

如何用R语言做量化分析

张丹，《R的极客理想》系列图书作者

前言

为什么用R语言做量化投资？

R、Python、Matlab比起来哪个更好？

目录

1. R语言适合做量化分析的5个理由
2. R语言的数据处理方法
3. 量化分析实例

个人经历

我是一个程序员，从2004年开始接触Java写了10多年的Java程序，期间还尝试过多种编程语言 PHP、Python、SAS、R、Nodejs，最后把自己确定在R, Nodejs 和 Java。

谈不上对每一种语言都有很深的理解，但是每种语言的特点还是有点心得。

从技术发展来看，编程开发变得越来越简单，10年前用JavaEE做一个简单的web项目至少要2人月，现在新人用Nodejs边学边搞只要10人天就能搞定。

1. R语言做量化的5个理由

1. R语言是统计语言，对数学计算和统计计算有良好的支持。
2. R语言是面向数据的，金融玩的就是数据。
3. R语言有完善第三方包体系，提供很多的量化工具包支持。
4. R语言本身很简单，易学易上手。
5. R语言有清晰的体系结构。

1.1 R语言是统计语言

R语言是统计语言，对数学和统计有良好的支持。

能干什么呢？

- 把论文中，高等**数学公式**变成算法模型。
- 检验数据是否符合某一种**统计分布**。
- 可以用线性代数的方法把循环计算**矩阵化**。
- 可以用R语言来解多元**线性方程组**。
- 用R语言来做**参数优化**的过程等。

1.1 解一个一元二次方程

一元二次方程： $y = 5x^2 + 11x - 12$ ，求 x ？

```
> f <- function (x, a, b, c) a*x^2+b*x+c
> a<-5;b<-11;c<- -12
> result <- uniroot(f,c(0,-5),a=a,b=b,c=c,tol=0.0001)
> result$root
[1] -2.999977
>
> result <- uniroot(f,c(0,5),a=a,b=b,c=c,tol=0.0001)
> result$root
[1] 0.7999958
```

1.1 解一个一元二次方程

作图： $y = 5x^2 + 11x - 12$



1.2 R语言是面向数据的

- 核心问题：数据处理的复杂性。
- ~~不是~~：代码的复杂性。

用R语言，你有什么数据，就对进行什么样的分析。

把数据变成算法模型需要的格式，作为输入参数传递到模型，不用考虑复杂的程序编写的问题。

1.2 金融数据

- 金融行业本身就是玩数据的行业。金融量化上，我们会有各种数据处理的需求。
- 技术指标，计算某个金融产品的5日移动平均线。



1.2 上证综指数数据集

- 上证综指数数据集包括2列，时间序列索引列和每日收盘价，已存储在SSE 变量中。

```
> head(Cl(SSE),20)
```

```
000001.ss.close
2016-01-04      3296.26
2016-01-05      3287.71
2016-01-06      3361.84
2016-01-07      3125.00
2016-01-08      3186.41
2016-01-11      3016.70
2016-01-12      3022.86
2016-01-13      2949.60
2016-01-14      3007.65
2016-01-15      2900.97
2016-01-18      2913.84
2016-01-19      3007.74
2016-01-20      2976.69
2016-01-21      2880.48
2016-01-22      2916.56
2016-01-25      2938.51
2016-01-26      2749.79
2016-01-27      2735.56
2016-01-28      2655.66
2016-01-29      2737.60
```

```
> summary(Cl(SSE))
```

```
Index
Min.   :2016-01-04
1st Qu.:2016-03-07
Median :2016-05-04
Mean   :2016-05-03
3rd Qu.:2016-07-04
Max.   :2016-08-26

000001.ss.close
Min.   :2656
1st Qu.:2864
Median :2953
Mean   :2945
3rd Qu.:3017
Max.   :3362
```

1.2 技术指标计算

- 计算5日平均线，10日平均线，合并到数据框

```
> library(TTR)
> # 计算5日移动平均线，合并到df数据框
> df<-cbind(close=Cl(SSE),SMA(Cl(SSE),5),SMA(Cl(SSE),10))
> names(df)<-c('close','ma5','ma10')
> head(df,20)
```

	close	ma5	ma10
2016-01-04	3296.26	NA	NA
2016-01-05	3287.71	NA	NA
2016-01-06	3361.84	NA	NA
2016-01-07	3125.00	NA	NA
2016-01-08	3186.41	3251.444	NA
2016-01-11	3016.70	3195.532	NA
2016-01-12	3022.86	3142.562	NA
2016-01-13	2949.60	3060.114	NA
2016-01-14	3007.65	3036.644	NA
2016-01-15	2900.97	2979.556	3115.500
2016-01-18	2913.84	2958.984	3077.258
2016-01-19	3007.74	2955.960	3049.261
2016-01-20	2976.69	2961.378	3010.746
2016-01-21	2880.48	2935.944	2986.294
2016-01-22	2916.56	2939.062	2959.309
2016-01-25	2938.51	2943.996	2951.490
2016-01-26	2749.79	2892.406	2924.183
2016-01-27	2735.56	2844.180	2902.779
2016-01-28	2655.66	2799.216	2867.580
2016-01-29	2737.60	2763.424	2851.243

1.3 完善第三方包体系

- 金融领域涉及范围是非常广的，包括银行业、保险业、信托业、证券业、租赁业等。**量化投资**是证券投资的一个很细分的专业领域。



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

[About R](#)
[R Homepage](#)
[The R Journal](#)

[Software](#)
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

[Documentation](#)
[Manuals](#)
[FAQs](#)
[Contributed](#)

CRAN Task View: Empirical Finance

Maintainer: Dirk Eddelbuettel
Contact: Dirk.Eddelbuettel at R-project.org
Version: 2016-05-09

This CRAN Task View contains a list of packages useful for empirical work in Finance, grouped by topic.

