

文章编号: 2095-2163(2020)08-0131-04

中图分类号: TP391

文献标志码: A

基于可视图模式的黄金价格波动性趋势预测

赵元英, 许洁林, 顾长贵

(上海理工大学 系统科学系, 上海 200093)

摘要: 传统黄金波动时间序列分析中使用基于统计理论的方法, 掩盖了序列中以模式形式存储的信息, 为此本文提出了一种基于可视图模式的时间序列分析方法。首先, 将黄金波动序列划分为一组大小相等的新时间序列; 其次, 对这些新序列进行图形化处理, 得到状态转移网络, 其中每个可视图模式作为网络节点, 可视图模式之间的转移频率作为节点之间连边权重。实验结果表明: 由黄金波动序列构建的状态转移网络中, 8号模式为特征模式, 具有强自相关特征, 并且这种强自相关特征在黄金波动序列中具有长期相关性, 说明黄金波动程度是趋向于增大的, 整体呈上升趋势, 但在市场其他因素影响下, 不会无限增长, 最终呈现出每过一段时间出现一个高点后迅速下降的现象。

关键词: 可视图模式; 状态转移网络; 时间序列分析; 长期相关性

Trend forecast of gold price volatility based on visibility graph pattern

ZHAO Yuanying, XU Jielin, GU Changgui

(Department of Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] The analysis based on statistical theory in traditional gold volatility series analysis will cover the information stored in the form of patterns in the series, therefore, a time series analysis method based on visibility-graph patterns is proposed. First, the gold volatility series is divided into a group of new time series with equal size. Then these new series are graphically processed to obtain state transition networks, where the nodes are visibility-graph patterns, and the transition frequency between visibility-graph patterns is the connection weight between nodes. The experimental results show that in the state transition network constructed by the gold volatility series, the No. 8 pattern is a characteristic pattern with strong autocorrelation characteristics, and this strong autocorrelation characteristic has a long-term correlation in the gold volatility series. The above results indicate that the degree of gold volatility tends to increase overall. However, under the influence of other factors in the market, it will not grow all the time, and eventually shows a phenomenon of rapid decline after a high point over a period of time.

[Key words] visibility graph pattern; state transition network; time series analysis; long-term correlation

0 引言

黄金价格的形成是多种因素综合作用的结果, 具有不确定性、随机性和非线性等特点, 因此其波动十分剧烈^[1]。对黄金价格进行有效的时序分析和预测, 不仅有利于黄金生产者和投资者发现经济运行的内在规律, 加深对黄金市场特点的了解, 为其提供可靠的决策依据, 还有利于提高市场风险防范能力, 对了解、评价和预测市场有潜在的用途^[2]。因此, 尽管人们已经对黄金价格和波动性进行了长期的研究, 预测黄金价格和波动性仍然是重要的研究课题。

近年来, 由于黄金价格预测受到了广泛的关注, 因此国内外学者开展了大量的相关研究。例如, 孙泽萍采用 EViews7 对数据进行多元回归分析, 实证分析了黄金交易价格的影响因素^[3]; 景志刚等利用小波分析以及 LS-SVM 与 ARIMA 模型对黄金价格

建模预测^[4]; Yu 等利用 ARFIMA 模型对黄金价格进行预测^[5]。这些研究关注的黄金价格序列的回归分析, 相关性检验等统计方法, 提取这些特征的方法, 都是建立在数理统计理论和概念之上, 在统计平均过程中时间序列以模式形式存储的信息被掩盖^[6]。

一般而言, 黄金市场是一个高维系统。黄金价格序列是从一个高维投影到一维的结果。在黄金价格序列中存在非平凡的模式, 这些模式并不是孤立的, 而是具有相互依赖性。一个有趣的想法是, 把模式看作节点, 模式之间存在的依赖关系看作节点间的连接, 时间序列就可以被映射到一个复杂的网络中, 通过网络结构, 可以研究从微观到宏观的模式以及模式之间的联系。由此, 构建复杂网络有效解决了在统计过程中时间序列中的结构信息被掩盖的问题。

