



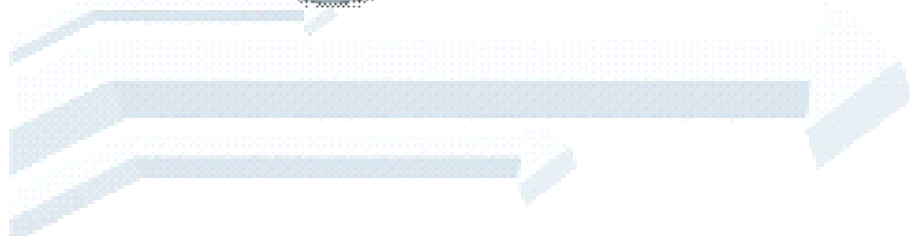


# 基于数据挖掘的 量化投资初探

# Contents

-  1 量化投资和对冲基金简介
-  2 基于SOM网络的股票聚类分析方法
-  3 基于关联规则的板块轮动
-  4 量化对冲的发展前景



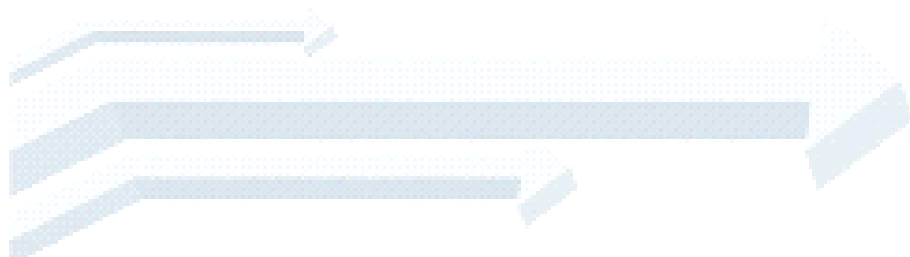
# 当前趋势——金融市场

20世纪90年代以来，数学、金融、计算机及全球经济呈现融合趋势。

——摘自Goodman和Stampfli的《金融数学》

美国最聪明的人都在华尔街做对冲基金。

——摘自丁鹏的《量化投资：策略与技术》



# 当前趋势——大数据

2012年10月的《哈佛商业评论》和2012年秋季的《MIT斯隆管理评论》都以“大数据”作为主题，畅想大数据时代的企业管理。

众多学者和企业家得出的结论是：在未来，数据将会像土地、石油和资本一样，成为经济运行中的根本性资源。数据科学家被认为是下一个十年最热门的职业。

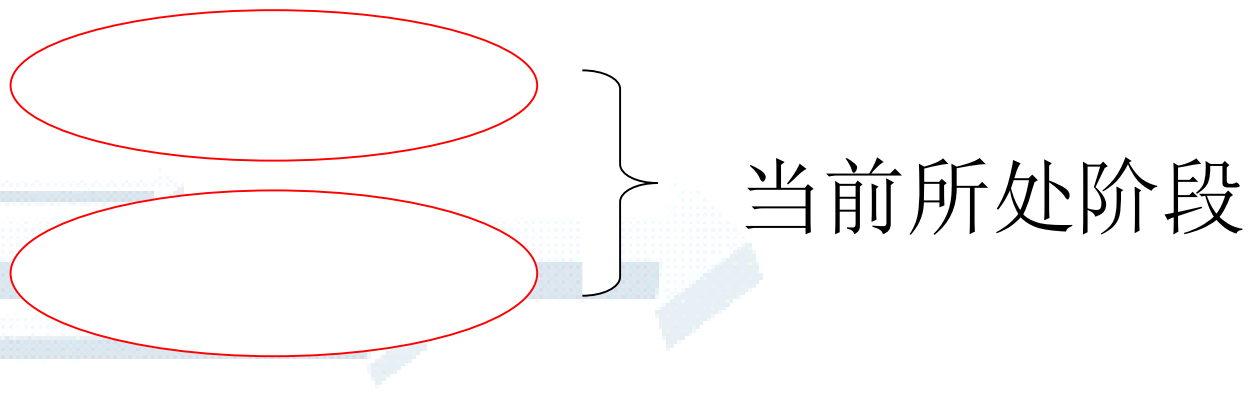
# 资本市场投资的五个阶段

1.投机者

2.职业经理人

3.学院派、科学化

——标志：Markowitz投资组合理论



# 金融市场传奇人物



詹姆斯·西蒙斯

连续**20**年，每年赚**60%**  
（为投资者获得超过**35%**  
的收益率），从来没有出现  
过亏损。

沃伦·巴菲特

连续**40**年，每年可以获得  
**20%**的复合稳定收益。

## 大奖章基金

大奖章基金历经数次金融危机。它在**1998**年俄罗斯债务危机，本世纪初的互联网泡沫和**2008**年的次贷危机期间，均获得了高额收益。

# 量化投资视频

宁夏卫视 量化投资再成机构投资模式风潮

<http://video.sina.com.cn/v/b/39218964-1154814654.html>

CCTV 证券资讯频道 解读量化投资 丁鹏

[http://v.youku.com/v\\_show/id\\_XMzMzODkzMjQw.html](http://v.youku.com/v_show/id_XMzMzODkzMjQw.html)



