A Short Introduction to R

Rossen Trendafilov and Brandon Vick

Fordham University

June 18, 2009

Outline

Introduction

- R is a free programming language and software environment for statistical computing and graphics.

Introduction

- R is a free programming language and software environment for statistical computing and graphics.
- R is highly extensible through the use of user-submitted packages for specific functions or specific areas of study.

Introduction

- R is a free programming language and software environment for statistical computing and graphics.
- R is highly extensible through the use of user-submitted packages for specific functions or specific areas of study.
- Major strength of R is its graphical facilities, which produce publication-quality graphs. R has its own LaTeX-like documentation format.

Outline

Getting Help

1. Within R - help()
   - help(command)
   - help (package = package name)
   - vignette (package = "package name"), vignette("vignette name")
   - ?? character string

Getting Help

1. Within R - help()
   - help(command)
   - help (package = package name)
   - vignette (package = "package name"), vignette("vignette name")
   - ?? character string
2. R web site
   - http://www.r-project.org/index.html

Getting Help

1. Within R - help()
   - help(command)
   - help (package = package name)
   - vignette (package = "package name"), vignette("vignette name")
   - ?? character string
2. R web site
   - http://www.r-project.org/index.html
3. Hands-on Intermediate Econometrics Using R, by H.D. Vinod

Getting Help

1. Within R - help()
   - help(command)
   - help (package = package name)
   - vignette (package = "package name"), vignette("vignette name")
   - ?? character string
2. R web site
   - http://www.r-project.org/index.html
3. Hands-on Intermediate Econometrics Using R, by H.D. Vinod
4. Professor Vinod's Notes on R Language
   - http://www.fordham.edu/economics/vinod/R-lang.doc

Getting Help

1. Within R - help()
   - help(command)
   - help (package = package name)
   - vignette (package = "package name"), vignette("vignette name")
   - ?? character string
2. R web site
   - http://www.r-project.org/index.html
3. Hands-on Intermediate Econometrics Using R, by H.D. Vinod
4. Professor Vinod's Notes on R Language
   - http://www.fordham.edu/economics/vinod/R-lang.doc
5. Journal of Statistical Software
   - http://www.jstatsoft.org/

Getting Help

1. Within R - help()
   - help(command)
   - help (package = package name)
   - vignette (package = "package name"), vignette("vignette name")
   - ?? character string
2. R web site
   - http://www.r-project.org/index.html
3. Hands-on Intermediate Econometrics Using R, by H.D. Vinod
4. Professor Vinod's Notes on R Language
   - http://www.fordham.edu/economics/vinod/R-lang.doc
5. Journal of Statistical Software
   - http://www.jstatsoft.org/
6. Google Searches
7. http://www.rseek.org/

Outline

Loading 3rd Party Data

- Tons of data available in R packages:
  - Packages: 'AER', 'lmtest', 'Ecdat'

Loading 3rd Party Data

- Tons of data available in R packages:
  - Packages: 'AER', 'lmtest', 'Ecdat'

R Code for Using 'lmtest' Data

```
> help(package=lmtest)
> library('lmtest')
> X=data.frame(USDistLag); attach(X)
```

Loading 3rd Party Data

- Tons of data available in R packages:
  - Packages: 'AER', 'lmtest', 'Ecdat'

R Code for Using 'lmtest' Data

```
> help(package=lmtest)
> library('lmtest')
> X=data.frame(USDistLag); attach(X)
```

- Read financial data from Yahoo and Oanda:
  - Package - 'tseries'

Loading 3rd Party Data

- Tons of data available in R packages:
  - Packages: 'AER', 'lmtest', 'Ecdat'

R Code for Using 'lmtest' Data

```
> help(package=lmtest)
> library('lmtest')
> X=data.frame(USDistLag); attach(X)
```

- Read financial data from Yahoo and Oanda:
  - Package - 'tseries'

R Code for Reading Online Financial Data

```
> library(tseries)
> get.hist.quote()
```

Loading Your Data

- Read and write from file:
  - read.table (read.csv) - Reads a file in table or .csv format.
  - write.table (write.csv) - prints its required argument x to a file.

Loading Your Data

- Read and write from file:
  - read.table (read.csv) - Reads a file in table or .csv format.
  - write.table (write.csv) - prints its required argument x to a file.

