

文献基本信息

- **题目**: Quantifying Trading Behavior in Financial Markets Using Google Trends (使用 Google 趋势量化金融市场中的交易行为)
- **发表期刊**: Nature----Scientific reports
- **日期**: 2013
- **IF**: 4.996 (Q2)
- **作者基本信息&研究方向**
 - **Tobias Preis** (华威商学院, 华威大学行为科学教授; 图灵研究所研究员)
 - [Computational Social Science](#)
 - [Data Science](#)
 - [Machine Learning](#)
 - [Artificial Intelligence](#)
 - [Forecasting](#)
 - **Helen Susannah Moat** (波士顿大学物理系, 华威大学行为科学教授; 艾伦图灵研究所研究员)
 - [computational social science](#)
 - [data science](#)
 - [online data](#)
 - [human behaviour](#)
 - [forecasting](#)
 - **H. Eugene Stanley**(波士顿大学物理学教授)
 - [Water](#)
 - [Networks](#)
 - [Econophysics](#) (经济物理学)
 - [Biophysics](#)
 - [Complex Systems](#)

文章研究内容

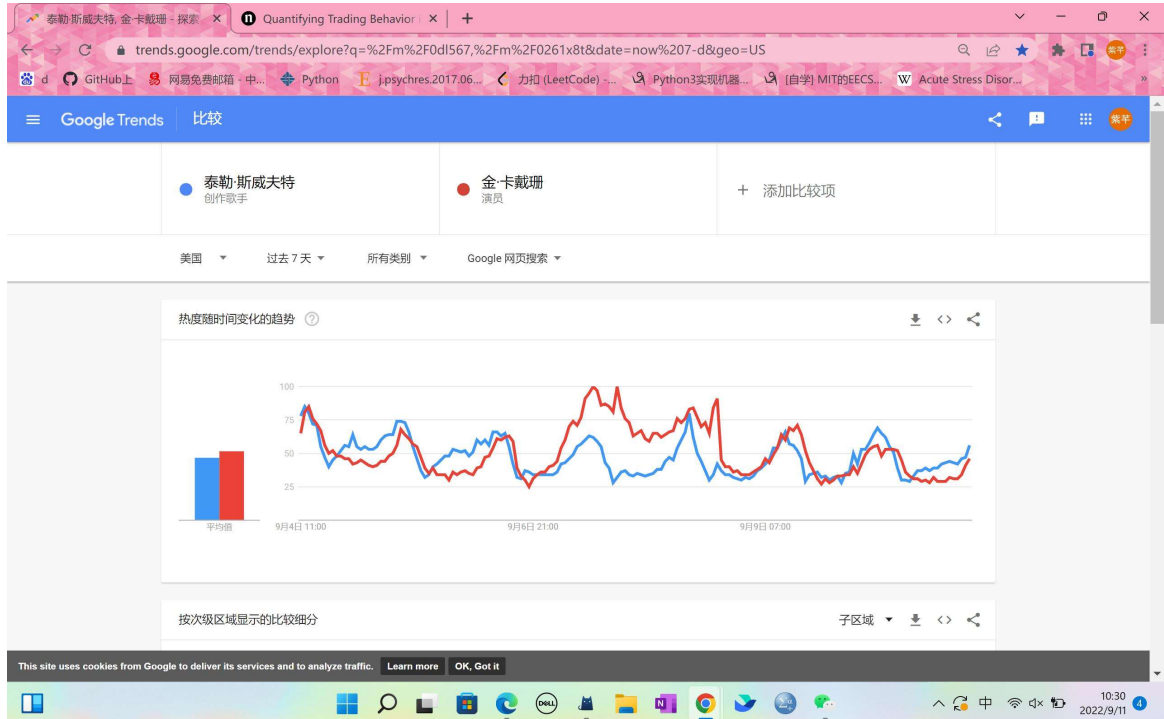
Abstract

- 与金融相关的谷歌搜索数据量的变化→股市走势“预警信号”模式
- 行为数据集→更好地理解人类行为

Introduction

本文主要研究内容

- 金融行业：收集信息→做出决策。如今，信息收集通常包括**搜索在线资源**。
- **谷歌趋势（google trends）**：谷歌其提供对不同搜索词的查询量以及这些量如何随时间变化的汇总信息的访问。