作者简介: 赵元英(1996-), 女, 硕士研究生, 主要研究方向: 数据挖掘、系统科学与复杂网络; 许洁林(1998-), 女, 本科生, 主要研究方向: 数据挖掘; 顾长贵(1982-), 男, 博士, 副教授, 上海市高校青年东方学者, 主要研究方向: 系统科学与复杂网络、非线性动力学。

通讯作者: 顾长贵 Email: gu_changgui@163.com

收稿日期: 2020-05-28

近年来许多学者对基于复杂网络的时间序列分析做了大量研究。例如,Zhang等第一次提出了将非周期时间序列映射到网络的方法,把时间序列中的周期提取出来,并将每个周期看作为节点。对于每一对节点,计算相关强度,得到节点之间相关强度矩阵。当强度大于一个阈值时,节点之间连边,进一步构造出一个复杂网络^[7];Xu等提出了一种将每个节点与其最近的k个邻居连接起来的方法^[8];有文献根据节点之间的关联强度来构建网络,是目前广泛使用的一种递归网络方法^[9]。

在上述方法中,时间序列被映射为一个静态网络,因此不能反映出复杂系统的动态行为。为了解决这个问题,McCullough等提出了一种基于可视化小图的时间序列分析方法,将固定大小的窗口沿时间序列滑动,并将得到的时间序列片段映射为小图,每个小图就作为对应时间段内的状态描述^[10]。穆图等利用可视性算法构建了可视图,并将各个可视图连接为网络,其中通过对分数布朗运动以及股票市场的研究表明,基于可视图模式的时间序列分析方法能够有效提取复杂系统中的动态特征^[11]。本文采用穆图等提出的基于可视图模式的时间序列分析方法来分析黄金价格序列。

1 数据和他方法

1.1 数据来源

本文数据来源于公共网站(<https://cn.investing.com>),包括黄金每日开盘、收盘、最高以及最低价格。本文以每日黄金的最高和最低价格为研究对象,时间跨度为1990年4月2日至2020年3月31日,共7624个最高(P_{\max})和最低(P_{\min})价格记录。黄金每日波动率序列用 $(\ln P_{\max} - \ln P_{\min})$ ^[12]来计算,并表示为序列 $x = \{x_1, x_2, x_3, \dots, x_N\}$,其中N表示序列长度,如图1所示。

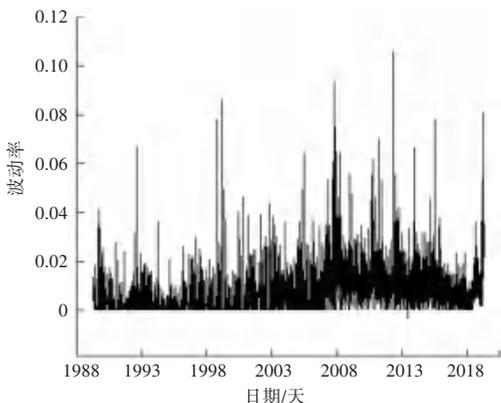


图1 黄金波动率时间序列

Fig. 1 Time series of gold volatility

1.2 可视图模式分析方法

本文采用穆图等提出的基于可视图模式的时间序列分析方法,具体步骤如下^[11]:

步骤1 将大小为s的窗口沿序列 $\{x_1, x_2, x_3, \dots, x_N\}$ 滑动,截取到序列片段,式(1):

$$P_m = \{x_m, x_{m+1}, \dots, x_{m+s-1}\},$$

$$m = \{1, 2, \dots, N - s + 1\}. \quad (1)$$

这些序列片段是对不同时间段里相应局部状态的描述。

步骤2 将 P_m 根据式(2)映射为可视图,每个可视图即为一个可视图模式。将每个数据值作为结点,如果两个结点可以看到对方,则将它们连接起来,即它们之间存在一条直线。例如,取任意两个数据值 x_a 和 x_b ,如果 x_a 和 x_b 之间的每个点 x_c 都满足式(2),则 x_a 和 x_b 是彼此可见。

$$x_c \leq x_a + (x_b - x_a) \cdot \frac{b - c}{b - a}. \quad (2)$$

通过这一规则,得到了第m个时间序列片段映射的可视图模式。用邻接矩阵表示第m个时间序列片段的可视图模式,即 x_a 和 x_b 是可见的,则相应的邻接矩阵元 $g_m(a, b) = 1$,若 x_a 和 x_b 是不可见的,则 $g_m(a, b) = 0$ 。从而得到第m段的一个 $s \times s$ 的矩阵。序列中所有片段用相应的邻接矩阵规则表示,得到 $G = \{g_1, g_2, \dots, g_{N-s+1}\}$ 。