# 三大主流投资分析方法

1. 基本面分析

2. 技术面分析

3. 量化投资

传统方法



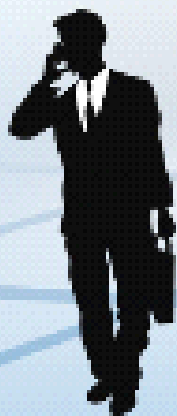
特点：  
定量化、  
精确化



# 什么是量化投资？

简单来讲，量化投资就是利用计算机科技并采用一定的数学模型去实现投资理念、实现投资策略的过程。

传统的投资方法主要有基本面分析法和和技术分析法这两种，与它们不同的是，量化投资主要依靠数据和模型来寻找投资标的和投资策略。



# 对冲基金

对冲基金的英文名称为Hedge Fund, 意为“风险对冲过的基金”, 起源于50年代初的美国。

其操作的宗旨, 在于利用期货、期权等金融衍生产品以及对相关联的不同股票进行买入卖空、风险对冲的操作技巧, 在一定程度上规避和化解投资风险。

# 对冲基金

在一类对冲操作中,基金管理人首先选定某类行情看涨的行业,买进该行业几只优质股,同时以一定比率卖出该行业中几只劣质股。

如此组合的结果是,如该行业预期表现良好,优质股涨幅必超过其他同行业的股票,买入优质股的收益将大于卖空劣质股的损失;如果预期错误,此行业股票不涨反跌,那么较差公司的股票跌幅必大于优质股,则卖空盘口所获利润必高于买入优质股下跌造成的损失。

# 量化对冲的策略类型

## 相对价值套利

- ✓ 股票市场(**beta**)中性策略
  - Alpha对冲
  - 成对交易 (Pairs Trading)
  - 新股长期弱势表现
- ✓ 可转化债券套利
- ✓ 统计套利：股指与ETF期现套利、跨期套利

## 事件驱动策略

- ✓ 兼并重组等特殊境况套利
  - 买入被兼并公司同时卖空实施兼并公司的股票
- ✓ 困境投资
  - 买入ST公司股票同时卖空同行业可比公司股票

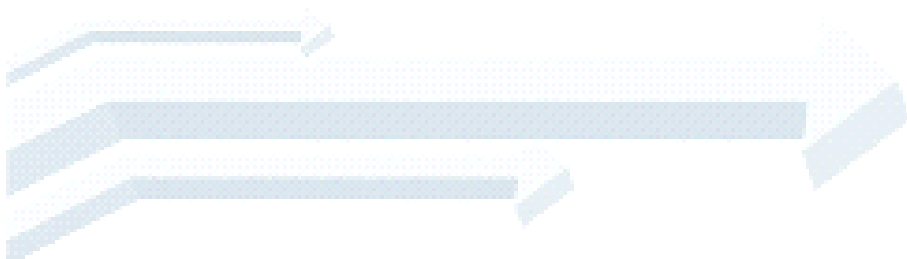
## 方向性策略

- ✓ 股票多空头 (**Equity Long/Short**)
  - 融资融券、杠杆交易、期权、选股能力
- ✓ 管理期货 (**Managed Futures**)
  - CTAs、趋势、全球分散化、程序化交易、模型同质
- ✓ 全球宏观对冲：索罗斯、量子基金

