



# An Empirical Analysis and Forecast Study Between the Baidu Index Based on the Amount of Web Search and Trading Volume on the Platform of P2P

**Zhang Xing**

School of Economics, Shanghai University, Shanghai, China

**Email address:**

18800201193@163.com

**To cite this article:**

Zhang Xing. An Empirical Analysis and Forecast Study Between the Baidu Index Based on the Amount of Web Search and Trading Volume on the Platform of P2P. *Science Innovation*. Vol. 5, No. 5, 2017, pp. 256-262. doi: 10.11648/j.si.20170505.12

**Received:** June 8, 2017; **Accepted:** July 2, 2017; **Published:** July 19, 2017

**Abstract:** There is a positive correlation between Baidu index based on the amount of web search and trading volume on the platform of P2P. In this paper, there are 184 daily data between July 1, 2016 to December 31, 2016 as sample data. I construct two regression models, one includes Baidu index and the other is included with lagged Baidu index to predict trading volume from pat platform. Therefore, I select randomly 20 daily data from January 1, 2017 to February 28, 2017. Compared with the true value and predicted value results show that the prediction effect, containing lag regression model of a Baidu index is better than the regression model with Baidu index. Taking into account the Baidu index factor P2P volume research, both in line with the development trend of Internet financial data, but also enrich the research content and research methods P2P.

**Keywords:** Baidu Index, Trading Volume, Forecast Result

---

## 网络搜索量与P2P平台成交量的实证分析及预测研究

**张星**

上海大学经济学院, 上海, 中国

**邮箱**

18800201193@163.com

**摘要:** 基于百度指数的网络搜索量对于P2P平台成交量有着正相关关系。本文以拍拍贷2016年7月1日至2016年12月31日共184天日数据为样本数据, 分别建立不含百度指数的回归模型和含有滞后一期的百度指数的回归模型对该平台成交量进行预测研究, 并随机抽取2017年20个日数据和预测值比较, 结果表明含有滞后一期百度指数的回归模型的预测效果好于不含百度指数的回归模型。把百度指数因素考虑到P2P成交量研究中, 既符合互联网金融大数据的发展态势, 也丰富了P2P的研究内容和研究方法。

**关键词:** 百度指数, 成交量, 预测效果

---

## 1. 引言

搜索引擎的发展为中国经济发展注入了活力,也改变了人们的行为方式。用户享受搜索引擎带给自己丰富的信息量的同时也把自己的行为记录在了搜索引擎上。大量的搜索引擎记录汇聚在一起,成为反映用户的行为信息和某些经济规律趋势的有力工具。近年来,互联网信息技术不断发展和普及,大数据时代改变了一些行业的商业模式,得数据者得天下,能够把沉淀的数据里的商业价值挖掘出来是很多学术界、商界乃至政界都在尝试的思想和行动指南。目前,基于网络搜索数据进行的研究主要表现在流感预测、失业率预测、旅游地客流量预测、房地产销售预测、汽车销量预测等方面。Ginsberg等[1]利用 Google 搜索引擎端的搜索数据检测流感的爆发;Askitas 等[2]运用搜索数据,研究德国的失业率;孔令硕[3]基于百度指数研究了互联网搜索量和大众汽车实际销售量之间的关系,发现加入百度指数后的模型预测精度有了很大的提高。

百度搜索引擎在中国搜索引擎市场上独占鳌头。据2016年4月中国互联网信息中心基于中国网民搜索行为调查而发布的《2015年中国网民搜索行为调查报告》显示,截至2015年12月,我国搜索引擎用户达到56623万人,用户规模较2014年年底增长4400万,增速为8.4%,使用率为82.3%。在搜索引擎市场中,仅百度一家公司,其综合搜索品牌渗透率占93.1%,手机端综合搜索品牌渗透率占87.5%,PC端综合搜索引擎品牌渗透率占92.4%。鉴于此,本文选择用百度公司旗下的百度指数来衡量网络搜索量。

互联网信息技术的发展也带动了互联网金融尤其是P2P平台的发展。根据网贷之家的统计数据显示,截至2017年2月,累积平台数量5882家,运营平台数量为2335家,其他平台或提现困难,或停业,或跑路,给投资者带来了不可挽回的损失。无论是投资者还是借款人都会在进行交易之前对平台进行搜索了解,搜索记录就成为了百度指数的一部分。本文主要针对网民的搜索行为对P2P平台的成交量展开研究。