步骤3 建立局部状态之间的状态转移网络。即在时间序列中,如果b时刻的一个状态紧接在a时刻的另一个状态之后,则构建从 g_a 到 g_b 的有向连接,从而得到状态链,式(3)

$$g_1 \rightarrow g_2 \rightarrow \dots \rightarrow g_{N-s+1}. \quad (3)$$

将状态链中相同的状态归并为一个可视图模式,并为每个可视图模式编号。即,若 $g_2 = g_5$,则 g_5 被 g_2 代替,按照此规则对状态链进行迭代。从状态链中可以得到模式之间的转移概率,即从一种可视图模式经过一个时间步转为另一个模式的频率作为其权重。由此,状态转移网络可以被定义,每个可视图模式被定义为状态转移网络的节点,可视图模式之间的转移频率为其对应的连接权重值。通过观察状态转移网络的结构特征,可以找到时间序列中隐含的动力学特征。

在实际计算中,一个关键问题是如何选择合适的窗口大小。由于状态模式的数量随窗口大小的增加呈指数增长,若窗口大小的值选取过大,会因为模式太多导致计算困难,也会对时间序列中的噪声产生较大影响,从而难于发现系统运行规律;若窗口大

小的值选取过小,则可表示的状态模式的数量有限,无法区分状态之间的差异。窗口大小选取的准则:对于现实中的时间序列,若该时间序列存在一个自然周期,则选择自然周期作为窗口大小^[13],例如股票市场中的一个星期或一个月。因为黄金期货的自然周期为五天,并且窗口为 5 时的模式数量比较适中,所以选择窗口大小 $s = 5$ 。图 2 给出了黄金波动序列中存在的所有唯一可视图模式,并为每个模式分配了一个特定的编号,显示在它的拓扑结构下。

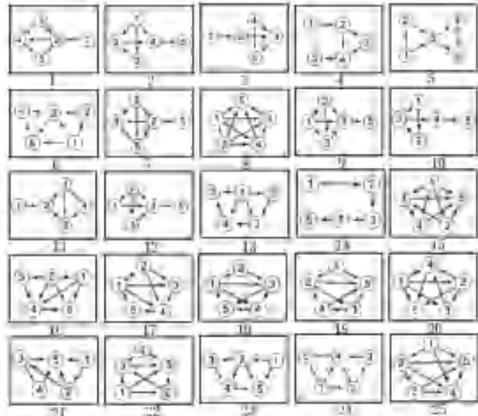


图 2 黄金波动序列中出现的所有可视图模式

Fig. 2 All visibility graph patterns in the gold volatility series

2 实验与结果分析

本文根据上述构建状态转移网络步骤,将黄金波动时间序列映射到网络中,映射的状态转移网络形式表示由黄金波动序列构建的原状态转移网络,如图 3 所示。为了方便观察网络的动力学特征,本文将网络中权重小于 0.01 的弱连接过滤掉,形成强状态转移网络(简称强网络),如图 4 所示。为了与无规律序列做对比,突出原始序列所隐藏的规律,本文对黄金波动序列的顺序打乱处理,重新构建打乱序列的状态转移网络,并过滤掉网络中权重小于 0.01 的弱连接,称为改组状态转移网络(简称改组网络),如图 5 所示。图 4 和图 5 中节点的大小表示相应可视化小图的出现频率,节点之间连接的宽度与其权重成正比。



