

如何用R语言开始量化投资

张丹



前言

用R语言开始量化投资

2

R语言在金融领域有着广阔的应用场景，是重要的量化投资工具。

金融本身是玩数据行业，R的最大的优势就是数据分析，把R语言与金融结合，用R来做量化投资的策略，真的很配，不仅顺手而且方便，用了你就会知道。

目录

用R语言开始量化投资

3

1. 为什么用R语言？
2. 跨界结合的思维模式
3. R语言量化工具包
4. 量化策略实战应用
5. 有理想的极客

为什么用R语言？

用R语言开始量化投资

4

编程语言之争？

Python, Java, PHP, Nodejs, **R**, Matlab, SAS

为什么用R语言？

用R语言开始量化投资

R语言是一门面向数据的编程语言，早期只是在统计领域，被统计学家所使用。

随着大数据技术的发展，R语言的应用领域，包括了互联网，数据科学，人工智能，机器学习，生物医疗，游戏，电子商务，全球地理科学，数据可视化等等。

[Bayesian](#)

[ChemPhys](#)

[ClinicalTrials](#)

[Cluster](#)

[DifferentialEquations](#)

[Distributions](#)

[Econometrics](#)

[Environmetrics](#)

[ExperimentalDesign](#)

[ExtremeValue](#)

[Finance](#)

Bayesian Inference

Chemometrics and Computational

Clinical Trial Design, Monitoring, ar

Cluster Analysis & Finite Mixture M

Differential Equations

Probability Distributions

Econometrics

Analysis of Ecological and Environn

Design of Experiments (DoE) & Ana

Extreme Value Analysis

Empirical Finance

[FunctionalData](#)

[Genetics](#)

[Graphics](#)

[HighPerformanceComputing](#)

[MachineLearning](#)

[MedicalImaging](#)

[MetaAnalysis](#)

[Multivariate](#)

[NaturalLanguageProcessing](#)

[NumericalMathematics](#)

[OfficialStatistics](#)

[Optimization](#)

[Pharmacokinetics](#)

[Phylogenetics](#)

[Psychometrics](#)

Functional Data Analysis

Statistical Genetics

Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization

High-Performance and Parallel Computing with R

Machine Learning & Statistical Learning

Medical Image Analysis

Meta-Analysis

Multivariate Statistics

Natural Language Processing

Numerical Mathematics

Official Statistics & Survey Methodology

Optimization and Mathematical Programming

Analysis of Pharmacokinetic Data

Phylogenetics, Especially Comparative Methods

Psychometric Models and Methods

为什么用R语言？

用R语言开始量化投资

我们每天都会跟钱打交道，会有消费的需求，月底会有工资收入，当收入大于消费，资金积累到一定阶段的时候，我们就会有投资的需求。

中国人的投资一般就几种：买房，买黄金，买股票，买保险，买理财。

投资时，大部分人是不理性的，心血来潮就买了，也不仔细观察行情，买完就被套住。像“中国大妈抢购黄金”，“中国大妈是救市的主力军”都一度被新闻媒体热炒。

为什么用R语言？

用R语言开始量化投资

7

在金融市场中最不缺的就是机会。

不管投资什么，如果能在交易的时候能多思考一下，说不定结果就会不一样。运用我们数据分析的知识，和对于金融市场的理解，通过R语言进行数据分析，抓住机会，让我们把辛辛苦苦赚到的钱，保值和增值。

目录

用R语言开始量化投资

1. 为什么用R语言？
2. 跨界结合的思维模式
3. R语言量化工具包
4. 量化策略实战应用
5. 有理想的极客

跨界结合的思维模式

用R语言开始量化投资

9

R语言与其他编程语言最大不同在于，R语言是面向数据的，这一点很重要，甚至影响了思维模式。

我写了10多年的JAVA程序，程序员思路在我的脑子里根深地固。

- 用面向对象思想来建模。
- 从架构设计到开发的，自上而下的设计思路。
- 需求驱动的敏捷开发，为业务随需而变的开发路径。

跨界结合的思维模式

用R语言开始量化投资

10

跳出原有的思维定式，通过多维的视角看世界、看自己。

通过R语言，帮我打开了通过数据看世界的窗口。能力提升不再是，如何把程序把写得得多得漂亮，在时间复杂度上用 $O(\log(N))$ 解决 $O(N^2)$ 的问题，而变成了怎么收集数据，怎么提高数据质量，用于解决实际的业务问题。

