Descriptive Statistics for Financial Time Series

Econ 424/Amath 540
Summer 2012
Eric Zivot
Updated: July 10, 2012

Data for Examples

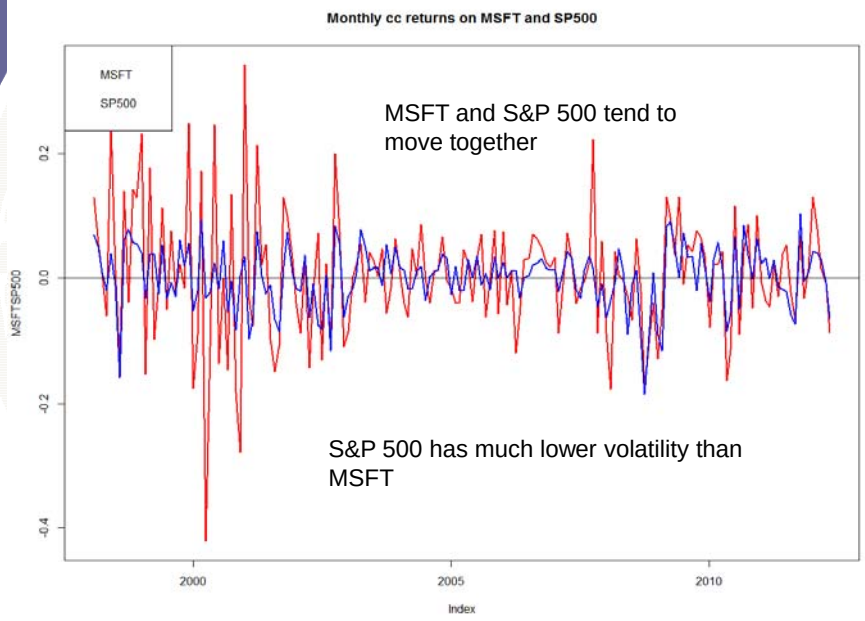
```
# Load libraries
> library(tseries)
> library(PerformanceAnalytics)

# Get adjusted closing price data from Yahoo!
> MSFT.prices = get.hist.quote(instrument="msft", start="1998-01-01",
+                              end="2012-05-31", quote="AdjClose",
+                              provider="yahoo", origin="1970-01-01",
+                              compression="m", retclass="zoo")

> SP500.prices = get.hist.quote(instrument="^gspc", start="1998-01-01",
+                               end="2012-05-31", quote="AdjClose",
+                               provider="yahoo", origin="1970-01-01",
+                               compression="m", retclass="zoo")
```

Price Data

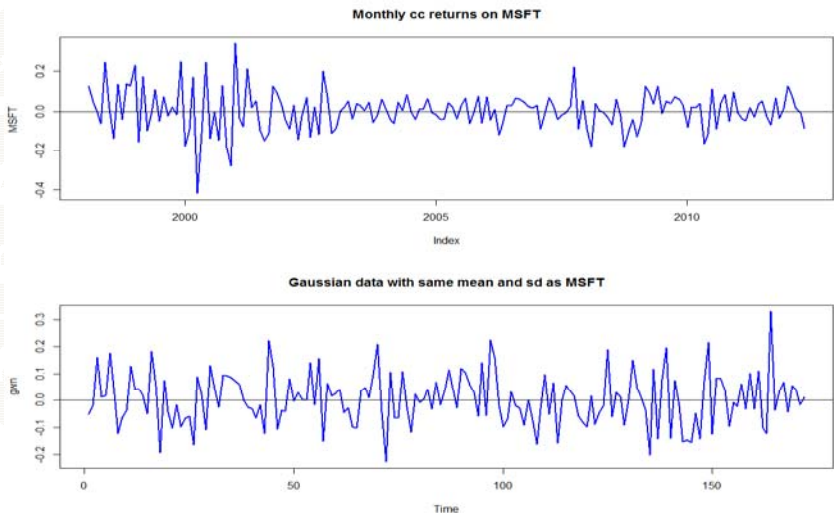

```
plot(MSFTSP500.prices, main="Adjusted Closing Prices",
     lwd=2, col="blue")
```


Monthly Continuously Compounded Returns

Q: What common features do you see?