图 3 原状态转移网络

Fig. 3 Original state transition network

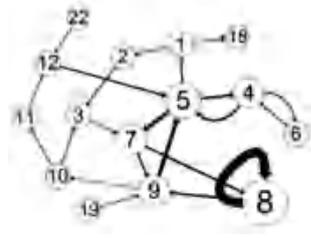


图 4 强状态转移网络

Fig. 4 Strong state transition network

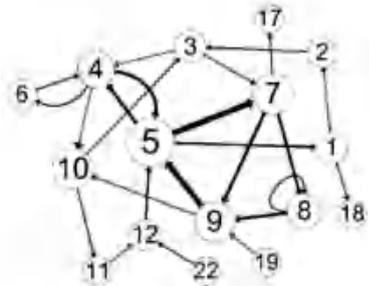


图 5 改组状态转移网络

Fig. 5 Shuffled state transition network

如图 3 所示,存在自连接 $14 \rightarrow 14$ 和 $8 \rightarrow 8$,其中自连接 $14 \rightarrow 14$ 在随后的强网络(图 4)和改组网络(图 5)中消失。由于黄金市场具有高度的随机动态性,并且受市场其他因素影响,使得这些网络中必然存在弱连接,因为弱连接反映了是黄金市场的不确定性和复杂性。通过观察图 4 与图 5,可以发现状态转移网络与改组网络中节点的大小和连接权重有显著的差异,这说明原始网络中的某些状态并非偶然,而是系统运行规律的反应。

状态转移网络与改组状态转移网络中存在的差异可以通过网络中节点的度比率来定量衡量。一个节点在时间序列内出现的频率在本文中称为度,度比率为原始状态转移网络中节点度与改组后状态转移网络中节点度的比值。黄金波动序列中所有状态的度比率,如图 6 所示。如果某个节点的度比率明显大于 1,即节点在原始时间序列中的出现频率明显大于打乱时间序列中的出现频率,则该节点称为 Motif^[10],可以作为时间序列的全局代表。可以找到一个 Motif,即编号为 8 的可视图模式为 Motif 8,其度比率为 3.449,远大于 1,这说明黄金波动序列中模式 8 的大量出现并不是偶然,而是由系统的性质决定的,一定程度上反应了系统的变化规律。

为进一步研究模式 8 在黄金波动序列中的出现是否具有长程相关性,本文提取了模式 8 在黄金波动序列中出现位置的间隔,形成一个位置间隔序列,并通过重标极差分析法(简称 R/S 分析法)计算赫斯特指数(简称 H)来说明。若 $H = 0.5$,则表明原

序列是随机游走过程;若 $0.5 < H < 1$, 则原序列为有偏随机游走, 并且序列中存在正效应;若 $0 < H < 0.5$, 说明序列中存在负效应。模式 8 的赫斯特指数为 0.69, 这说明模式 8 的间隔序列为有偏随机游走过程, 并且存在正效应, 会大概率发生聚集出现的现象, 同时揭示了模式 8 存在大量自转移的原因, 充分说明了模式 8 的自转移现象并不是偶然, 如图 7 所示。

