）假定一个时间序列 P_i，长度（数据个数）为 M，逐一计算序列的对数比，这样就可以生成一个新的对数序列 R_i，其长度为 N=M-1，这样做的目的是消除序列的短期自相关性，以满足 R/S 分析对观测对象独立的要求，新的时间序列为：

$$R_i = Log(\frac{P_{i+1}}{P_i}), \qquad i = 1,2,3,\cdots,M-1 \tag{9}$$

（2）以长度为 A 均分这个序列为 n 个相邻的子区间，这样 A·n=N。任一子区间表示为 I_a，a=1，2，3，…，n。在 I_a 中的元素表示为 N(k,m)，k=1，2，3，…，n，m=1，2，3，…，A。I_a 的均值为：

$$e_a = \frac{1}{A}\sum_{k=1}^{A} N_{k,a} \tag{10}$$

e_a=子区间 I_a 中 N_i 的均值。

（3）每个子区间 I_a 对于均值的累积截距(X_{ka})时间序列定义为：

$$X_{k,a} = \sum_{i=1}^{k}(N_{i,a} - e_a) \qquad k = 1,2,3,\cdots,n \tag{11}$$

（4）极差定义为

$$R_{I_a} = \max(X_{k,a}) - \min(X_{k,a}) \qquad 1 \le k \le n \tag{12}$$

（5）子区间 I_a 的标准差为：

$$S_{I_a} = \sqrt{\frac{\sum_{k=1}^{A}(N_{k,a} - e_a)^2}{A}} \tag{13}$$

（6）每一个 R_{I_a} 均由对应的 S_{I_a} 进行标准化。则 R/S 定义为：

$$(R/S)_n = \frac{1}{n}\sum_{a=1}^{n}\frac{R_{I_a}}{S_{I_a}} \tag{14}$$

（7）不断增加 A 长度，并重复 1——6 步，直到 A=(M-1)/2。以 Log(n)为解释变量，Log(R/S)为被解释变量，进行线性回归，Log(R/S)=Log(c) + H·Log(n)+ε。计算得到的方程中斜率就是赫斯特指数 H 的估计值。

1.3 R/S 分析对循环的探测

通过观察我们通常认为价格的波动具有周期性的规律，但这并不是我们一般所说的周期，周期是一个可以严格用正弦波来表达的概念，而通常经济指标波动表现出来的周期性波动并没有精确的周期长度，不同波的相位和振幅经常发生偏移。近期的研究成表明，存在着非周期循环这个概念，这些循环有一个平均持续周期，换言之，一个未来的循环是不确定的，但它大致位于平均的循环周期附近。这正好符合我们对经济指标变动的周期的观测。通过 R/S 分析，我们就可以找到这个平均的循环长度。

在一个时间序列里，R/S 分析不但能够发现持久性长期记忆还能估计周期性或非周期性循环的长度。而且这一循环对于噪声是稳定的，这对于经济分析是十分有效的。这里非周期循环是指这个循环没有绝对的频率，但有一个平均的频率，在相空间中表现为一个有限集，混沌理论中称作奇异吸引子。以往在时间序列分析中，我们的焦点一直放在规则、周期循环上。在傅立叶分析中，是假定形成时间序列不规则形状的是许多正弦波的集合组成，它们各自有不同的频率和振幅。而谱分析是试图用不明显的循环来分割时间序列，使其成为正弦波的集合。然而这些方法均不能是分析结果较好的符合实际观测。而非周期循环却可以对资本市场的循环做出解释。

