## Rapid Trade Development

Leveraging $R$ with xts and quantmod for quantitative trading

Jeffrey A. Ryan<br>jeffrey.ryan@insightalgo.com

Presented on March 27, 2009
Department of Statistics and Mathematics
Vienna University of Economics and Business Administration

## What is quantitative trading?

## What is quantitative trading?

A programmatic/automated process to execute trades based on mathematical or statistical logic.

## Quant Trading Components



## Quant Trading Components



## It can in R!

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN) Strong academic research usage

Open source - customization and validation
Fasy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN) Strong academic research usage

Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)

## Strong academic research usage

Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Fasy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

## Powerful language out of the box

Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to R

Powerful language out of the box
Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to $\mathbf{R}$

Powerful language out of the box
Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## Advantages to $\mathbf{R}$

Powerful language out of the box
Huge financial community use and contributions
Vector-based data and processing (time-series)
Unequaled statistical toolkit (base + CRAN)
Strong academic research usage
Open source - customization and validation
Easy to integrate with external data stores
Extensibility with C, Fortran, C++, etc
Flexible graphics on-screen and off

## What was missing for trading?

A trader-oriented workflow
Standard financial visualization tools
Native tools for managing [large] time series quickly
Connections to real-time data feeds (Bloomberg, IB, DTN)
Connections to industry data stores (Vhayu, OneTick, Kdb)
Wide-spread adoption \& visibility

## What was missing for trading?

## A trader-oriented workflow

## Standard financial visualization tools

Native tools for managing [large] time series quickly
Connections to real-time data feeds (Bloomberg, IB, DTN)
Connections to industry data stores (Vhayu, OneTick, Kdb)
Wide-spread adoption \& visibility

## quantmod

## What was missing for trading?

## A trader-oriented workflow

Standard financial visualization tools
Native tools for managing [large] time series quickly
Connections to real-time data feeds (Bloomberg, IB, DTN)
Connections to industry data stores (Vhayu, OneTick, Kdb)
Wide-spread adoption \& visibility

## XtS

## What was missing for trading?

A trader-oriented workflow focus
Standard financial visualization tools
Native tools for managing [large] time series quickly
Connections to real-time data feeds (Bloomberg, IB, DTN)
Connections to industry data stores (Vhayu, OneTick, Kdb)
Wide-spread adoption \& visibility

## xts: extensible time series

Most everything in trading involves a time series

> Regular data (positions data, P\&L)
> Irregular data (market data, book data, trades)
$R$ has many ways to manage this...

## xts: extensible time series

## Data Classes

## xts: extensible time series

fts

## Data Classes

matrix

tframe mts
data.table

## data.frame ZOO

its

## zooreg

vectors

## xts: extensible time series

fts

## Data Classes

matrix mts tframe data.table data.frame ZOO its

## timeSeries $_{\text {zooreg }}^{\text {ts }}$

vectors

## Time Classes

## xts: extensible time series


data.table

## data.frame ZOO

 its
## timeSeries ${ }_{\text {zooreg }}^{\text {ts }}$

## xts: extensible time series

## fts <br> Data Classes

matrix

## mts

data.table

Time Classes

character
yearmon
Date POSIXIt numeric yearqtr timeDate

## xts: extensible time series

The "solution" ?

## xts: extensible time series

add one more class of course...

## xts: extensible time series

## Motivation (c. 2007)

Avid user of zoo

- Natural R-like interface
- Flexible and complete methods
- S3!

I still wanted a few features for trading...

- Additional series metadata
- Require time-based indexing
- Conversion/reconversion tools


## xts: extensible time series

## Significant design requirements for xts:

- Preserve zoo behavior
- Utilize time-based indexing
- Allow for arbitrary attributes to be cleanly attached
- ISO 860 I subsetting by time strings
- Lossless conversion utilities to hide messy details


## xts: extensible time series

What's inside an xts object?