R Code for Reading/Writing or Transforming Data

```
> x = read.table("c:/data/data.csv",header=T, strip.white=TRUE)
> write.csv(x, file = "c:\\data.csv")
```

Loading Your Data

- Read and write from file:
  - read.table (read.csv) - Reads a file in table or .csv format.
  - write.table (write.csv) - prints its required argument x to a file.

R Code for Reading/Writing or Transforming Data

```
> x = read.table("c:/data/data.csv",header=T, strip.white=TRUE)
> write.csv(x, file = "c:\\data.csv")
```

- Transform data into matrix, time series, or data frame

Loading Your Data

- Read and write from file:
  - read.table (read.csv) - Reads a file in table or .csv format.
  - write.table (write.csv) - prints its required argument x to a file.

R Code for Reading/Writing or Transforming Data

```
> x = read.table("c:/data/data.csv",header=T, strip.white=TRUE)
> write.csv(x, file = "c:\\data.csv")
```

- Transform data into matrix, time series, or data frame

R Code for Reading/Writing or Transforming Data

```
> X1 = as.matrix(X)
> X2= ts(X)
> X3 = as.data.frame(X)
```

Viewing Your Data

- View the head/tail of the data
- View summary statistics
- More Basic stats

Viewing Your Data

- View the head/tail of the data
- View summary statistics
- More Basic stats

R Code for head/tail

```
> head(X,n=3); tail(X,n=3)

  consumption     gnp
1      522.67  832.57
2      550.36  876.32
3      578.81  929.40
   consumption     gnp
18      933.27 1473.97
19      942.77 1502.58
20      951.60 1475.41
```

Viewing Your Data

R Code for basicStats

```
> library(fBasics); basicStats(X); # Use xtable(basicStats(X)) for presentation below
```

	consumption	gnp
nobs	20.00	20.00
NAs	0.00	0.00
Minimum	522.67	832.57
Maximum	951.60	1502.58
1. Quartile	642.89	1046.46
3. Quartile	868.77	1386.94
Mean	747.83	1197.21
Median	752.48	1208.74
Sum	14956.59	23944.23
SE Mean	30.90	47.82
LCL Mean	683.16	1097.13
UCL Mean	812.50	1297.29
Variance	19094.74	45725.84
Stdev	138.18	213.84
Skewness	0.04	-0.04
Kurtosis	-1.40	-1.36

Outline

OLS Regression

- OLS Regression are very simple in R
- Output statistics and confidence intervals are easy to view and present

OLS Regression

- OLS Regression are very simple in R
- Output statistics and confidence intervals are easy to view and present

R Regression Output

```
> reg = lm(consumption ~ gnp)
> xtable(summary(reg)) # Creates Output Table
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-24.0889	12.3920	-1.94	0.0677
gnp	0.6448	0.0102	63.23	0.0000

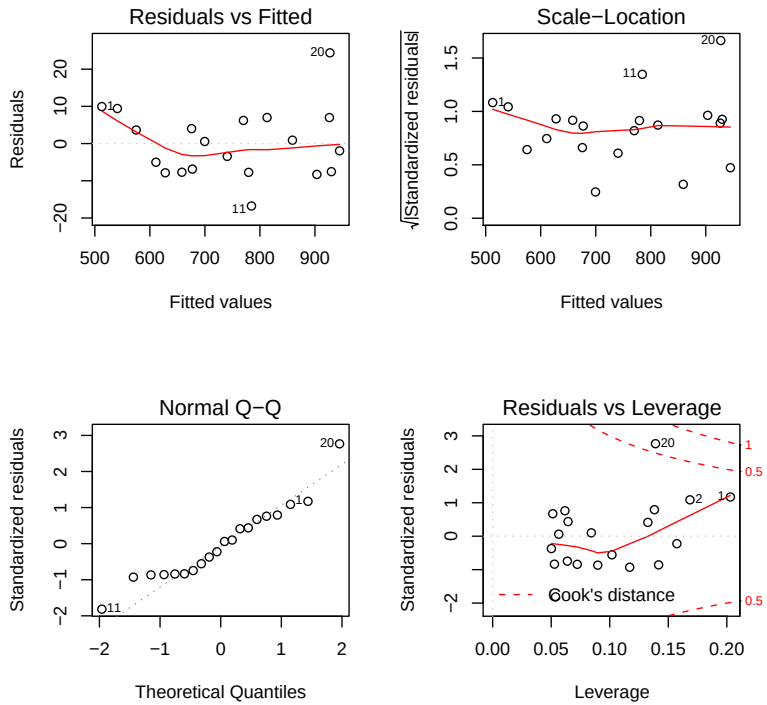
```
> xtable(confint(reg))
```

	2.5 %	97.5 %
(Intercept)	-50.12	1.95
gnp	0.62	0.67

OLS Regression Plots

Four Diagnostic Plots are Provided

