



网络搜索与股票市场相关性研究



报告人：刘颖

报告时间：2011年9月22日

主要内容

- 选题背景、研究内容及意义
- 主要文献回顾
- 网络搜索与股票市场关联机制分析
- 网络搜索指数合成方法
- 搜索指数与股票市场关系的实证检验
- 基于搜索指数的投资者情绪非对称效应研究
- 不足及展望



选题背景

- 股价调整是投资者接收信息、更新预期、执行交易的过程：
 - ▶ 对投资者行为预期的度量，以往多采取调查数据或代理变量方法
 - 调查数据是一种直接测度法，较为真实有效，但样本量有限，并存在统计延迟；
 - » 美国个人投资者协会指数、投资者智慧指数
 - » 央视看盘指数、好谈指数
 - » 分析师预期指数、投资者信心指数
 - 代理变量是一种间接测度法，虽然客观及时，但也难免存在一定偏差；
 - » 封闭基金折价率 (Zweig, 1973)
 - » 成交量或换手率 (Baker M, 2004)
 - 更高质量的数据源对研究结论的有效性具有重要意义 (Simon, 1984)



选题背景

- 网络技术与搜索引擎的发展为行为研究开辟了新的领域：
 - ▶ 网络搜索与社会经济行为相关性研究成为一个新的热点
 - 搜索数据基于全体网民“准自然”采样，具有普遍性与及时性
 - 反应用户关注与需求，映射现实生活中的行为趋势和规律，可以作为行为过程直接、及时的度量指标
 - 搜索数据作为传统数据的良好补充，对某些典型的社会、经济行为存在一定的预测能力
 - ▶ 本文以中文搜索数据为基础，选择中国股票市场为研究对象，探讨二者之间关系的存在性及表现特征
 - 股票市场与互联网联系紧密，市场总交易量中网上交易的比例已经达到80%（马光悌，2009），说明搜索数据在股票交易行为中具有一定普遍意义



相关概念界定

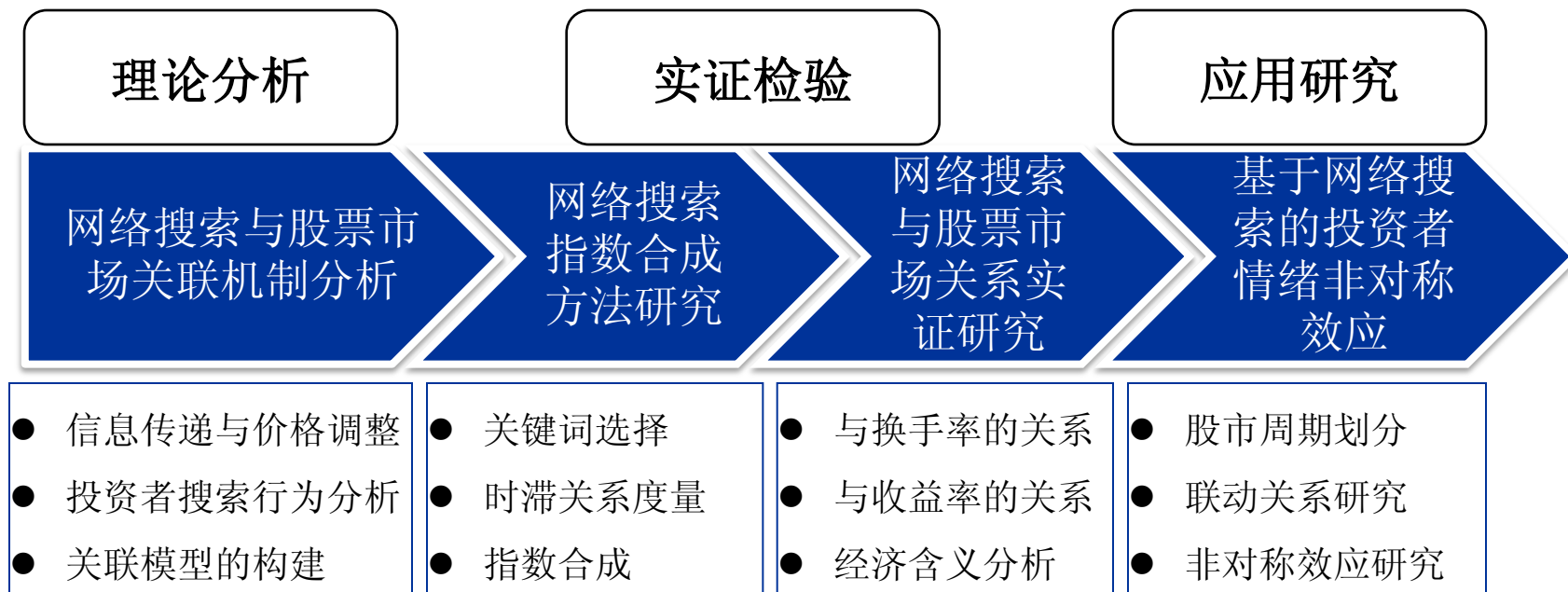
● 概念界定

- ▶ **搜索关键词：**指用户使用搜索引擎查找相关信息时输入搜索框中的文字，如“上证指数”、“股市行情”、“600159”等。
- ▶ **搜索数据：**即互联网上的搜索频次，指在某一时间范围内某些关键词的标准化搜索量。
- ▶ **搜索指数：**一般来讲，一个关键词仅能从一个侧面反映某个事件的特征，衡量其总体趋势需要由尽可能全面的关键词共同反映。将有效关键词按照某种算法组合成连续的时间序列数据，称之为搜索指数。



主要研究内容

● 研究内容:



主要文献回顾：网络搜索行为研究的主要领域

医药卫生领域

■ 搜索数据对流感疫情的监测

- Johnson和Heather (2004) 研究网站日志发现医学网站访问次数以及流感文章搜索次数与流感发病人数之间具有很强的相关性。
- Ginsberg等 (2009) 发现Google中与流感类关键词搜索量与流感看诊量有很强的相关性，由此构建搜索数据的监测模型，该模型比传统方法提前2周测算出流感的爆发趋势。该文在Nature的发表引发学者广泛关注。



主要文献回顾： 网络搜索行为研究的主要领域

医药卫生领域研究汇总

学者	数据来源	主要结论
Heather A(2004); Johnson(2004)	医学类网站有关流感的信息访问次数	相关信息访问次数与CDC公布的流感人数存在相关性
Eysenbach(2006) Polgreen(2008); Philip (2008); Hulth(2009); Wilson(2009) Jurgen(2009)	网民利用搜索引擎对流行性感冒的相关关键词的搜索次数	搜索指数与流感发病率存在长期稳定关系，利用指数可以提前预测出流感发病情况及流感的死亡率
Jeremy Ginsberg(2009)	谷歌日志中的5000万个关键词中与流感发病率拟合最好的45个关键词	利用搜索指数可以比传统方法提前1-2周预测流感发病率，且可以分地区精确监测
Joshua Ritterman(2009)	利用 蜘蛛 程序 在 UGC(User Generate Content)媒体抓取的关键词	利用流感的历史数据和关键词指标组合预测的效果最好



主要文献回顾：网络搜索行为研究的主要领域

宏观经济领域

■ 失业率的预测

- Askitas (2009) 对网络搜索与德国失业率之间的关系做出了实证研究，结果表明就业类词汇搜索量与当月失业率有着较强的相关性；
- Francesco (2009) 以意大利、美国的失业率为研究对象，用搜索数据与经济数据（工业生产指数、就业预期指数）的预测相比较，发现搜索数据预测失业率效果优于传统经济数据。



主要文献回顾：网络搜索行为研究的主要领域

宏观经济领域

■ 居民消费的预测

- Torsten和Simeon (2009)认为基于统计调查的消费者信心指数是宏观因素的反映，并未体现出预期与决策之间的关系；而网络搜索数据可以弥补该不足；对比发现，搜索指数对消费的预测精度好于消费者信息指数，且搜索指数能够预测2008年12月的消费拐点；
- Nicol á s (2009)， Marta (2009)， 也得到类似的结论



主要文献回顾：网络搜索行为研究的主要领域

宏观经济领域-股票市场

- Zhi Da, Joseph and Pengjie, 2009, In Search of Attention. Finance Department, Mendoza College of Business, University of Notre Dame.
- 样本: Russell 3000 stock
- 投资者注意力代理变量: 换手率、超预期收益、新闻量、广告花费等
- 而搜索数据可以作为注意力的直接度量指标
- 搜索数据与以往代理变量的关系: 搜索数据 \Longleftrightarrow 代理变量
- 结论1: 搜索的增加导致交易量的增加; 并且对散户的影响大于机构
- 结论2: 高关注度导致高的上涨动力;