- 在本研究中，作者调查分析了来自 Google 趋势的搜索查询数据的有趣可能性，以提供对股票市场数据中记录的交易决策之前的**信息收集过程**的新见解。

之前相关的研究

- 来自特定国家/地区的搜索结果的点击次数与**该国家/地区的投资额**相关。
(Mondria, J., Wu, T. & Zhang, Y, 2010)
- 利用谷歌趋势数据的时间维度的进一步研究表明，选定搜索词的查询量变化反映了当前**流感病例数量** (Ginsberg, J. et al, 2009) 和**股票市场交易量的变化** (Choi, H. & Varian, H, 2012) 。
- **股票市场交易量和搜索量**之间的联系也已使用 **Yahoo!** 进行了复现。 (Bordino, I. et al, 2012)
- 来自 Google Trends的数据可以与各种经济指标的当前值相关联，包括**汽车销售、失业申请、旅游目的地规划和消费者信心**。 (Choi, H. & Varian, H, 2012)
- 来自**人均 GDP 较高国家**的互联网用户比过去几年更有可能搜索**有关未来年份的信息**。 (Preis, T., Moat, H. S., 2012)

总结本文假设以及一些研究结果讨论

- 本文假设，在分析的时间段内，**谷歌趋势数据不仅反映了股票市场的当前状态，而且还可能能够预测某些未来趋势。**
- 作质的研究结果与一个有趣的提议是一致的，即**金融市场的显著下跌之前**是投资者关注的时期，在一时期，投资者可能会在最终决定买入或卖出之前搜索有关市场的更多信息。
- 研究结果表明，在 2004 年到 2011 年期间，谷歌趋势中某些术语的搜索查询量可以用于预测更有利的交易策略 **(搜索量→预测交易策略)**。

Results (重点关注数据分析)

搜索词组的确定 (search terms)

- 分析了98个搜索词的性能 (搜索量的变化→预测金融危机的能力)
- 搜索词的确定:
- **Google Sets**: 从一组事物集合中输入一些词组, **Google Sets** 将尝试预测集合中的其他词。**作者使用与股票市场概念相关的术语, 其中一些由 Google Sets 服务建议。**
- 参考术语**债务 (debt)**, 一个与最近的金融危机有明显语义联系的关键词, 以及在分析中表现最好的术语, 根据它们搜索量的变化来解释策略。

数据的选择 (揭示特定术语的搜索查询量与交易者决策的总体方向之间的关系)

- **道琼斯工业平均指数 (DJIA)** 在第 t 周的第一个交易日的收盘价 $p(t)$
道琼斯工业平均指数 (Dow Jones Industrial Average, DJIA, Dow (/ˈdaʊ/), 简称“道指”) 是由《**华尔街日报**》(The Wall Street Journal) 和**道琼斯公司**联合创始人**查尔斯·道**创造, 并以他和他的商业伙伴暨统计学家**爱德华·琼斯**的名字命名的**股票市场指数**。**道琼斯工业平均指数不是加权算术平均值, 并不代表其组成公司的市值, 而是每个组成公司的一股股票价格总和后的平均值。是在美国证券交易所上市的 30 家著名公司的价格加权衡量股票市场指数。**
- 使用**Google Trends (谷歌趋势)** 来确定在第 $t - 1$ 周内针对**特定搜索词** (例如**债务**) 进行了多少次**搜索 $n(t - 1)$** 。

信息搜索行为的量化

- 使用**搜索量的相对变化**来量化 (t 以周为单位测量) :