```
> layout(matrix(1:4, nrow = 2, ncol = 2)) # Presents 4 Plots as 1
> plot(reg)
> layout(matrix(1:1, nrow = 1, ncol = 1)) # Sets Print Layout to Default
```



Regression Tests in R

R provides a great number of tools to deal with problems/issues that typically arise in regression analysis:

Issue	Test	Package	Command
Autocorrelation	Durbin-Watson	'car'	durbin.watson(reg)
Unit Root	Aug. Dickey-Fuller	'urca'	adf.test(gnp)
Non-causality	Granger	'MSBVAR'	granger.test(X2, p=1)
Specification	Ramsey Reset Hausman	'lmtest' 'systemfit'	resettest(reg) hausman.systemfit( )
Multicollinearity	Eigenval. Decomp. Ridge Regression	'fEcofin' 'RXshrink'	eigen( ) RXridge( )
Heteroscedasticity	Goldfeld-Quandt Breusch-Pagan-Godfrey	'lmtest'	gqtest(reg) bgtest(reg)

Other Regression Models

R allows straightforward implementation of other regression models:

Regression	Package	Command
2SLS	'systemfit'	systemfit("2SLS", ...)
3SLS	'systemfit'	systemfit("3SLS", ...)
Probit	'stats'	glm(Y X, family=binomial(link="probit"))
Logit	'stats'	glm(Y X, family=binomial(link="logit"))
Heckit	'sampleSelection'	heckit2fit(Y X, Z W)

Outline

Value at Risk

- Value at Risk Calculation from Log-Likelihood Fit of General Pareto Distribution - not too much sweat:)

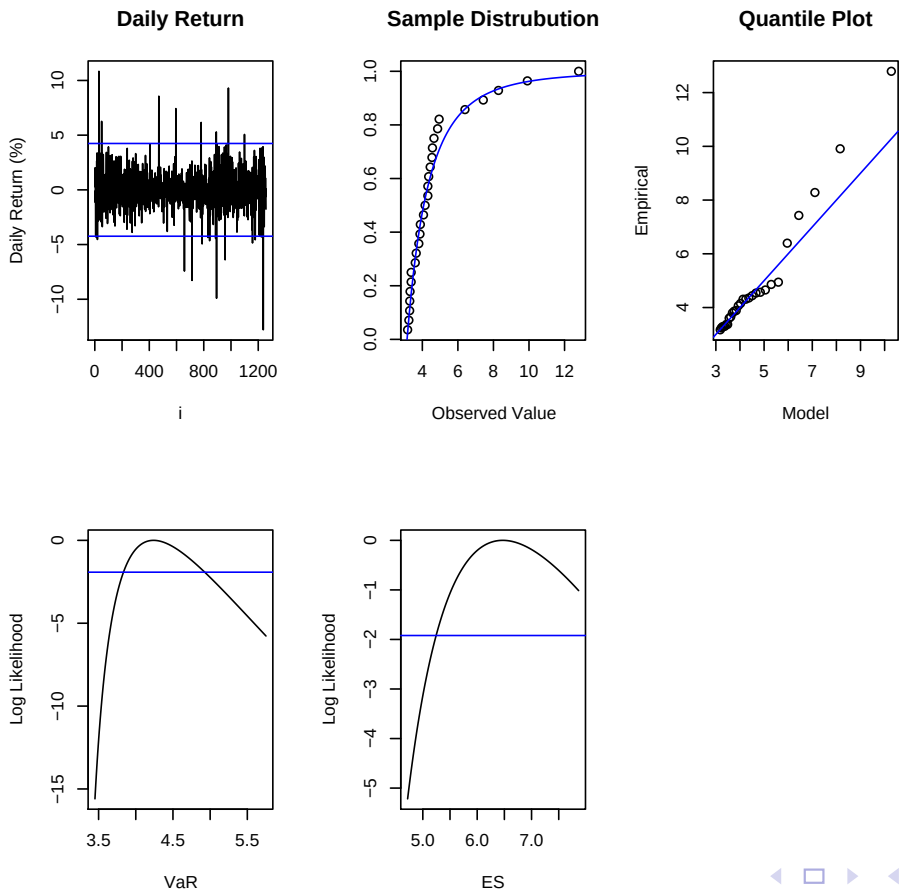
R Code for VaR