问题被转化了！

跨界结合的思维模式

用R语言开始量化投资

我创业时，经常被投资人问到，做量化投资的项目有多难？招几个金融背景的，再招几个IT背景的，捣鼓个项目，很快就做出来了。

其实不然，不同背景的人是很难交流和沟通，尤其是金融和IT。

- 一种是封闭不愿意分享，一种是开放毫无保留；
- 一种是先把话说出来，一种是先把事情做起来；
- 一种是所有人都要为我服务，一种是只专注于自己的兴趣。

所以，如果没有能够跨界通吃的人，是做不好量化的项目的。

跨界结合的思维模式

用R语言开始量化投资

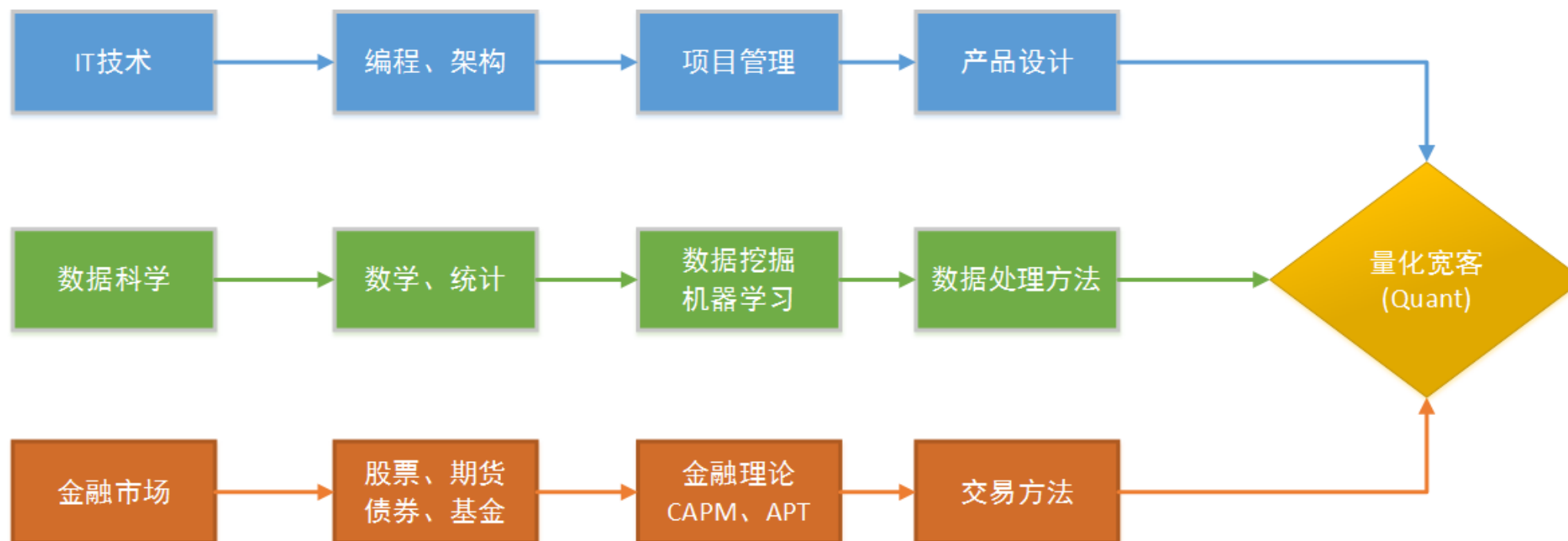
12

我们就要立志做个跨着学科的人，这才是最核心的点，是不可被替代的。

- **IT技术：**
你需要扎实的编程功底架构思想，科学的项目管理方法，严谨的产品设计逻辑；
- **基础学科：**
统计、数学等基础学科的知识，数据挖掘与机器学习的数据处理经验；
- **业务理解：**
你要能结合业务，不管是量化投资或其他的领域，理解市场的规则。