主要文献回顾：网络搜索行为研究的主要领域

典型行业与商品的预测

■ 研究对象：零售品、汽车等行业

- H Choi, H Varian, 2009, Predicting the Present with Google Trends [C], Technical Report, Google Inc.
- 结论：用每月前两周的搜索数据预测当月销量，发现加入搜索项后预测精度明显的改善
- 研究方法：时间序列模型（自回归+搜索项）

■ 研究对象：美国房地产市场的销量和价格

- Lynn Wu, and Erik Brynjolfsson, 2009, The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales, MIT Sloan School of Management.
- 结论：搜索指数对未来的销量和价格具有预测能力，搜索指数每增长1个百分点，未来一季度的销量将增长67220套。
- 方法：自回归+搜索项



主要文献回顾：网络搜索行为研究的主要领域

典型行业与商品的预测

- 研究对象：旅游客流量预测
- Varian (2009) 研究香港旅游客流量时，将搜索数据作为新解释变量加入到季节性AR模型中，结果表明，搜索数据变量在5%的显著水平上通过检验，对客流量存在显著影响。
- 彭赓等，基于网络搜索数据的上海世博会日客流量预测方法研究，2011；
- 研究结论：网络搜索指数对上海世博会日客流量具有领先性，因此，搜索指数可以作为历史数据的良好补充，有效地预测出未来七天的上海世博会日客流量。



主要文献回顾：文献评述

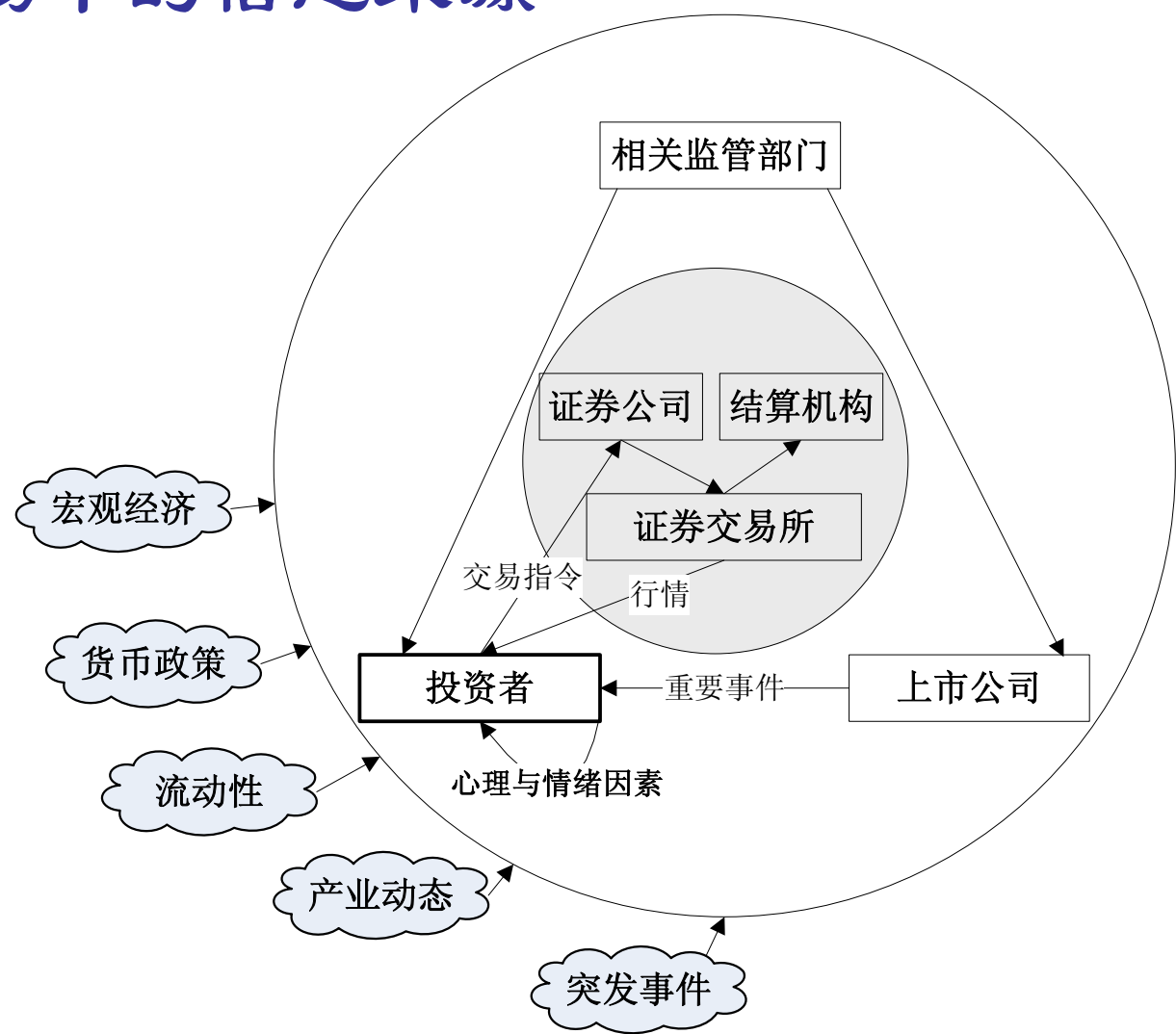
现有研究评述及不足

- 网络搜索数据与社会经济行为之间存在较强的关联性，在传统的预测模型的基础上，加入搜索指数新变量能有效地提高预测精度。
- 现有研究的一些不足
 - ▶ 理论层面，对相关性背后的机理分析得不够充分
 - ▶ 研究方法上，如何收集、处理和分析网络搜索数据，目前国内外尚未形成一套成熟的方法，尤其是在关键词选择、搜索指数合成等方面需要更深入的研究
 - ▶ 能否应用中文搜索数据预测中国社会经济行为，目前缺乏系统性的研究成果