## 2. 文献综述

伴随着搜索引擎的发展,许多学者开始研究搜索量和市场实际需求量之间的关系,发现二者呈显著的正相关关系,并且在预测模型中引入互联网搜索量有助于提高模型的预测精度。目前市场上用的比较多的是谷歌公司提供的谷歌趋势和百度公司提供的百度指数,虽然其提供的是经过一定加工后的包含搜索量信息的数据,但是也可以反映一定的趋势。考虑到百度在搜索引擎领域市场渗透率较高,本文采用百度指数作为网络搜索量的代理变量。

曹小琳,牟红[4]基于百度指数构建回归模型对全国商品房销售面积进行拟合和预测。李山、邱荣旭[5]基于百度指数研究国内66个5A级旅游景区网络搜索量及其实际客流量的关系。刘颖等[6]以谷歌趋势上的搜索量为基础,研究网络搜索量和股票指数之间的关系,其对于上证指数的收益率也具有很好的预测效果。刘伟江,李映桥[7]运用主成分分析方法合成搜索指数对台湾地区消费者信息指数进行预测,发现预测数据比官方公布的数据提前一个月左右。

互联网信息技术的发展带动了学者们对网贷平台的研究,关于P2P平台的实证研究也在不断涌现。王丹,张洪潮[8]参照企业与银行的信用风险等级评价指标分析框架,运用层次分析法等方法构建了P2P网贷平台信用风险评价指标体系。欧阳慧[9]以宜人贷为例,对P2P网络借贷面临的主要风险进行识别与分析,构建了P2P网络借贷风险预警理论框架。叶夏菁[10]基于BP神经网络与半监督学习构建网贷平台信用评估模型。骆建文等[11]介绍了大数据背景下的互联网融资平台信用评级。王程龙、陈程[12]基于决策树建立P2P网贷平台信用评级体系研究。李敏芳、田晨君[13]采用因子分析法对我国31家P2P网贷平台进行评级。

综上所述,P2P平台确实存在风险问题,监管有待完善,国内外学者对于网贷P2P平台的研究多是进行评级或者风险预警。从百度指数方面研究P2P平台的文章很少,本文尝试借鉴前人的经验从代表网络搜索量的百度指数角度对P2P平台成交量进行研究。

## 3. 背景介绍和变量选取

### 3.1. 拍拍贷背景介绍

拍拍贷全称是上海拍拍贷金融信息服务有限公司,是中国第一家网络信用借贷平台。它成立于2007年6月,是P2P平台的私营企业的代表。鉴于此,本文拟选取拍拍贷作为论文选取对象。

### 3.2. 数据及样本情况

#### 3.2.1. 百度指数和成交量的关系

百度指数是以百度海量网民行为数据为基础的股份平台,它包含趋势研究、需求图谱、舆情洞察和人群画像四个方面。其中,百度整体搜索指数=移动搜索指数+PC搜索指数。从图3可以初步判断,平台的成交量和百度指数的走势基本是一致的,即二者呈现正相关关系,但是正相关关系不是同期的,而是滞后的正相关关系。投资人数和借款人数对平台的关注越多,其百度指数会上升,成交量也会上升。平台的成交量越高,越会吸引更多的人来关注它。同时,人们如果有投资或借款动机,会先关注某一P2P平台,再会根据关注结果采取投资不投资和借款不借款的行动,所以滞后期的百度指数和P2P平台成交量呈现正相关关系。

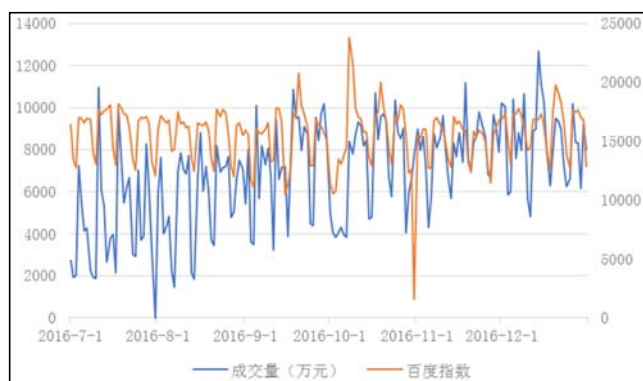


图1 拍拍贷的成交量和百度指数走势图(2016年7月1日—2017年2月28日)。

### 3.2.2. 平台交易数据及变量的选择

本文中所涉及的平台交易数据全部来自网贷之家。考虑到本文主要研究的是影响平台成交量的因素，所选取的变量依次为百度指数、成交量(万元)、预期收益率(%)、投资人数(人)、借款人数(人)、平均借款期限(月)、资金净流入(万元)和借款标数(个)八个变量并截取网贷之家中关于拍拍贷2016年7月1日至2016年12月31日共184天的日数据作为样本数据，随机选取2017年1月1日至2017年2月28日的20日数据进行预测值和真实值的比较，表1是关于变量名称和变量的说明。

