

基于互联网搜索指数的多因素集成下 人民币汇率预测

王 轩¹, 杨海珍^{2*}

(1. 中国工商银行博士后工作站, 北京 100140;

2. 中国科学院大学经济与管理学院, 北京 100190)

摘要: 针对人民币短期汇率预测问题, 构建了基于互联网大数据信息的互联网搜索指数, 改进了人民币汇率短期预测 CAR 模型. 首先, 基于涵盖丰富信息的互联网大数据, 构建了反映大众关于汇率波动观点、情绪及预期的互联网搜索指数, 并运用热最优路径(TOP)方法、去趋势交叉相关分析(DCCA)方法对指数的有效性进行了检验, 结果表明该指数可以有效地表征出大众情绪和预期的变化. 进一步地, 利用所编制的互联网搜索指数对人民币汇率的短期预测模型进行改进, 构建了融合互联网搜索指数的 CAR 模型. 实证结果表明, 融合互联网搜索指数的人民币汇率短期 CAR 预测模型可以有效提升短期汇率的预测精度.

关键词: 汇率预测; 互联网搜索指数; 互联网大数据; 人民币汇率

中图分类号: TP273

文献标识码: A

文章编号: 1000-5781(2017)03-0360-10

doi: 10.13383/j.cnki.jse.2017.03.007

Research on the forecasting of RMB exchange rate based on multi factor integration and web-searching index

Wang Xuan¹, Yang Haizhen^{2*}

(1. Postdoctoral workstation, Industrial and Commercial Bank of China, Beijing 100140, China;

2. School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: This paper improves the exchange rate prediction model by integrating the web-searching index based on internet big data. First, this paper builds a web-searching index which includes the investor's expectations and emotions. The TOP and DCCA methods are employed to test the validity of the index, and it is shown that the index can measure the investor's expectations and emotions effectively. Then, a CAR model is established by incorporating the web-searching index. The results indicate that the improved CAR model with the web-searching index is an effective method to predict short-term exchange rate and can improve the forecasting accuracy.

Key words: exchange rate prediction; web-searching index; big data; RMB exchange rate

1 引 言

汇率预测一直是各国学者研究和关注的重点问题^[1]. 随着中国汇率市场化改革的推进及人民币国际化

收稿日期: 2016-07-14; 修订日期: 2017-02-09.

基金项目: 国家自然科学基金资助项目(71273257; 71532013).

*通信作者

程度不断提高, 人民币汇率的波动频率和波动幅度都有显著的增加. 从 2015 年 7 月汇改到 2016 年 1 月, 人民币兑美元名义汇率累计升值 20.3%, 年均波动幅度达到 1.8%, 在 2007 年甚至达到 5%. 中国企业、金融机构以及国际投资者所面临的外汇风险与日俱增. 因此, 探索一种更加精确有效的短期汇率预测方法, 更为准确把握和预判人民币汇率的走势具有十分重要的现实意义. 早期的汇率预测方法是以宏观经济理论为基础, 但这类方法往往无法解释市场中普遍存在的汇率剧烈波动现象和市场汇率脱离汇率决定理论等现象. Lyons^[2]和丁晖等^[3]的研究表明, 市场上汇率的波动是因为掌握不同程度信息的外汇交易者在市场交易中博弈所导致的, 而充分掌握信息的一方的可以凭借其信息优势使自己的利益达到最大化. 尽管理论得到进一步完善, 但基于传统汇率决定理论的短期汇率预测效果并不显著. 随着汇率预测研究的发展, 基于时间序列方法进行短期汇率预测的优点逐步显现. Box 等^[4]提出了著名的预测模型-自回归移动平均方法即 ARIMA 模型, Engle^[5]在前人研究基础上进一步改进并提出自回归条件异方差即 ARCH 模型, Bollerslev^[6]又在改进 Engle^[5]研究成果的基础上放宽假设条件, 提出了广义 ARCH 模型. 也有学者进一步打破计量模型的线性假设, 针对金融时间序列的非线性特征, 尝试使用人工神经网络等非线性算法进行建模预测^[7-10]. 通过时间序列分析方法, 无论汇率制度如何, 均能利用汇率自身变化所传递出的时间序列信息进行预测, 可以获得较好的预测效果.