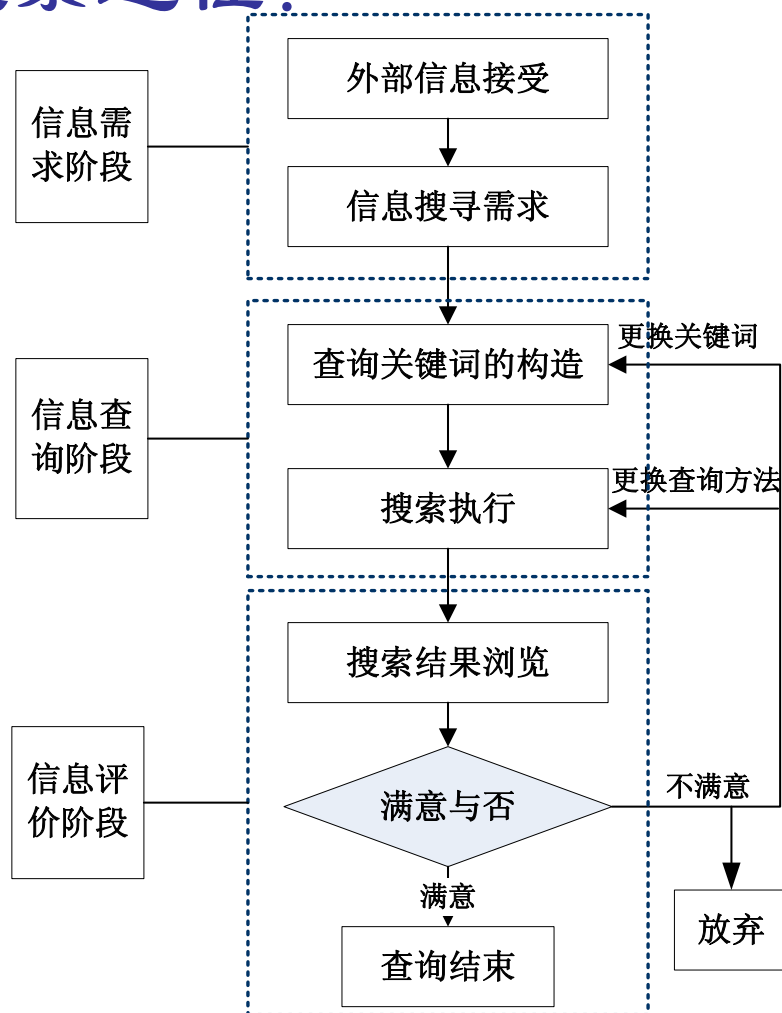
网络搜索与股票市场关联机制分析

股票市场中的信息来源



网络搜索与股票市场关联机制分析

投资者信息搜索过程：

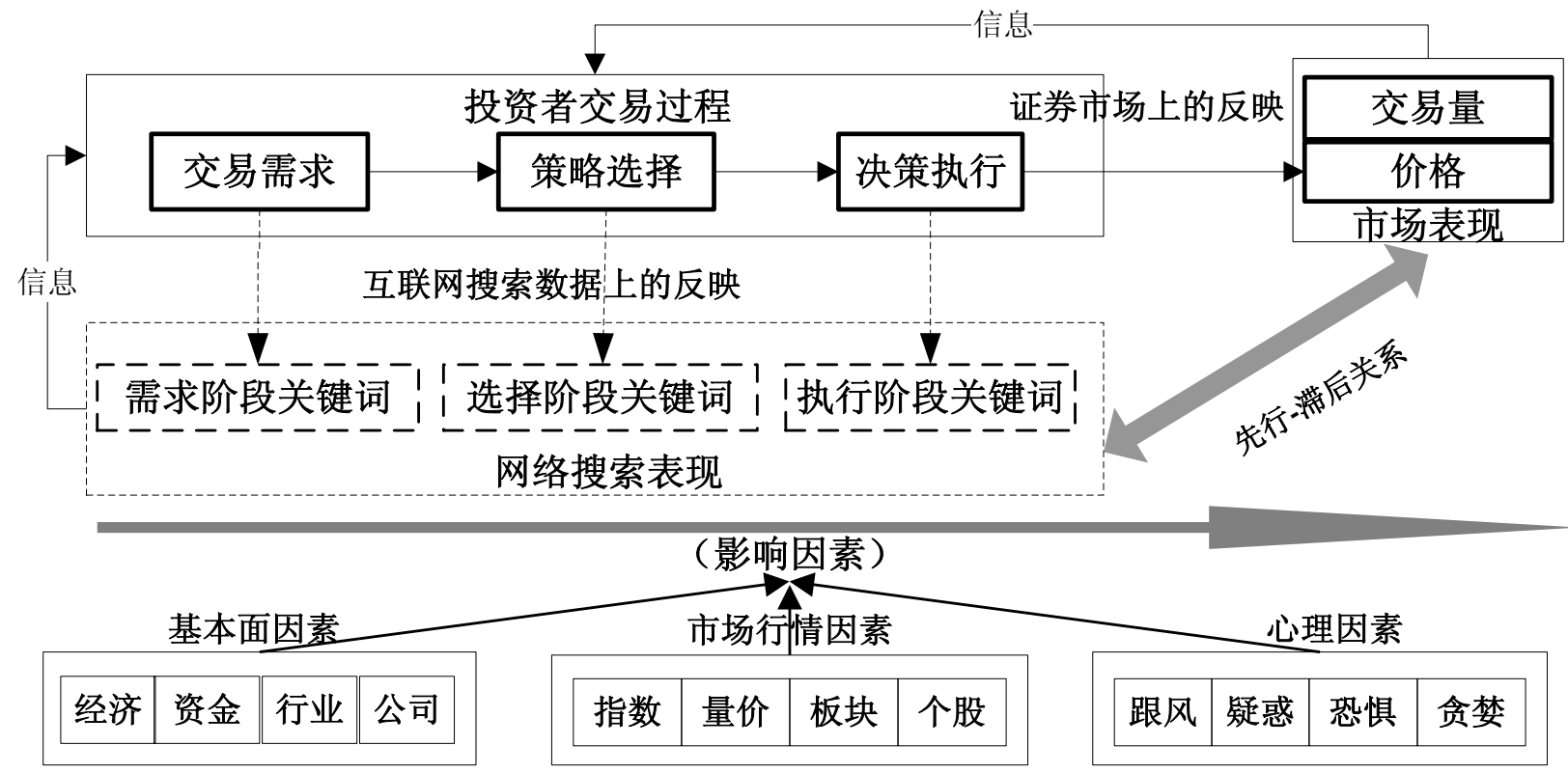


信息搜寻行为的一般模式(Kalervo和Wilson,2003模型的改进)



网络搜索与股票市场关联机制分析

网络搜索与股票市场的关联模型



注：实线代表投资者行为在股票市场的反映，虚线代表投资者行为在网络搜索的反映

股票市场与网络搜索关系的概念模型

网络搜索与股票市场关联机制分析

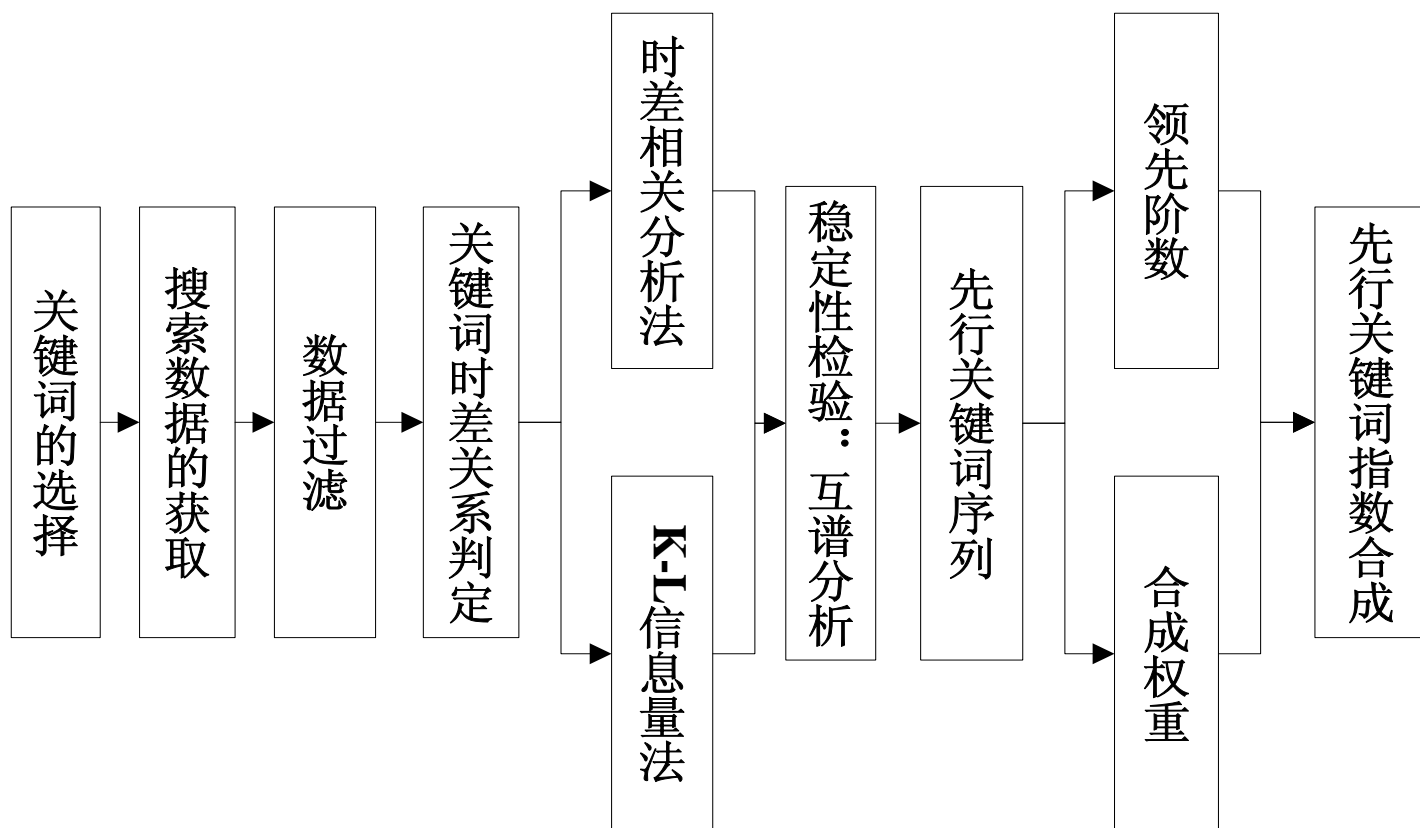
领先性搜索数据的特点：

- 搜索关键词所代表的经济含义具有预期、指示性的作用，这类搜索数据一定程度上反映了投资者的行为趋势
- 搜索关键词所代表的行为活动在传导流程上处于股票量价调整的前端，并具有一定的时间间隔，行为活动间隔股票量价调整越长，这类搜索数据的领先性越明显。