# 数据挖掘在量化投资中的应用

1. 基于SOM网络的股票聚类分析方法

2. 基于关联规则的板块轮动

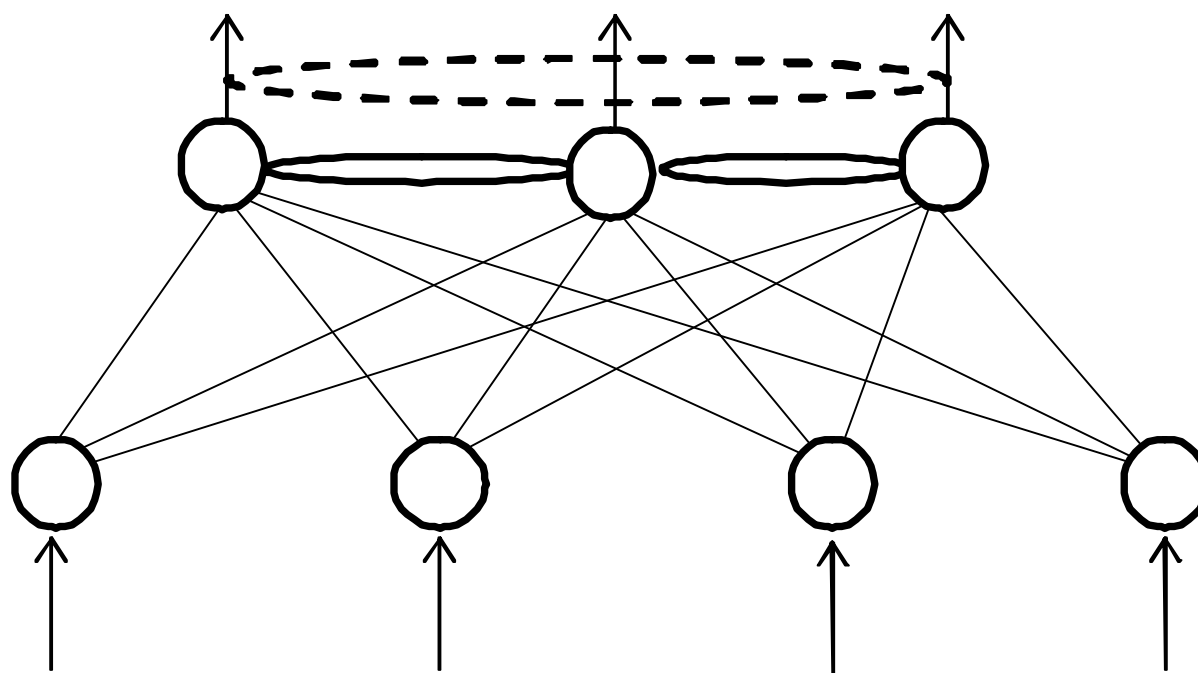


# 自组织特征映射神经网络（SOM）

1981年芬兰Helsinki大学的T. Kohonen教授提出一种自组织特征映射网（Self-Organizing Map），简称SOM网，又称Kohonen网。

Kohonen认为：一个神经网络接受外界输入模式时，将会分为不同的对应区域，各区域对输入模式具有不同的响应特征，而且这个过程是自动完成的。自组织特征映射正是根据这一看法提出来的，其特点与人脑的自组织特性相类似。

# 自组织特征映射神经网络 (SOM)



**SOM**神经网络的拓扑结构由两层节点构成：输入层和输出层（竞争层）。

# SOM网的生物学基础

生物学研究的事实表明，在人脑的感觉通道上，神经元的组织原理是有序排列。因此当人脑通过感官接受外界的特定时空信息时，大脑皮层的特定区域兴奋，而且类似的外界信息在对应区域是连续映象的。