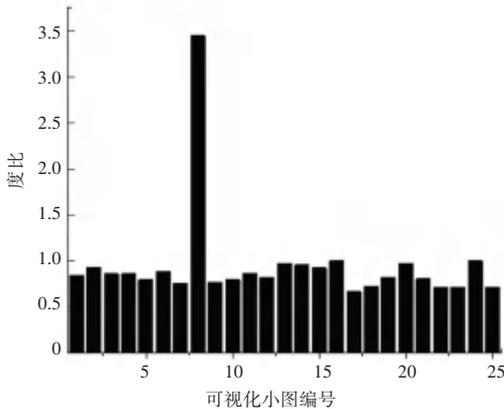


图 6 黄金波动序列中所有状态的度比率

Fig. 6 Degree ratios of allstates in the gold volatility series

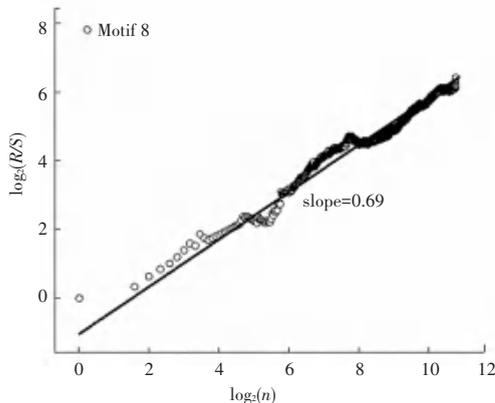


图 7 R/S 分析法分析位置时间序列的结果

Fig. 7 Results of R/S analysis on location time series

通过上述分析得出两点结论:(1)模式 8 的大量出现并不是偶然, 而是系统运行规律的反应;(2)模式 8 的出现具有正效应, 会出现大量自转移现象。由图 2 可知, 编号为 8 的可视图模式是一个全连接的状态, 由可视图模式的定义可以得出, 8 号模式代表的是一种二阶导数大于零的情况, 结合上述两点结论说明黄金波动序列中经常出现并且连续出现二阶导数大于零的情况, 这一结论可以在图 1 中得到印证。

3 结束语

在传统的黄金价格序列预测问题的研究中, 使用基于统计学理论和概念的方法意味着对序列进行

了一个统计平均的过程, 虽然取得了一些具有启发性的结果, 但是序列中存在的模式及其结构信息不能够好的展现。因此, 本文采用了一种基于可视图模式的时间序列分析方法来研究黄金波动序列。

实验结果表明, 黄金波动序列中存在两种比较显著的规律:(1)模式 8 具有很高的出现频率;(2)模式 8 的出现具有正效应, 存在大概率的自转移。这说明黄金波动序列中经常出现并且连续出现二阶导数大于零的情况, 即黄金波动程度是趋向于增大的, 整体上是呈上升趋势;但是另一方面市场受到大量因素影响, 波动程度不会无限制的增长。所以在这两种作用的影响下, 经过一段时间的发展, 出现一个较高的峰值, 但在市场的调控下迅速下跌, 进入新一轮酝酿峰值的阶段。

由于黄金价格序列具有分形性质, 在未来的研究中, 要进一步探索黄金价格在其他时间尺度运行的规律, 希望揭示出黄金价格运行的更深刻规律。

参考文献

- [1] CHEN L, ZHANG X. Gold Price Forecasting Based on Projection Pursuit and Neural Network [J]. Journal of Physics Conference, 2019, 1168(6):062009.
- [2] YANG Y. Gold Price Forecast based on ESMD Multi-Frequency Combination Model [J]. Iop Conference, 2018, 466(1):012031.
- [3] 孙泽萍. 黄金交易价格变动的影响因素分析 [J]. 时代金融(中旬), 2017, (2):46-47, 52.
- [4] 景志刚, 施国良. 基于小波分析的 LS-SVM-ARIMA 组合模型的金价格预测 [J]. 黄金, 2017, 38(5):5-8, 14.
- [5] LIN Yu, KONG Liuliu, LIU Pei. Gold Price Prediction Based on ARFIMA Model [J]. Journal of Nanhua University (Social Science Edition), 2010, 11(1):36-39.
- [6] WANG Y, WENG T, DENG S, et al. Sampling frequency dependent visibility graphlet approach to time series [J]. Chaos An Interdisciplinary Journal of Nonlinear Science, 2019, 29(2).
- [7] ZHANG J, SMALL M. Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics [J]. PHYSICAL REVIEW LETTERS, 2006, 96(23):238701-0.
- [8] X, Xu, J. Superfamily phenomena and motifs of networks induced from time series [J]. Proceedings of the National Academy of Sciences, 2008.
- [9] STEPHEN M, GU C, YANG H. Visibility Graph Based Time Series Analysis [J]. Plos One, 2015, 10(11):e0143015.
- [10] McCullough M, Small M, Stemler T, et al. Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems [J]. Chaos, 2015, 25(5):053101.
- [11] STEPHEN M, GU C, YANG H. Visibility Graph Based Time Series Analysis [J]. Plos One, 2015, 10(11):e0143015.
- [12] BARIVIERA, AURELIO F. The inefficiency of Bitcoin revisited: A dynamic approach [J]. Economics Letters, 2017, 161:1-4.
- [13] LACASA L, LUQUE B, BALLESTEROS F, et al. From time series to complex networks: The visibility graph [J]. Proceedings of the National Academy of sciences, 2008, 105(13):4972.