网络搜索指数合成方法

搜索数据预处理流程：



网络搜索指数合成方法

关键词的选择：

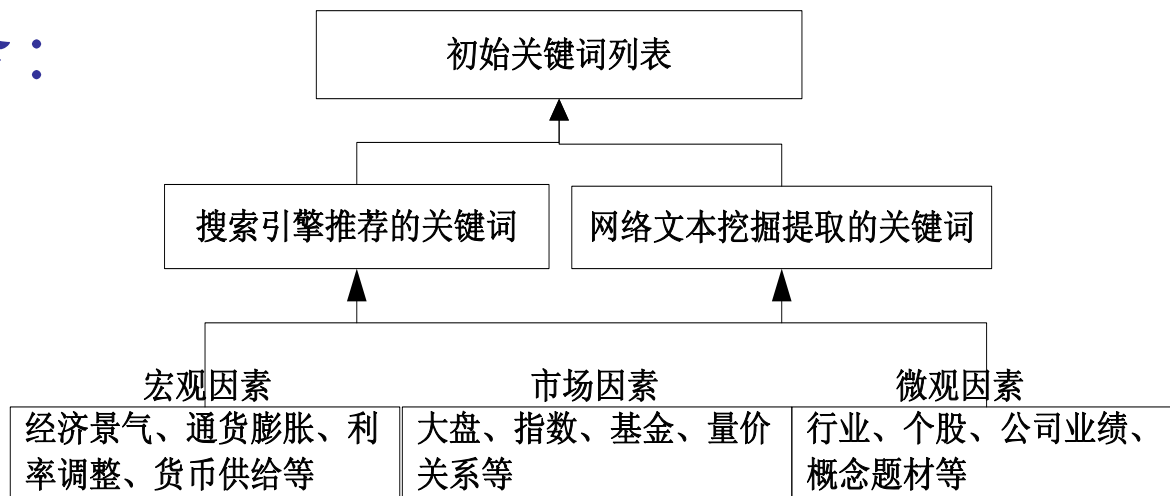
作者	取词归类	具体方法
Ginsberg（2009）	技术取词法	利用800台计算机对5000万个词进行统计检验。
Joshua（2009）	技术取词法	利用蜘蛛程序结合文本挖掘方法在网络UGC资源提取。
Hulth（2009）	范围取词法	分析与流感相关的专业类网站频率最高的20个词，从这20个词中调选拟和最好的。
Konstantin(2009);Tanya(2009) ； Brynjolfsson(2009); Varian(2009) ； Nicolás(2009)	范围取词法	在搜索引擎提供的与主题相关的热点推荐词中挑选。
Askitas(2009) ； Francesco（2009） ； Philip(2008)等	直接取词法	根据先验知识直接选择与特定事件相关的关键词。

以往研究中的关键词选择方法



网络搜索指数合成方法

关键词的选择：



- 关键词来源的拓展，以尽可能保证全面性
 - ▶ 论坛：分时段收集并保存各大论坛中的相关帖子
 - ▶ 博客：分时段收集并保存相关博文
 - ▶ 网络新闻：分时段收集并保存相关重要新闻
- 采用文本挖掘进行关键词的提取
 - ▶ 中文分词工具（开源）：[ICTCLAS](#)



网络搜索指数合成方法

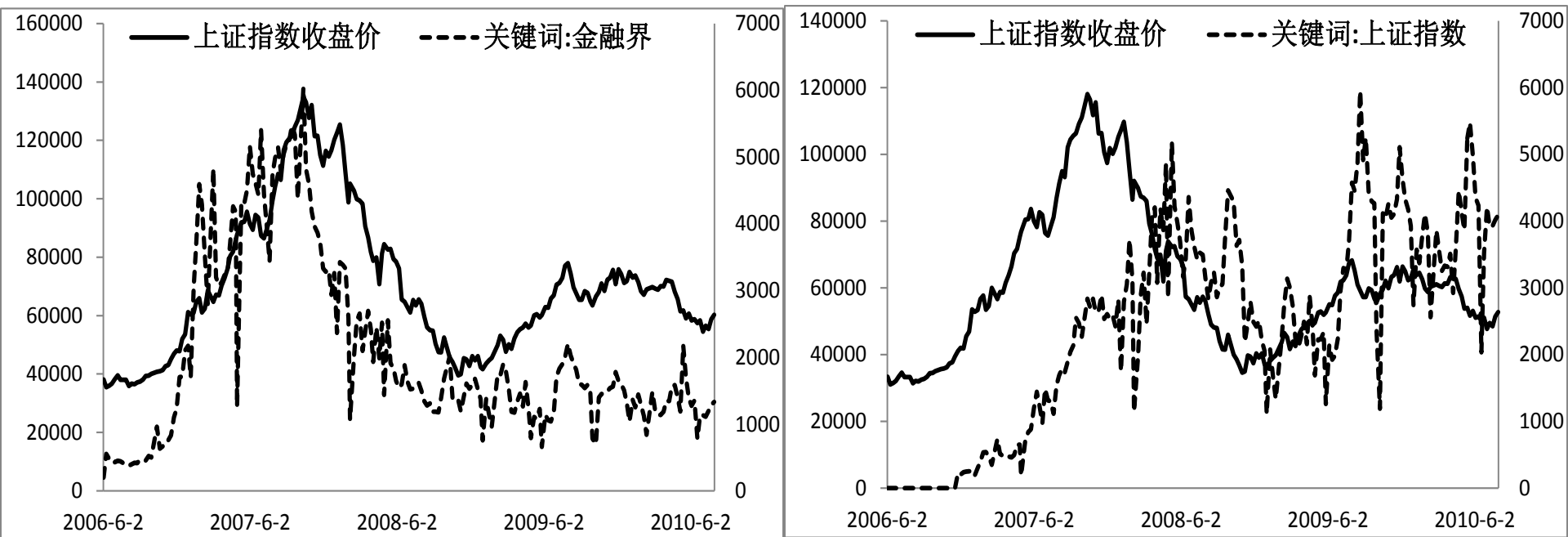
初始关键词列表：

关键词来源	个数
搜索引擎推荐	500
网络新闻关键词提取	300
网络论坛关键词提取	300
网络博客关键词提取	300
成分股代码及名称	1200
总计	2600
去重后总计	2103