跨界结合的思维模式

用R语言开始量化投资



目录

用R语言开始量化投资

14

1. 为什么用R语言？
2. 跨界结合的思维模式
3. R语言量化工具包
4. 量化策略实战应用
5. 有理想的极客

R语言量化工具包

用R语言开始量化投资

15

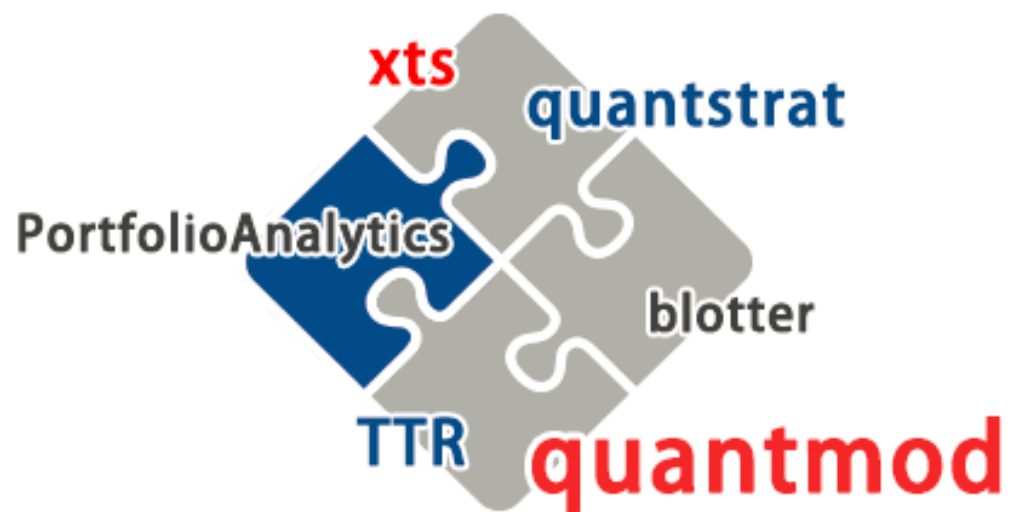
做量化投资，R语言的**优势**在哪？

- 编程难度小，20-30行代码，能做很多事情。
- 面向数据，有方便数据处理操作。
- 数学、统计等基础学科有丰富的函数支持
- 多种数据挖掘和机器学习的算法库。

R语言量化工具包

用R语言开始量化投资

R包，投资研究包，回测分析包，金融产品定价包，投资组合优化包，风险管理包，涵盖了量化投资的方方面面。



R语言量化工具包

用R语言开始量化投资

	第三方独立R包	Rmetrics体系R包
数据管理	zoo, xts, RQuantLib, WindR RJDBC, rhadoop, SparkR, rhive, rredis, rmongodb	timeSeries, timeDate, fimport
指标计算	TTR, TSA, urca,	fArma, fAsianOptions, fBasics, fCopulae, fExoticOptions, fGarch, fNonlinear, fOptions, fRegression, fTrading, fUnitRoots
回测交易	FinancialInstrument, quantstrat, blotter	fTrading
投资组合	PortfolioAnalytics, stockPortfolio	fAssets
风险管理	PerformanceAnalytics	fPortfolio, fExtremes

目录

用R语言开始量化投资

1. 为什么用R语言？
2. 跨界结合的思维模式
3. R语言量化工具包
4. 量化策略实战应用
5. 有理想的极客

量化策略实战应用

用R语言开始量化投资

用R语言，做一个简单的交易策略，基于MACD，针对全球指数的投资方案。



量化策略实战应用

用R语言开始量化投资

20

下面是一个简化的量化策略的研究过程，分为如下的6步进行。

1. 用quantmod包下载数据。
2. 用zoo包和xts包进行数据格式标准化。
3. 用TTR包进行模型计算。
4. 用PerformanceAnalytics包进行指标风险。
5. 用ggplot2包进行可视化输出。
6. 最后，进行结果分析。

量化策略实战应用

用R语言开始量化投资

通过quantmod包，从Yahoo财经下载数据的。
我选择了全球的5个市场的指数进行比较，代码和名称对应关系。

指数名称	指数代码(Yahoo财经代码)	简称
标普500	^GSPC	GSPC
日经225	^N225	N225
恒生指数	^HSI	HSI
富时新加坡STI	^STI	STI
上证综指	000001.SS	SSE

量化策略实战应用

用R语言开始量化投资

下面通过R语言代码，我们来完成这个交易策略模型的构建。

```
# 加载程序库
> library(quantmod)
> library(TTR)
> library(PerformanceAnalytics)
> library(ggplot2)
> library(scales)

# 从Yahoo财经下载各全球指数数据
> options(stringsAsFactors = FALSE)
> symbols<-c("^GSPC", "^N225", "^HSI", "^STI", "000001.SS")
> suppressWarnings(getSymbols(symbols,src = "yahoo",from="2012-01-01"))
[1] "GSPC"      "N225"      "HSI"       "STI"       "000001.SS"

# 取指数价格调整后的数据，合并数据集
> df<-merge(GSPC$GSPC.Adjusted,HSI$HSI.Adjusted,N225$N225.Adjusted,STI$S
[1] "GSPC"      "N225"      "HSI"       "STI"       "000001.SS"

# 对列重命名
> names(df)<-c("GSPC", "HSI", "N225", "STI", "SSE")
```

```
# 查看数据前6行
```

```
> head(df)
```