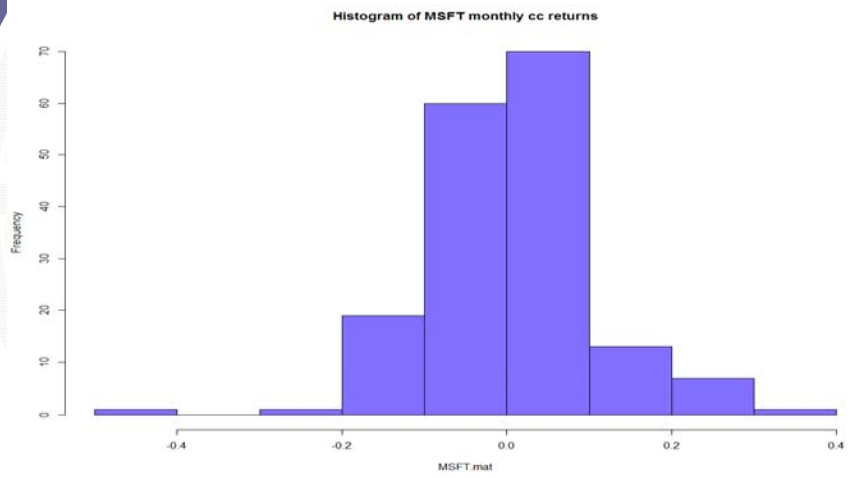


Are Monthly Returns Gaussian White Noise?

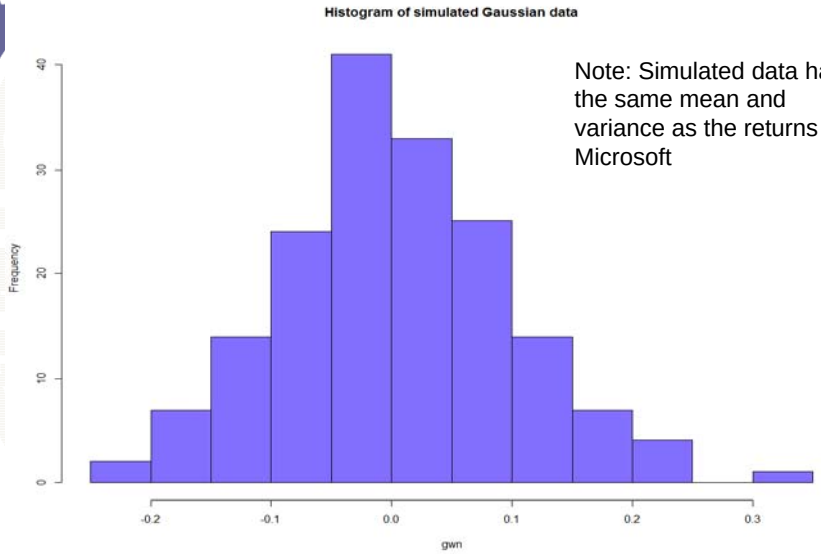



```
> set.seed(123)
> gwn = rnorm(length(MSFT),mean=mean(MSFT),sd=std(MSFT))
```

Estimating the pdf: Histogram

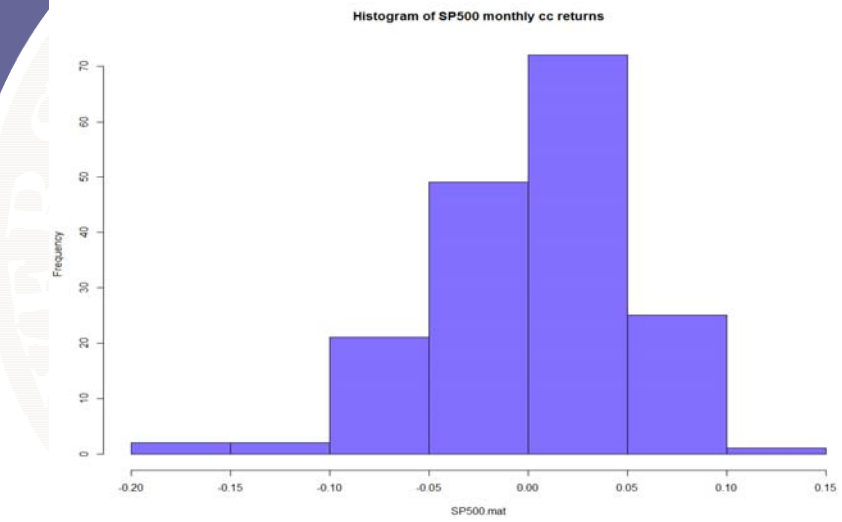



```
> hist(MSFT,main="Histogram of MSFT monthly cc returns“,
+      col=“slateblue1”)
```