Besides these packages, a very wide variety of functions suitable for empirical work in Finance is provided by both the basic R system (and its set of recommended core packages), and a number of other packages on the Comprehensive R Archive Network (CRAN). Consequently, several of the other CRAN Task Views may contain suitable packages, in particular the [Econometrics](#), [Multivariate](#), [Optimization](#), [Robust](#), [SocialSciences](#) and [TimeSeries](#) Task Views.

Contributions are always welcome, and encouraged. Since the start of this CRAN task view in April 2005, most contributions have arrived as email suggestions. The source file for this particular task view file now also reside in a GitHub repository (see below) so that pull requests are also possible.

Standard regression models

- A detailed overview of the available regression methodologies is provided by the [Econometrics](#) task view. This is complemented by the [Robust](#) which focuses on more robust and resistant methods.
- Linear models such as ordinary least squares (OLS) can be estimated by `lm()` (from by the `stats` package contained in the basic R distribution). Maximum Likelihood (ML) estimation can be undertaken with the standard `optim()` function. Many other suitable methods are listed in the [Optimization](#) view. Non-linear least squares can be estimated with the `nls()` function, as well as with `nls0` from the [nlme](#) package.
- For the linear model, a variety of regression diagnostic tests are provided by the `car`, `lmtest`, `strucchange`, `urca`, and `sandwich` packages. The `Rcmdr` and `Zelig` packages provide user interfaces that may be of interest as well.

Time series

- A detailed overview of tools for time series analysis can be found in the [TimeSeries](#) task view. Below a brief overview of the most important methods in finance is given.
- Classical time series functionality is provided by the `arima()` and `KalmanLike()` commands in the basic R distribution.
- The `dse` and `timesac` packages provide a variety of more advanced estimation methods; `fracdiff` can estimate fractionally integrated series; `longmemo` covers related material. The `fractal` provide fractal time series modeling functionality.
- For volatility modeling, the standard GARCH(1,1) model can be estimated with the `garch()` function in the [tseries](#) package. `Rmetrics` (see below) contains the `rfGarch` package which has additional models. The `rugarch` package can be used to model a variety of univariate GARCH models with extensions such as ARFIMA, in-mean, external regressors and various other specifications; with methods for fit, forecast, simulation, inference and plotting are provided too. The `rugarch` builds on it to provide the ability to estimate several multivariate GARCH models. The `betatsegarch` package can estimate and simulate the Beta-t-EGARCH model by Harvey. The `bayesGARCH` package can perform Bayesian estimation of a GARCH(1,1) model with Student's t innovations. For multivariate models, the `cogarch` package can estimate (multivariate) Conditional Correlation GARCH models whereas the `gogarch` package provides functions for generalized orthogonal GARCH models. The `gets` package (which was preceded by a related package `AutoSEARCH`) provides automated general-to-specific model selection of the mean and log-volatility of a log-ARCH-X model. The `GEVStableGarch` package can fit ARMA-GARCH or ARMA-APARCH models

1.3 量化交易体系

- 如果我们想用R构建自己的量化交易系统，你需要用到5方面的R语言工具包：数据管理、指标计算、回测交易、投资组合、风险管理。

	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, flmport
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

1.3 量化交易体系

- 引用国外量化的教材上的一张图，国外用R来研究量化交易已经体系。

Quantitative analysis package hierarchy

Application Area	R Package
Performance metrics and graphs	PerformanceAnalytics - Tools for performance and risk analysis
Portfolio optimization and quantitative trading strategies	PortfolioAnalytics - Portfolio analysis and optimization
	quantstrat - Rules-based trading system development
	blotter - Trading system accounting infrastructure
Data access and financial charting	quantmod - Quantitative financial modeling framework
	TTR - Technical trading rules
Time series objects	xts - Extensible time series
	zoo - Ordered observation

图片摘自Introduction to Trading Systems，作者Guy Yollin

1.4 R语言易学易上手

- R语言把**使用者**和**包开发者**，完全分成2个难度。
- 使用者：找到目标场景的第三方包，直接调用即可。代码通常也就是20-30行，就算死记硬背都可以了。
- 开发者：难度大。如果你想开发出一个比较好的包，你需要考虑代码的优雅性，程序计算性能的高效性，函数API封装的方便性，跨平台支持的兼容性，业务场景的明确性，以及帮助手册的完整性。