网络搜索指数合成方法

关键词时差关系判断：



关键词搜索趋势与上证指数走势的对比

网络搜索指数合成方法

关键词时差关系判断：

- 关键词评价参数：

- ▶ 领先阶数：度量与基准指标的领先性，小于0表示先行关系
- ▶ 相关系数：表示关键词的重要性，搜索曲线与基准指标相似程度

- 相关方法：

- ▶ 时差相关分析法
- ▶ K-L信息量法
- ▶ 峰谷对比分析法

- 关键词稳定性检验：

- ▶ 互谱分析法



网络搜索指数合成方法

搜索指数合成：

- 领先阶数错位对齐：

- ▶ 选择先行关键词进入合成列表
- ▶ 根据领先阶数交错对齐

- 合成权重：

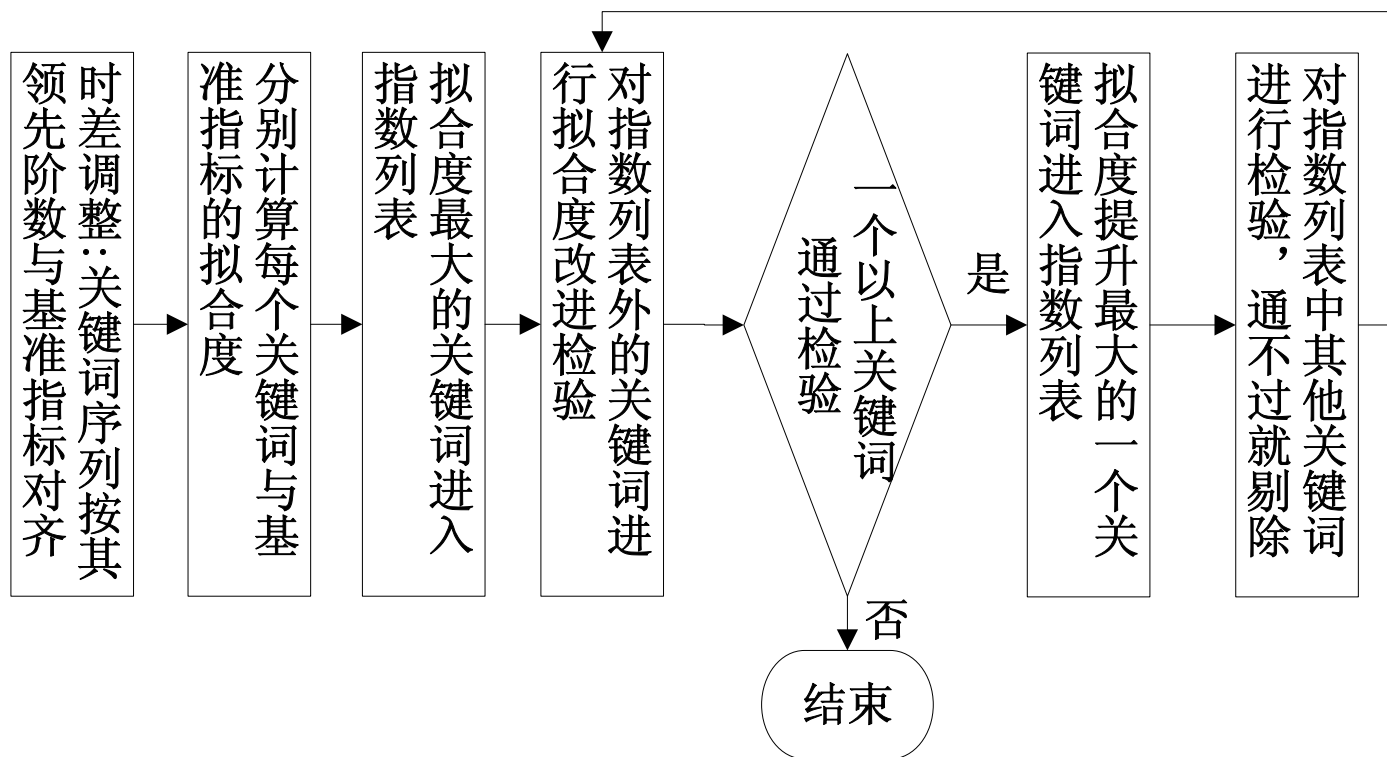
- ▶ 根据重要性赋权，定义为相关系数的函数，即 $w=f(r)$

- 合成方法：

- ▶ 简单相加
- ▶ 逐步合成法



逐步合成法



逐步合成法的流程图

搜索指数与股票市场关系的实证检验

检验所回答的问题：

- 网络搜索指数在一定程度上能够体现市场内外部影响因素作用于投资者心理后，所引发的投资者预期与情绪的变化，那么，搜索指数与以往文献中提到的市场流动性、投资者情绪等指标是否存在关联性？
- 网络搜索指数与股票市场中的被解释变量（股价、收益率）是否存在关联关系？与其他股票市场中的解释变量（成交量、换手率）相比，网络搜索指数的解释力如何？



搜索指数与股票市场关系的实证检验

关键词的分类：

- 基本面类关键词：宏观形势、政策、交易需求等词汇
 - ▶ 加息、升值、准备金、债务危机、地产政策、投资股市等
 - ▶ 反映对基本面信息的关注度、股票交易的需求
- 市场行情类关键词：
 - ▶ 上证指数、沪深300、新兴行业板块、绩优股、个股代码名称等
 - ▶ 反映对市场行情的关注度、投资策略的选择
- 投资者行动倾向类关键词
 - ▶ 利好/空、放量上涨/下跌、调整、建仓、反弹
 - ▶ 带有一定倾向性，反映投资者之间心理和情绪的相互影响



搜索指数与股票市场关系的实证检验

- 以加权合成方法，将关键词合成为三类搜索指数：
 - ▶ 基本面指数：X1
 - ▶ 市场行情指数：X2
 - ▶ 行动倾向指数：X3
- 股票市场指标的选择：
 - ▶ 股票价格：sp
 - ▶ 收益率： $rsp = \ln(sp_t) - \ln(sp_{t-1})$
 - ▶ 成交量：v
 - ▶ 换手率：turn
- 与收益率对应，对指标进行变化率转换：
 - ▶ rx1、 rx2、 rx3、 rv、 rturn



搜索指数与股票市场关系的实证检验

● 假设提出

- ▶ **H1:** 网络搜索指数与投资者情绪指标存在一定关联性，并且网络搜索指数包含了比成交量、换手率指标更多的信息量。
- ▶ **H2:** 网络搜索指数与股票市场中的股价或收益率等结果性变量存在显著的关联关系，并且与其他指标（成交量、换手率）相比，网络搜索指数的解释力更强。



搜索指数与换手率关系的实证检验

平稳性检验：

搜索指数、股票市场变量的单位根检验

变量	符号表示	ADF检验	MacKinnon临界值			P值	ADF检验结果
		t-Stat	1%	5%	10%		
上证指数收盘价	sp	-0.60	-2.58	-1.94	-1.62	0.46	I (1)
	Δ sp	-3.29	-2.58	-1.94	-1.62	0.00	
上证指数成交量	v	-0.48	-2.58	-1.94	-1.62	0.51	I (1)
	Δ v	-8.09	-2.58	-1.94	-1.62	0.00	
上证指数换手率	turn	-1.24	-2.58	-1.94	-1.62	0.20	I (1)
	Δ turn	-6.57	-2.58	-1.94	-1.62	0.00	
基本面指数	x1	-1.08	-2.58	-1.94	-1.62	0.25	I (1)
	Δ x1	-2.31	-2.58	-1.94	-1.62	0.02	
市场行情指数	x2	-0.31	-2.58	-1.94	-1.62	0.57	I (1)
	Δ x2	-1.96	-2.58	-1.94	-1.62	0.05	
行动倾向指数	x3	-0.13	-2.58	-1.94	-1.62	0.64	I (1)
	Δ x3	-4.02	-2.58	-1.94	-1.62	0.00	
上证指数收益率	r	-3.28	-2.58	-1.94	-1.62	0.00	I (0)
成交量变化率	rv	-8.10	-2.58	-1.94	-1.62	0.00	I (0)
换手变化率	rturn	-6.47	-2.58	-1.94	-1.62	0.00	I (0)
基本面指数变化率	rx1	-2.41	-2.58	-1.94	-1.62	0.02	I (0)
市场行情指数变化率	rx2	-2.22	-2.58	-1.94	-1.62	0.03	I (0)
行动倾向指数变化率	rx3	-10.07	-2.58	-1.94	-1.62	0.00	I (0)



搜索指数与换手率关系的实证检验

协整检验：

$$\log(turn_t) = c + \beta_1 \log(x1_t) + \beta_2 \log(x2_t) + \beta_3 \log(x3_t) + \varepsilon$$
$$\log(x_t) = c + \beta_1 \log(turn_t) + \beta_2 \log(v_t) + \varepsilon$$

		因变量换手率	因变量x1	因变量x2	因变量x3
c		6.64***	19.55***	9.98***	11.20***
β_1		0.79***	0.23**	0.12*	0.04
β_2		1.59***	0.24***	0.25***	0.03
β_3		0.36***	——	——	——
修正R ²		0.51	0.06	0.16	0.004
F值		66.16	6.98	18.4	0.62
AIC		1.19	1.53	0.44	0.73
残差平稳性	ADF值	-3.8	-2.39	-2.59	-2.57
	1%临界	-2.61	-2.58	-2.58	-2.57
	5%临界	-1.95	-1.94	-1.94	-1.94
	10%临界	-1.61	-1.62	-1.62	-1.61
	结论	平稳	平稳	平稳	平稳
结论		协整	协整	协整	协整

搜索指数与换手率关系的实证检验

- 协整检验结果

- ▶ 搜索指数与换手率之间存在长期均衡关系
- ▶ 换手率为因变量时的R方**0.51**，明显高于搜索指数为因变量时的R方，说明搜索指数对换手率的解释力远强于换手率对搜索指数的解释力，支持H1，搜索指数比换手率包含了更多的信息量。



搜索指数与上证指数关系的实证检验

协整检验：

$$\log(sp_t) = c + \beta_1 \log(x1_t) + \beta_2 \log(x2_t) + \beta_3 \log(x3_t) + \varepsilon$$
$$\log(sp_t) = c + \beta_1 \log(turn_t) + \beta_2 \log(v_t) + \varepsilon$$

方程		自变量：搜索指数		自变量：换手率、成交量	
		系数	P值	系数	P值
c		-5.36	0.00	11.93	0.00
β_1		0.27	0.00	0.16	0.00
β_2		0.54	0.00	-0.14	0.01
β_3		0.22	0.00	——	——
修正R ²		0.91		0.10	
F值		648.03		10.80	
AIC		-1.95		0.30	
残差平稳性	ADF值	-4.54		-1.95	
	1%临界	-2.58		-2.58	
	5%临界	-1.94		-1.94	
	10%临界	-1.62		-1.62	
	结论	平稳		平稳	
结论		协整		协整	

搜索指数与收益率关系的实证检验

收益率检验：
$$rsp_t = c + \beta_1 rx1_t + \beta_2 rx2_t + \beta_3 rx3_t + \varepsilon$$

$$rsp_t = c + \beta_1 rturn_t + \beta_2 rv2_t + \varepsilon$$

方程	自变量:搜索指数变化率		自变量: 换手与成交变化率	
	系数	P值	系数	P值
c	-0.002	0.10	0.002	0.47
β_1	0.15	0.00	0.03	0.07
β_2	0.46	0.00	0.08	0.00
β_3	0.07	0.00	——	——
修正R ²	0.47		0.22	
F统计量	55.68		26.97	
F检验P值	0.00		0.00	
AIC	-4.81		-4.43	



搜索指数与收益率关系的实证检验

● 协整检验结果

- ▶ 搜索指数与上证指数之间存在长期均衡关系
- ▶ 搜索指数为自变量时的R方**0.91**，明显高于换手率成交量为因变量时的R方；搜索指数变化率为自变量时的R方也明显高于换手成交变化为自变量时的R方
- ▶ 说明搜索指数对上证指数的解释能力明显强于换手率和成交量的解释力，由此证实了H2中的推论。