	GSPC	HSI	N225	STI	SSE
2012-01-03	1277.06	18877.41	NA	2688.36	NA
2012-01-04	1277.30	18727.31	8560.11	2711.02	2169.39
2012-01-05	1281.06	18813.41	8488.71	2713.02	2148.45
2012-01-06	1277.81	18593.06	8390.35	2715.59	2163.40
2012-01-09	1280.70	18865.72	NA	2691.28	2225.89
2012-01-10	1292.08	19004.28	8422.26	2719.83	2285.74

```
# 查看数据最后6行
```

```
> tail(df)
```

	GSPC	HSI	N225	STI	SSE
2017-02-24	2367.34	23965.70	19283.54	3117.03	3253.43
2017-02-27	2369.73	23925.05	19107.47	3108.62	3228.66
2017-02-28	2363.64	23740.73	19118.99	3096.61	3241.73
2017-03-01	2395.96	23776.49	19393.54	3122.77	3246.93
2017-03-02	2381.92	23728.07	19564.80	3136.48	3230.03
2017-03-03	2383.12	23552.72	19469.17	3122.34	3218.31

```
# 查看数据类型，为xts
```

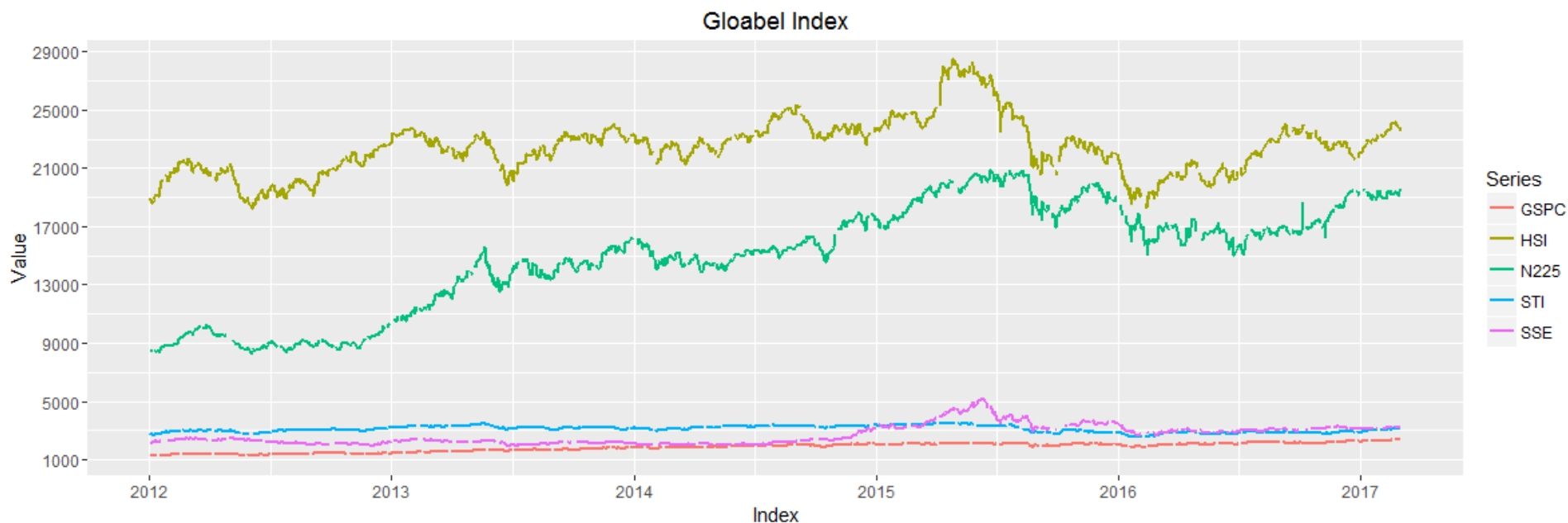
```
> class(df)
```

```
[1] "xts" "zoo"
```