## xts: extensible time series

## What's inside an xts object?



## xts: extensible time series

## What's inside an xts object?



## xts: extensible time series

## What's inside an xts object?



## xts: extensible time series

## What's inside an xts object?



## xts: extensible time series

## Index as numeric? That isn't "time-based"!!

- Internally all index values are represented in POSIX time (seconds since the epoch)
- Coercion happens at object creation or upon index replacement
- index() converts back to user-level class
- .index() access the raw seconds in the index
- .indexCLASS, indexFORMAT and .indexTZ attributes
- Rationale? Simplicity for $C$ level code, removal of multiple conversion in most instances, more consistent behavior
- All details are hidden from the user


## xts: extensible time series

## Time-based indexing in xts (ISO 8601)

- Date and time organized from most significant to least significant: CCYY-MM-DD HH:MM:SS[.s]
- Fixed number of digits
- Separators can be omitted e.g. CCYYMMDDHHMMSS
- Reduced accuracy forms are valid: e.g. CCYY-MM
- Fractional decimal time is supported
- Intervals can be expressed e.g. 2000-05/200I-04


## xts: extensible time series

## Create an xts object



## xts: extensible time series

## Create an xts object



## xts: extensible time series

## Indexing by time



## xts: extensible time series

## Indexing by time



## xts: extensible time series

## Indexing by time



## xts: extensible time series

## Indexing by time

> x["2009"]
[, I]
2009-03-24 0.3554788 2009-03-25 I.28I2633 2009-03-26 0.1268833 2009-03-27-0.6945146 2009-03-28-0.3936148 2009-03-29-0. 1938840 2009-03-30 0.2368576 2009-03-3I - .2152293 2009-04-01 0.8100493 2009-04-02 I.4I52439

> x["2009030I/200903"]

## All of April 2009

## xts: extensible time series

## Indexing by time

> x["2009"]
[, I]
2009-03-24 0.3554788 2009-03-25 I.28I2633 2009-03-26 0.1268833 2009-03-27-0.6945146 2009-03-28-0.3936148 2009-03-29 -0. 1938840 2009-03-30 0.2368576 2009-03-3I - .2152293 2009-04-01 0.8100493 2009-04-02 I.4I52439

> x["2009030I/200903"]

## [, I]

2009-03-24 0.3554788 2009-03-25 I.28I2633 2009-03-26 0.1268833 2009-03-27-0.6945146 2009-03-28-0.3936148 2009-03-29 -0. 1938840 2009-03-30 0.2368576 2009-03-3I -I.2I52293

## xts: extensible time series

## All subsetting is via a binary search algorithm. F-A-S-T!

```
> str(xIOm) # 10 million observations
An 'xts' object from 2009-03-23 16:I9:00 to 2009-07-17 10:05:39 containing:
Data: int [I:I0000000, I] I 2 3456789 IO ...
Indexed by objects of class: [POSIXt,POSIXct] TZ:America/Chicago
xts Attributes:
NULL
> str(xI00k) # 100 thousand observations
An 'xts' object from 2009-03-23 16:19:00 to 2009-03-24 20:05:39 containing:
    Data:int [l:I00000, I] | 23456789 IO ...
Indexed by objects of class: [POSIXt,POSIXct] TZ:America/Chicago
xts Attributes:
NULL
> system.time(xIOm['20090323'])
    user system elapsed
0.006 0.00| 0.006
> system.time(xl00k['20090323'])
    user system elapsed
    0.006 0.001 0.006
> system.time(x10m[index(x|0m) >= as.POSIXct('2009-03-23 16:19:00') & index(x10m) <=
    as.POSIXct('2009-03-23 23:59:58')])
    user system elapsed
    |.457 |. 372 2.832
```


## xts: extensible time series

All subsetting is via a binary search algorithm. F-A-S-T!

```
> str(x|0m) # 10 million observations
An 'xts' object from 2009-03-23 16:I9:00 to 2009-07-17 10:05:39 containing:
    Data:int [l:|0000000, |] | 2 3456789 |0
    Indexed by objects of class: [POSIXt,POSIXct] TZ:America/Chicago
    xts Attributes:
    NULL
> str(xl00k) # l00 thousand observations
An 'xts' object from 2009-03-23 I6:I9:00 to 2009-03-24 20:05:39 containing:
    Data: int [I:I00000, I] | 2 345678 9 IO
    Indexed by objects of class: [POSIXt,POSIXct] TZ:America/Chicago
    xts Attributes:
NULL
> system.time(xIOm['20090323'])
    user system elapsed
    0 . 0 0 6 ~ 0 . 0 0 1 ~ 0 . 0 0 6 ~
> system.time(xl00k['20090323'])
    user system elapsed
    0.006 0.001 0.006
> system.time(xIOm[index(x|0m) >= as.POSIXct('2009-03-23 16:19:00') & index(x10m) <=
    as.POSIXct('2009-03-23 23:59:58')])
    user system elapsed
    |.457 |. 372 2.832
```