表1 变量名称及变量说明。

| 变量名称      | 变量编号  | 变量说明   |
|-----------|-------|--|
| 成交量(万元)   | Y     | 该平台在某一时间段内满标并通过复审成交的总金额(包括债券转让;不包括新手标、体验标)   |
| 百度指数      | $X_1$ | 百度指数以网民在百度的搜索量为数据基础,以关键词为统计对象,科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权和。根据搜索来源的不同,搜索指数分别PC搜索指数和移动搜索指数。 |
| 预期收益率(%)  | $X_2$ | 指平台标的的预期收益率  |
| 投资人数(人)   | $X_3$ | 平台某个时间段内发生过投资的人数。  |
| 借款人数(人)   | $X_4$ | 拍拍贷平台在某个时段端内发生过借款的人数。  |
| 平均借款期限(月) | $X_5$ | 平台在某一时间段内已成交标的的按成交量加权的加权平均数。   |
| 借款标数(个)   | $X_6$ | 平台实际发生借款标的的数量。   |
| 资金净流入(万元) | $X_7$ | 平台当天待收金额减去前一天待收金额。   |

资料来源:网贷之家

## 4. 实证分析

### 4.1. 检验数据的平稳性(单位根检验)

由于样本数据为时间序列数据,为了保证序列的平稳性和避免伪回归现象,在建立计量经济学模型之前要对序列进行单位根检验。本文单位根检验采取ADF检验方法,只要ADF值小于三个临近值中的任何一个,那么该序列为平稳序列。表2是对8个变量分别进行的ADF检验,结果表明8个变量序列均是平稳序列。

表2 单位根检验(ADF检验)结果。

| 变量        | ADF值      | 1%临近值     | 5%临近值     | 10%临近值    | 结论 |
|-----------|-----------|-----------|-----------|-----------|----|
| 成交量(万元)   | -3.693005 | -4.010740 | -3.435413 | -3.141734 | 平稳 |
| 百度指数      | -8.755009 | -3.466580 | -2.877363 | -2.575284 | 平稳 |
| 预期收益率(%)  | -7.383912 | -4.008987 | -3.434569 | -3.141237 | 平稳 |
| 投资人数(人)   | -10.67626 | -3.466176 | -2.877186 | -2.575189 | 平稳 |
| 借款人数(人)   | -11.32585 | -4.008987 | -3.434569 | -3.141237 | 平稳 |
| 平均借款期限(月) | -11.13682 | -4.008987 | -3.434569 | -3.141237 | 平稳 |
| 借款标数(个)   | -10.75446 | -4.008987 | -3.434569 | -3.141237 | 平稳 |
| 资金净流入(万元) | -12.92920 | -4.008987 | -3.434569 | -3.141237 | 平稳 |

### 4.2. 格兰杰因果检验

格兰杰因果检验能够考察变量之间是否具有预测能力。检验过程中,由于格兰杰因果关系检验对滞后期的选取有较大的敏感性,本文根据赤池信息量(AIC)准则和贝叶斯信息(SC)准则以及残差序列是否存在序列相关性来确定最优滞后期。本文主要考察的是和百度指数( $X_1$ )相关的格兰杰因果检验,考察百度指数是否对于其他变量具有预测能力。表4是各变量滞后一期的结果。