但是, 汇率的短期波动是受到经济基本面因素、外汇市场因素、投资者情绪及预期因素等因素的共同影响, 以往研究中普遍使用的时间序列模型依赖于历史价格数据进行预测, 难以涵盖基本面、市场因素以外的投资者情绪及预期等因素, 而这些因素在汇率的波动过程中起到至关重要的作用, 因而影响模型预测精度. 互联网大数据技术的发展为预期、情绪因素的测度提供了便捷、精确的方法. 公众的互联网行为、言论数据构成了互联网大数据, 其中涵盖着公众的喜好、忧虑、预期及疑问等丰富的信息. 搜索引擎是互联网获取信息最为常用的工具, 是公众获取资源和信息的渠道和载体, 因此, 网络搜索数据就记录并能体现出用户搜索行为背后的行为趋势和规律^[11]. 人们为了满足其特定的目的或者需求而使用网络搜索引擎进行信息查询^[12], 而这种信息的搜索需求正是受到外界条件、环境的改变而产生的^[13], 因而网民在搜索平台上进行的关键词搜索可以反映出外部的条件、环境和要素的变化^[14]. 随着互联网大数据的发展, 许多学者开始利用网络搜索数据对宏观经济问题开展研究. 如文献[15-20]的研究表明, 基于互联网搜索数据对 CPI、GDP、通货膨胀率、失业率、私人消费比率以及消费者信心指数等经济指标的监测和预警具有良好的效果, 加入网络搜索数据的预测模型精度明显高于传统的时间序列模型. Preis 等^[21]进一步将互联网搜索指数应用于对微观市场的预测, 并发现网络搜索指数与股票市场量化交易行为之间具有显著的相关性, 借助网络搜索数据可以更准确地把握投资者的偏好和行为特点从而预测股票市场的涨跌. Kristoufek^[22]则在应用层面利用网络搜索数据来指导和建立股票市场投资策略, 研究表明基于网络搜索指数的投资策略可以有效降低股市投资的风险、获得稳定收益. 刘颖等^[11]针对中国的网络搜索数据与股票市场的相关关系进行研究, 发现网络搜索指数对上证指数收益率具有显著的预测能力, 网络搜索指数可以有效地预测股市表现. 但目前的研究中对于搜索数据的利用, 往往基于对某个或某几个关键词的搜索数据, 覆盖面较窄, 针对搜索指数影响目标变量的机理分析较少, 并且, 对于搜索指数的编制效果缺乏科学、有效的检验.

本文在总结相关文献的基础上, 利用网络搜索数据, 针对短期汇率预测模型进行改进, 以期突破因市场信息不完全所带来的预测瓶颈. 首先, 本文将在理论分析层面阐明搜索指数对改进汇率预测的作用机理; 其次, 通过文本挖掘的方法获取大量与汇率波动高度相关的关键词, 根据所有关键词的搜索数据编制一个综合稳健的互联网搜索指数, 并对其有效性进行检验; 最后, 将互联网搜索指数纳入短期汇率预测模型, 以期获得优于传统预测模型的效果.

2 互联网搜索指数构建与有效性检验

本部分对基于互联网搜索指数的短期汇率预测机理进行阐述, 基于文本挖掘等技术构建了互联网搜索指数, 并运用热最优路径(TOP)方法、去趋势交叉相关分析(DCCA)方法对指数的刻画效果进行检验.

2.1 基于互联网搜索指数的短期汇率预测机理

从机理来看,投资者是否做出某项交易决策,本质上是源于投资者的价值判断和心理预期,而心理预期是投资者内心的主观价值判断,这一主观价值判断既受投资者个体价值判断的影响,也受市场情绪的影响(即投资者具有从众效应),难以直接精确量化。但是,大众价值判断的心理预期可以通过其集体行为表现出来,当大众对于一项资产价格的预期是看涨时,一定会伴随着大众集体上对这项资产的持续、高频率的关注表现。经济学家的相关评论、与汇率变动相关的新闻和财经专栏、人们发布的微博信息以及论坛跟帖都会充分及时地体现出大众预期的变化,因此,可以通过网络搜索数据捕获大众的集体预期、情绪等因素的变化和趋势,进而把时间序列模型与网络搜索指数融合,以有效解决因信息不充分导致的预测精度不足问题。

2.2 互联网搜索指数构建

互联网搜索指数构建的步骤如下:第一步,构造语料库。本文以外汇市场中交易最为频繁的人民币兑美元汇率为例进行实证研究,从财经新闻报道、财经专栏、财经论坛以及财经评论人员微博中摘录搜集了与人民币兑美元汇率相关的新闻资料作为筛选关键词的语料库。第二步,分词与关键词提取。本文使用 NLPPIR 汉语分词系统对语料文本进行语义分词,通过 TextRank 算法选出与汇率波动具有较高关联度的关键词。表 1 展示了本文所选取的部分网络搜索关键词及其与汇率数据的相关性水平。第三步,获取关键词搜索指数。本文选取了既对搜索数据提供开源下载又在本土化方面较为领先的 360 搜索指数。为了获得较多的样本,本文选择搜索平台所提供的全部时间区间即 2013-11-01~2015 的时间段,并对筛选出的关键词的日度搜索指数进行下载。第四步,搜索指数的合成。为了避免单个词语的搜索数据存在异常值、波动剧烈等现象,本文通过合成的方法将多个关键词的搜索数据合成一个稳健的网络搜索指数(INDEX)。表 1 为本文所选取的部分网络搜索关键词及其与汇率数据的相关性水平。

表 1 部分网络搜索关键词及其与汇率数据的相关性

Table 1 Partial network search keywords and their correlation to exchange rate data

序号	关键词	相关系数	序号	关键词	相关系数
1	美国	0.808	11	汇率	0.602
2	中心	0.76	12	购买	0.601
3	全球	0.75	13	退出	0.598
4	美元	0.707	14	状态	0.586
5	国际	0.68	15	出口	0.550
6	进口	0.67	16	指数	0.529
7	影响	0.632	17	账户	0.528
8	欧元	0.623	18	市场	0.526
9	人民币	0.617	19	通道	0.515
10	分析	0.607	20	效果	0.505