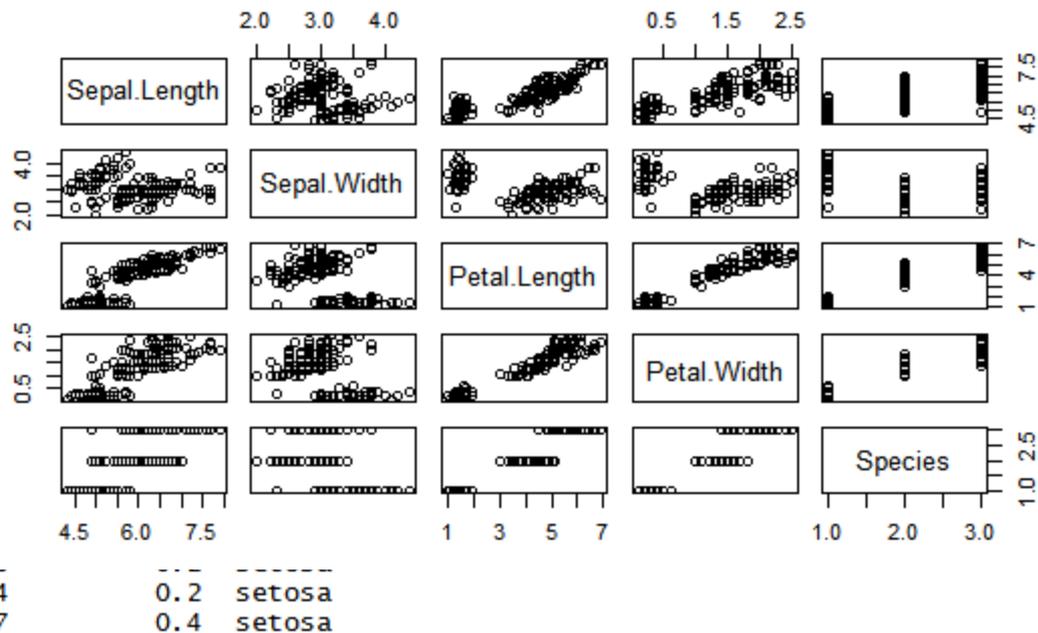
1.4 R的语法

- R语言语法。

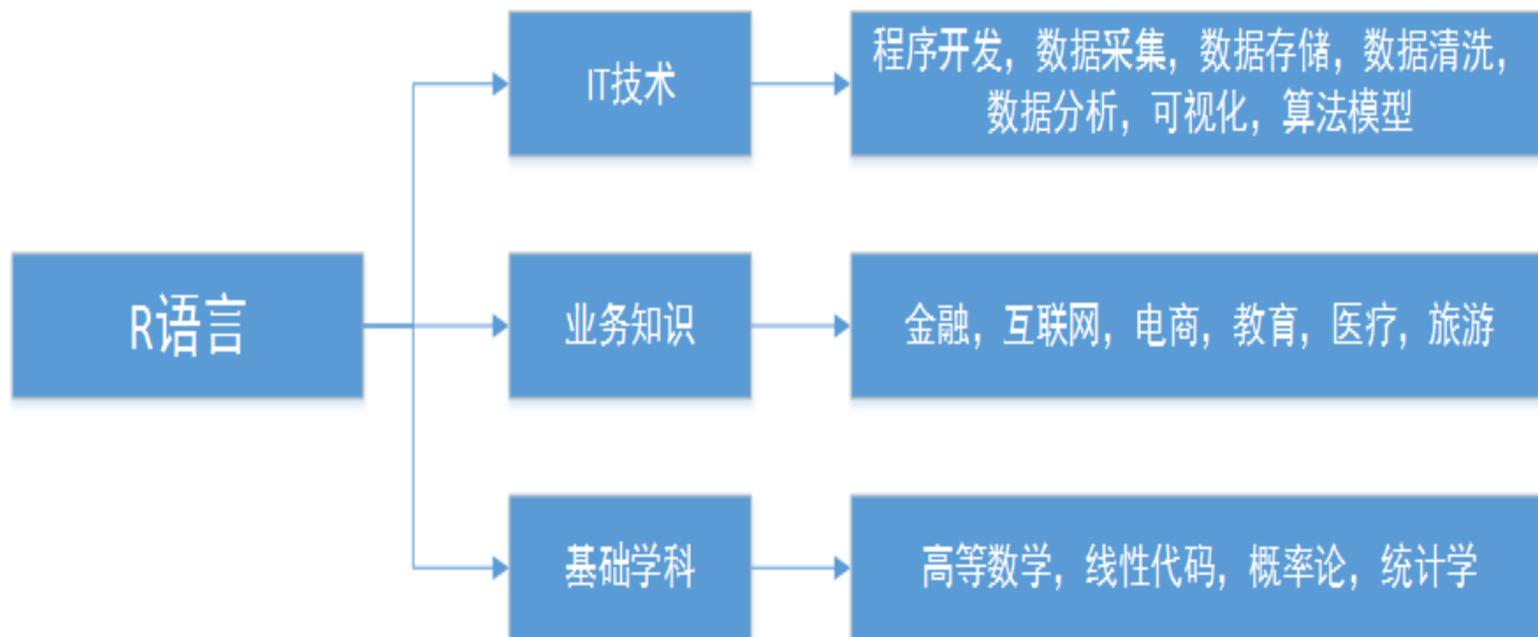
```
> a1<-c(1,2,3,4)->a2
> a1
[1] 1 2 3 4
> a2
[1] 1 2 3 4
> rnorm(10)
[1] 2.01748489 0.24053792 -2.
[6] 0.67800188 0.15310508 0.
```

- 用R画鸢尾花的数据集的散

```
> data(iris) #加载数据集
> head(iris) #查看前6行数据集
  Sepal.Length Sepal.width Petal.Length
1           5.1           3.5           1.
2           4.9           3.0           1.
3           4.7           3.2           1.
4           4.6           3.1           1.
5           5.0           3.6           1.4
6           5.4           3.9           1.7
> plot(iris) #画图
```