```
> rm(list = ls())
> library(tseries)
> library(VaR)
> y <- get.hist.quote(instrument = "DELL", provider = "yahoo",
+     quote = "AdjClose", start = "2003-01-01", end = "2007-12-31")
> y = data.frame(y)
> y1 = y[, 1]
> z <- VaR.gpd(y1, p = 0.01, p.tr = 0.97, drift.appx = TRUE, init = c(1,
+     0.3), cflevel = 0.95)
> z$VaR

[1] 4.239154

> z$VaR.interval

[1] 3.834567 4.927063

> z$ES

[1] 6.470641

> z$ES.interval

[1] 5.244788 8.554813
```

Diagnostic Plots

R Code for VaR plots

```
> par(mfrow = c(3, 2))
> VaR.gpd.plots(z)
> par(mfrow = c(1, 1))
```



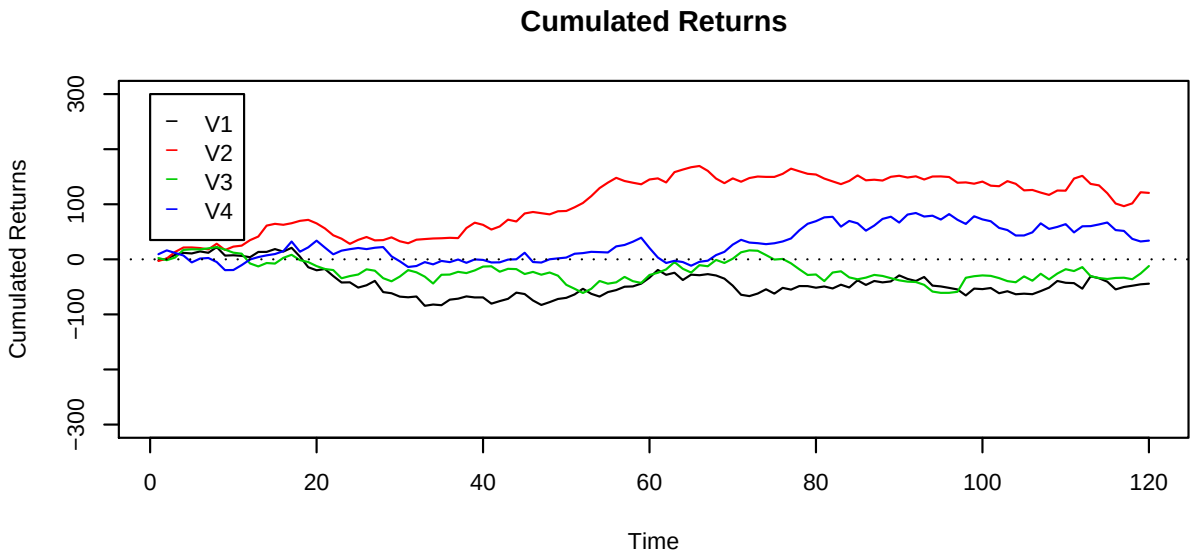
Value at Risk

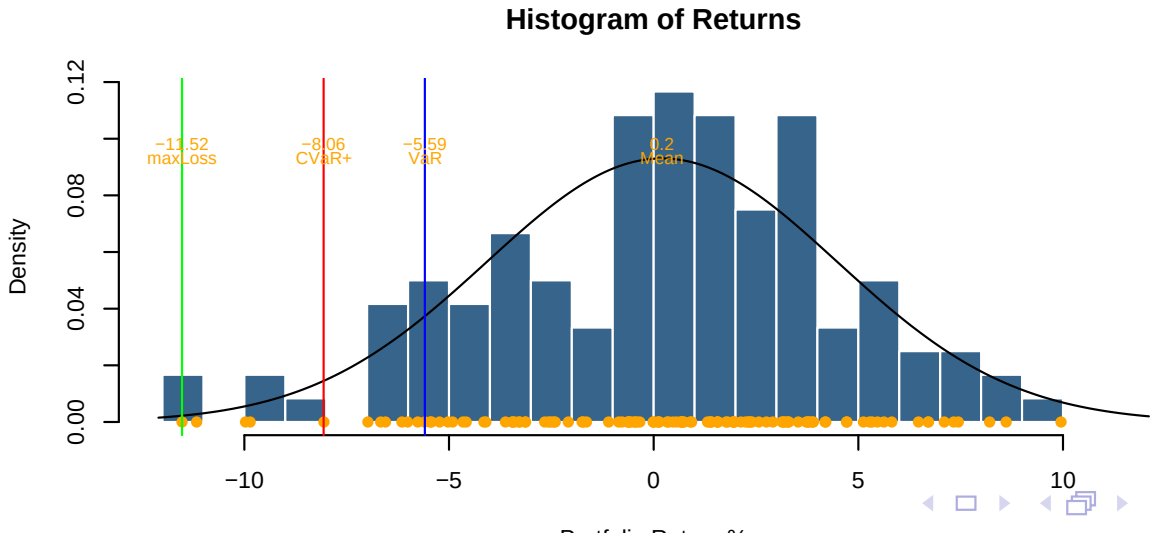
- Value at Risk Calculation from Normal distribution and some fancy plots

R Code for VaR

- fAssets package
  - pfolioVaR
  - pfolioCVaRplus
  - pfolioCVaR
  - pfolioMaxLoss

Value at Risk





Other Related Functions

R allows easy implementation many other measures and models:

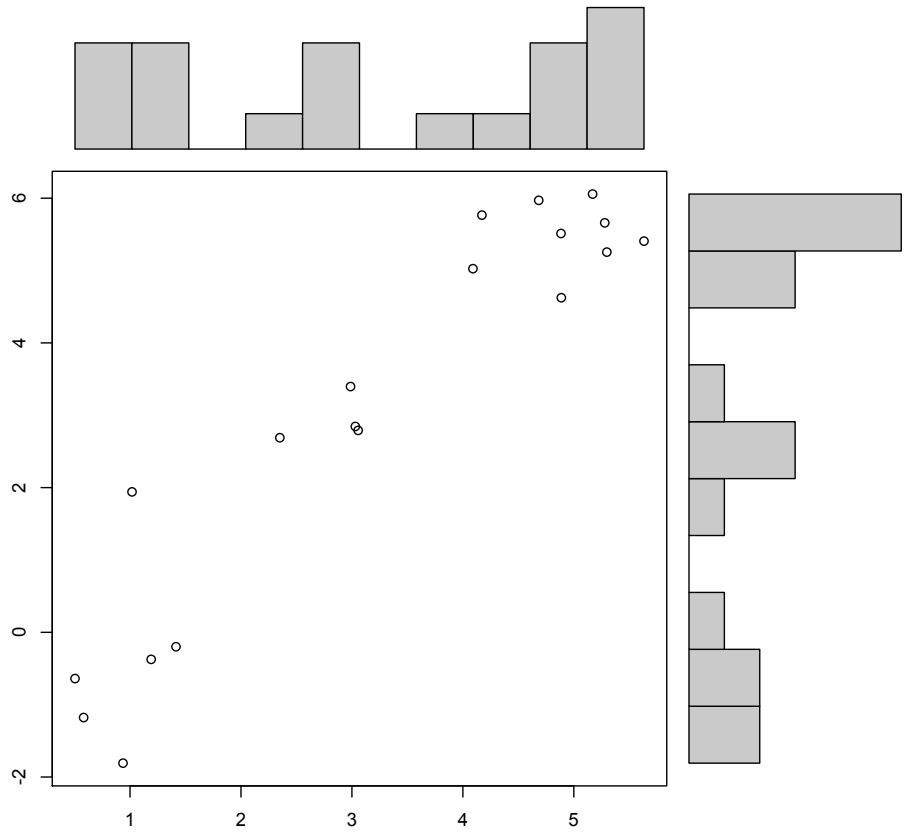
Issue	Package	Command
Threshold	"fExtremes"	findThreshold(...)