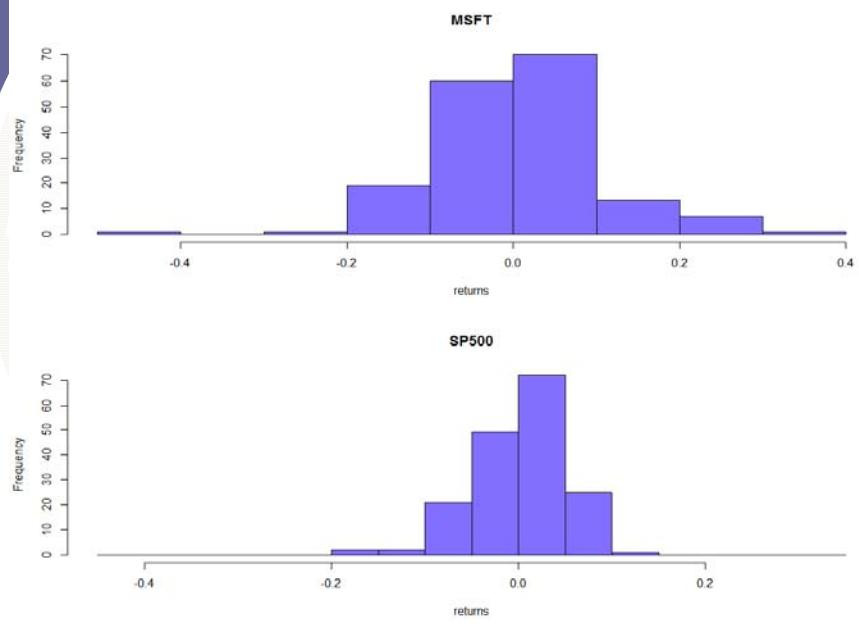
Note: Simulated data has the same mean and variance as the returns on Microsoft

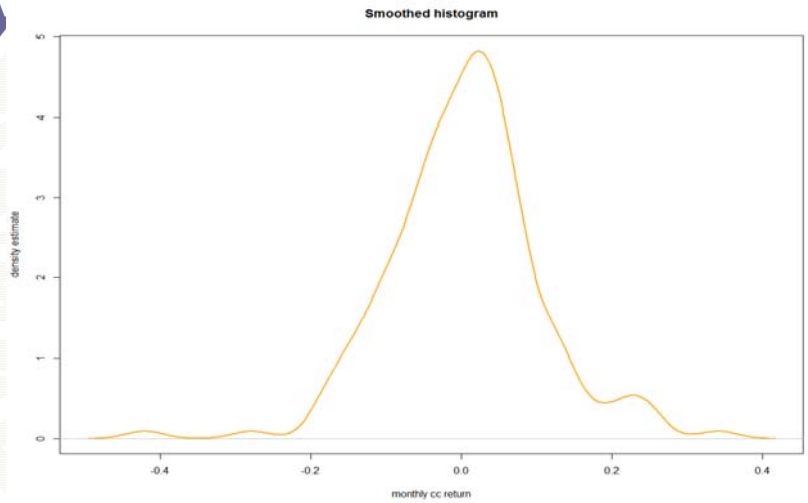
```
> hist(gwn,main="Histogram of simulated Gaussian data“,
+      col=“slateblue1”)
```