本文构建的互联网搜索指数如图 1 所示。分析可知,本文针对人民币汇率所构造的互联网搜索指数与美元兑人民币中间价格数据之间存在较强的相关性,两列数据基本实现峰谷对应,呈现出十分明显的负相关关系。

2.3 指数有效性检验

本部分运用金融物理学中的热最优路径(TOP)方法、去趋势交叉分析(DCCA)方法对所构建指数的刻画效果进行检验。

2.3.1 互联网搜索指数与汇率波动的动态相关性检验

由于使用传统相关性分析无法测度出在每个时点上,样本序列之间动态的领先滞后关系和相关关系,因此,本文采用热最优路径(TOP)方法对搜索指数的动态相关性进行检验。TOP 方法是一种用于识别和量化两个时间序列之间时变动态领先滞后关系的一种无参数估计方法。热最优路径方法通过构建距离矩阵的方式将两个时间序列之间的经济问题进行转化,将经济问题转化为统计物理学问题中的经典概率传递模型,通过配分函数的递归运算,最终得到两组时间序列之间的动态领先滞后关系^[23]。研究表明,热最优路径法可以

较好地对阶数的变化进行捕捉, 只在阶数发生变动时有一定的波动, 但很快就恢复平稳, 能够稳定在本身的领先滞后阶数上^[24,25].

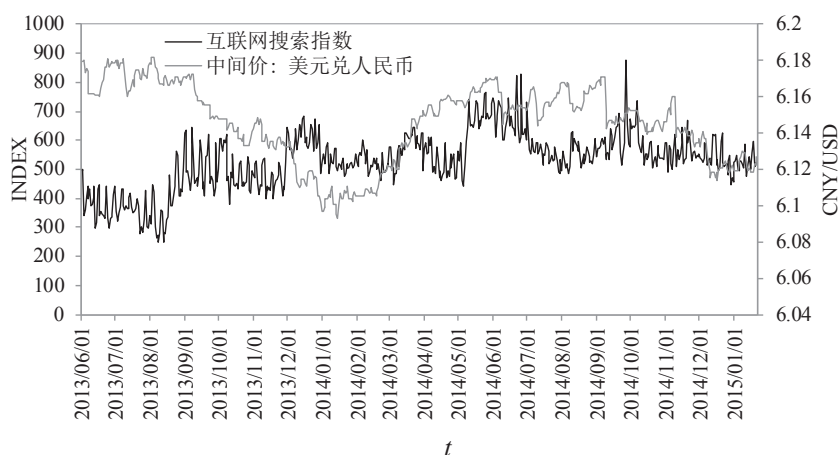


图1 互联网搜索指数与人民币对美元名义汇率

Fig. 1 The internet search index and the RMB nominal exchange rate against the U.S. dollar

由于构建网络搜索指数的目的是准确预测汇率走势、指导外汇交易, 因此需要确定网络搜索数据与汇率数据之间的领先滞后关系. 本文通过热最优路径分析法确定网络搜索指数和美元兑人民币汇率数据之间的动态相关关系. 通过 TOP 分析, 可以得到如下结果(如图 2 所示): 本文构建的网络搜索指数与美元兑人民币汇率之间存在长期稳定的动态领先关系; 网络搜索指数稳定领先于美元兑人民币汇率; 领先阶数稳定在 1~2 阶之间, 平均领先阶数为 1.2 阶.

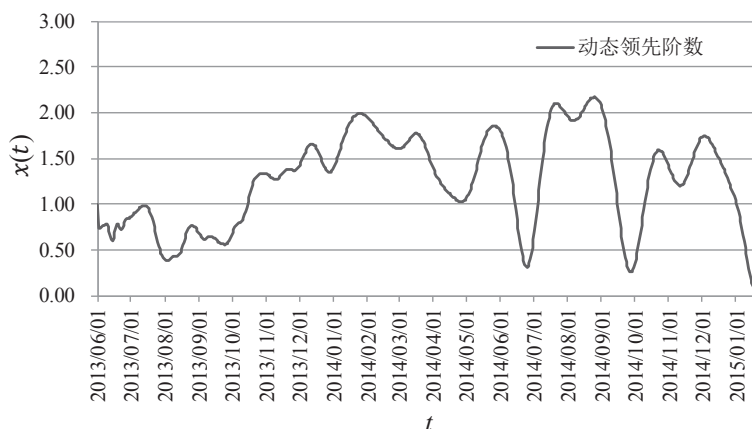


图2 动态领先阶数

Fig. 2 The dynamic leading order

为了检验这种动态相关关系是否显著, 通过对样本数据 Bootstrap 双侧检验检验搜索指数与汇率之间的动态相关关系是否显著. 检验结果如图 3 所示, 其中 $U(t)$ 和 $L(t)$ 为 Bootstrap 双侧检验曲线, 如果动态相关阶数的曲线突破检验曲线则说明领先(滞后)关系显著. 从检验结果可以看出, 在样本区间内搜索指数对汇率的领先关系较为显著, 仅在 2014 年底显著性水平有所下降, 其余基本稳定领先 1~2 阶. 表明互联网搜索指数对汇率的预测作用稳定、有效.