1.4 V 统计量的计算及非周期循环的判断

1.4.1 V 统计量的计算

V 统计量计算的主要目是找到非周期循环的长度。计算公式为：

$$V_n = \frac{(R/S)_n}{\sqrt{n}} \tag{15}$$

其中 n 为观测的天数。

1.4.2 非周期循环的判断

在 V 统计中，若 V 线表现为水平，则序列为随机。反之，如果有上升趋势，则表现为序列的长期记忆，图形的拐点处即为序列的非周期循环的长度。

2 Hurst 指数的计算程序

2.1 Hurst 指数的程序设计

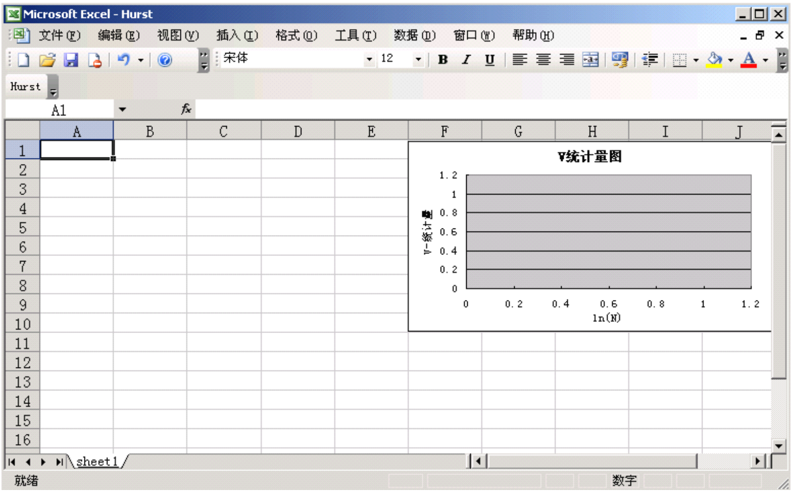
现有常用的统计软件中可以计算赫斯特指数的非常少，而根据前面所阐述的理论基础可以看出赫斯特指数计算量相当大，如果单纯依靠手工计算是很难做到的，为了方便更多的统计学者在实际的操作过程中可以计算赫斯特指数，我们决定编写程序来解决赫斯特指数复杂的计算问题。

众所周知,如果在 Microsoft Excel 中经常重复某项任务，可以使用宏将其变为可自动执行的任务。宏就是贮存在 Visual Basic 模块中的一系列命令和函数，并且在需要执行该项任务时可随时运行的程序。在 Excel 中记录宏就如同用磁带录音机录制音乐，然后运行宏使其重复执行或"回放"这些命令。可以使用 Visual Basic 编辑宏程序，也可以利用 Microsoft Excel 自带的 Microsoft Visual Basic 宏编辑器直接编制宏。由于宏程序可以方便灵活的加载到 Microsoft Excel 中，而 Excel 又是统计中最常用的软件。我们选择使用 VBA 语言编写计算赫斯特指数的宏程序。

2.2 Hurst 指数程序的特点

本程序实际上是一个 Excel 文档，用户只需将逐一计算对数比后的时间序列数据输入到 Excel 的 A 列中，然后直接点击左上角的“Hurst”按钮，即可得出该序列的赫斯特指数和相应的 V 统计量图，为了使程序更加方便实用，我们将程序和所有附带文件打包为标准的安装软件，成功安装后即可直接在 Excel 中计算赫斯特指数，详细使用方法请参考程序使用说明。（源程序请参见附录），

软件界面如下图：



3 外汇市场 R/S 实证分析

3.1 数据来源

我们采用 1998 年 9 月 19 日——2006 年 5 月 18 日的欧元对美元汇率的收盘价（共 2799 个数据点）为分析对象。由于数据量已超过了 2000 这个经验值，因此样本数量是足够。