表3 各变量的格兰杰因果检验结果。

| Null Hypothesis                                      | Obs | F—Statistic | Prob   | 结论  |
|--|-----|-------------|--------|-----|
| X <sub>1</sub> does not Granger Cause Y              | 184 | 3.77968     | 0.0247 | 拒绝  |
| Y does not Granger Cause X <sub>1</sub>              | 184 | 1.35503     | 0.2606 | 不拒绝 |
| X <sub>1</sub> does not Granger Cause X <sub>2</sub> | 184 | 0.48492     | 0.6166 | 不拒绝 |
| X <sub>2</sub> does not Granger Cause X <sub>1</sub> | 184 | 0.51725     | 0.5971 | 不拒绝 |
| X <sub>1</sub> does not Granger Cause X <sub>3</sub> | 184 | 3.41140     | 0.0352 | 拒绝  |
| X <sub>3</sub> does not Granger Cause X <sub>1</sub> | 184 | 1.41984     | 0.2445 | 不拒绝 |
| X <sub>1</sub> does not Granger Cause X <sub>4</sub> | 184 | 3.50494     | 0.0322 | 拒绝  |
| X <sub>4</sub> does not Granger Cause X <sub>1</sub> | 184 | 1.21537     | 0.2991 | 不拒绝 |
| X <sub>1</sub> does not Granger Cause X <sub>5</sub> | 184 | 0.04767     | 0.9535 | 不拒绝 |
| X <sub>5</sub> does not Granger Cause X <sub>1</sub> | 184 | 0.22926     | 0.7954 | 不拒绝 |
| X <sub>1</sub> does not Granger Cause X <sub>6</sub> | 184 | 4.48378     | 0.0126 | 拒绝  |
| X <sub>6</sub> does not Granger Cause X <sub>1</sub> | 184 | 1.07982     | 0.3419 | 不拒绝 |
| X <sub>1</sub> does not Granger Cause X <sub>7</sub> | 184 | 0.58326     | 0.5591 | 不拒绝 |
| X <sub>7</sub> does not Granger Cause X <sub>1</sub> | 184 | 2.03979     | 0.1331 | 不拒绝 |

由表3可知：

（1）百度指数与成交量存在单项的因果关系，百度指数是成交量的格兰杰原因的概率是73.94%，百度指数的变化会引起成交量的变化。含有滞后期的百度指数有助于提高预测成交量的能力和精准度。

（2）百度指数和预期收益率存在双向的格兰杰因果关系，百度指数是预期收益率的格兰杰原因的概率是40.29%，百度指数的变化会引起预期收益率的变化。不论是投资者还是借款者，都会非常关心P2P平台的预期收益率，同等条件下预期收益率高的P2P平台更受投资者的青睐。同时，P2P平台只是一个提供投资和借款的理财平台，其赚取的收益多是从投资者和借款者的利息差额获取，预期收益率高的平台意味着借款者要偿还更高的借款利息。

（3）百度指数和投资人数存在单项的因果关系。百度指数是投资人数的格兰杰原因的概率是75.55%，百度指数的变化会引起投资人数的变化。出于投资收益的考虑，投资者更加关注本金的安全性及收益性，含有滞后期的百度指数有助于提高对投资人数的预测能力。

（4）百度指数和借款人数存在双向的格兰杰因果关系，百度指数是借款人数的格兰杰原因的概率是70.09%，百度指数的变化会引起借款人数的变化。该平台百度指数的趋势可能更多的是靠投资者驱动。

（5）百度指数和平均借款期限存在双向的格兰杰因果关系，百度指数是平均借款期限的格兰杰原因的概率是20.46%，百度指数的变化会引起平均借款期限的变化。

（6）百度指数和借款标数存在单向的格兰杰因果关系，百度指数是借款标数的格兰杰原因的概率是65.81%，百度指数的变化会引起借款标数的变化。借款者关注越多，采取的借款行动可能性越大，借款标数也会随之增长。含有滞后期的百度指数有助于提高对借款标数的预测能力。

（7）百度指数和资金净流入存在双向的格兰杰因果关系，百度指数是资金净流入的格兰杰原因的概率是86.69%，百度指数的变化会引起资金净流入的变化。资金净流入的多少可以反映该平台支付投资者本金和收益的能力，也可以向借款者传递有能力借款的信号，吸引更多的投资者和借款者的关注。一旦投资者和借款者对于某平台关注度高，该平台成交量上升的同时，资金的净流入也会随之上升。滞后期的百度指数有助于提高资金净流入的预测能力。

综上所述，百度指数的滞后项有助于提高对其他变量的预测能力。根据以往经验及不同滞后性模型的判定，考虑到内生性问题，最终决定用滞后一期的百度指数对成交量预测研究。

4. 3. 回归分析及预测结果

为了比较说明百度指数对于成交量预测的作用，本文先后建立两个线性回归模型，分别为不含百度指数的回归模型和含有滞后一期的百度指数的回归模型，进行比较研究，表4和表5分别是两个模型的回归结果。