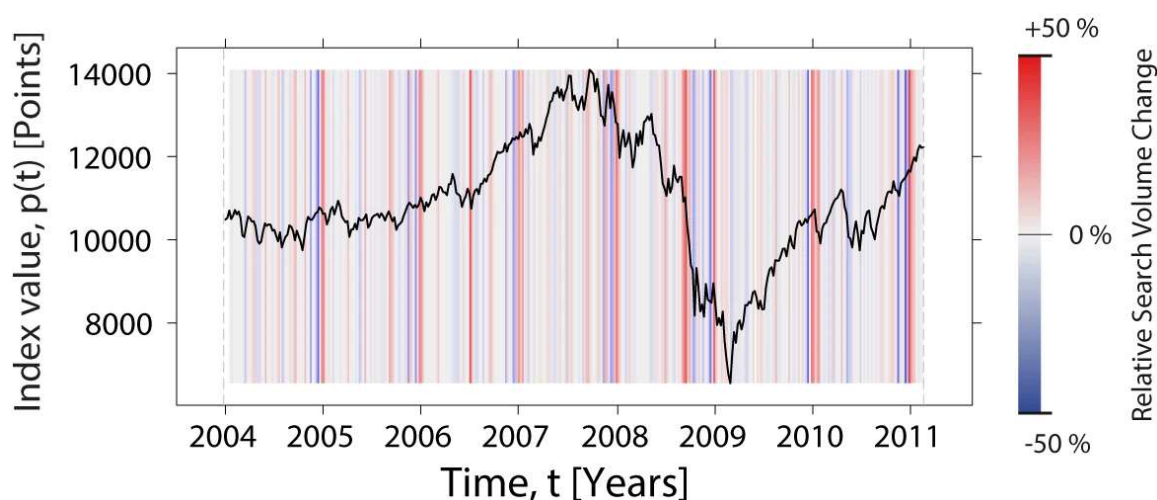
$$\Delta n(t, \Delta t) = n(t) - N(t-1, \Delta t))$$

其中：

$$N(t-1, \Delta t) = \frac{(n(t-1) + n(t-2) + \dots + n(t-\Delta t))}{\Delta t}$$

数据的描述性统计

- **搜索量与股市走势**：2004 年 1 月 5 日至 2011 年 2 月 22 日期间，每周交易第一天道琼斯工业平均指数 (DJIA) 收盘价 $p(t)$ 的时间序列。不同颜色对应搜索词债务 (debt) 相对搜索量的变化， $\Delta t = 3$ 周。（搜索量数据仅限于美国本地用户的请求）



Google Trends strategy(为了验证假设，即搜索量与股市走势存在关系，并且搜索量可以预测，作者进行了投资实验，不是现实世界的)

首先了解一些股市术语

long position：又称多头，好仓，市价看涨（股票术语）；采取多头市场；对市场价格持看涨观点。在证券中拥有“多头”头寸意味着**拥有该证券**。投资者维持“多头”证券头寸，预期该股未来会升值。“多头”头寸的反面是“空头”头寸。投资者对股市看好，预计股价会看涨，于是会在股票价低的时候买进，等到股票上涨到一定价位时再卖出，以此获得差额收益。**这种投资方式以先买进为基础，投资者在未售出前手头多有了一只股票或商品，故称为多头。**

short position：指的是“空头，淡仓，市场看跌，证券或外汇交易中未补抛空差额，欠缺的头寸”。**空头和多头是相对的**，投资者和股票商认为现时段的股价虽然比较高，但是对股市的前景不看好，预计股价将会下跌。于是把股票卖出，希望在股票跌价以后再买回或补进，以此获得差额收益。**投资者在卖出后至未买回或补进前，手中并无实物，称为空头。**

- 研究谷歌趋势数据捕获的**信息收集行为**的变化是否与 2004 年至 2011 年期间**股价的后期变化**有关，作者使用**搜索量数据**设计了投资策略（**Google Trends strategy**），分别为：

- 当 $\Delta n(t-1, \Delta t) > 0$ 时，即搜索量增加，第 t 周的第一个交易日以收盘价 $p(t)$ 卖出 DJIA，并在下周第一个交易日以价格 $p(t+1)$ 买入（即**空头**，之前的假设：搜索量增加时，意味着经济形势不好，交易者会在做出决策前在网上搜寻相关信息）。
- 当 $\Delta n(t-1, \Delta t) < 0$ 时，即搜索量降低，在第 t 周的第一个交易日以收盘价 $p(t)$ 买入 DJIA，并在第 t 周结束时以 $p(t+1)$ 的价格卖出。（即**多头**，预期股价会上涨，经济形势较好）