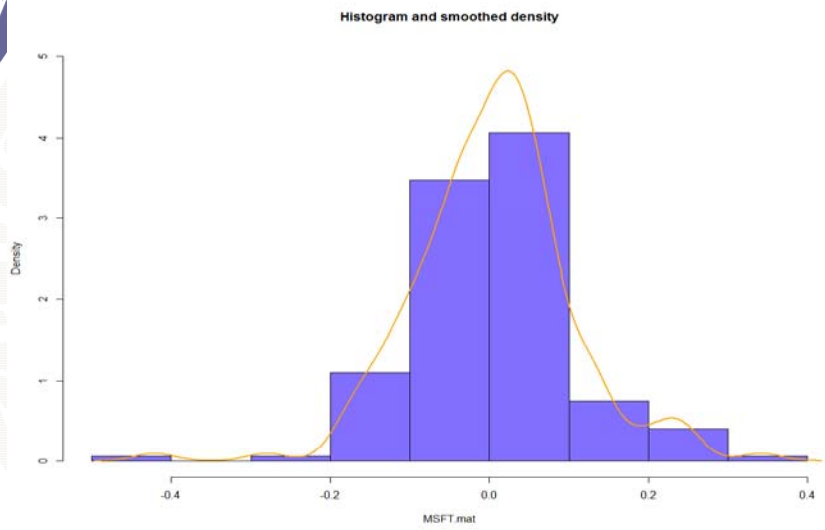

```
> hist(SP500,main="Histogram of SP500 monthly cc returns“,
+       col=“slateblue1”)
```


Note: MSFT has larger SD (volatility) than S&P 500






```
> MSFT.density = density(MSFT)
> plot(MSFT.density,type="l",xlab="monthly return",
+      ylab="density estimate",main="Smoothed histogram for MSFT
+      monthly cc returns", col="orange", lwd=2)
```


```
> hist(MSFT,main="Histogram and smoothed density of MSFT
+      monthly returns", probability=T, col="slateblue1",
+      ylim=c(0,5))
> points(MSFT.density, type="l", col="orange", lwd=2)
```

Computing quantiles

```
> quantile(MSFT.mat)
       0%       25%       50%       75%      100%
-0.420864 -0.050131  0.008387  0.055450  0.341592

# 1% and 5% empirical quantiles
> quantile(MSFT.mat,probs=c(0.01,0.05))
     1%      5%
-0.2106 -0.1474
# compare to 1% and 5% normal quantiles
> qnorm(p=c(0.01,0.05), mean=mean(MSFT.mat), sd=sd(MSFT.mat))
[1] -0.2290 -0.1607

# SP500 empirical and normal quantiles
> quantile(SP500.mat,probs=c(0.01,0.05))
      1%       5%
-0.12846 -0.08538
> qnorm(p=c(0.01,0.05), mean=mean(SP500.mat), sd=sd(SP500.mat))
[1] -0.11107 -0.07804
```

1% and 5% quantiles are used for Value-at-Risk calculations

Monthly VaR Using Empirical Quantiles

```
> q.01 = quantile(MSFT.mat, probs=0.01)
> q.05 = quantile(MSFT.mat, probs=0.05)
> q.01
     1%
-0.2106
> q.05
     5%
-0.1474

# Monthly VaR on $100,000 investment
> VaR.01 = 100000*(exp(q.01) - 1)
> VaR.05 = 100000*(exp(q.05) - 1)
> VaR.01
     1%
-18992

> VaR.05
     5%
-13708
```

Summary Statistics

```
> mean(MSFT.mat)
    MSFT
0.004124

> var(MSFT.mat)
         MSFT
MSFT 0.01004

> sd(MSFT.mat)
  MSFT
0.1002

> skewness(MSFT.mat)
[1] -0.09033

> kurtosis(MSFT.mat)
[1] 2.023
```