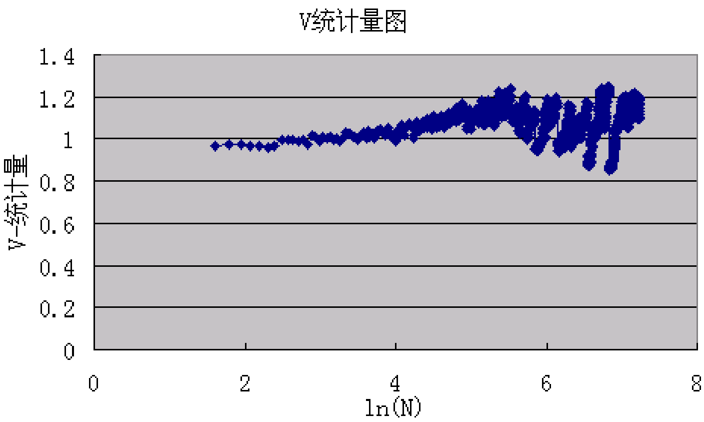
3.2 数据初步处理

根据 R/S 分析的要求对数据进行对数一阶差分（逐日对数收益）处理，消除序列的短期记忆。计算公式为：

$$X_t = \ln(\frac{P_t}{P_{t-1}}) \tag{16}$$

3.3 赫斯特指数（H）的计算

将 2799 个经过处理的欧元对美元的汇率数据代入程序中进行计算，最后结果是：Hurst=0.625657443267641。V 统计量图如下所示：



3.4 计算结果的非参数显著性检验

为了验证计算所得 H 值的可靠性，我们打乱原时间序列的顺序，重新计算 H。而这时的 H 应显

著值地接近 0.5。通过计算检验，结果证明 H 值是显著的。

为了和传统单位根检验结果进行比较，我们对原序列进行 ADF 检验，检验结果如下：

Null Hypothesis: X has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=27)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-56.73592	0.0001
Test critical values:	1% level	-3.432497	
	5% level	-2.862374	
	10% level	-2.567259	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(X)

Method: Least Squares

Date: 05/20/06 Time: 17:13

Sample (adjusted): 2 2798

Included observations: 2797 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-1.070497	0.018868	-56.73592	0.0000
C	6.27E-06	0.000120	0.052360	0.9582

R-squared	0.535248	Mean dependent var	-2.47E-08
Adjusted R-squared	0.535082	S.D. dependent var	0.009281
S.E. of regression	0.006328	Akaike info criterion	-7.286861
Sum squared resid	0.111932	Schwarz criterion	-7.282616
Log likelihood	10192.67	F-statistic	3218.964

Durbin-Watson stat	1.998557	Prob(F-statistic)	0.000000

ADF 检验是一个被最早提出来的检验单位根零假设的检验，而且是实际中最常用的检验。不过，其他的检验随后被相继提出来，他们中的许多都比 ADF 检验有更高的功效，因为一个功效更强的检验有能力更好的将单位 AR 根个数据接近于 1 但小于 1 的根区分开来。Elliott，Rothenberg 和 Stock 于 1996 年提出了 DF-GLS 检验，它能更好的区别单位根的零假设和平稳性的备则假设。试验证明假设 Y 实际上是一个平稳的 AR（1），其自回归系数为 0.95，且有两百个观测值，计算一个没有时间趋势的单位根检验，那么 ADF 检验在 5%的显著性水平下真确拒绝零假设的概率约为 31%，而 AF-GLS 检验相应的概率为 75%。因此，为了保证计算结果的准确性我们同时对原数据进行 DF-GLS 检验，结果如下：

Null Hypothesis: X has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=27)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-56.28992
Test critical values:	1% level	-2.565789
	5% level	-1.940937
	10% level	-1.616623

*MacKinnon (1996)

DF-GLS Test Equation on GLS Detrended Residuals

Dependent Variable: D(GLSRESID)

Method: Least Squares

Date: 05/20/06 Time: 17:16

Sample (adjusted): 2 2798

Included observations: 2797 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.