2.3.2 互联网搜索指数与汇率波动的长程相关性检验

本部分将通过检验网络搜索指数与人民币汇率之间是否存在长期稳定的均衡关系, 来进一步说明网络搜索数据对汇率变动存在长期稳定的影响. 由于金融时间序列数据往往存在非平稳的问题, 针对这一问题,

学者们从物理学领域借鉴提出了 DCCA、DFA 等分析方法来监测两个非平稳序列的长程相关性(及其分形特征)^[26-28]. DCCA 方法是 DFA 方法的重要扩展, DFA 方法是将一个特定的时间序列进行分段趋势拟合后再减去趋势, 再对得到的序列进行相关关系的计算^[29]. 经由 Podobnik 等^[30]拓展, 提出了 DCCA 方法, 用来分析两个时间序列的长期相关性.

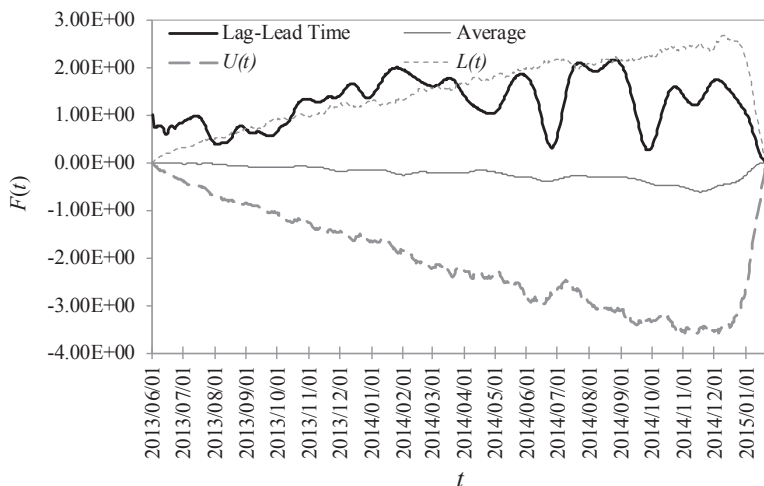


图 3 搜索指数与汇率动态相关关系的 Bootstrap 双侧检验

Fig. 3 Bootstrap two sided test for dynamic relationship between search index and exchange rate

据此, 本文对关注度指数波动率和汇率数据的波动率分别进行分析, 检验结果如图 4 所示, 根据 DFA 曲线可以发现这两个时间序列均表现出幂律自相关, 其中标度指数 α 等于 0.653 8 和 0.554 2, 表明两列时间序列数据的正向相关.

同时, 从图 4 中的 DCCA 曲线可以发现两列数据之间还存在着幂律交叉相关, 相关性为正相关. 进一步计算 DCCA 标度指数, 得到 $\beta = 0.81$, 指数大于 0.5 小于 1 且不等于 0.5, 这意味着数据序列服从幂律分布并且存在长程相关性, 进一步说明搜索指数与汇率波动之间确实存在比较紧密的长期相关关系.

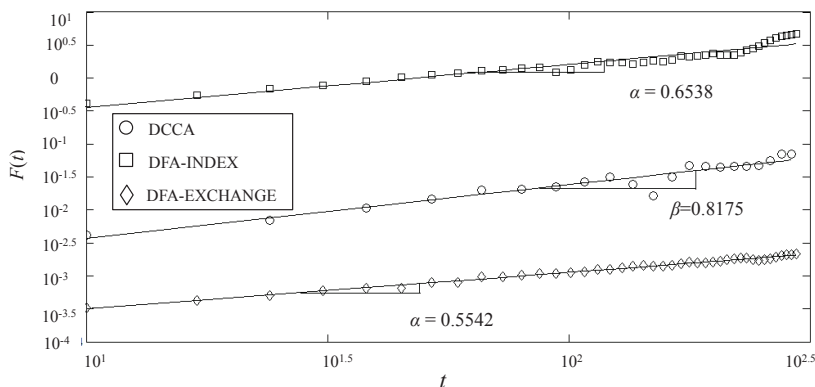


图 4 网络搜索指数与人民币对美元名义汇率波动率的 DFA 及 DCCA 检验

Fig. 4 The DFA and DCCA inspection of web search index and the RMB nominal exchange rate against the dollar

3 融合互联网搜索指数的人民币短期汇率预测

本节通过对互联网大数据充分挖掘, 构建涵盖投资者情绪、预期等丰富信息的互联网搜索指数, 并对该指数的有效性进行研究. 在此基础上, 将互联网搜索指数融入汇率预测的时间序列模型, 并对改进模型的预

测效果进行横向对比和评价. 进一步地, 针对互联网搜索指数对汇率突发波动的识别效果进行实证研究.

