

文章编号: 1005-0523(2009)05-0079-05

基于 ARIMA 模型的外汇汇率时间序列预测研究

张奕韬

(华东交通大学 软件学院, 江西 南昌 330013)

摘要:利用数据挖掘技术分析外汇汇率时间序列,从时间序列中获得正确的、隐含的、潜在的信息对于金融领域研究具有重要的现实意义。通过数据挖掘中的 ARIMA 模型,以某银行的外汇汇率时间序列为研究对象,采用差分方法和建模规则,对外汇的卖出价进行了建模与预测。通过与逐步自回归预测模型相比较,ARIMA 模型对外汇汇率时间序列数据具有很强的预测能力。

关键词: 外汇汇率; 时间序列; ARIMA 模型; 预测
中图分类号: TP274+.2 **文献标识码:** A

利用计算机数据挖掘技术,对外汇汇率时间序列的趋势预测分析,给出短期外汇预测结果,具有很重要的现实意义^[1]。目前,通过离散化的方法对外汇汇率时间序列抽取静态属性^[2],把时间序列分成不同的模式,利用相邻模式的相关属性组建新的数据库^[3],并且使用粗糙集的方法进行属性和值的简约化,生成决策规则,达到一定的预测效果。然而,实际的外汇汇率时间序列是以小时、分钟甚至秒为抽样频率的高频数据,而且数据频率越高,获得的汇率信息就越多,那么预测的结果将会更精确。离散化就破坏了时间的连续性,减少了信息量,降低了预测的精度^[4]。同时,在国外有学者通过 BP 神经网络的方法和自回归的方法以高频数据为样本对股指的收益进行预测^[5]。但是,在对样本的选取中就必须考虑到样本的代表性和学习过程的真实性,否则就会导致预测数据的不准确。

基于以往研究的不足,本文以外汇汇率时间序列数据为对象,利用 ARIMA 模型中的差分过程可以有效地提取序列中所蕴含的确定性信息。另外,ARIMA 模型的分析方法通过等间距单变量时间序列数据、转移函数数据以及干预数据,将响应时间序列中的值作为自己的过去值,过去误差以及其他时间序列当前和过去值的组合实现预测,实现汇率市场的有效风险控制和管理。

1 ARIMA 模型

ARIMA 模型的实质就是差分运算与 ARMA 模型的组合。它的基本思想是将预测对象随时间推移而形成的数据序列视为一个随机序列,用一定得数学模型来近似描述这个序列。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。一般简记为 $ARIMA(p, d, q)$ 模型,其中 p 为自回归项, d 为时间序列成为平稳时所做的差分次数, q 为移动平均项数。

ARIMA 模型预测的具体步骤归纳如下:

- (1) 根据时间序列的散点图、自相关函数和偏相关函数图以 ADF 单位根检验其方差、趋势及其季节性变化规律,对序列的平稳性进行识别。一般外汇汇率时间序列就是不平稳序列;
- (2) 对非平稳序列进行平稳化处理,如果数据序列是非平稳的,并存在一定得增长或下降趋势,则需要对数据进行差分处理,如果数据存在异方差,则需要对数据进行技术处理,直到处理后的数据的自相关函数值和偏相关函数值无显著的异于零;
- (3) 根据时间序列模型的识别规则,建立相应的模型。若平稳序列的偏相关函数是截尾的,可断定序列适合 AR 模型;若平稳序列的偏相关函数是拖尾的,而自相关函数是截尾的,则可断定序列适合 MA 模

收稿日期: 2009-06-23

作者简介: 张奕韬(1984-),女,江西南昌人,硕士,助教,研究方向为数据挖掘。

型;若平稳序列的偏相关函数和自相关函数均是拖尾的,则序列适合 ARMA 模型;

- (4) 进行参数估计,检验是否具有统计意义;
- (5) 进行假设检验,诊断残差序列是否为白噪声;
- (6) 利用已经通过检验的模型进行预测分析。

2 外汇汇率时间序列的 ARIMA 建模和预测

以 2009 年 1 月 1 日 0:00:00 到 6 月 9 日 7:28:22 人民币兑换英镑收盘价的汇率为时间序列,一共 3 912条记录,采用 ARIMA 模型以及结合外汇汇率时间序列的特殊性,对以上的时间序列差分处理、白噪声检验和建立预测模型。