基于搜索指数的投资者情绪非对称效应研究

● 投资者情绪定义

- ▶ 投资者情绪是指投资者对某种资产未来收益和风险的一种预期和信念，既受到市场环境的影响，也与投资者的教育经历、社会背景、理念经验、个人偏好等方面密切相关
- ▶ 本节将投资者情绪定义为一种综合的预期和评估，它既反映了影响市场的客观信息，也反映了投资者的主观因素。
- ▶ 本节以网络搜索指数作为投资者情绪的综合度量指标

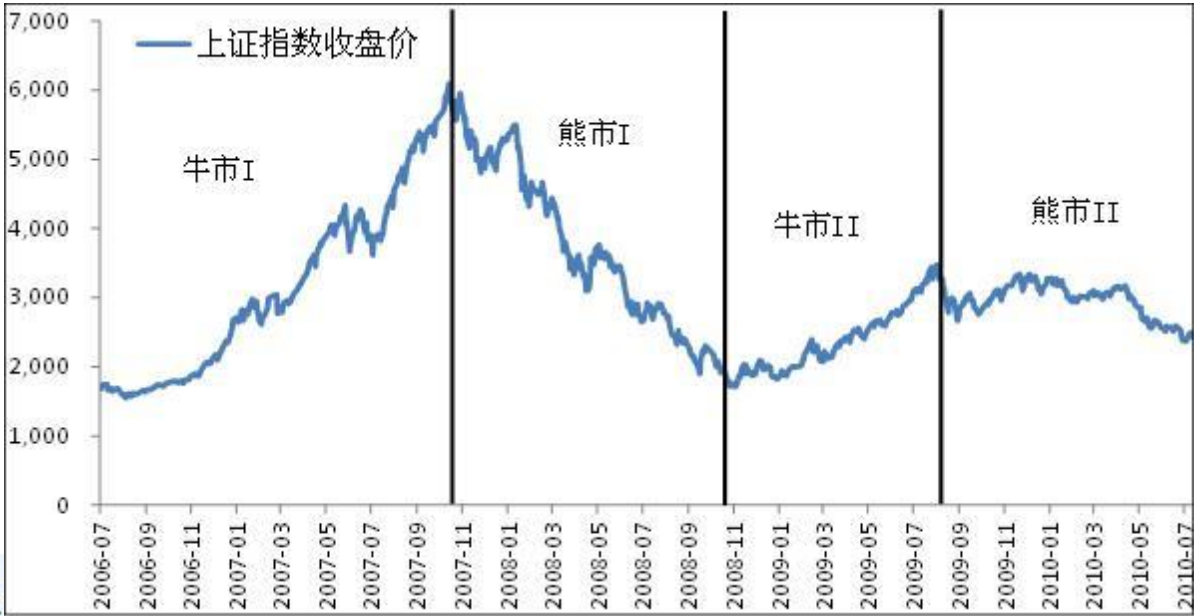
- 以往关于投资者情绪的研究大多没有区分市场形态（牛市、熊市），然而在不同市场状态下，情绪可能表现为不同特征，引发不同投资行为，因此，有必要对这种非对称效应进行研究。



基于搜索指数的投资者情绪非对称效应研究

市场周期划分：借鉴Pagan和Sossounov（2003）的方法

波段区间	所处状态	时间	上证指数收盘价
牛市I	谷值1	2006-07-04	1682
	峰值1	2007-10-16	6092
熊市I	峰值2	2007-10-16	6092
	谷值2	2008-09-18	1896
牛市II	谷值3	2008-09-18	1896
	峰值3	2009-08-04	3471
熊市II	峰值4	2009-08-04	3471
	谷值4	2010-07-15	2424



基于搜索指数的投资者情绪非对称效应研究

VECM（向量误差修正模型）

平稳性检验

变量	ADF检验	MacKinnon临界值			ADF检验 结果
	t-Stat	1%	5%	10%	
$\log(sp)$	0.45	-2.57	-1.94	-1.62	I (1)
$\Delta\log(sp)$	-31.30	-2.57	-1.94	-1.62	I (0)
$\log(x)$	-1.88	-2.57	-1.94	-1.62	I (1)
$\Delta\log(x)$	-12.86	-2.57	-1.94	-1.62	I (0)

股价与投资者情绪Johansen协整检验结果

原假设	特征根	迹统计量(p值)	最大特征值统计量(p值)
0个协整向量	0.0520	15.49 (0.03)	14.26 (0.02)
至少1个协整向量	0.0001	3.84 (0.89)	3.84 (0.89)



基于搜索指数的投资者情绪非对称效应研究

VECM（向量误差修正模型）检验结果：

- 脉冲响应分析：

- ▶ 从投资者情绪对股价影响的持续周期上看，牛市持续时间可以达到**30**天，熊市持续时间为**10**天；牛市影响周期明显长于熊市

- 方差分解分析：

- ▶ 从投资者情绪对上证指数变化的贡献率上看，大幅震荡时期的贡献率高于小振幅时期贡献率，其中，大振幅的熊市中投资者情绪对股价影响的贡献率最大，超过**50%**。



基于搜索指数的投资者情绪非对称效应研究

Egarch模型

- 以往文献中大多用Egarch模型验证“好消息”或者“坏消息”对市场波动影响的非对称性，其原理是用均值方程的扰动项方差代表信息对市场的影响作用
- 设计两步检验法：
 - ▶ 第一步，建立股价波动的Egarch模型
 - 均值方程： $\log(sp_t) = a + b\log(sp_{t-1}) + u_t$
 - 方差方程： $\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \alpha \left| \frac{u_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{u_{t-1}}{\sigma_{t-1}} + \theta \ln(x_{t-1})$
 - ▶ 第二步，在原均值方程加入投资者情绪变量x
 - 均值方程： $\log(sp_t) = a + b\log(sp_{t-1}) + c\log(x_t) + u_t$
 - 扰动项就表示：剔除投资者情绪后，信息对股价波动的影响



基于搜索指数的投资者情绪非对称效应研究

● 两步检验法:

▶ 对比第一步和第二步检验结果的差异，做如下几种情形的推理

- 情形1：若第二步检验结果不存在非对称性，而第一步检验结果存在非对称性，则说明当扰动项剔除投资者情绪因素时不存在非对称效应，包含投资者情绪因素时存在非对称效应，可以证明**投资者情绪是引发股价波动非对称效应的主要因素**。
- 情形2：若两步检验结果都不存在非对称性，说明无论投资者情绪存在与否，股价波动非对称效应都不存在，可以认为**投资者情绪不能引发股价波动的非对称效应**
- 情形3：若第二步存在非对称性，而第一步不存在非对称性，则说明剔除投资者情绪因素后，信息的影响存在非对称效应，而包含投资者情绪因素时非对称效应却变得不显著，可以认为**投资者情绪有平抑股价波动非对称效应的作用**
- 情形4：若两步检验结果都存在非对称性，那么需要对比显著性大小再做进一步判断。



基于搜索指数的投资者情绪非对称效应研究

Egarch模型：两步检验结果

变量系数	第一步方程形式				第二步方程形式			
	牛市I	熊市I	牛市II	熊市II	牛市I	熊市I	牛市II	熊市II
a	-0.010	-0.045	0.038	0.031	0.121	1.647	0.647	0.229
	(0.737)	(0.596)	(0.574)	(0.733)	(0.003)	(0.000)	(0.006)	(0.004)
b	1.002	1.005	0.996	0.996	0.988	0.845	0.942	0.970
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	——	——	——	——	0.006	0.102	0.042	-0.001
	——	——	——	——	(0.000)	(0.000)	(0.004)	(0.925)
ω	-5.528	-1.565	-9.213	-3.879	-0.225	-1.006	-12.226	-1.381
	(0.000)	(0.543)	(0.009)	(0.114)	(0.064)	(0.000)	(0.002)	(0.000)
α	-0.040	0.007	-0.388	-0.459	0.129	-0.559	-0.650	-1.061
	(0.728)	(0.955)	(0.098)	(0.004)	(0.006)	(0.000)	(0.008)	(0.000)
γ	-0.316	-0.049	-0.539	-0.067	0.039	-0.199	-0.481	-0.114
	(0.000)	(0.582)	(0.000)	(0.554)	(0.245)	(0.000)	(0.002)	(0.070)
β	0.188	0.739	0.033	0.394	0.974	0.802	-0.177	0.327
	(0.342)	(0.075)	(0.869)	(0.068)	(0.000)	(0.000)	(0.540)	(0.000)
θ	0.200	0.080	-0.318	0.227	0.018	0.002	-0.639	0.828
	(0.009)	(0.537)	(0.583)	(0.553)	(0.138)	(0.738)	(0.331)	(0.000)