3.1 样本描述

本文选取中国银行美元兑人民币外汇中间价的日度数据作为人民币汇率的代理变量, 其中非交易日的数据设置为前后两个交易日的算术平均价格. 由于 360 搜索平台提供的网络搜索指数的数据开始时间为 2013-01-11, 因此样本区间为 2013-01-11~2015-01-21, 其中 2013-01-11~2015-01-14 的样本数据作为构建模型估计参数所用样本, 并对未来 7 天进行预测, 2015-01-14~2015-01-21 的数据用于对照模型的预测结果测算误差.

由于时间序列分析中要求时间序列数据应为平稳序列, 因此对构建的搜索指数和美元兑人民币汇率进行平稳性检验即单位根检验. ADF 检验的结果如表 2 所示, 美元兑人民币汇率和网络搜索指数均为非平稳序列, 取对数后各个变量序列在 1 % 显著水平下为平稳序列, 说明这两列数据均为一阶单整序列.

表 2 变量的平稳性检验
Table 2 ADF test of variable

变量	ADF 检验(P 值)	MacKinnon 临界值			ADF 检验结果
		1 %	5 %	10 %	
y	0.094 4	-2.568 151	-1.941 260	-1.616 406	非平稳
index	0.729 8	-2.568 151	-1.941 260	-1.616 406	非平稳
$\ln(y)$	8.483 e-38	-2.568 151	-1.941 260	-1.616 406	平稳
$\ln(\text{index})$	5.255 e-16	-2.568 151	-1.941 260	-1.616 406	平稳

3.2 融合模型构建

将网络搜索指数融合进传统时间序列模型中, 可以使以往仅以价格历史信息作为单一自变量的预测模型更加多元化. 多元化后的模型, 也就是将传统时间序列模型融合网络搜索指数后的模型, 将是多变量时间序列模型.

对于多变量时间序列的建模方法有很多种, 如: 自回归滑动(CARMA)模型和多变量自回归(CAR)模型等. 但由于多变量自回归滑动模型的建模十分复杂且建模效果不佳, 所以本文采用多变量自回归(CAR)模型^[6]. 根据需要, 本文建立两变量的 n 阶 CAR 模型.

本研究采用取对数后的数据序列进行 CAR 模型的建模, 分别记为 $\ln(y_t)$ 和 $\ln(\text{index}_t)$. 通过计算网络搜索指数序列与当期和滞后几期汇率序列的皮尔逊相关系数, 确认因变量网络搜索指数 $\ln(\text{index}_t)$ 的领先阶数为 3 阶; 同时, 观察自相关图发现 $\ln(y_t)$ 存在 1 阶自相关, 因此选定被解释变量的滞后 1 期值纳入模型.

$$\ln(y_t) = c_0 + c_1 \ln(\text{index}_{t-3}) + c_2 \ln(y_{t-1}) + u_t. \quad (1)$$

即, 通过残差的自相关函数图和偏自相关函数图, 发现残差序列存在 8 阶序列自相关. 同时, 也通过反复对模型进行估计, 比较不同模型的变量对应参数的显著性来交叉对比模型阶数. 回归结果如表 3 所示.

表 3 CAR 模型的建模分析
Table 3 Modeling analysis of CAR model

模型	系数	P 值	
C	0.019 491*	0.003 8	
$\ln(\text{index}(-3))$	-4.65E-05*	0.057 7	
$\ln(y(-1))$	0.989 323***	0.000 0	
AR(8)	-0.087 177 **	0.019 7	
R-squared	0.995 247	Adjusted R-squared	0.995 227
F-statistic	50 184.73	Prob(F-statistic)	0.000 000
Akaikeinfo criterion	-12.253 69	Schwarz criterion	-12.228 33
Durbin-Watson stat	2.016 464		

注: *, **, ***分别代表 10%, 5%, 1% 的显著性水平

回归方程的 R^2 为 0.995 说明所建模型可以较好地拟合汇率的变动, DW 统计量等于 2 表明不存在序列

相关, 自变量的显著水平均在 95 % 的置信区间内显著, 表明所建立的搜索指数对人民币兑美元汇率中间价的变动有显著的影响效果.

可得回归方程如下

$$\ln(y_t) = 0.019\,491 - 4.65 \times 10^{-5} \ln(\text{index}_{t-3}) + 0.989\,323 \ln(y_{t-1}) + u_t,$$