2.1 数据处理

通过 SAS Base 对某银行公布的外汇汇率时间序列数据进行处理,由于 SAS 系统中对时间的格式定义有着严格的要求,所以必须把时间参数通过 SAS 编程转换成“日期 + 时间”的格式,例如 2009 年 6 月 9 日 0 点 0 分 8 秒的日期时间表示成的 SAS 日期时间格式为“09JUN09:00:00:08”。另外整理后的 SAS 数据集包含的属性如表 1 所示。在实际应用中,设定 mcj(卖出价)为预测对象。

表 1 数据集属性

名称	类型	长度	输出	输入
name(币种名称)	字符	8	\$ 8.0	\$ 8.0
xhmr(现汇买入价)	数值	8	Best 12.0	F12.0
xcmr(现钞买入价)	数值	8	Best 12.0	F12.0
mcj(卖出价)	数值	8	Best 12.0	F12.0
jjz(基准价)	数值	8	Best 12.0	F12.0
zhzs(折算价)	数值	8	Best 12.0	F12.0
dt(日期时间)	日期	8	Best 12.0	F12.0

2.2 差分平稳化

在预测模型的步骤进行数据差分,最终对原序列做 1 阶,12 步差分,使得差分之后的序列呈现类似平稳状态。差分后序列的自相关性按照滞后数(Lag)显示,如表 2 所示。自相关图显示延迟 12 阶自相关系数(-0.485)显著大于 2 倍标准差范围 $[-2 \times 0.016, 2 \times 0.016]$,这说明差分后序列中仍然含有非常显著的季节效应,延迟 1 阶的子相关系数也大于 2 倍标准差,这说明差分后序列还具有短期相关性。差分后序列的偏自相关性如表 2 所示,偏自相关图显示延迟 12 阶子相关系数(-0.485)显著大于 2 倍标准差范围 $[-2 \times 0.016, 2 \times 0.016]$,这说明差分后序列中仍然含有非常显著的季节效应。

表 2 自相关和偏自相关系数

滞后数	协方差	自相关系数	标准差	偏自相关系数
1	-0.345	-0.039	0.016	-0.040
2	-0.201	-0.023	0.016	-0.025
3	-0.074	-0.008	0.016	-0.010
4	-0.233	-0.267	0.016	-0.028
5	-0.137	-0.016	0.016	-0.019
6	-0.022	-0.003	0.016	-0.005
7	0.234	0.027	0.016	0.025
8	-0.116	-0.013	0.016	-0.013
9	0.162	0.019	0.016	0.018
10	0.018	0.002	0.016	0.003
11	0.164	0.019	0.016	0.021

滞后数	协方差	自相关系数	标准差	偏自相关系数
12	-4.199	-0.485	0.016	-0.485
13	0.257	0.029	0.019	-0.005
14	0.287	0.033	0.019	0.013
15	-0.020	-0.002	0.019	-0.005
16	0.423	0.049	0.020	0.025
17	0.048	0.006	0.020	-0.001
18	0.073	0.008	0.020	0.009
19	-0.144	-0.017	0.020	0.009
20	0.184	0.021	0.020	0.005
21	-0.247	-0.028	0.020	-0.016
22	0.007	0.001	0.020	-0.002
23	0.069	0.008	0.020	0.019
24	-0.304	-0.035	0.020	-0.350

根据差分后序列的自相关图和偏自相关系数的结果,按照模型设定原则采取 ARMA 模型。

3.3 模型检验

把处理好的数据输入模型,生成的白噪声自相关性检验如表 3 所示。发现针对不同的滞后数(Lag)它的 P 值($Pr > ChiSq$)大部分 < 0.05 ,非常显著,模型效果良好。

表 3 白噪声自相关性检验

白噪声自相关检验								
滞后数	χ^2 检测	显著水平	自相关性					
6	12.39	0.0538	-0.040	-0.023	-0.009	-0.027	-0.016	-0.003
12	937.48	< 0.0001	0.027	-0.013	0.019	0.002	0.019	-0.485
18	954.98	< 0.0001	0.030	0.033	-0.002	0.049	0.006	0.008
24	966.09	< 0.0001	-0.017	0.021	-0.028	0.001	0.008	-0.035