表4 不含百度指数的回归模型结果。

| Variable       | Coefficient | Std.Error | t-Statistic | Prob.  |
|----------------|-------------|-----------|-------------|--------|
| C              | -2270.501   | 537.4608  | -4.224497   | 0.0000 |
| X <sub>2</sub> | 146.3455    | 24.43947  | 5.988080    | 0.0000 |
| X <sub>3</sub> | 0.149996    | 0.023086  | 6.497305    | 0.0000 |
| X <sub>4</sub> | -0.017200   | 0.022290  | -0.771633   | 0.4414 |
| X <sub>5</sub> | -193.8081   | 28.91256  | -6.703250   | 0.0000 |
| X <sub>6</sub> | 0.253461    | 0.022446  | 11.29193    | 0.0000 |
| X <sub>7</sub> | 0.000652    | 0.007402  | 0.088134    | 0.9299 |

|                    |           |                       |          |
|--------------------|-----------|-----------------------|----------|
| R-squared          | 0.954112  | Mean dependent var    | 6888.598 |
| Adjusted R-squared | 0.952557  | S.D. dependent var    | 2445.879 |
| S.E of regression  | 532.7476  | Akaike info criterion | 15.43127 |
| Sum squared resid  | 50236135  | Schwarz criterion     | 15.55358 |
| Log likelihood     | -1412.677 | Hannan-Quinn criter   | 15.48085 |
| F-statistic        | 613.3755  | Durbin-watson stat    | 1.800768 |
| Prob(F-statistic)  | 0.000000  |                       |          |

表5 含有百度指数的回归模型结果。

| Variable            | Coefficient | Std.Error             | t-Statistic | Prob.  |
|---------------------|-------------|-----------------------|-------------|--------|
| C                   | -2602.865   | 582.6845              | -4.467024   | 0.0000 |
| X <sub>1</sub> (-1) | 0.028634    | 0.016959              | 1.688398    | 0.0931 |
| X <sub>2</sub>      | 146.3173    | 24.64729              | 5.936445    | 0.0000 |
| X <sub>3</sub>      | 0.142623    | 0.023693              | 6.019694    | 0.0000 |
| X <sub>4</sub>      | -0.015659   | 0.022207              | -0.705143   | 0.4817 |
| X <sub>5</sub>      | -192.2745   | 28.78464              | -6.679760   | 0.0000 |
| X <sub>6</sub>      | 0.251815    | 0.022348              | 11.26785    | 0.0000 |
| X <sub>7</sub>      | 0.001823    | 0.007411              | 0.245984    | 0.8060 |
| R-squared           | 0.954364    | Mean dependent var    | 6911.377    |        |
| Adjusted R-squared  | 0.952538    | S.D. dependent var    | 2432.940    |        |
| S.E of regression   | 530.0340    | Akaike info criterion | 15.42649    |        |
| Sum squared resid   | 49163801    | Schwarz criterion     | 15.56680    |        |
| Log likelihood      | -1403.524   | Hannan-Quinn criter   | 15.48336    |        |
| F-statistic         | 522.8087    | Durbin-watson stat    | 1.818032    |        |
| Prob(F-statistic)   | 0.000000    |                       |             |        |

由表4和表5可得：不含百度指数的回归模型结果为：

$$\hat{Y}_1 = -2270.50 + 146.3455X_2 + 0.149996X_3 - 0.017200X_4 - 193.8081X_5 + 0.253461X_6 + 0.000652X_7$$

含有百度指数的回归模型结果为：

$$\hat{Y}_2 = -2602.865 + 0.028634X_1(-1) + 146.3173X_2 + 0.142623X_3 - 0.015659X_4 - 192.2745X_5 + 0.251815X_6 + 0.001823X_7$$

模型整体对成交量变动的解释效果较好，在回归模型中加入滞后一期的百度指数后， $R^2$  由95.41%上升到95.43%，无自相关。

(1) 百度指数每增长1%，该平台1天后的日成交量将增长2.86%；平均收益率每增长1%，该平台的日成交量将增长146.32%；投资人数每增长1%，该平台的日成交量将增长14.26%；借款人数每增长1%，该平台的日成交量将下降1.57%；平均借款期限每增长1%，该平台的日成交量将下降192.27%；借款标数每增长1%，该平台的日成交量将增长25.18%；资金净流入每增长1%，该平台的日成交量将增长0.18%。由此可得，投资者倾向于收益率高、借款时间短的产品。

(2) 图2是回归模型的预测图，RMSE=518.3191，而不含百度指数的回归预测图中，RMSE=522.5155。所以，含有滞后一期的百度指数的回归模型的预测效果更好。

(3) 真实值与预测值的比较

下表是随机截取的2017年1月11日至2017年1月30日共20日的成交量数据真实值与拟合值的比较。其中，Y表示该平台日成交量的真实值， $Y_1$ 表示不含百度指数的回归模型的预测值， $Y_2$ 表示含有滞后一期的百度指数的回归模型的预测值， $E_1 = (Y_1 - Y)^2$ ， $E_2 = (Y_2 - Y)^2$ ，它们分别表