$$u_t = -0.087\,177 u_{t-8} + \varepsilon_t.$$

模型的拟合效果和残差的平稳性水平如图 5 所示. 进一步观察回归方程的残差图, 并对比方程残差的自

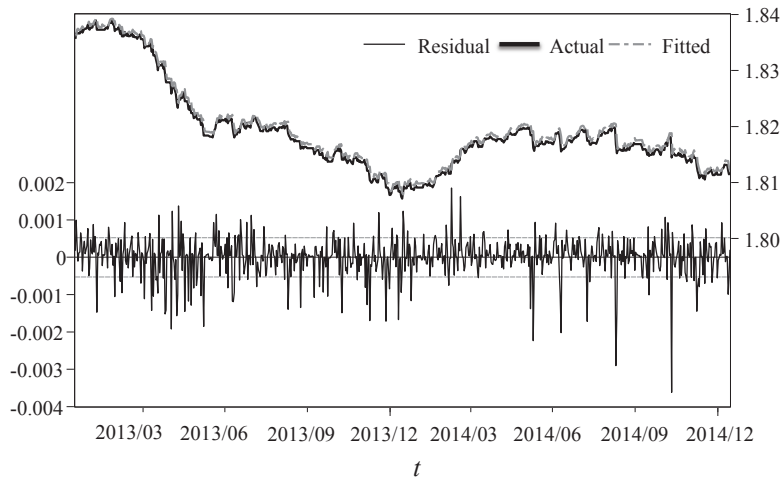


图 5 搜索指数与汇率水平的拟合效果图

Fig. 5 Fitting effect diagram of search index and exchange rate level

相关系数、偏自相关系数和 Q 统计量可以发现, 各阶滞后的自相关和偏自相关系数值均接近零值, Q 统计量的 P 值较大, 表明残差序列不存在序列自相关.

3.3 预测精度对比分析

本部分使用改进的 CAR 模型对美元对人民币汇率中间价进行预测, 并将改进的 CAR 模型与传统汇率预测中较为常用的 AR 模型、ARIMA 模型及 GARCH 模型进行对比, 模型拟合结果如表 4 所示.

表 4 模型拟合结果比较
Table 4 Comparison among model fitting results

模型	融合搜索数据的 CAR 模型	ARIMA 模型	AR 模型	GARCH 模型
C	0.019 491*, (2.899 744)	1.813 961***, (401.488 6)	1.812 691***, (362.005 3)	1.811 655***, (237.817 2)
ln(index(−3))	−4.65E−05*, (−1.900 932)			
ln(y(−1))	0.989 323***, (269.429)			
AR(1)		0.993 455***, (312.673 0)	0.995 032***, (403.260 2)	0.996 202***, (411.684 5)
AR(8)	−0.087 177 **, (−2.336 841)			
MA(2)		0.237 488***, (931.523 3)		
C				2.34E−08*, (1.698 615)
RESID(−1) ²				0.014 272*, (0.095 2)
GARCH(−1)				0.902 016***, (16.162 96)
R-squared	0.995 247	0.995 150	0.995 482	0.995 481
Log likelihood	4 433.708	4 510.215	4 536.465	4 538.891
AIC	−12.253 69	−12.181 66	−12.255 31	−12.253 76
SC	−12.228 33	−12.162 99	−12.242 86	−12.222 63
DW	2.016 464	2.029 912	1.998 579	2.018 485

注: *, **, ***分别代表 10 %, 5 %, 1 % 的显著性水平

为了更加客观、准确的衡量不同模型预测结果的精度, 本文采用国内外相关研究中通用的误差指标来

度量模型的预测精度. 如表 5 所示, 融合互联网搜索指数的 CAR 模型, 其预测结果无论是绝对误差还是相对误差, 均明显小于其他预测模型. 在预测区间的绝对误差总计这一项指标中可以看出融合搜索数据的 CAR 模型的预测误差总计与其他模型相比有明显的优势, 平均绝对百分误差(MAPE)的对比结果更加明显, 其他模型的平均绝对百分误差显著高于融合网络搜索数据后的 CAR 模型, 可以看出在所有误差评价指标下融合了搜索数据的 CAR 模型的预测效果显著优于其它模型.

表 5 汇率突发波动下的模型预测精度

Table 5 Model prediction accuracy under sudden fluctuation of exchange rate

模型	绝对误差总计	MAPE	MAE	Thiel
融合搜索数据的 CAR 模型	0.001 920	1.32 %	0.000 240	0.000 093
ARIMA 模型	0.004 513	3.11 %	0.000 564	0.000 194
AR 模型	0.008 567	1.92 %	0.000 348	0.000 137
GARCH 模型	0.003 181	2.19 %	0.000 398	0.000 173

注: 绝对误差 = |模型预测值 - 实际观测值|; 相对误差 = |模型预测值 - 实际观测值| / 实际观测值 × 100 %;

MAPE(平均绝对百分误差) = $\sum (|预测值 - 实际值| \times 100 / 实际值) / 样本数量$; MAE(平均绝对误差) =

$\sum (|预测值 - 实际值|) / 样本数量$; Thiel(不相等系数)是介于(0,1)之间的一个系数, 如果该系数等于 1 时表明

模型预测能力差; 当系数为 0 时表明模型预测值与实际值完全吻合, 预测能力好.

3.4 CAR 模型对于突发波动的拟合分析

为了进一步检验模型的稳健性和对于突发波动的捕捉预测能力, 使用 2015-08-08~2015-08-12, 人民币汇率政策性贬值的区间, 作为测试样本进行检验. 图 6 为突发波动情况下, 网络搜索指数及汇率预测结果的变化情况, 分析可知, 本研究所构建的搜索指数显著领先于汇率波动 2~3 d, 能够在突发的汇率波动中稳定准确的捕捉市场参与者情绪、预期的变化, 并能继续有效地改进预测结果.