## xts: extensible time series

All subsetting is via a binary search algorithm. F-A-S-T!

```
> str(xI0m) # 10 million observations
An 'xts' object from 2009-03-23 16:I9:00 to 2009-07-17 10:05:39 containing:
    Data:int [I:I0000000, I] | 2 345678 9 |0
    Indexed by objects of class: [POSIXt,POSIXct] TZ:America/Chicago
    xts Attributes:
NULL
> str(xI00k) # 100 thousand observations
An 'xts' object from 2009-03-23 16:I9:00 to 2009-03-24 20:05:39 containing:
Data:int [I:|00000, I] | 2 345678 9 IO
Indexed by objects of class: [POSIXt,POSIXct] TZ:America/Chicago
xts Attributes:
NULL
> system.time(x 0m['20090323'])
    user system elapsed
    0.006 0.001 0.006
> system.time(xl00k['20090323'])
    user system elapsed
    0.006 0.00| 0.006
> system.time(x10m[index(xI0m) >= as.POSIXct('2009-03-23 16:19:00') & index(x10m) <=
    as.POSIXct('2009-03-23 23:59:58')])
    user system elapsed
    l.457 I. 372 2.832
```


## xts: extensible time series

## xts $+C$

- Moving [.xts to $C$ dramatically decreased subsetting costs
- Highest cost basic operation in R was merge. Prime C candidate
- Implemented optimized sort-merge-join in $C$ with custom algorithm
- Additional C based routines followed...
xts now has 3000+ lines of $C$


## xts: extensible time series

...in development

## xts: extensible time series

## Binary .xd files

Representation of xts objects on disk
Seekable for disk-based subsetting
Future time-series database structure

xtsDB

## xts: extensible time series

## Parallel processing

period.apply<br>runSum, runCov, runSD, etc.

## xts: extensible time series

## Parallel processing

period.apply runSum, runCov, runSD, etc.

Multiple index support index lists

## xts: extensible time series

## Parallel processing

period.apply<br>runSum, runCov, runSD, etc.

# Multiple index support index lists 

Tighter zoo integration
Backport C code into zoo


## quantmod

## quantmod

quantmod was envisioned to be a rapid prototyping environment in R to facilitate quantitative modeling, testing, and trading

## quantmod

Data. Visualization. Modeling.

## quantmod

## Data. Visualization. Modeling.

Trading requires lots of different types of data, from many different sources. quantmod aims to hide the details of the data source, to make using data a priority

## quantmod

## Data. Visualization. Modeling.

Trading requires lots of different types of data, from many different sources. quantmod aims to hide the details of the data source, to make using data a priority

## getSymbols

## quantmod

Data. Visualization. Modeling.

## getSymbols

| csv | Rdata | MySQL |
| :---: | :---: | :---: |
| SQLite | google | yahoo |
| Interactive <br> Brokers | FRED | oanda |

## Data. Visualization. Modeling.

## getSymbols

getSymbols is the top level function that dispatches to custom methods based on user direction

setSymbolLookup<br>getSymbolLookup<br>saveSymbolLookup<br>loadSymbolLookup

## quantmod

## Data. Visualization. Modeling.

## getSymbols

getSymbols behave like base::load by assigning objects into the user's workspace (.GlobalEnv)

## Data. Visualization. Modeling.

## getSymbols

getSymbols behave like base::load by assigning objects into the user's workspace (.GlobalEnv)

Rationale: when dealing with potentially dozens of symbols interactively, it is redundant to have to manually assign each. Also facilitates multiple requests.