经过筛选,模型使用最大似然法估计方法,通过实验结果对比最终设定最大迭代次数为 1 000 次,可以生成如表 4 所示的最大似然参数估计图。从表中可以发现参数 p 和 q (分别用 AR 和 MA 表示)的不同取值,模型的显著水平($Pr > |t|$)呈现不同的值,但是总体上满足显著水平(< 0.05)。因而,模型的效果显著。

表 4 最大似然参数估结果

最大似然估计					
参数	估计值	标准差	t 值	显著水平	滞后数
MU	0.001	0.0157	0.06	0.9507	0
MA(1,1)	-1.777	0.1340	-13.31	< 0.0001	1
MA(1,2)	-0.123	0.0860	-1.34	0.1520	2
MA(1,3)	1.610	0.1210	13.24	< 0.0001	3
MA(1,4)	0.955	0.1470	6.52	< 0.0001	4
AR(1,1)	-1.640	0.0180	-90.75	< 0.0001	1
AR(1,2)	-0.120	0.0400	-2.97	0.0030	2
AR(1,3)	1.197	0.0400	29.97	< 0.0001	3
AR(1,4)	0.632	0.0170	36.34	< 0.0001	4

2.4 预测结果

最终,根据上述模型可以生成如图 1 所示的卖出价(mcj)预测曲线图。

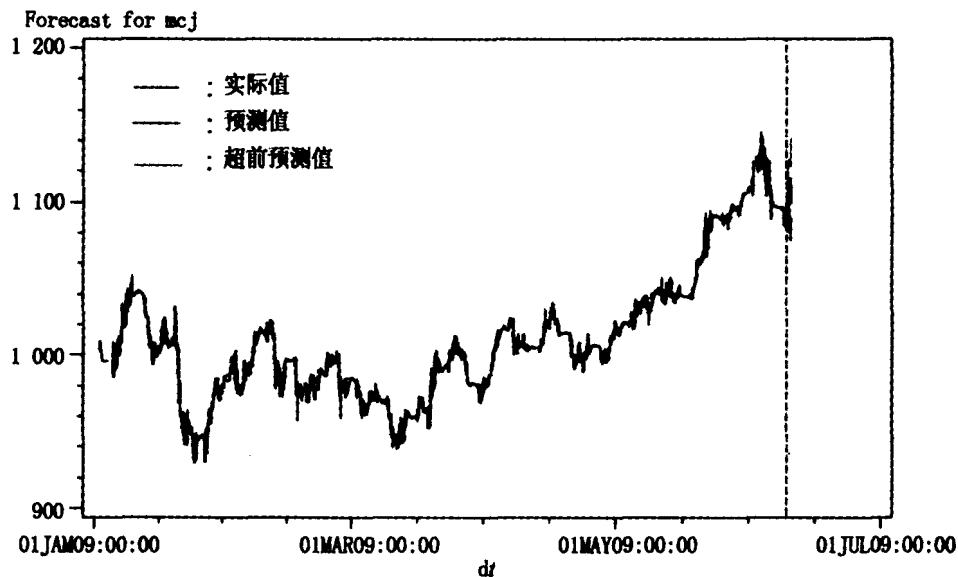


图1 卖出价(mcj)预测图

分析上述的预测结果图,发现预测曲线与实际值基本上是吻合的,红色线的超前预测范围设定为最后记录数据的后24小时预测,即预测范围是09JUN09:08:00:00至10JUN09:07:00:00。对于预测图而言,预测的精确性主要取决于极值的预测准确度。从数据集中找到6月上旬存在极小点(08JUN09:15:39:47),此时的卖出价(mcj)取值是1085.94,而相应的预测值是1086.01,误差是0.07,误差率是 6.45×10^{-5} ,误差率是相当小的。同样,观察24小时后的预测值,与实际值相对比发现其误差率也是相当小的。可见,该模型对于外汇汇率时间序列预测效果良好。