- 定义累积收益**：the cumulative return，简写为：R。

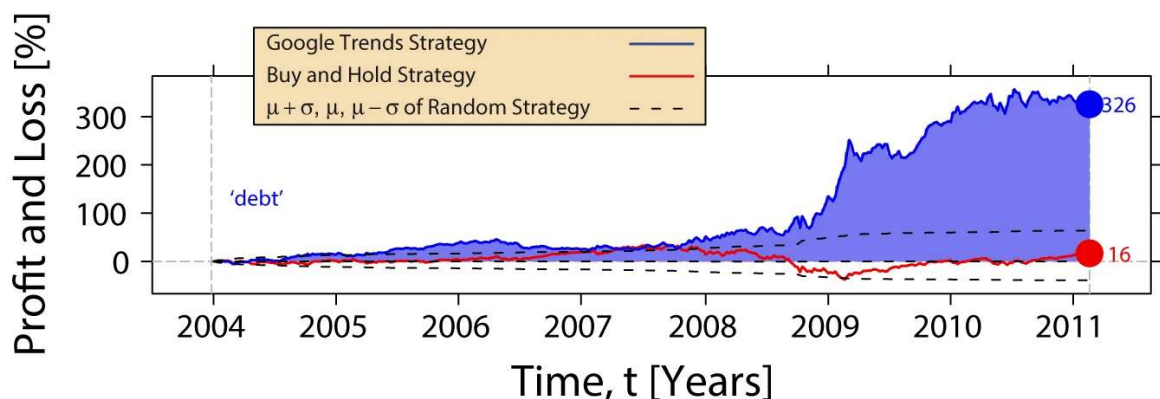
- 如果持有“空头”，以收盘价 $p(t)$ 卖出并以 $p(t+1)$ 价格回购，那么累积收益R会改变：

$$\Delta R = \log(p(t)) - \log(p(t+1))$$

- 如果持有“多头”，收盘价 $p(t)$ 买入并以 $p(t+1)$ 价卖出，那么累积收益R会改变：

$$\Delta R = \log(p(t+1)) - \log(p(t))$$

- 模拟结果**：以词debt（债务）的搜索量制定投资策略为例



- 蓝色线：表示**依据数据搜索量制定的投资策略**，使用基于搜索词debt（债务）和 $\Delta t = 3$ 周的 Google 趋势策略，可以将投资组合的价值**提高 326%**
- 虚线：**随机投资策略**，表示以不相关的随机方式买卖市场指数的策略的累积回报的标准偏差，标准差来自对随机投资策略的 10,000 次独立实现的模拟。
- 红色线：表示**买入持有策略**（buy and hold），字面意思，是简单的长期投资策略，**忽略短期金融市场的波动及衰退，专注于长期成长与收益率。**