1.5 R语言的体系结构



- 图中我将R语言知识体系结构分为3个部分：
- IT技术 + 业务知识 + 基础学科

目录

1. R语言适合做量化分析的5个理由
2. 如何用R语言做量化分析
3. 量化分析实例

2. 如何用R语言做量化分析

- 金融行业有很多**背景知识**和**金融市场规则**需要我们了解。
- **量化关注的数据：**
- 交易数据，财务数据，上市公司的各种事件数据，基本面数据，宏观数据，舆情数据，互联网数据等
- 量化投资是跨学科应用的结合，有2个关键点。
 - **数据：**用R语言来解决数据的问题。
 - **金融知识：**就是我们对金融行业的行业积累。

2.1 R语言数据处理 – data.frame

- **数据框**(data.frame)类型，是R语言内置的一种数据类型，与关系型数据库中表的结构是类似的，是一种二维的数据结构。

```
> data.frame(A=1:6,B=LETTERS[1:6])
  A B
1 1 A
2 2 B
3 3 C
4 4 D
5 5 E
6 6 F
```

- **数据库**读取数据，导入**CSV**数据时，与R语言有了一个很好的映射关系，直接加载到R语言的内存中变成标准化数据格式。

2.1 R语言数据处理 - sqldf

- sqldf包，使用SQL语句对数据进行数据变换。

```
> library(sqldf)
> sqldf('select * from iris limit 6')
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1          3.5          1.4          0.2  setosa
2          4.9          3.0          1.4          0.2  setosa
3          4.7          3.2          1.3          0.2  setosa
4          4.6          3.1          1.5          0.2  setosa
5          5.0          3.6          1.4          0.2  setosa
6          5.4          3.9          1.7          0.4  setosa
```

- R语言默认的函数方法。

```
> iris[1:6,]
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1          3.5          1.4          0.2  setosa
2          4.9          3.0          1.4          0.2  setosa
3          4.7          3.2          1.3          0.2  setosa
4          4.6          3.1          1.5          0.2  setosa
5          5.0          3.6          1.4          0.2  setosa
6          5.4          3.9          1.7          0.4  setosa
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1          3.5          1.4          0.2  setosa
2          4.9          3.0          1.4          0.2  setosa
3          4.7          3.2          1.3          0.2  setosa
4          4.6          3.1          1.5          0.2  setosa
5          5.0          3.6          1.4          0.2  setosa
6          5.4          3.9          1.7          0.4  setosa
```

2.1 R语言数据处理

R语言做数据处理，要注意的地方

- 性能上，R比C++/Java等语言要慢不少。
- 语法上，使用避免使用for循环语句，使用apply族函数代替。
- 设计上，用向量计算或矩阵计算，代替循环计算。

2.2 海量数据集

中等规模数据：1G , 10G , 100G

- 数据处理包：data.table, bigmemory, ff等
- 并行计算包：snow, snowfall, parallel, doParallel, Rmpi等
- 加速R语言在单机上的计算的性能。

海量数据处理方法：

超过**100GB**这个量级，不只是R语言，每种语言都会遇到计算性能的瓶颈。
把数据放到分布式系统中，比如Hadoop。

R语言与大数据平台的通信接口：

RHadoop , rhive, rhbase, rmongodb, rCassandra, SparkR, sparklyr等

2.3 金融数据

金融市场每日产生的交易数据，以GB的量来增长，跟互联网比起来不是很快。但对于你程序加载10年的数据，也要GB或TB的一个量级。

离线模型：把海量数据变成离线的方式，放到hadoop或spark计算，用海量数据进行模型的训练。

在线模型：实时数据，一般就是一天或几天的数据，每天从开盘到收盘可能也就1-2GB，放到内存中进行计算。同时配合Docker容器化技术，实现并行化市场扫描。

2.3 时间序列数据

对于金融的数据处理，一般我会把它变成标准的时间序列类型的数据，R语言中基本的时间序列的类型为 **zoo** 和 **xts**类型

```
> df<-data.frame(A=1:6,B=rnorm(6))
> xdf<-xts(df,order.by=as.Date('2016-01-01')+1:6);xdf
      A      B
2016-01-02 1 -0.5849855
2016-01-03 2  2.0607084
2016-01-04 3  1.0555526
2016-01-05 4  1.2986014
2016-01-06 5 -0.9232844
2016-01-07 6  0.5505789
> class(xdf)
[1] "xts" "zoo"
```

2.3 时间序列数据模型

时间序列：以时间作为数据的索引列，把数据以**时间维度进行对齐**。

模型设计：你设计了一个股票交易策略和一个期货交易策略，由于股票是T+1交易，今天买了明天才能卖；而期货是T+0交易，今天买了马上就可以卖出。

如果这两个策略是**对冲**的，那么我们就需要把它们以时间维度进行对齐，才能进行实现对策略模型对冲的准确计算。

2.4 金融行业知识

从金融交易分析的角度，可以从3个维度进行。

- **基本面**：指对宏观经济、行业和公司基本情况的分析，包括公司经营理念策略、公司报表等的分析。长线投资一般用基本面分析，通过基本面可以判断是否值去交易。
- **技术面**：指通过技术指标变化，判断股票走势形态，进行K线组合等，通过技术面可以判断如何进行交易。
- **消息面**：指上市公司发布的利好和利空的消息，通过消息面可以判断市场的情绪。