对于某一图形或某一频率的特定兴奋过程，神经元的有序排列以及对外界信息的连续映象是自组织特征映射网中竞争机制的生物学基础。



# 基于SOM网络的股票聚类分析方法

## ①数据处理

为了更好地说明问题而不受个别变量单位的影响，在聚类分析计算之前需要对原始数据进行标准化处理。

$$x_{ij} = (x_{ij} - \bar{x}_j) / S_j$$

# 基于SOM网络的股票聚类分析方法

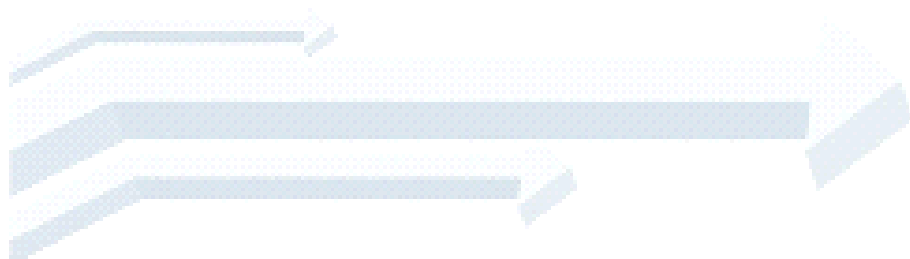
$$\bar{x}_j = 1/n \sum x_{ij} \quad S_j = \sqrt{1/n \sum (x_{ij} - \bar{x}_j)^2}$$

其中， $i$ 为股票上市公司的数量， $j$ 为上市公司的综合盈利能力指标； $x_{ij}$ 为标准化后的数据， $\bar{x}_j$ 为某指标的均值， $S_j$ 为某指标的均方差。

# 基于SOM网络的股票聚类分析方法

## ②实验结果

本案例所选取的数据来自于2005年80家上市公司的年报信息，分析上市公司的每股收益、每股净资产、净资产收益率、每股经营性现金流量及净利润5项反应上市公司综合盈利能力的指标作为主要研究对象。



# 基于SOM网络的股票聚类分析方法

邻域函数:  $\alpha(t) = \tau + \alpha(t-1) * (1 - 0.001 * t)$

学习速率因子函数:  $\eta(t) = \eta(0) * \exp(-t / GAP)$

在本案例中, 选取的参数:

$\eta(0) = 0.1$        $\alpha(0) = 5$        $\tau = 0.5$

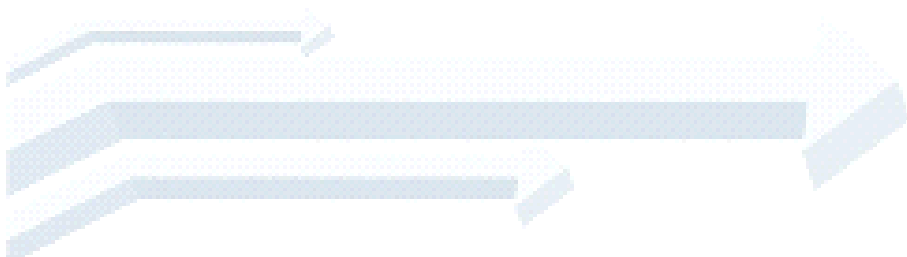
迭代次数GAP=14

# 基于SOM网络的股票聚类分析方法

类别	每股收益	每股净资产	净资产收益率	每股经营性现金流量	净利润
9	0.22	2.49	7.64	0.47	8374.7
8	0.012	1.9	1.16	-0.08	1591.3
7	-0.16	3.11	-7.99	0.45	-3993
6	-0.32	2.16	-31.4	-1.14	-7149
5	0.43	1.89	65.7	0.087	8932
4	0.63	0.46	159.8	0.019	12155
3	-0.65	2.81	-26.9	1.48	-24552
2	-0.001	1.09	1.09	-0.001	109
1	-1.5	1.09	-193.6	-0.087	-63583