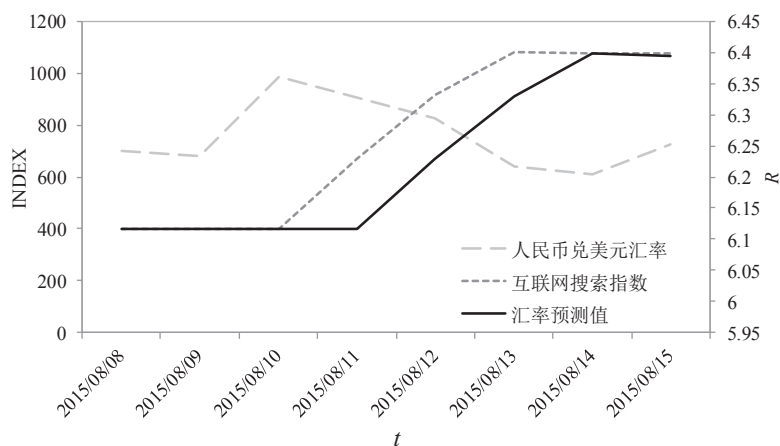


图 6 汇率突发波动下的搜索指数变化与模型预测值

Fig. 6 The search index change and model prediction value under sudden fluctuation of exchange rate

本文进一步比较在汇率突发波动情况下, 融合搜索数据的 CAR 模型与其他模型(AR 模型、ARIMA 模型和 GARCH 模型)预测误差情况. 结果如表 6 所示.

融合搜索数据的 CAR 模型可以有效捕捉和预测汇率波动过程中的突发波动, 其预测结果的几种主要误差度量指标均显著低于其它预测模型.

综上, 本研究构建的融合网络搜索指数的 CAR 模型预测准确率显著高于其它模型, 表现出明显的优势. 并且, 融合网络搜索指数的 CAR 模型在汇率发生突发波动的情况下依然能够稳定、准确的对汇率波动进行拟合和预测. 结果表明, 通过编制互联网搜索指数的方式, 将传统时间序列模型难以涵盖的投资者情绪、预期等信息量化并纳入预测模型, 可以显著改善传统时间序列模型的预测效果.

表 6 汇率突发波动下模型预测精度对比

Table 6 Comparison of model prediction accuracy under sudden fluctuation of exchange rate

模型	MAPE	MAE	Thiel
融合搜索数据的 CAR 模型	0.517 4	0.032 637	0.004 519
ARIMA 模型	0.586 3	0.036 987	0.004 555
AR模型	0.576 9	0.036 392	0.004 758
GARCH 模型	0.576 2	0.036 345	0.004 755

4 结 束 语

本文从涵盖了公众情绪、预期等丰富信息的互联网大数据入手,编制了反映投资者关于汇率波动观点、情绪、预期的互联网搜索指数,并运用金融物理学中的热最优路径(TOP)方法、去趋势交叉相关分析(DCCA)方法对指数的有效性进行检验.进一步地,本文利用互联网搜索指数有效改进了人民币汇率的短期预测模型.研究表明,本文构造的互联网搜索指数与人民币兑美元汇率之间存在长期、稳定的相关关系.互联网搜索指数波动稳定领先于人民币兑美元汇率波动,平均领先阶数为一阶,表明互联网搜索指数可以有效地表征出既往历史数据预测中没有涵盖的汇率波动因素.实证研究表明,基于互联网搜索指数改进的 CAR 模型预测精度显著高于其他模型;同时,对于汇率的突发波动能够及时、准确地进行拟合,表现出明显的优势.因此,通过本文构造的互联网搜索指数对传统时间序列模型进行改进,可以有效改进短期汇率的预测精度.

参考文献:

- [1] 战 毅,安 佳. ARIMA 模型在汇率预测中的应用: 基于人民币汇率的验证. 中国证券期货, 2013(5): 74–75.
Zhan Y, An J. Application of ARIMA model in exchange rate prediction based on the experience of RMB. Securities & Futures of China, 2013(5): 74–75. (in Chinese)
- [2] Lyons R K. The Micro Structure Approach to Exchange Rates. Cambridge: MIT Press, 2001, 1–70.
- [3] 丁 晖,谢 赤,陈 琼,等. 人民币汇率波动对国际企业价值影响实证研究. 经济问题, 2008 (8): 102–105.
Ding H, Xie H, Chen Q, et al. The empirical research on the affection of RMB exchange rate fluctuations to the value of international business. On Economic Problems, 2008 (8): 102–105. (in Chinese)
- [4] Box G E P, Jenkins G M, Reinsel G C, et al. Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015, 137–191.
- [5] Engle R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 1982, 50(4): 987–1007.
- [6] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 1986, 31(3): 307–327.
- [7] 张冬青, 宁宣熙, 刘雪妮, 等. 基于 RBF-HMM模型的时间序列在线预测. 系统工程学报, 2010, 25(1): 19–25.
Zhang D Q, Ning X X, Liu X N, et al. Time series on-line prediction bases on RBF-HMM model. Journal of Systems Engineering, 2010, 25(1): 19–25. (in Chinese)
- [8] Santos A A P, Costa N C A. Computational intelligence approaches and linear models in case studies of forecasting exchange rates. Expert Systems with Applications, 2007, 33(4): 816–823.
- [9] Davis J T, Episcopos A, Wettimuny S, et al. Predicting direction shifts on Canadian-US exchange rates with artificial neural networks. Intelligent Systems in Accounting Finance & Management, 2001, 10(2): 83–96.
- [10] 王晓辉, 张卫国, 刘玉芳, 等. 融合 ICA 的 BP 网络在人民币汇率预测中的应用. 系统工程学报, 2014, 29(3): 344–352.
Wang X H, Zhang W G, Liu Y F, et al. Application of BP neural networks model combining with ICA in RMB exchange rate prediction. Journal of Systems Engineering, 2014, 29(3): 344–352. (in Chinese)
- [11] 刘 颖, 吕本富, 彭 赓, 等. 网络搜索对股票市场的预测能力: 理论分析与实证检验. 经济管理, 2011(1): 172–180.
Liu Y, Lü B F, Peng G, et al. Web search on the ability to predict the stock market theoretical analysis and empirical test. Economic Management Journal, 2011(1): 172–180. (in Chinese)
- [12] Wilson T D. Human information behavior. Informing Science, 2000, 3(2): 49–56.