## Data. Visualization. Modeling.

## getSymbols

getSymbols("AAPL") getSymbols("AAPL;SBUX") getSymbols("USD/EUR",src= "oanda")

## quantmod

## Data. Visualization. Modeling.

Interactive, highly customizable financial charting in R

## quantmod

## Data. Visualization. Modeling.

${ }^{\wedge} \mathrm{gspc}$


Basic OHLC chart from tseries

## quantmod

Data. Visualization. Modeling.

candleChart(GSPC, subset='200902/', theme='white',TA=NULL)

## Data. Visualization. Modeling.

## Requirements

- Fast rendering (base plotting tools)
- Interactive and scriptable
- Work with all timeseries classes
- Minimal commands
- Highly customizable
- Full technical analysis support (via TTR)


## quantmod

Data. Visualization. Modeling.

## The Basics



## quantmod

Data. Visualization. Modeling.

## The Basics



## quantmod

Data. Visualization. Modeling.

## The Basics



## quantmod

## Data. Visualization. Modeling.

## The Basics


> reChart(subset= "2009",theme= "white", type= "candles")

## quantmod

## Data. Visualization. Modeling.

## Inside chartSeries



## quantmod

## Data. Visualization. Modeling. <br> Inside chartSeries

chartSeries



## quantmod

## Data. Visualization. Modeling.

## Inside chartSeries



## quantmod

## Data. Visualization. Modeling.

## Inside chartSeries



Drawn by chartSeries.chob

## quantmod

## Data. Visualization. Modeling.

## Extending chartSeries

## Data. Visualization. Modeling.

GMMA<br>Guppy Multiple Moving Average<br>(with newTA)

## quantmod

## Data. Visualization. Modeling.

```
> # create a function that returns our GMMA
> GMMA <- function(x) {
+ fastMA <- c(3,5,8,10,12,15)
    slowMA <- c(30,35,40,45,50,60)
    x <- sapply(c(fastMA,slowMA),
    function(xx) EMA(x,xx))
    return(x)
+ }
```


## quantmod

## Data. <br> Visualization. <br> Modeling.



```
# create an addGuppy function with newTA
addGuppy <- newTA(FUN=GMMA,
    preFUN=Cl,
    col=c(rep(3,6),
    rep("#333333",6)),
    legend="GMMA")
class(addGuppy)
[I] "function"
```


## quantmod

## Data. Visualization. Modeling.

## candleChart(AAPL); addGuppy()



## quantmod

Data. Visualization. Modeling.


## quantmod

## Data. Visualization. Modeling. chartSeries3d

Yield Curve 2008 --- Daily


## quantmod

## Data. Visualization. Modeling.

## chartSeries + chartSeries3d



## quantmod

## Data. Visualization. Modeling.



## attachSymbols

New functionality to extend upon getSymbols

## attachSymbols

New functionality to extend upon getSymbols

Create a demand based database system using getSymbols that allows for implicit loading of an entire universe of symbols

## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

```
> search()
[l] ".GlobalEnv" "package:quantmod" "package:Defaults"
[4] "package:xts" "package:zoo" "package:stats"
[7] "package:graphics" "package:grDevices" "package:utils"
[IO] "package:datasets" "package:methods" "Autoloads"
[I3] "package:base"
```


## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

```
> search()
    [I] ".GlobalEnv" "package:quantmod" "package:Defaults"
    [4] "package:xts" "package:zoo" "package:stats"
    [7] "package:graphics" "package:grDevices" "package:utils"
    [I0] "package:datasets" "package:methods" "Autoloads"
    [I3] "package:base"
    > attachSymbols(DB=DDB_Yahoo(), pos=2, prefix="E.")
        /
```


## attachSymbols

## Example:All US Equity symbols on demand.

```
> search()
    [I] ".GlobalEnv" "package:quantmod" "package:Defaults"
    [4] "package:xts" "package:zoo" "package:stats"
    [7] "package:graphics" "package:grDevices" "package:utils"
    [IO] "package:datasets" "package:methods" "Autoloads"
    [I3] "package:base"
    > attachSymbols(DB=DDB_Yahoo(), pos=2, prefix="E.")
> search()
    [I] ".GlobalEnv" "DDB:Yahoo" "package:quantmod"
    [4] "package:Defaults" "package:xts" "package:zoo"
    [7] "package:stats" "package:graphics" "package:grDevices"
    [IO] "package:utils" "package:datasets" "package:methods"
    [I3] "Autoloads" "package:base"
```