2.4 金融行业知识

- **散户行为**，听到一个消息就跟着风的买卖股票，或者凭自己感觉大盘该涨了就跟进去。
- 量化的方法，主要是以技术指标为主。通过数据的基础，建立信心，而不是完全拍脑袋的事儿。
- 常用量化指标有MACD、KDJ、Boll、RSI、CCI等。但并不是单独使用，多个指标通过变换组合使用，通过变换再生成新的衍生指标，找到市场上赚钱的机会

目录

1. R语言适合做量化分析的5个理由
2. R语言的数据处理方法
3. 量化分析实例

3. 量化分析实例

用quantmod包构建量化模型

quantmod是一个R语言的金融量化分析包，主要用于数据下载，数据处理、金融建模、交易回测和可视化。

案例实战：

分析IBM的股价和上证综指(000001.ss)的关系，那么可以用quantmod包来完成整个投资研究的操作过程，包括数据下载，数据处理，数据建模，交易回测，可视化的部分。

3.1 实例一：IBM股价和上证指数

从yahoo财经，下载IBM股票数据。

International Business Machines Corporation (IBM) [☆ Add to watchlist](#)
NYSE - NYSE Real Time Price. Currency in USD

159.54 +0.66 (+0.42%)

At close: 4:02 PM EDT

Summary Conversations Statistics Profile Financials Options Holders **Historical Data** Analysts

Open	158.32	Market Cap	152.5B
Prev Close	158.88	P/E Ratio (ttm)	12.97
Bid	158.74 x 100	Beta	0.82
Ask	160.15 x 100	Volume	2,324,213
Day's Range	158.10 - 159.62	Avg Vol (3m)	3,318,447
52wk Range	116.90 - 164.95	Dividend & Yield	5.60 (3.52%)
1y Target Est	156.10	Earnings Date	Oct 17, 2016

Trade prices are not sourced from all markets

1D 5D 1M 6M YTD 1Y 2Y 5Y **10Y** MAX [↗ Interactive chart](#)