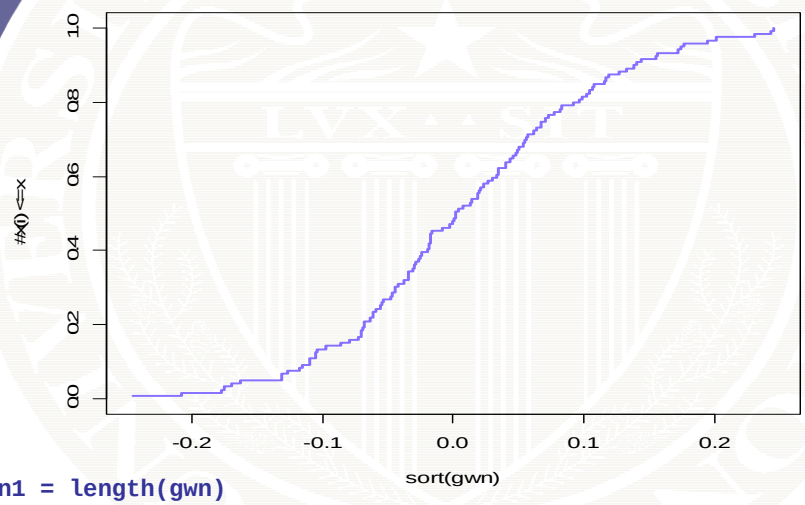
`Skewness()` function is in package PerformanceAnalytics

`kurtosis()` function is in package PerformanceAnalytics and computes *excess kurtosis*

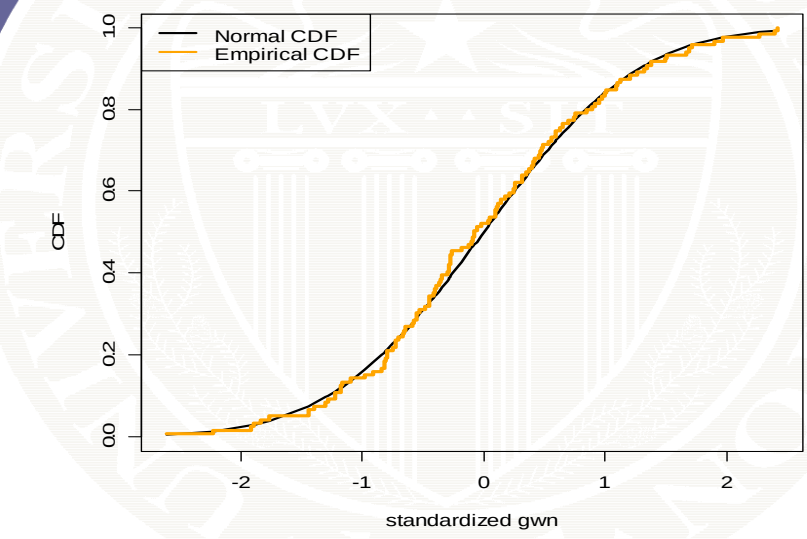
Summary Statistics by Column

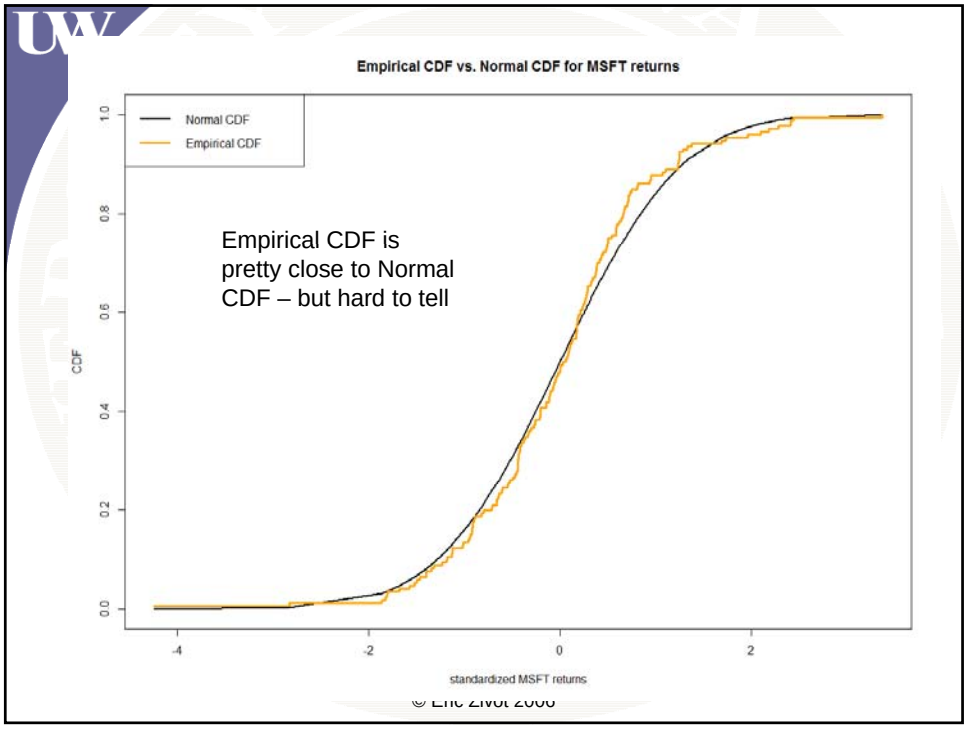
```
> apply(MSFTSP500.mat, 2, mean)
    MSFT    SP500
0.004124 0.001687

> apply(MSFTSP500.mat, 2, sd)
   MSFT   SP500
0.10022 0.04847

> apply(MSFTSP500.mat, 2, skewness)
    MSFT    SP500
-0.09033 -0.73344

> apply(MSFTSP500.mat, 2, kurtosis)
 MSFT SP500
2.023 1.021
```