示预测值与真实值的差距，差距越小，预测效果越好。 $F = E_1 - E_2$ ，逻辑值表示如果 $F > 0$ ，则表示为1；如果 $F < 0$ ，则表示为0。在随机截取的20日数据中，逻辑值为1的有17个，逻辑值为0的有3个。综上所述， $Y_1$ 比 $Y_2$ 更接近于真实值Y，即含有滞后一期的百度指数的回归模型的预测效果好于不含百度指数的回归模型。

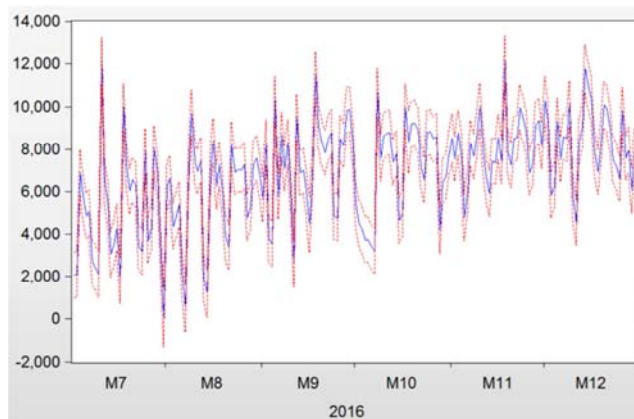


图2 回归模型预测图。

表7 真实值与预测值的比较。

| 日期        | Y        | Y <sub>1</sub> | Y <sub>2</sub> | F=E <sub>1</sub> -E <sub>2</sub> | 逻辑值 |
|-----------|----------|----------------|----------------|----------------------------------|-----|
| 2017/1/11 | 12924.49 | 9503.646723    | 12932.79434    | 11702099.8                       | 1   |
| 2017/1/12 | 9946.62  | 9445.855392    | 9384.824543    | -64848.943                       | 0   |
| 2017/1/13 | 10642.78 | 6456.211286    | 9396.428506    | 15973965.6                       | 1   |
| 2017/1/14 | 7509.25  | 6300.588731    | 6435.569097    | 308071.381                       | 1   |
| 2017/1/15 | 6414.18  | 16908.72822    | 6277.372183    | 110116826                        | 1   |
| 2017/1/16 | 18470.5  | 9329.27761     | 16778.99149    | 80700745.7                       | 1   |
| 2017/1/17 | 9460.95  | 9317.005033    | 9217.85889     | -38373.135                       | 0   |
| 2017/1/18 | 8939.41  | 10706.02801    | 9209.594469    | 3047939.53                       | 1   |

| 日期        | Y        | Y <sub>1</sub> | Y <sub>2</sub> | F=E <sub>1</sub> -E <sub>2</sub> | 逻辑值 |
|-----------|----------|----------------|----------------|----------------------------------|-----|
| 2017/1/19 | 11688.59 | 13920.05193    | 10573.77022    | 3736599.19                       | 1   |
| 2017/1/20 | 15741.17 | 9174.427796    | 13839.19319    | 39504587.4                       | 1   |
| 2017/1/21 | 9100.32  | 9665.191163    | 9198.309718    | 309477.446                       | 1   |
| 2017/1/22 | 8775.64  | 11084.08671    | 9670.550799    | 4528060.89                       | 1   |
| 2017/1/23 | 11079.61 | 9984.376357    | 11095.96653    | 1199269.2                        | 1   |
| 2017/1/24 | 9507.81  | 16211.39082    | 10014.00886    | 44681758.6                       | 1   |
| 2017/1/25 | 17538.57 | 13363.23617    | 16287.81314    | 15869019.9                       | 1   |
| 2017/1/26 | 16153.93 | 3993.050649    | 13509.94445    | 140896327                        | 1   |
| 2017/1/27 | 3964.87  | 3721.921372    | 4235.054206    | -13975.47                        | 0   |
| 2017/1/28 | 3812.94  | 3198.471431    | 3994.282832    | 344686.4                         | 1   |
| 2017/1/29 | 3533.74  | 3327.397901    | 3434.714495    | 32771.0112                       | 1   |
| 2017/1/30 | 3580.16  | -2402.341098   | 3554.267529    | 35789649                         | 1   |