# 基于SOM网络的股票聚类分析方法

上述实验结果表明，利用SOM网络进行股票分析所得结果基本与公司的实际情况相符，为股票的分析 and 选择提供了较好的依据。



# 基于关联规则的板块轮动

股票市场交易规律显示：在股价波动的过程中，整个市场并不是经常性地普涨普跌，而是呈现出板块轮动、涨跌不一的状况。

其实金融风险并不可怕，如果能够发现股票板块的运动规律及其相互的联动关系，那么金融风险在一定程度上可以通过风险流动得以规避。

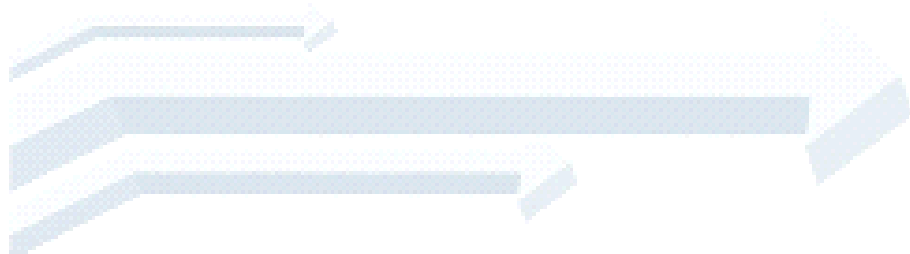
# 基于关联规则的板块轮动

本案例使用的数据，来自2007年1月4日至2008年3月31日A股市场的21个板块指数的日收盘值，取得295条日收盘数据记录共6195个观察值进行分析。



# 基于关联规则的板块轮动

这21个板块指数分别为采掘指数、地产指数、食品指数、金属指数、批零指数、石化指数、运输指数、制造指数、机械设备指数、服务指数、金融指数、水电指数、医药指数、建筑指数、造纸指数、IT指数、农林牧渔指数、纺织指数、传播指数、电子指数及木材指数。