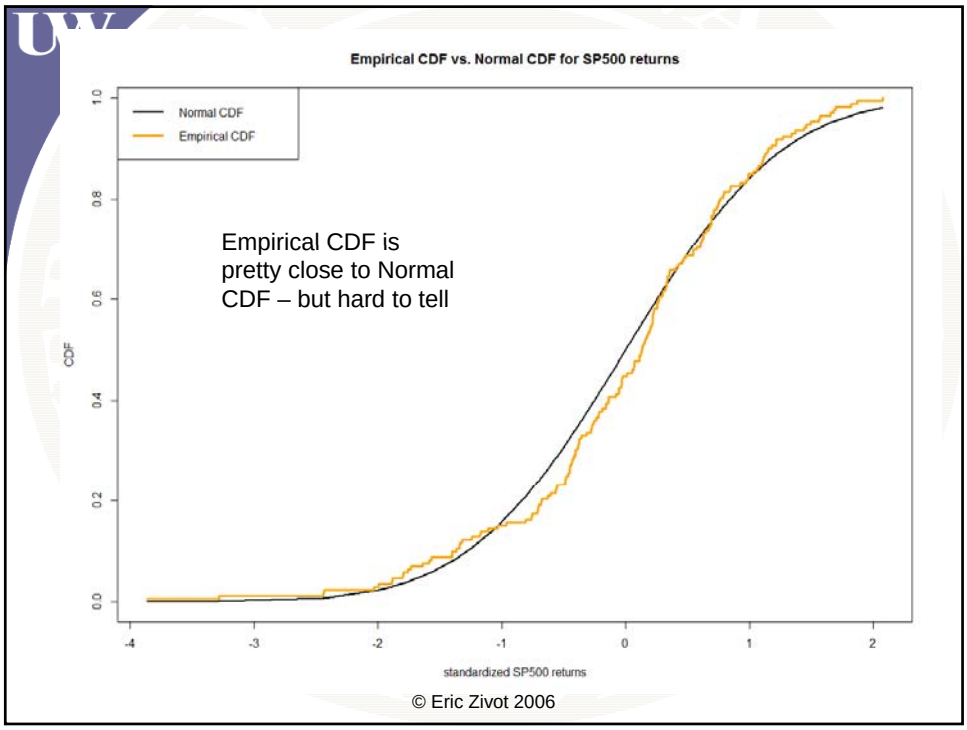
Note: MSFT has a higher mean and higher SD than SP500