5. 结论与建议

本文探讨了基于百度指数的网络搜索量与P2P平台成交量、预期收益率、投资人数、借款人数、平均借款期限、借款标数和资金净流入的关系，并以拍拍贷为例，创建两个回归预测模型，结果表明含有滞后一期的百度指数的回归模型比不含有百度指数的回归模型预测效果更好。主要结论如下：

第一，搜索引擎的快速发展为投资者关注互联网金融尤其是P2P平台提供了契机。百度是中国搜索引擎市场的方向标，其旗下的百度指数也能够最大程度的反映投资者和借款者对于P2P平台的关注。本文虽然验证了百度指数与投资人数和借款人数均呈正相关关系，但是百度指数的趋势更多是有投资者驱动还是借款者驱动是没法验证的。互联网金融时代，用户关注和流量都对平台的发展作用不可小觑，甚至是平台能否在互联网金融的浪潮中生存下去的决定因素。在研究P2P平台成交量时，引入百度指数，是符合互联网金融发展的趋势，具有时代性和创新性。

第二，百度指数与P2P平台成交量是正相关关系，同时具有滞后性。本文选取的是滞后一期的百度指数进行分析研究，发现含有滞后一期的百度指数的模型比不含百度指数的模型的决定系数高，并且对于P2P平台成交量有更好的预测能力。不可否认的是近年来，互联网金融尤其是P2P平台呈现爆发式的增长态势，但是监管层的法律法规及相应措施仍处在发展和完善之中。投资者更倾向于在预期收益率高和平均借款期限较短的平台进行投资，如何控制P2P平台的风险非常重要。投资者具有盲目性和羊群效应，一旦关注某一平台的人数骤增，就要时刻警惕随时可能发生的风险。从百度指数的研究视角监测P2P平台可能的风险，无论是对于投资者、借款者还是平台本身的发展，都具有重大的意义。从监管者角度来看，鉴于百度指数与P2P平台成交量的正相关关系和百度指数的前瞻性，监管者可以分析P2P平台的百度指数来检测和控制P2P平台的风险，维护广大投资者、借款者和P2P平台的切身利益。

第三，大数据时代已经到来，研究P2P平台也要加入大数据的时代印记。基于网络搜索量而形成的百度指数不仅为研究互联网金融领域提供新的数据来源，同时利用百

度指数建立预测P2P平台成交量模型，进行风险监测，丰富了P2P领域的研究内容和研究方法。在运用此方法的时候，要注意到本文首先初步判断百度指数和P2P平台成交量存在正相关关系后运用拍拍贷的数据做实证研究，如果二者之间不成正相关关系，能否研究还有待商榷。除此之外，本文所选取的拍拍贷百度指数只是搜索了拍拍贷，但和该平台相关的搜索并没有涉及，在之后的研究中，可以考虑对百度指数进行研究，而并不是简单的运用百度指数的结果。

第四，本文主要探讨了基于网络搜索量的百度指数与P2P平台成交量的关系。平台背景主要有民营系、银行系、上市公司系和国资系四种，其对投资者对P2P平台的选取意义重大。其次，随着互联网技术的不断进步，投资者可获取信息的渠道越来越广泛，信息披露程度也会影响平台成交量。百度指数与平台背景和平台信息披露程度之间的关系在本篇论文中尚未得到验证和解释。此外，P2P平台众多，据网贷之家数据显示，截至2017年2月，运营平台数量为2335家。本文仅仅选取了拍拍贷一家平台进行实证研究，有一定偶然性因素存在，或许由于各种原因，也有某些P2P平台与百度指数存在其他关系，缺乏与其他平台的对比性研究是本文的局限性之一。

第五，百度指数分为移动搜索指数和PC搜索指数，仅就百度指数而言，不同客户端的搜索指数可以侧面反映用户行为。如何把不同客户端的搜索指数纳入到P2P平台成交量预测模型中刻画用户行为，为P2P平台经营者和监管者提供研究支撑是以后可以研究的方向。此外，无论是百度指数研究还是P2P平台研究，本文都没有分地区进行研究，把不同地区百度指数所刻画出来的用户的年龄、性别、收入等信息与P2P平台相应成交量结合在一起，来探索互联网金融时代投资者行为，也许是个不错的研究方向。

参考文献

[1] Krugman, P. How Did Economists Get It So Wrong? [N]. New York Times, 2009.

[2] Simon, Herbert A. On the Behavioral and Rational Foundations of Economic Dynamics [J]. Journal of Economic Behavior and Organizations, 1984, 5(1): 35-55.