# 基于关联规则的板块轮动

在进行关联规则挖掘之前，必须对所收集的数据进行预处理，转化为二进制数据。

日期	采掘指数	地产指数	食品指数	金属指数
2007-1-4	0	0	0	0
2007-1-5	1	1	1	1
2007-1-8	1	1	1	1

“1”代表上涨，“0”代表下跌。

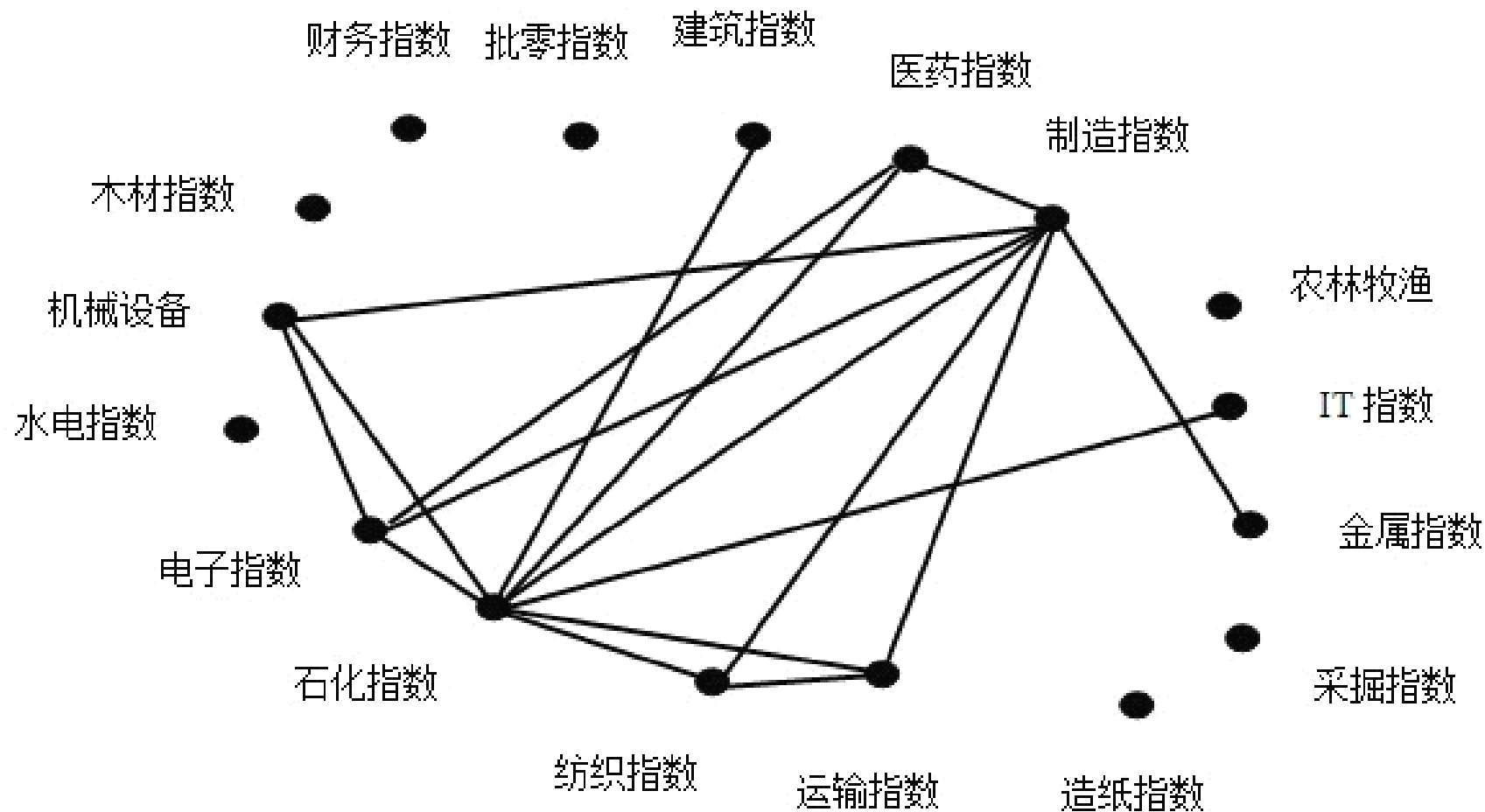
# 基于关联规则的板块轮动

数据预处理后，根据股票板块指数“涨”、“跌”情况，采用二值型关联规则算法进行挖掘，这里采用Apriori算法来实施关联分析。

设定置信度 $1-\alpha$ 的区间，可以从模型中找到许多有意义的强关联规则。

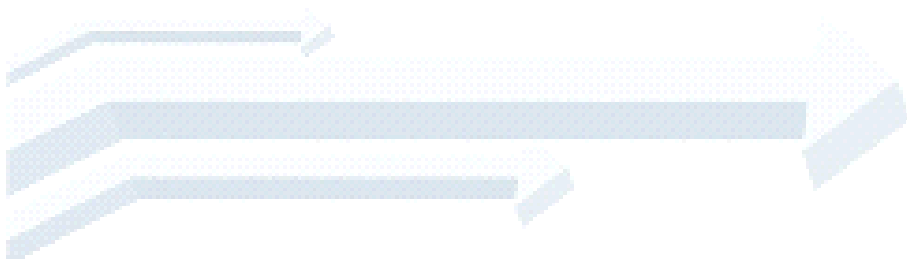
例如，当金属、机械设备、石化3种指数上涨时，100%的情况下，制造指数也在上涨。

# 基于关联规则的板块轮动



# 基于关联规则的板块轮动——结论

1. 制造、石化指数作为关联规则的核心，与其他板块指数（除金融、地产、食品与传播4种指数之外）之间有着密切的关系。
2. 模型结果显示，金融、地产、食品与传播4种指数与其他板块指数不存在关联规则。



# 量化对冲的发展前景

- 过去10年间，全球对冲基金资产规模年复合增长率超过15%，对冲基金数量年复合增长率近10%。
- 根据对冲基金研究公司(Hedge Fund Research, Inc., HFR)报告，截止2011年底，全球对冲基金数量达9,523家，管理的总资产规模达2.02万亿美元。

自2008年金融危机以来，全球对冲基金的年收益率为4.1%，波动率为8.5%；而全球股市只有0.4%的年收益率，但波动率却高达20.4%。

# 量化对冲的发展前景

- 2011年底，国内阳光私募基金中量化对冲类产品不超过100亿元，券商集合理财的量化对冲产品规模不足50亿元。
- 对冲工具缺乏、融资融券渠道限制、政策监管严格、准入门槛过高等市场现状是阻碍国内对冲基金发展的主要因素。



# 量化对冲的发展前景

- 2012年5月券商创新大会之后，证监会推出一系列新政，券商、基金资产管理业务的监管政策大松绑，期货公司资管业务也将开放，保监会也陆续推出13条新规以促进保险资管机构的转型。
- 各种金融衍生工具的推出、市场准入门槛的降低、投资范围的扩大、投资渠道的多样化等等，都将为量化对冲的蓬勃发展提供历史性机遇。



# 对冲基金视频

金价下滑重创对冲基金大鳄鲍尔森

<http://v.ku6.com/show/3Su27iFdJuGH9Jz0.html>

何为对冲基金

[http://v.youku.com/v\\_show/id\\_XMzI30DMwMDA4.html](http://v.youku.com/v_show/id_XMzI30DMwMDA4.html)

华尔街对冲基金大鳄内线做多AMD股票

<http://v.ifeng.com/f/200911/ec888b9f-e3a6-4a4b-a680-5dc690084400.shtml>



谢谢！