Multivariate normal,t dist.	"fAssets"	assetsFit(...)
Generalized Pareto dist.	"fExtremes"	dgpd(...)
Sharpe Ratio	"fTrading"	sharpeRatio (...)
Efficient Portfolio	"fPortfolio"	efficientPortfolio(...)

Outline

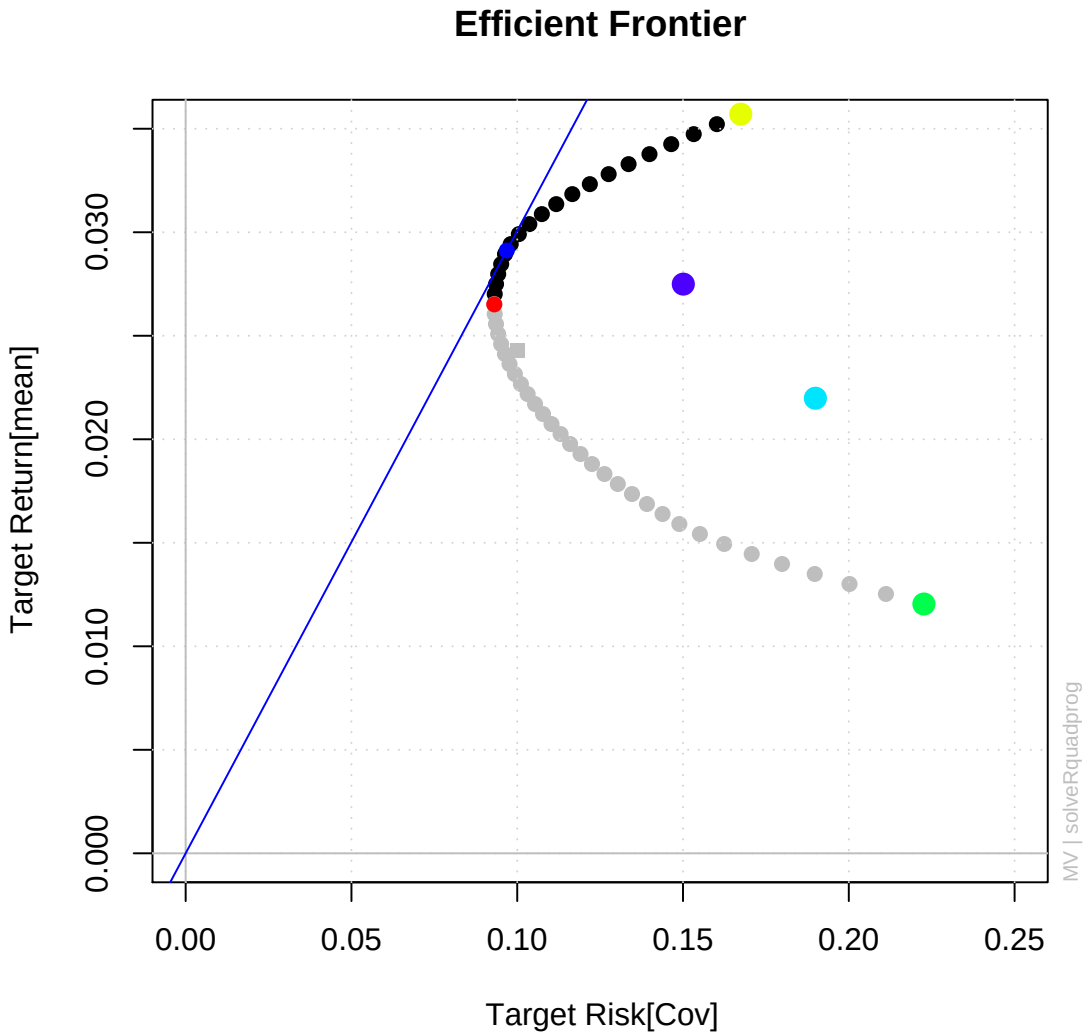
More Graphs

GNP Growth on Consumption Growth

```
> help(layout) # Scatterplot with marginal histograms
```



More Graphs


Efficient Frontier
Target Return[mean]
0.000
0.010
0.020
0.030
0.00
0.05
0.10
0.15
0.20
0.25
Target Risk[Cov]
MV | solveRquadprog