- [3] Wilson T D. Human Information Behavior [J]. Informing Science, 2000, 3(2): 49-55.
- [4] Moe W W, Fader P S. Dynamic Conversion Behavior at E-Commerce Sites [J]. Management Science, 2004, 50(3): 326-355.
- [5] Heather A. Johnsona, Michael M. Wagnera, William R. Hogana, Wendy Chapmana, Robert T Olszewskia, John Dowlinga, Gary Barnasb. Analysis of Web Access Logs for Surveillance of Influenza [C]. M. Fieschi. MEDINFO 2004, Amsterdam: IOS Press, 2005: 1202-1206.
- [6] Heather L R, Tierney B Pan. A Poisson Regression Examination of the Relationship between Website Traffic and Search Engine Queries [R]. Working Paper, 2010.
- [7] WE RTHNER H, KLEIN S. Information Technology and Tourism: A Challenging Relationship [M]. Austria: Springer, 1999.
- [8] S KADBERG Y X, SKADBE RG A N, KIM M E L J R. Flow experience and its impact on the effectiveness of a tourism website [J]. Information Technology & Tourism, 2005, 7: 147-156.
- [9] 孔令硕. 基于互联网搜索量的大众途观汽车销量预测研究[J]. 时代金融, 2015(10): 222-226.
- [10] 曹小琳, 牟红. 基于网络搜索的商品房成交量预测研究[J]. 建筑经济, 2016, 37(2): 73-77.
- [11] 刘佳霖. 引入互联网搜索量的市场需求预测模型研究[D]. 北京: 北京邮电大学, 2013.
- [12] 王丹, 张洪潮. P2P网贷平台信用风险评级模型构建[J]. 财会月刊, 2016, (9): 76-80.
- [13] 隋婷婷, 张友棠, 张雅倩. P2P网贷平台资金风险预警指数模型构建—基于三维视角[J]. 财会通讯, 2016(17): 81-85.
- [14] 杨树新. 基于网络关键词搜索的房地产价格影响因素研究[J]. 新疆财经大学学报, 2013(3).
- [15] 樊国虎. 网络搜索数据与消费者信心指数的相关性研究[D]. 长春: 吉林大学, 2014.
- [16] 刘伟江, 李映桥. 基于网络搜索数据的消费者信心指数预测研究——以台湾地区为例[J]. 浙江学刊, 2015(2): 180-186.
- [17] 宋双杰、曹晖、杨坤. [投资者关注与IPO异象——来自网络搜索量的经验证据[J]. 经济研究, 2011(增1) 145-155.
- [18] 朱清香, 王莉. 互联网金融民营系P2P网贷平台的风险预警[J]. 河北大学学报, 2016(4): 123-130.
- [19] 欧阳慧. P2P网络借贷风险预警体系研究——以宜人贷为例[D]. 上海: 华东政法大学, 2015.
- [20] 叶夏菁. 基于BP神经网络与半监督学习的网贷平台信用评估模型[D]. 杭州: 浙江大学, 2015.
- [21] 肖肖, 骆建文. 基于大数据的互联网金融平台信用评级[J]. 现代管理科学, 2015(01):66-68.
- [22] 王程龙, 陈程. 基于决策树的P2P网贷平台信用评级体系研究[J]. 商业银行, 2016(12): 45-50.
- [23] 李敏芳, 田晨君. 基于因子分析法的我国P2P网贷平台评级研究[J]. 湖北经济学院学报, 2015(03): 37-40.
- [24] 孙毅, 吕本富. 网络搜索与经济行为相关性研究综述[J]. 管理评论, 2011(07): 72-77.
- [25] 王博永, 杨欣. 基于网络搜索的房地产政策调控效果研究[J]. 管理评论, 2014(09):78-88.
- [26] 中国互联网络信息中心. 2015年中国搜索引擎市场调查报告[R]. 2015.
- [27] 路紫色, 赵亚红, 吴世峰等. 旅游网站访问者行为的时间分布及导引分析[J]. 地理学报, 2007, 62(6): 621-630.
- [28] 王烁, 宁一鉴, 贾建民. 基于网络搜索的销量与市场份额预测: 来自中国汽车市场的证据[J]. 管理工程学报, 2015(04): 56-64.
- [29] 王烁, 贾建民. 突发性灾害事件风险感知的动态特征——来自网络搜索的证据. 管理评论, 2014(05): 78-88.
- [30] 网络搜索与经济行为相关性研究综述[J]. 孙毅, 吕本富. 管理评论, 2011(07): 72-77.