基于搜索指数的投资者情绪非对称效应研究

- Egarch模型的验证结论表明：

- ▶ 我国股票市场价格波动存在显著的集中效应，并且这种集中效应在熊市中比牛市中更为明显。
- ▶ 非对称项系数小于0说明存在信息影响的非对称性，即利空信息比等量利好信息更能引起股价的波动。
- ▶ 进一步分析投资者情绪在股价波动非对称性中的作用发现，牛市中投资者情绪存在增强信息非对称效应的作用，而熊市中投资者情绪则存在平抑非对称效应的作用。

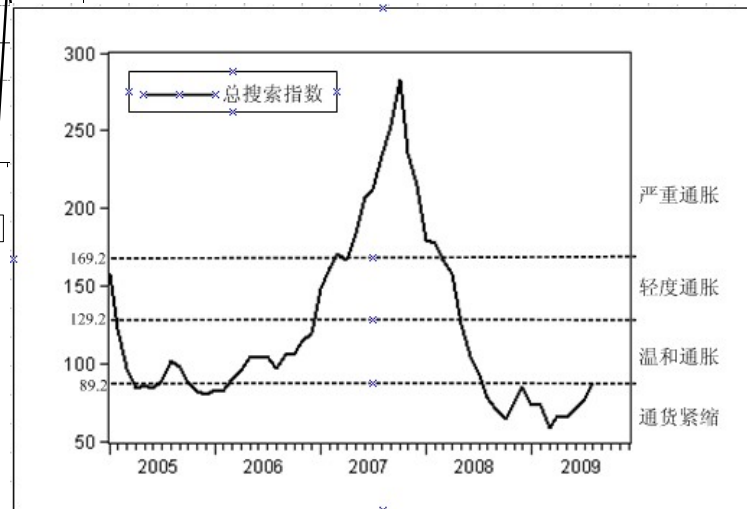
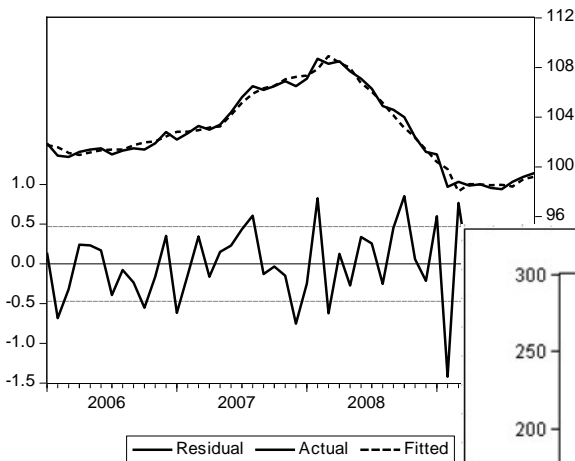


- 涉及的研究领域:
 - 中国股票市场
 - 中国消费者物价指数(CPI)
 - 世博会客流量预测
 - 中国旅游客流量预测
 - 中国汽车行业销量预测
 - 中国房地产价格指数预测

课题组相关研究成果

- 网络搜索数据与CPI的相关性研究，管理科学学报，已录用；
- 搜索数据对CPI预测绝对误差为0.48，宏观形势指数和供求搜索指数相对于CPI的先行周期分别为五个月和两个月；同时模型比国家统计局的数据发布提前一个月左右；与传统的预测方法相比，模型还具备一定的转折点预测能力

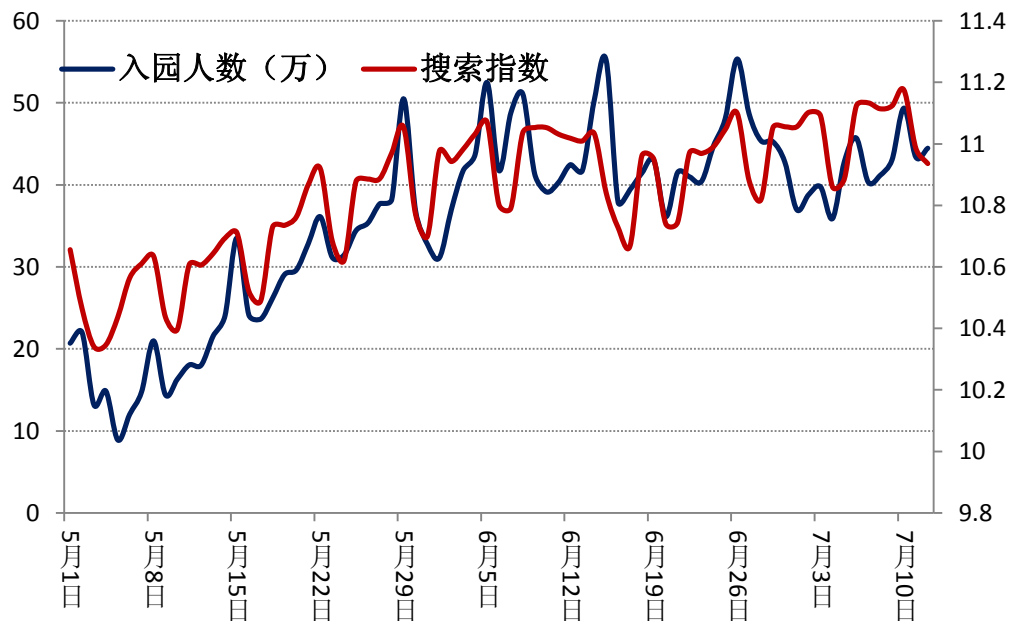
	宏观形势指数(X1)	相关系数	供求关系指数(X2)	相关系数
1	货币政策	0.621	粮食	0.47
2	中行外汇牌价	0.830	花生油价格	0.48
3	存款利率	0.604	按摩价格	0.55
4	准备金	0.670	旅游景点	0.49
5	加息	0.698	携程网	0.84
6	物价指数	0.845	减肥药	0.59
7	投资	0.727	冰箱价格	0.613
8	中国经济	0.583	液晶电视价格	0.432



课题组相关研究成果

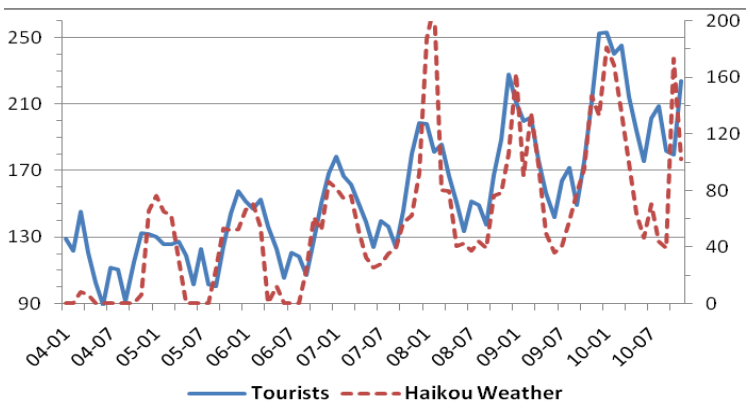
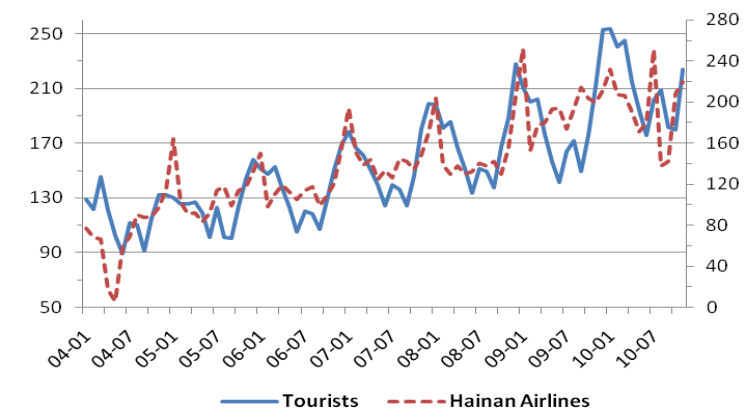
- 基于网络搜索数据的上海世博会日客流量预测方法研究，2011;
- 研究结论: 网络搜索指数对上海世博会日客流量具有领先性，因此，搜索指数可以作为历史数据的良好补充，有效地预测出未来七天的上海世博会日客流量。

关键词	关键词
上海地铁时刻表	上海特产食品
上海世博会攻略	上海世博会 攻略
上海酒店预订	上海世博攻略
上海地铁图	上海市地图查询
上海世博地图	上海世博游
上海住宿	上海天气预报7天
上海地铁票价	上海世博园附近酒店
上海酒店预订网	上海住宿预订