3.1 实例一：IBM股价和上证指数

画出K线图

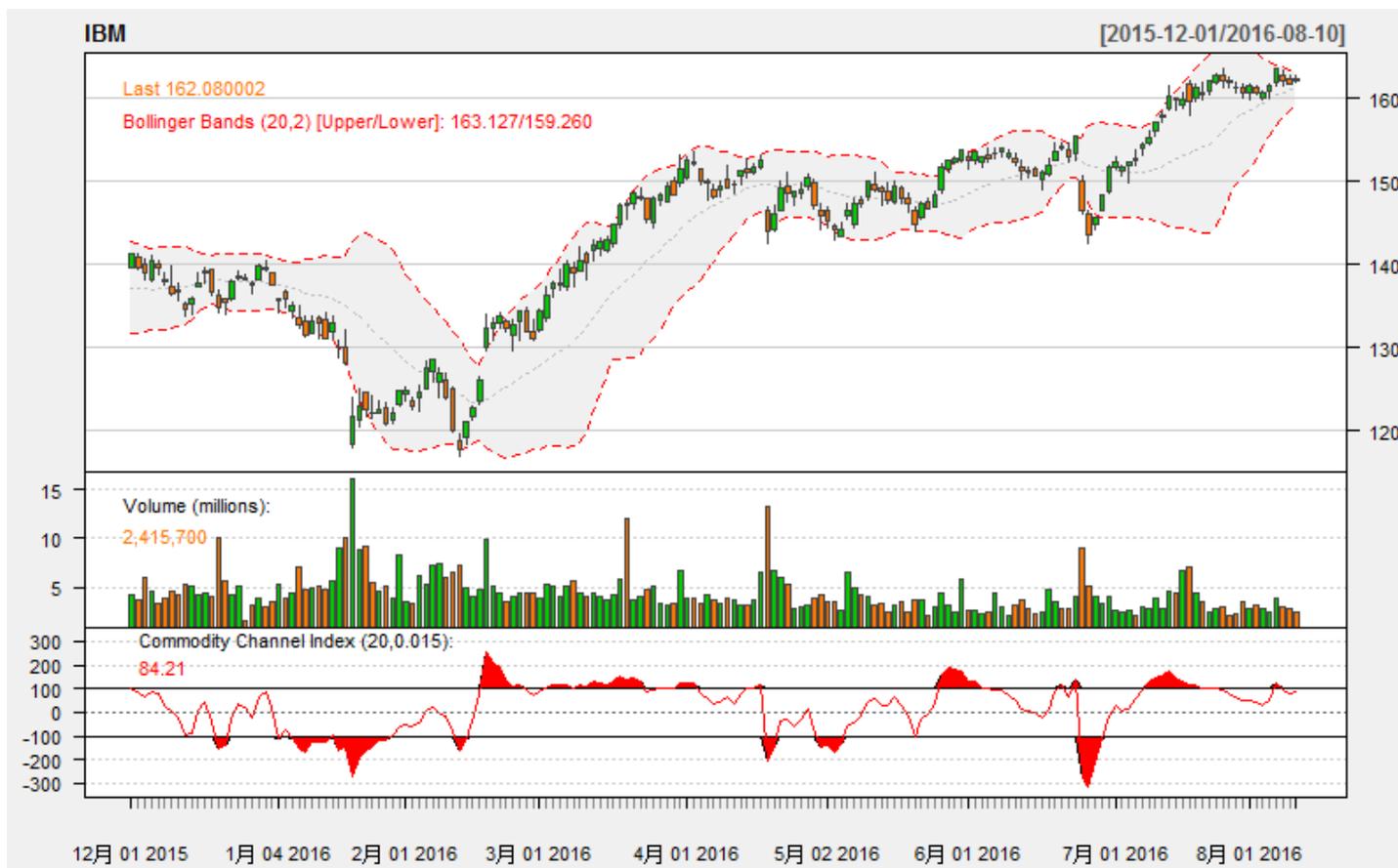
```
> chartSeries(IBM)
```



3.1 实例一：IBM股价和上证指数

调整数据集，样式，并增加技术指标

```
> chartSeries(IBM,subset='2015-12::2016-08',theme="white",TA="addVo();addBBands();addCCI()")
```



3.1 实例一：IBM股价和上证指数

下载上证综指(000001.SS)的数据。

对IBM股价和上证综指，建立线性关系。

```
> SSEC<-getsymbols("000001.SS",auto.assign = FALSE, from = '2010-10-10')
warning message:
In download.file(paste(yahoo.URL, "s=", symbols.name, "&a=", from.m, " :
  downloaded length 78582 != reported length 200
> dat <- specifyModel(Cl(IBM) ~ Cl(SSEC))
> dat

quantmod object:      Build date:

Model specified:
  Cl(IBM) ~ Cl(SSEC)

Model Target:  Cl.IBM      Product:  IBM
Model Inputs:

Fitted Model:

None Fitted
> head(modelData(dat))
      Cl.IBM Cl.SSEC
2010-10-11 139.66 2806.94
2010-10-12 139.85 2841.41
2010-10-13 140.37 2861.36
2010-10-14 141.50 2879.64
2010-10-15 141.06 2971.16
2010-10-18 142.83 2955.23
```

3.1 实例一：IBM股价和上证指数

构建一元线性回归模型，并进行回归检验。

```
> model <- buildModel(dat,method='lm',training.per=c('2014-08-01','2016-08-01'))
> summary(model)
```

```
quantmod object:   lm1472823992.75013   Build date:   2016-09-02 21:46:32
```

```
Model specified:
  cl(IBM) ~ cl(SSEC)
```

```
Model Target:  cl.IBM           Product:  IBM
Model Inputs:  cl.SSEC
```

```
Fitted Model:
```

```
Modelling procedure:  lm
Training window:  456 observations from 2014-08-01 to 2016-08-01
```

```
Call:
lm(formula = quantmod@model.formula, data = training.data)
```

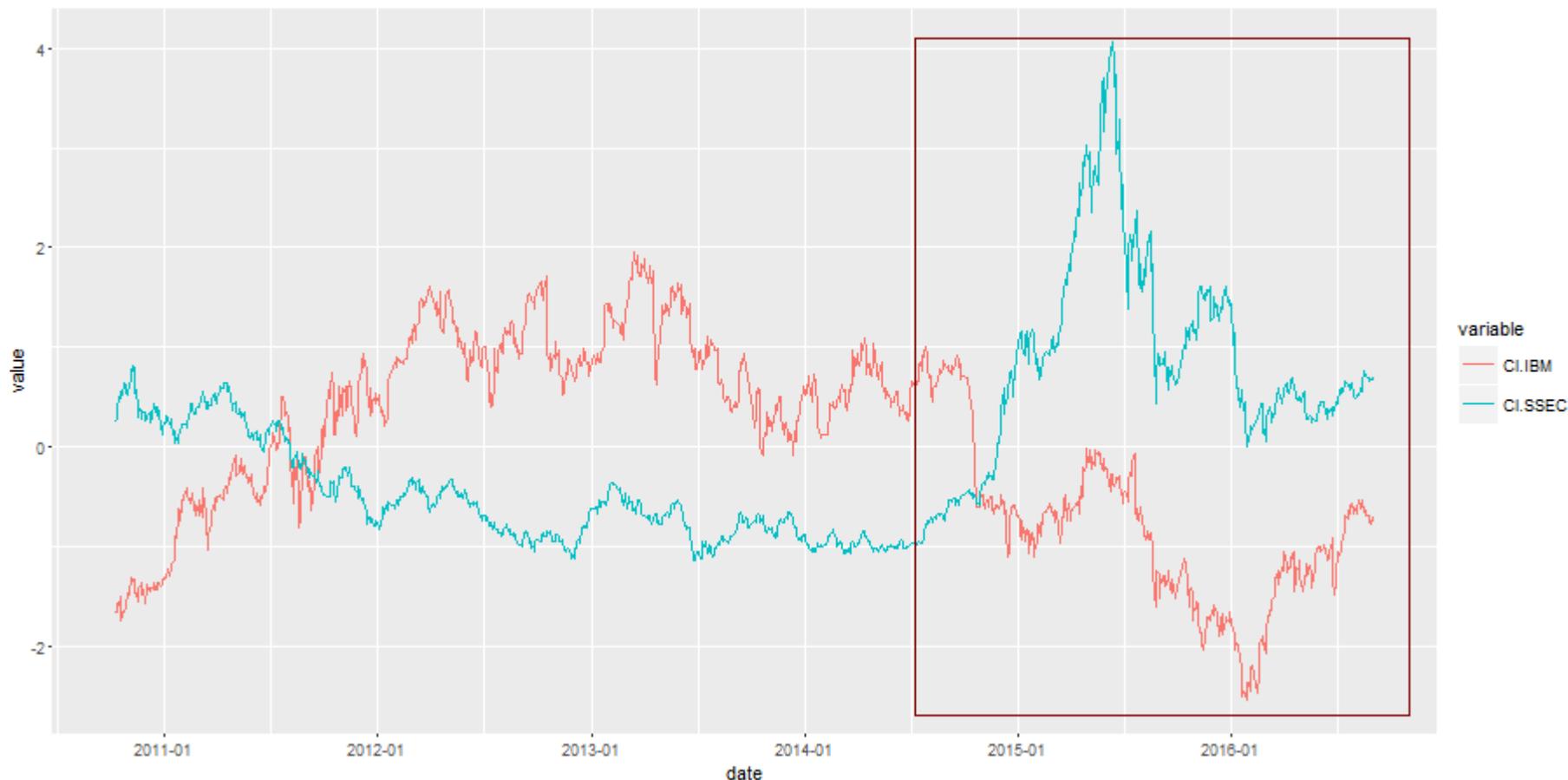
```
Residuals:
   Min     1Q  Median     3Q    Max
-36.845 -10.759   0.560   7.567  35.457
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 162.774814   3.750804  43.397 <2e-16 ***
cl.SSEC      -0.001817   0.001136  -1.599   0.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 16.13 on 454 degrees of freedom
Multiple R-squared: 0.005602, Adjusted R-squared: 0.003412
F-statistic: 2.558 on 1 and 454 DF, p-value: 0.1105
```