量化策略实战应用

用R语言开始量化投资

整个数据集是从2012年01月03日到2017年03月03日之间数据，数据为xts类型，xts类型是R语言中专用的时间序列类型。



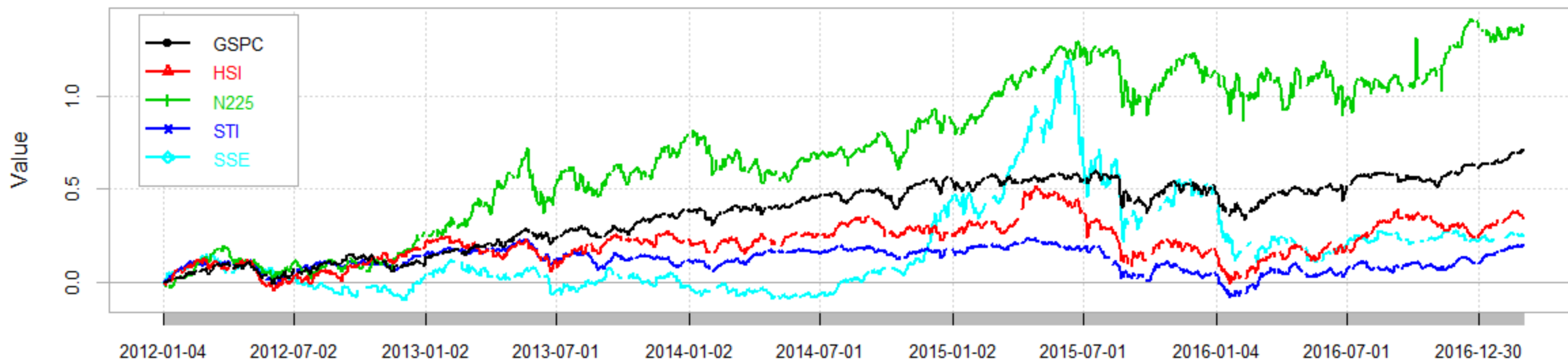
量化策略实战应用

24

用R语言开始量化投资

用收益率进行度量，这样所有的标的都是在同一个维度了。

Cumulative Daily Returns for Gloabel Index



量化策略实战应用

用R语言开始量化投资

25

计算指数的平均年化收益率，如果我们把钱一直投资在这些看指数上面，那么我们每年的平均回报是多少呢？

```
> Return.annualized(ret_df)
              GSPC      HSI      N225      STI      SSE
Annualized Return 0.1133813 0.0619811 0.1927681 0.03696703 0.04817027
```

这里日经指数(N225)年化回报率是最高的，有每年平均有19.28%的年化收益率，而中国的上证综指(SSE)仅有4.82%的年化收益率，回报不算高，但波动太大了。综合来看，标普500(GSPC)其实是投资的首选，11.34%年化收益率，平均波动率也不是太高，所以风险和收益都是相对不错的。

量化策略实战应用

用R语言开始量化投资

我们写了一个MACD的策略函数，就相当于建模的过程，函数的输出为策略的收益率。然后，我们把指数数据传给MACD()函数，经过计算输出策略收益率。把策略收益率与纯指数率，放到一起来进行对比。

策略	GSPC	HSI	N225	STI	SSE
指数收益	0.1133813	0.0619811	0.1927681	0.03696703	0.04817027
策略收益	0.2137435	0.2406476	0.2261996	0.01869112	0.2817241

量化策略实战应用

用R语言开始量化投资

27

要搭建一套量化交易的系统，主要就是完成这些部分。

对于个人玩家来说，按照上面R语言代码的思路，就已经可以实现了。



目录

用R语言开始量化投资

1. 为什么用R语言？
2. 跨界结合的思维模式
3. R语言量化工具包
4. 量化策略实战应用
5. 有理想的极客

有理想的极客

用R语言开始量化投资

- 本文是新书 **《R的极客理想-量化投资篇》** 的开篇内容，希望大家跟着我的思路，通过R语言入手，体会R语言建模的乐趣，并真正地把技术变现。我认为R语言是非常好的一门语言，从量化投资的角度，我会一直推动R语言在中国的发展，当成一项事业来做。
- 同时，本书也会译成**英文版**在全球发行，让中国的技术走出去。加油！！



作者介绍

《R的极客理想》系列图书

张丹，《R的极客理想》系列图书作者，民生银行金融大数据分析师，前况客创始人兼CTO。

12年IT编程背景，4年量化投资经验，金融大数据专家。精通R, Java, Nodejs 编程语言，曾获得10项SUN及IBM技术认证。丰富的互联网应用开发架构经验，掌握大数据处理、数据挖掘机器学习等核心技术，熟悉金融二级市场、交易规则和投研体系。

博客<http://fens.me>，Alexa全球排名70k。



Thank you!

本文地址: <http://blog.fens.me/r-quant-start/>

张丹

《R的极客理想》系列图书作者