- [13] Moe W W, Fader P S. Dynamic conversion behavior at e-commerce sites. *Management Science*, 2004, 50(3): 326–335.
- [14] 孙 毅. 网络搜索与通货膨胀相关性研究. 中国科学院研究生院, 2010: 10–22.
Sun Y. The Research on The Correlation of Web Search and Inflation. Graduate School of Chinese Academy of Sciences, 2010: 10–22. (in Chinese)
- [15] D’Amuri F, Marcucci J. The predictive power of google searches in forecasting unemployment. *Social Science Electronic Publishing*, 2012, 891(1): 5–49.
- [16] Alesina A. Macroeconomic policy in a two-party system as a repeated game. *Quarterly Journal of Economics*, 1987, 102(3): 651–678.
- [17] Vosen S, Schmidt T. Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, 2009, 30(6): 565–578.
- [18] 张 崇, 吕本富, 彭 庚, 等. 网络搜索数据与 CPI 的相关性研究. *管理科学学报*, 2012, 15(7): 50–59.
Zhang C, Lü B F, Peng G, et al. Correlation of internet search and CPI data. *Journal of Management Sciences in China*, 2012, 15(7): 50–59. (in Chinese)
- [19] 孙 毅, 吕本富, 陈 航, 等. 基于网络搜索行为的消费者信心指数构建及应用研究. *管理评论*, 2014, 26(10): 117–125.
Sun Y, Lü B F, Chen H, et al. The construction of consumer confidence index based on internet search behavior. *Management Review*, 2014, 26(10): 117–125. (in Chinese)
- [20] Askitas N, Zimmermann K F. Google econometrics and unemployment forecasting. *General Information*, 2009, 55(2): 107–120.
- [21] Preis T, Moat H S, Stanley H E, et al. Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, 2013, 3(7446): 542–542.
- [22] Kristoufek L. Can Google Trends search queries contribute to risk diversification. *Scientific Reports*, 2013, 3(3): 1–5.
- [23] Sornette D, Zhou W X. Non-parametric determination of real-time lag structure between two time series: The ‘optimal thermal causal path’ method. *Quantitative Finance*, 2005, 5(6): 577–591.
- [24] Zhou W X, Sornette D. Non-parametric determination of real-time lag structure between two time series: The “optimal thermal causal path” method with applications to economic data. *Journal of Macroeconomics*, 2006, 28(1): 195–224.
- [25] Zhou W X, Sornette D. Lead-lag cross-sectional structure and detection of correlated – anticorrelated regime shifts: application to the volatilities of inflation and economic growth rates. *Physica A: Statistical Mechanics and its Applications*, 2007, 380(7): 287–296.
- [26] Kristoufek L. Multifractal height cross-correlation analysis: A new method for analyzing long-range cross-correlations. *Europhysics Letters*, 2011, 95(6): 525–604.
- [27] Alessio E, Carbone A, Castelli G, et al. Second-order moving average and scaling of stochastic time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2002, 27(2): 197–200.
- [28] Gu G F, Zhou W X. Detrending moving average algorithm for multifractals. *Physical Review E*, 2010, 82(1): 1859–1860.
- [29] Peng C K, Buldyrev S V, Havlin S, et al. Mosaic organization of DNA nucleotides. *Physical Review E*, 1994, 49(2): 1685–1689.
- [30] Podobnik B, Stanley H E. Detrended cross-correlation analysis: A new method for analyzing two non-stationary time series. *Physical Review Letters*, 2008, 100(8): 38–71.

作者简介:

王 轩(1988—), 男, 河北廊坊人, 博士, 研究方向: 金融管理, Email: wang.xuan@hotmail.com;

杨海珍(1965—), 女, 宁夏平罗人, 博士, 博士生导师, 研究方向: 国际资本流动与金融稳定, 汇率制度, 资本急停与货币危机, 货币危机机理, Email: haizheny@ucas.ac.cn.