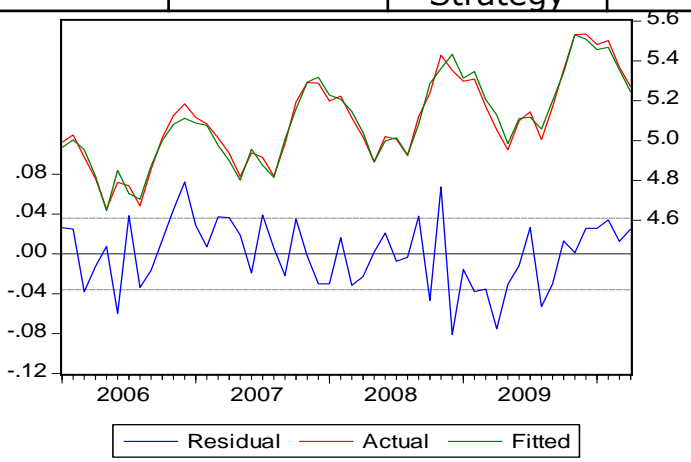


课题组相关研究成果

- A Prediction Study on Tourist Amount Based on Web Search Data—A Case from Hainan, 2011;



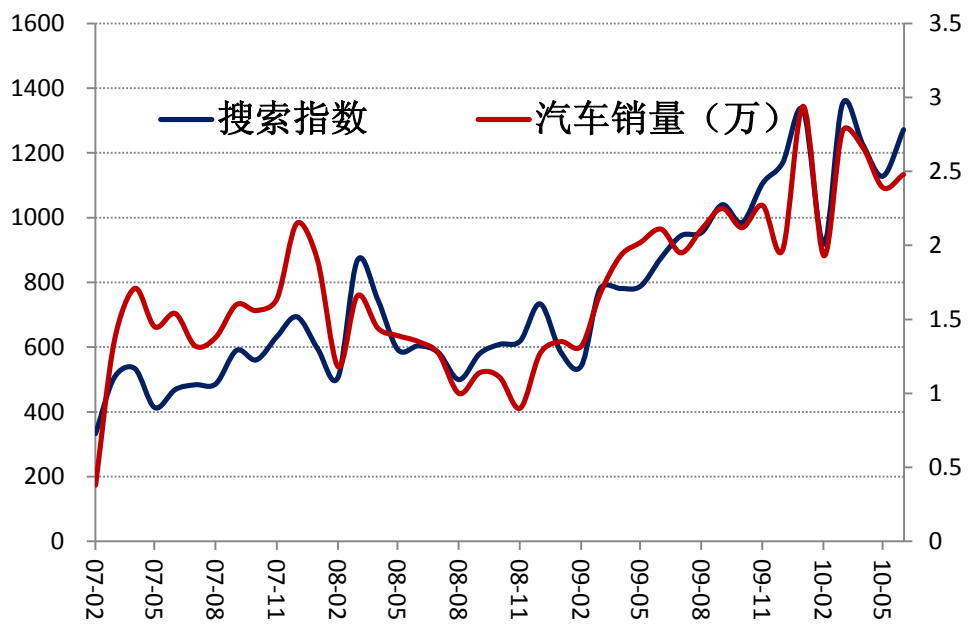
Search terms	Corr. Coef.	Search terms	Corr. Coef.
Air China	0.81	Hainan Sanya Spot	0.80
Eastern Miles	0.80	Hainan Tourist Attractions	0.80
Hainan Airlines	0.80	Sanya Tourist Attractions	0.78
Hainan Sanya Weather	0.77	Hainan Airlines Ticket Queries	0.78
Haikou Weather	0.80	Hainan Tourism Strategy	0.81



课题组相关研究成果

- A Prediction Study on the Car Sales Based on Web Search Data , 2011;

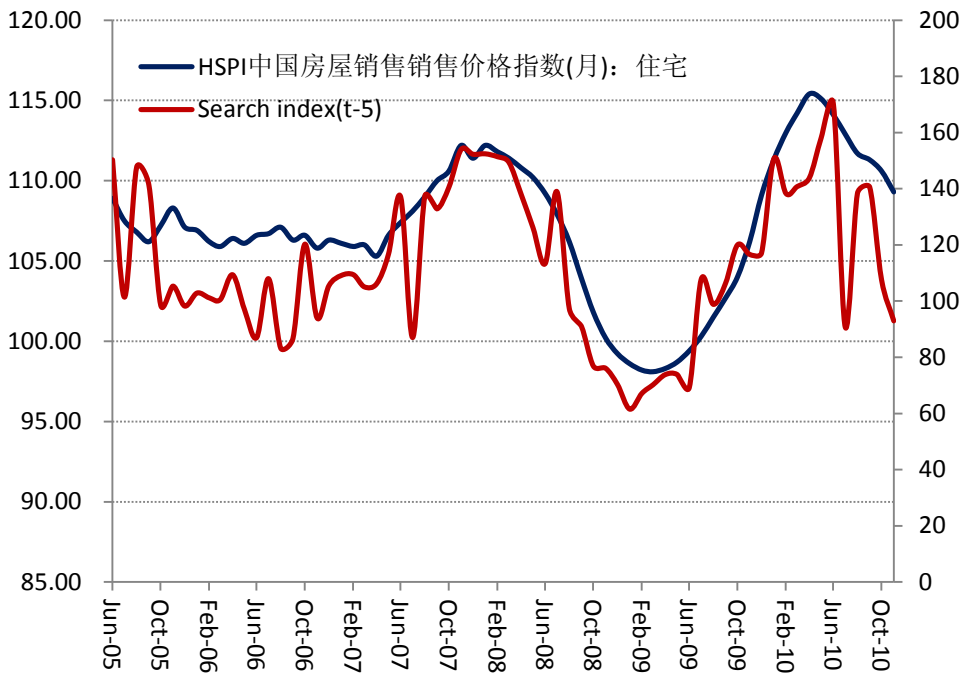
汽车评价	汽车4S店	汽车维修	汽车论坛
汽车配置	汽车优惠	汽车保养	汽车网
汽车报价	汽车试驾	车保险	用车心得
一汽大众汽车	陪练	上牌流程	汽车之家
一汽轿车	汽车降价	汽车改装	搜狐汽车
丰田汽车	驾照	百公里耗油	新浪汽车
.....



课题组相关研究成果

- A Prediction Study of Real Estate Price Index Based on Web Search Data, 2011;

房价	户型装修
出租信息	房产网
二手房房产网	房源
二手房信息	装修材料
开盘	



研究不足与局限性

- 网络搜索数据虽具有普遍性，但并非是投资者获取信息的唯一渠道，因此可能会遗漏部分有价值的信息。网络搜索数据与其他互补数据源相结合，可能会对基准指标有更充分地解释。
- 研究还应该区分投资者类型，根据不同投资者的行为特征建立合适的理论演绎模型，这种一般性规律的探讨在本文中尚存在欠缺。另外，实证上也可以对不同类型投资者行为进行检验。
- 实证研究还可以尝试应用不同的研究方法，与更多的经济调查数据相对比，才能进一步增强研究结论的可靠性。



研究展望

- 尝试将网络搜索数据与互联网其他指标相结合（如网络浏览、网络评论等）、以及互联网指标与传统市场指标相结合，并采用不同的研究方法进行验证，提高研究结论的可靠性。
- 通过博弈论、经济模拟等研究方法，抽象出股票市场中不同类型投资者的行为特征及策略，并构建理论演绎模型求解微观搜索行为与宏观市场现象的一般性规律。
- 网络搜索行为研究的边界与适应性问题，即哪些社会经济行为适合应用网络搜索数据来刻画？研究这一问题有助于我们更深刻地理解网络搜索及互联网数据的内涵，从而能够更好地利用互联网数据资源来监测社会经济中的典型事件、研判某些典型的社会经济行为的发展态势。



主要参考文献

1. Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine Query data [J]. Nature, 2009, 457, 1012-1014.
2. N Askitas and KF Zimmermann. Google Econometrics and Unemployment Forecasting [J]. Applied Economics Quarterly, 2009, 55(2), 107-120.
3. Choi, H. and H. Varian. Predicting initial claims for unemployment benefits [R]. Technical report, 2009a, Google Inc.
4. Choi, H. and H. Varian. Predicting the present with Google trends [R]. Technical report, 2009b, Google Inc.
5. Carneiro, HA, and E Mylonakis. Google Trends: A Web based tool for real time surveillance of disease outbreaks [J]. Clinical Infectious Diseases, 2009, 49:1557-1564.
6. D'Amuri, F. Predicting unemployment in short samples with internet job search Query data. MPRA Paper 18403, University Library of Munich, Germany, 2009.
7. D'Amuri, F. and J. Marcucci. "Google it!" Forecasting the US unemployment rate with a Google job search index. MPRA paper, University Library of Munich, Germany, 2009.
8. Doornik, J. A. Improving the timeliness of data on influenza-like illnesses using Google search data [R]. Technical report, University of Oxford, 2009.
9. Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. What can search predict [R]. Technical report, Yahoo Research, 2010.



感谢各位老师！
请各位老师提出宝贵意见！