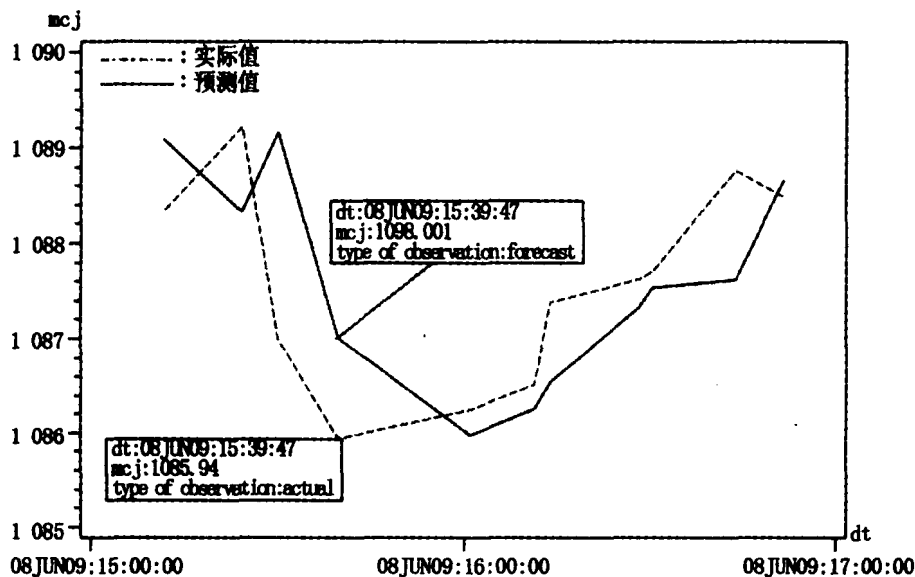


图2 逐步自回归方法的mcj极值预测图

3 结果分析

利用生成的sas数据集,采用逐步自回归方法作为预测模型算法,该方法首先对数据序列拟合时间趋势模型,并计算每个值和估计趋势之间的差值;然后用自回归模型来拟合其余的变差。按照数据记录的单位数为时间间隔,预测卖出价(mcj)在单位数时间间隔的预测数值。截取其中极值点的预测情况,即6月

上旬极小点(08JUN09:15:39:47)结果如图2所示。观察生成的结果可以得出利用逐步自回归方法在该极值点上的预测值为1 087.001,误差是1.06,误差率是 9.76×10^{-4} ,相比较上述的预测结果发现利用逐步自回归方法的预测误差较大,而且该方法不能得到24小时后的预测值,预测效果没有ARIMA模型预测效果好。

4 结束语

从数据挖掘的角度出发,对外汇汇率时间序列通过ARIMA模型进行建模和预测,结合外汇汇率时间序列的特殊性,以2009年1月1日0:00:00到6月9日7:28:22人民币兑换英镑收盘价的汇率为时间序列为实例,一共3 912条记录,对时间序列差分处理、白噪声检验和建立预测模型,并生成相应的预测曲线图。通过与逐步自回归预测模型相比较,基于ARIMA模型的外汇汇率时间序列预测模型很好的预测性能。

参考文献:

- [1] 陈凤娟.数据挖掘在外汇市场中的预测作用[J].数据库与信息管理,2008,1(4):33-34.
- [2] 宋旭东,朱伟红,宁涛.外汇汇率时间序列静态属性的抽取[J].大连交通大学学报,2007,4(28):55-58.
- [3] Chen M S, Han J, Yu P S. Data Mining, an overview from a database perspective[J]. IEEE Tran, on Knoeledge and Data Mining, 1996, 8(6):866-883.
- [4] Last M, Klen Y. Knowledge discovery in time series database[J]. IEEE Trans, on System, Man and Cyberneticpart, 2001, 31(1):160-169.
- [5] 叶银龙,黄晓莉,刘干.基于BP神经网络的股指收益率预测研究[J].统计教育,2009,4(115):38-42.

Prediction of Foreign Exchange Rate Time Series Based on ARIMA Model

ZHANG Yi-tao

(School of Software, East China Jiaotong University, Nanchang 330013, China)

Abstract: Data mining are used to analyze the foreign exchange rate time series and acquire the correct, implicated and hidden information, which has practical significance in the financial field. In this paper, the ARIMA model, which is one of the data mining technologies, is used to study the foreign exchange rate time series of certain bank. In this model, difference and modeling rules are adopted to model and forecast the selling rate. The results show that ARIMA model is capable of predicting accurately data of foreign exchange rate time series.

Key words: foreign exchange rate; time series; ARIMA model; forecast

(责任编辑:王建华)