## attachSymbols

## Example:All US Equity symbols on demand.

```
> search()
    [I] ".GlobalEnv" "package:quantmod" "package:Defaults"
    [4] "package:xts" "package:zoo" "package:stats"
    [7] "package:graphics" "package:grDevices" "package:utils"
[IO] "package:datasets" "package:methods" "Autoloads"
[I3] "package:base"
> attachSymbols(DB=DDB_Yahoo(), pos=2, prefix="E.")
> search()
    [I] ".GlobalEnv" "DDB:Yahoo" "package:quantmod"
    [4] "package:Defaults" "package:xts" "package:zoo"
    [7] "package:stats" "package:graphics" "package:grDevices"
[IO] "package:utils" "package:datasets" "package:methods"
[I3] "Autoloads" "package:base"
> str(Is("DDB:Yahoo"))
chr [I:7406] "E.A" "E.AA" "E.AAC" "E.AACC" "E.AAI" "E.AAII" ...
```


## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

```
> search()
[I] ".GlobalEnv" "package:quantmod" "package:Defaults"
[4] "package:xts" "package:zoo" "package:stats"
[7] "package:graphics" "package:grDevices" "package:utils"
[IO] "package:datasets" "package:methods" "Autoloads"
```


## 7406 symbols are available

```
[I] ".GlobalEnv" "DDB:Yahoo" "package:quantmod"
[d] "package:Defaults" "package:xts" "package:zoo"
[\ "package:stats" "package:graphics" "package:grDevices"
[IO] "package:utils" "package:datasets" "package:methods"
[I3] Autoloads" "package:base"
> str(Is.\DDDB:Yahoo"))
chr [I:7406] "E.A" "E.AA" "E.AAC" "E.AACC" "E.AAI" "E.AAII" ...
```


## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

```
> str(E.A)
An 'xts' object from 2007-01-03 to 2009-03-23 containing:
    Data: num [I:559, I:6] 35 34.3 34.3 34 34.1 ...
    - attr(*, "dimnames")=List of 2
    ..$:NULL
    ..$ : chr [I:6] "A.Open" "A.High" "A.Low" "A.Close" ...
    Indexed by objects of class:[POSIXt,POSIXct] TZ:America/
Chicago
    xts Attributes:
List of 2
$ src :chr "yahoo"
$ updated: POSIXct[I:I], format: "2009-03-24 10:59:I4"
```


## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

```
> str(E.A)
An 'xts' object from 2007-01-03 to 2009-03-23 containing:
    Data: num [I:559, I:6] 35 34.3 34.3 34 34.1 ...
    - attr(*, "dimnames")=List of 2
    ..$:NULL
    ..$ : chr [I:6] "A.Open" "A.High" "A.Low" "A.Close" ...
    Indexed by objects of class:[POSIXt,POSIXct] TZ:America/
Chicago
    xts Attributes:
List of 2
$ src :chr "yahoo"
$ updated: POSIXct[I:I], format: "2009-03-24 10:59:I4"
```


## First access loads data

## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

> system.time(E.AKAM)<br>user system elapsed<br>$0.032 \quad 0.004 \quad 0.267$

## quantmod

## attachSymbols

## Example:All US Equity symbols on demand.

```
> system.time(E.AKAM)
    user system elapsed
    0.032 0.004 0.267
> system.time(E.AKAM)
    user system elapsed
        0 0 0
```


## attachSymbols

## Two Methods to Cache



## Memory

after first access objects remain in
memory
delayedAssign

## quantmod

## attachSymbols

## Custom DDB methods

## example: DDB:Yahoo

> DDB_Yahoo()
> \# creates DDB of all US Equity Symbols

## quantmod

## attachSymbols

## Custom DDB methods

## example: DDB:Yahoo

$>$ attachSymbols()
> \#"binds" symbols to functions to load/reload

## quantmod

## attachSymbols

## Custom DDB methods

## example: DDB:Yahoo

All symbols are attached to new environment and accessible on demand.

## Future Work

Integration of trading/testing with blotter package More data methods, easier to extend specifyModel - buildModel - tradeModel work

## More Information

www.insightalgo.com

## www.quantmod.com



Presented by Jeffrey A. Ryan jeffrey.ryan@insightalgo.com
www.quantmod.com/Vienna2009