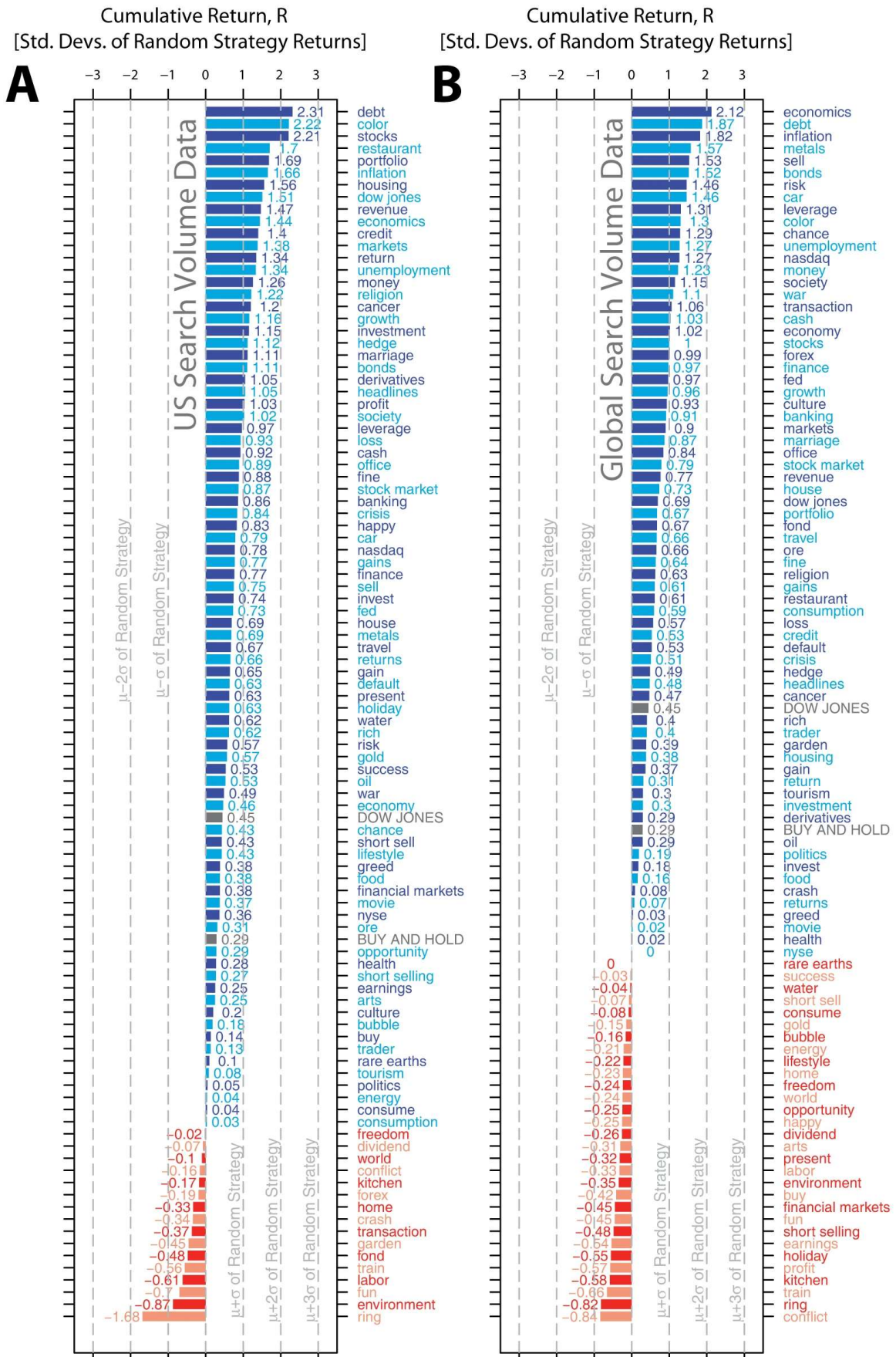
• 模拟结果 (98个词组) :

- 根据交易表现 (即累计回报R) 对98个搜索词进行排名, 其中图A: 仅使用美国用户的搜索数据; 图B: 全球生成的搜索量。
- 为了确保结果的稳健性, 基于给定搜索词的策略的整体交易表现被确定为 $\Delta t = 1 \dots 6$ 周获得的六次回报的平均值。
- 随机投资策略**产生的最终投资组合价值分布接近**对数正态分布**。从这些投资组合值的对数得出的**随机投资策略的累积回报因此服从正态分布**, 平均值为 $R(\text{RandomStrategy}) = 0$ 。
- 时间: 2004年1月5日-2011年2月22日; 两种**蓝色**表示**正回报**, 使用两种**红色**表示**负回报**; **灰色块: 买入持有策略和道琼斯策略** (使用道琼斯工业平均指数 (DJIA) 的每周收盘价预测走势)
 - 买卖持有策略:** 在开始时买入并在持有期结束时卖出, 该策略产生 16% 的利润, 相当于 DJIA 在 2004 年 1 月至 2011 年 2 月期间的整体价值增长。
 - 道琼斯策略:** 该策略在 $\Delta t = 3$ 周时也仅产生 33% 的利润, 或者当确定为 $\Delta t = 1 \dots 6$ 周获得的六个回报的平均值时, 为不相关随机投资策略的累积回报的 **0.45** 个标准差。
- 使用美国用户数据时, 单样本T检验:** 谷歌趋势策略的回报总体上显著高于随机策略的回报 ($R(\text{US}) = 0.60$; $t = 8.65$, $df = 97$, $p < 0.001$) 。
- 使用全球用户数据时, 相关样本T检验,** 将搜索量分析范围扩大到全球用户会**降低**谷歌趋势交易策略在美国市场取得的整体回报。 ($R(\text{US}) = 0.60$, $R(\text{Global}) = 0.43$; $t = 2.69$, $df = 97$, $p < 0.01$) ; **单样本T检验:** 回报仍然**显著高于**随机投资策略的平均回报 ($R(\text{Global}) = 0.43$; $t = 6.40$, $df = 97$, $p < 0.001$) 。