3.1 实例一：IBM股价和上证指数

标准化数据，并可视化显示。



3.2 实例二：黑色系期货品种

商品期货，黑色系品种是具有产业链上下游的关系。

- **铁矿石**是炼钢的原材料。
 - **焦煤**和**焦炭**是炼钢的能源资源。
 - **热卷**是以板坯为原料经加热后制成的钢板。
 - **螺纹钢**是表面带肋的钢筋。
-
- **预测螺纹钢的价格？**

3.2 实例二：黑色系期货品种

定义模型：多元线性回归

$$y = a + b * x1 + c * x2 + d * x3 + e * x4 + \epsilon$$

- y ，为因变量，螺纹钢
- $x1$ ，为自变量，焦煤
- $x2$ ，为自变量，焦炭
- $x3$ ，为自变量，铁矿石
- $x4$ ，为自变量，热卷
- a ，为截距
- b,c,d,e ，为自变量系数
- ϵ ，为残差，是其他一切不确定因素影响的总和，其值不可观测。假定 ϵ 服从正态分布 $N(0, \sigma^2)$ 。

3.2 实例二：黑色系期货品种

数据集为2016-03-15的5个品种的1分钟线数据。

- x1, 焦煤
- x2, 焦炭
- x3, 铁矿石
- x4, 热卷
- y, 螺纹钢

```
# 数据集已存在df变量中
> head(df, 20)
           x1    x2    x3    x4    y
2016-03-15 09:01:00 754.5 616.5 426.5 2215 2055
2016-03-15 09:02:00 752.5 614.5 423.5 2206 2048
2016-03-15 09:03:00 753.0 614.0 423.0 2199 2044
2016-03-15 09:04:00 752.5 613.0 422.5 2197 2040
2016-03-15 09:05:00 753.0 615.5 424.0 2198 2043
2016-03-15 09:06:00 752.5 614.5 422.0 2195 2040
2016-03-15 09:07:00 752.0 614.0 421.5 2193 2036
2016-03-15 09:08:00 753.0 615.0 422.5 2197 2043
2016-03-15 09:09:00 754.0 615.5 422.5 2197 2041
2016-03-15 09:10:00 754.5 615.5 423.0 2200 2044
2016-03-15 09:11:00 757.0 616.5 423.0 2201 2045
2016-03-15 09:12:00 756.0 615.5 423.0 2200 2044
2016-03-15 09:13:00 755.5 615.0 423.0 2197 2042
2016-03-15 09:14:00 755.5 615.0 423.0 2196 2042
2016-03-15 09:15:00 756.0 616.0 423.5 2200 2045
2016-03-15 09:16:00 757.5 616.0 424.0 2205 2052
2016-03-15 09:17:00 758.5 618.0 424.0 2204 2051
2016-03-15 09:18:00 759.5 618.5 424.0 2205 2053
2016-03-15 09:19:00 759.5 617.5 424.5 2206 2053
2016-03-15 09:20:00 758.5 617.5 423.5 2201 2050
```