GLSRESID(-1)	-1.062462	0.018875	-56.28992	0.0000

R-squared	0.531231	Mean dependent var	-2.47E-08
Adjusted R-squared	0.531231	S.D. dependent var	0.009281
S.E. of regression	0.006354	Akaike info criterion	-7.278968
Sum squared resid	0.112899	Schwarz criterion	-7.276846
Log likelihood	10180.64	Durbin-Watson stat	1.997645

结果表明，在 1%的置信度下，两种检验同时拒绝单位根假设，因此 ADF 检验认为序列是随机的。但通过 Hurst 指数分析发现，由于其指数大于 0.5，因此原时间序列并不是真正意义上的随机序列，而是具有长期的相关性。结果出现这样的矛盾源自于在进行 ADF 检验时，我们对于辅助模型的滞后期一般设定为 28，这就会使得不能对更为长期的相关性做出准确判断。

这里需要我们注意的是，hurst 指数的计算对时间序列数据量要求较大，因此对于数据较少的时间序列它并不能发挥其特长，甚至产生错误的判断，这时就必须使用类似 ADF 检验等较为常用的方法进行研究。

参考文献

[1]H.E. Hurst (1951), “Long-term storage of reservoirs”, *Trans. Amer. Soc. Civil Eng.* 116 770-808

[2]B. Mandelbrot, J.R. Wallis (1969), “Robustness of the rescaled range R/S in the measurement of non-cyclic long run statistical dependence”, *Water Resour. Res.* 5 909

[3]Hurst H.E,Black R.P,Simaika Y.M.Long –term storage capacity of reservoirs[J]. Transactions of the American Society of Civil Engineer 116, 1951

[4]Moody J.Improved estimates for Rescaled Range and Hurst exponents[A].

Neural Networks in Financial Engineering.London,October 1995

[5] Bill Jelen.Excel 2003 VBA 与宏[美] 电子工业出版社 2005

[6] 埃得加 E 彼得斯.分形市场分析[美] 经济科学出版社 2002

[7] 黄睿.Excel VBA 应用程序专业设计实用指南 电子工业出版社 2005

[8] 李子奈，叶阿盅.高级计量经济学[M].北京:清华大学出版社.2000

附录 1:

以下便是利用 Excel 里的 VBA 计算 Hurst 指数的源代码:

```
Sub Hurst()

    '变量和数组的定义

    Dim Data()

    Dim Array1()

    Dim Array2()

    Dim R()

    Dim S()

    Dim Result()


    Dim NoOfDataPoints As Integer

    Dim NoOfPlottedPoints As Integer

    Dim NoOfPeriods

    Dim PeriodNo


    Dim n As Integer

    Dim A As Integer

    Dim i As Integer

    Dim m

    Dim e

    Dim RS


    '验证 A 列中是否输入数据

    If Worksheets("Sheet1").Range("A1").Value = 0 Then MsgBox ("请在 A 列输入数据!"): Exit Sub

    '清空主要的单元格

    Worksheets("Sheet1").Range("B3").Value = "Hurst = "

    Worksheets("Sheet1").Range("C3").Value = Null


    '统计数据的个数

    i = 1
```

```
Do While i < 10000
i = i + 1
If Worksheets("Sheet1").Cells(i, 1).Value = 0 Then Exit Do
Loop
NoOfDataPoints = i - 1

ReDim Data(NoOfDataPoints)

'验证 A 列的数据后将其加载到数组中
i = 1
counter = 1
Do While counter <= NoOfDataPoints
    Set curCell = Worksheets("Sheet1").Cells(i, 1)
    If Application.WorksheetFunction.IsNumber(curCell.Value) Then
        Data(counter) = curCell.Value
        counter = counter + 1
    End If
    i = i + 1
Loop

'运行以下代码则可以直接输入原数据
'i=2
'Do While i <= NoOfDataPoints
    'Data(i - 1) = Log(Data(i) / Data(i - 1))
    'i = i + 1
'Loop


ReDim Result(NoOfDataPoints / 2, 2)

 '进入主循环
```

```
A = 2
Do While A <= NoOfDataPoints / 2
 '再次定义数组变量
   NoOfPeriods = NoOfDataPoints / A
   ReDim Array1(Int(NoOfPeriods))
   ReDim Array2(A, NoOfPeriods)
   ReDim S(Int(NoOfPeriods))
   ReDim R(Int(NoOfPeriods))
   RS = 0

   '求得各个子区间均值
   i = 1
   Do While i <= NoOfPeriods
      e = 0
      For PeriodNo = 1 To A
         e = e + Data(PeriodNo + (i - 1) * A)
      Next PeriodNo
      Array1(i) = e / A
   i = i + 1
   Loop

   '求得各个子区间的累积截距和极差
   i = 1
   Do While i < NoOfPeriods
      m = 0
      e = 0
      For PeriodNo = 1 To A
        m = m + ((Data(PeriodNo + (i - 1) * A) - Array1(i)) ^ 2)
        e = e + (Data(PeriodNo + (i - 1) * A) - Array1(i))
        Array2(PeriodNo,i) = e
      Next PeriodNo
```

```
        '比较最大值与最小值
        Maxi = Array2(1,i)
        Mini = Array2(1,i)
        For n = 1 To A
            If Array2(n,i) > Maxi Then Maxi = Array2(n,i)
            If Array2(n,i) < Mini Then Mini = Array2(n,i)
        Next n
        '求得 R/S 值
        R(i) = Maxi - Mini
        S(i) = Sqr(m / A)
        RS = RS + R(i) / S(i)
    i = i + 1
    Loop

      '将 V 统计量表的数据输出到 Excel 表格中
    Worksheets("sheet1").Cells(A + 2, 5).Value = (RS / NoOfPeriods) / Sqr(A)
    Worksheets("sheet1").Cells(A + 2, 6).Value = Log(A)

    '将计算结果装入 Result（）数组中
    Result(A, 1) = Log(A)
    Result(A, 2) = Log(RS / NoOfPeriods)

    A = A + 1
    Loop

'对方程 Log(R/S)=Log(c) + H · Log(n)+ε进行线性回归，估计出斜率 H 就是 Hurst 指数
sumx = 0
Sumy = 0
Sumxy = 0
Sumxx = 0
```

```
    NoOfPlottedPoints = NoOfDataPoints / 2


    For i = 2 To NoOfPlottedPoints
        sumx = sumx + Result(i, 1)
        Sumy = Sumy + Result(i, 2)
        Sumxy = Sumxy + (Result(i, 1)) * (Result(i, 2))
        Sumxx = Sumxx + (Result(i, 1)) * (Result(i, 1))
    Next i


    H = (Sumxy - ((sumx * Sumy) / NoOfPlottedPoints)) / (Sumxx - ((sumx * sumx) / NoOfPlottedPoints))
    Worksheets("sheet1").Range("C3").Value = H

End Sub
```

需要说明的是:

①本程序最多只能计算 20000 个数据。

②本程序的默认输入列是 A 列，请将数据直接输入 A 列中，从 A1 开始输入，请不要空格。

③请先将原始数据进行逐一计算的对数比，然后再输入程序进行计算。否则得到的将不是正确结果。

Hurst Index and The Realization in Excel

Han Haibo

(Statistics School of Lanzhou, Business School of Lanzhou, Lanzhou Gansu 730020)

Abstract: Hurst index is a key indicator of description of long-period of non-function. It is different from the traditional unit root test, and it can find the existence of long time series periodicity, can be used to determine the market risk, but rather tedious calculations, a separate calculation using Excel time-consuming and laborious. The authors fully understand the meaning and application of Hurst index of based on the use of Excel's macro language VBA easily prepared Hurst index calculation, also hope that through this work will enable Hurst index can be widely used.

Key words: Hurst index; R / S analysis; Excel; VBA

收稿日期: 2006-05-14

作者简介: 韩海波，兰州商学院统计学院，主要从事经济计量分析研究。