说明: 投资者更喜欢在国内市场进行交易, 这表明迄今为止分析中使用的仅针对美国用户的搜索数据应该比针对全球谷歌用户的数据更好地捕捉美国股市参与者的信息收集行为

- 结果表明, **谷歌趋势策略的表现因选择的搜索词而异**。作者想是否可以使用**不同术语与财务相关程度的指标来部分解释这些性能差异**, 将每个搜索词的谷歌点击次数标准化 (将其分级, 计算与财务的相关程度, 作者标准化的方法: 通过计算 2004 年 8 月至 2011 年 6 月金融时报在线版 (<http://www.ft.com>) 中每个搜索词的频率来量化财务相关性) 。使用 Kendall's tau 等级相关系数, **与给定搜索词相关的回报与财务相关性指标** ($\text{Kendall's tau} = 0.275$, $z = 4.01$, $N = 98$, $p < 0.001$) 相关。

在统计学中, Kendall等级相关系数, 通常称为Kendall的tau系数 (在希腊字母 τ 之后), 是用于测量两个测量量之间的序数关联的统计量



• 数据分析结果与假设

- 实证结果与两部分假设一致：即DJIA价格的主要上涨之前是某些金融相关术语的搜索量下降，相反，DJIA价格的主要下跌之前某些与财务相关的术语的搜索量增加。

- 交易策略可以分解为两个策略组成部分：搜索量的减少促使买入；搜索量的增加促使卖出。
 - 为了验证这两个策略组成部分在结果中都发挥了重要作用，以便我为该假设的两个部分提供证据，作者进行了两次不同策略，分别为搜索量减少后做多头（买入），但从不做空头（卖出）和在搜索量增加后做空头（卖出），但从不做多头（买入）。
 - **单样本T检验**结果如下：
 - 多头策略：R (USLong) = 0.41; $t = 11.42$, $df = 97$, $p < 0.001$
 - 空头策略：R (USShort) = 0.19; $t = 5.28$, $df = 97$, $p < 0.001$
 - 谷歌趋势策略的两个组成部分的回报总体上显著高于随机投资策略的回报。
 - **Kendall's tau相关分析**
 - 财务相关性指标与多头策略回报之间存在正相关关系 (Kendall's $\tau = 0.242$, $z = 3.53$, $N = 98$, $p < 0.001$)
 - 财务相关性指标与空头策略回报之间存在正相关关系 (Kendall's $\tau = 0.275$, $z = 4.01$, $N = 98$, $p < 0.001$)。

Discussion

- 结果与假设一致，**谷歌趋势数据不仅反映了当前经济状况的各个方面，而且可能还提供了对经济行为者行为未来趋势的一些洞察。**使用 2004 年 1 月至 2011 年 2 月期间的历史数据，发现在股市下跌之前，谷歌搜索与金融市场相关的关键字的搜索量有所增加。**结果表明，搜索量数据中的这些警告信号可以被用于构建有利的交易策略。**
- **Herbert Simon 的决策模型（有限理性模型）**的背景下对结果提供了一种可能的解释。谷歌趋势数据和股市数据可能反映投资者决策过程中的两个后续阶段。在金融市场上以较低价格出售的趋势可能会出现一段时间的担忧。在这样的担忧时期，人们可能倾向于收集更多关于市场状况的信息，这种行为可能反映在谷歌趋势财务相关性术语搜索量的增加上。
- 基于美国用户搜索量数据的策略在美国市场上比使用全球搜索量数据的策略更成功。假设美国互联网用户人口中包含的美国市场交易者比例高于全球互联网用户人口所包含的比例。
- 在这项工作中，作者**量化了搜索量变化与股票市场价格变化之间的关系。**未来的工作将需要对**导致人们在以较低价格出售股票之前搜索债务等术语的潜在心理机制进行解释。**
- 结果表明，将金融交易数据等大型行为数据集与搜索查询量数据相结合，**可能会为大规模集体决策的不同阶段提供新的见解。**这些结果进一步说明了**新的大数据集提供的令人兴奋的可能性，以促进对社会中复杂集体行为的理解。**