3.2 实例二：黑色系期货品种

```
> summary(lm1)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.9648	-1.3241	-0.0319	1.2	4.9648

Coefficients:

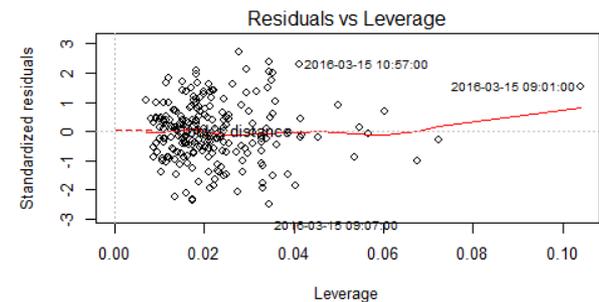
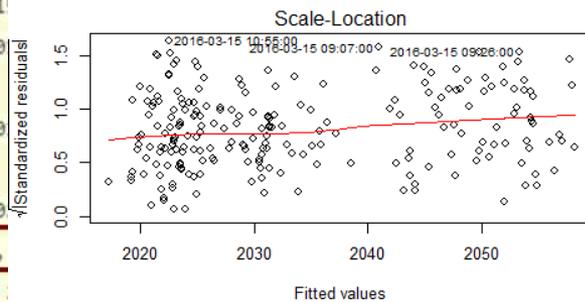
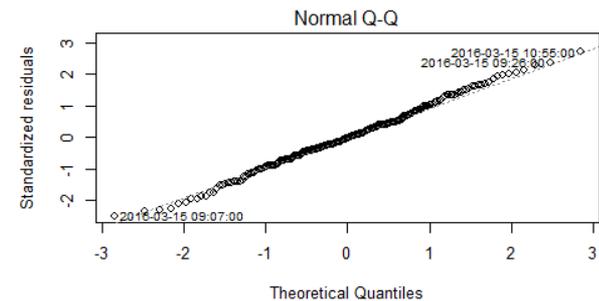
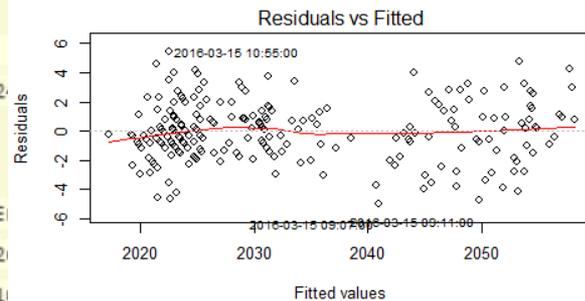
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	212.87796	58.2	3.658	0.00036
x1	0.85423	0.1	8.542	<0.0001
x2	0.66724	0.1	6.672	<0.0001
x3	-0.66741	0.1	-6.674	<0.0001
x4	0.48214	0.0	4.821	<0.0001

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.0

Multiple R-squared: 0.9725,

F-statistic: 1956 on 4 and



结论，没有明显的异常点，残差符合假设条件。

3.2 实例二：黑色系期货品种

螺纹钢 = $212.87796 + 0.85423 * \text{焦炭} + 0.66724 * \text{焦煤} - 0.66741 * \text{铁矿石} + 0.48214 * \text{热卷}$



3.3 个性化量化建模

量化交易的过程是复杂地，不是简简单单的几行代码，就能保证我们能够市场上赚到钱的，中间还涉及市场规则，交易规则等等很多方面的问题。

就算回测和模拟盘表现的再好，到了实盘中也可能背道而驰，让你血本无归。所以，我们其实更多的时候，还要发现市场特有的规则，进行个性化的程序开发。

- 二条均线打天下 (<http://blog.fens.me/finance-stock-ma/>)
- 均值回归，逆市中的投资机会 (<http://blog.fens.me/finance-mean-reversion/>)
- R语言构建追涨杀跌量化交易模型 (<http://blog.fens.me/finance-chase-sell/>)
- R语言构建配对交易量化模型 (<http://blog.fens.me/finance-pairs-trading/>)

总结

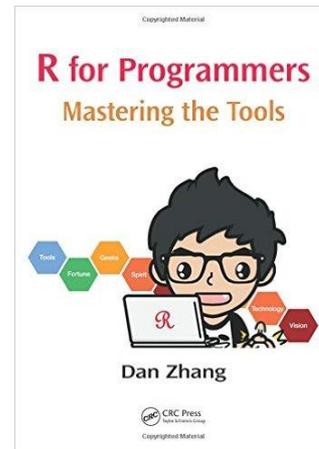
2015年我在创业，希望能推动R语言在金融量化领域的发展，但是由于种种原因项目没有持续发展。接下来，我还会以个人的方式继续努力，继续推动R在金融领域的发展。

R语言，对我的影响和改变是非常大的，我认识R语言是非常好的一门语言，我会把推动R的发展，当成一项事业来做。

希望也能认识各位业界朋友，一起努力，把这份事业做下去！！

作者介绍

《R的极客理想》系列图书作者



联系方式：

Blog：<http://fens.me>

Email：bsspirit@gmail.com

QQ群：383275651

参考资料

R的极客理想系统列文章 (<http://blog.fens.me/series-r/>)
用IT技术玩金融系列文章 (<http://blog.fens.me/series-it-finance/>)

R语言为量化而生 (<http://blog.fens.me/r-finance/>)
R语言量化投资常用包总结 (<http://blog.fens.me/r-quant-packages/>)

R语言时间序列基础库zoo (<http://blog.fens.me/r-zoo/>)
可扩展的时间序列xts (<http://blog.fens.me/r-xts/>)

超高性能数据处理包data.table (<http://blog.fens.me/r-data-table/>)
掌握R语言中的apply函数族 (<http://blog.fens.me/r-apply/>)

R语言解读一元线性回归模型 (<http://blog.fens.me/r-linear-regression/>)
R语言解读多元线性回归模型 (<http://blog.fens.me/r-multi-linear-regression/>)
R语言解读自回归模型 (<http://blog.fens.me/r-ar/>)

Q&A

THANK YOU