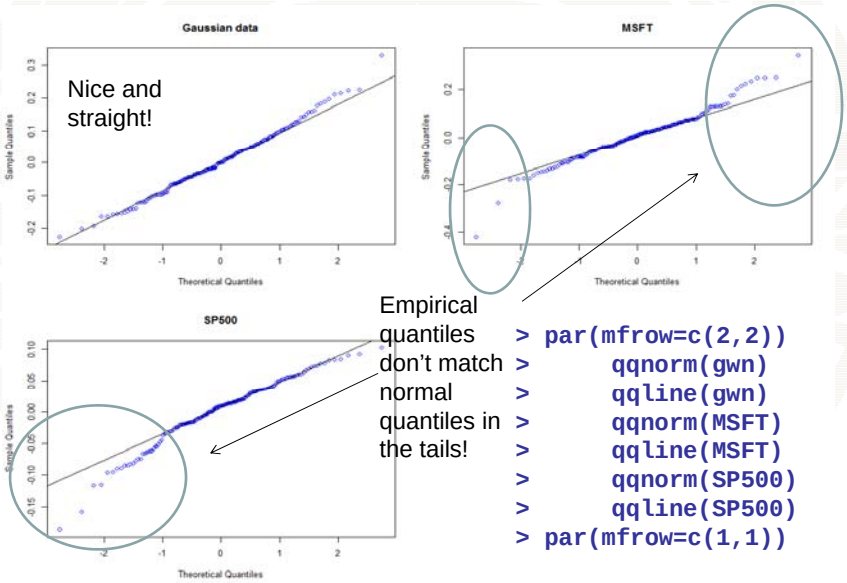

```
> n1 = length(gwn)
> plot(sort(gwn),(1:n1)/n1,type="s",ylim=c(0,1), col="slateblue1"
+ main="Empirical CDF of Gaussian data", ylab="#x(i) <= x")
```





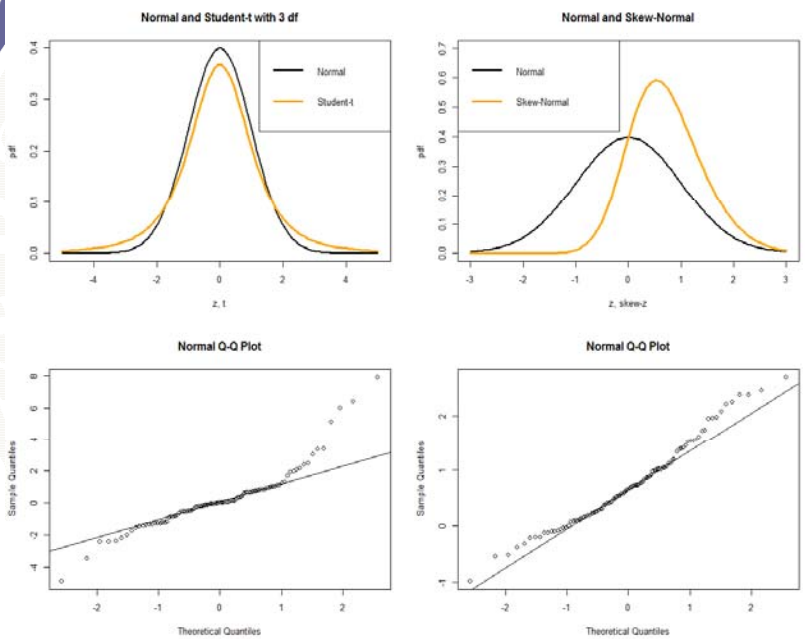
QQ-plots against Normal Distribution



Nice and straight!

Empirical quantiles don't match normal quantiles in the tails!

```
> par(mfrow=c(2,2))
>     qqnorm(gwn)
>     qqline(gwn)
>     qqnorm(MSFT)
>     qqline(MSFT)
>     qqnorm(SP500)
>     qqline(SP500)
> par(mfrow=c(1,1))
```

Normal QQ-Plot Shapes

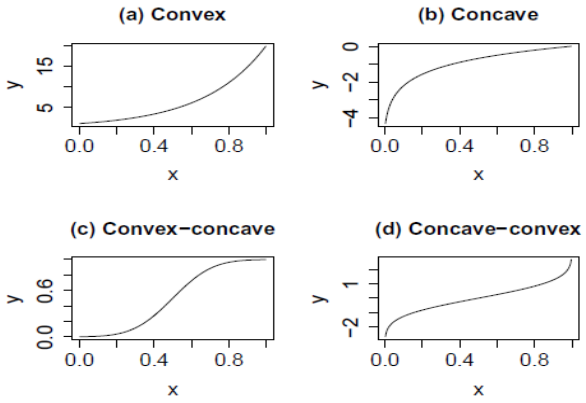
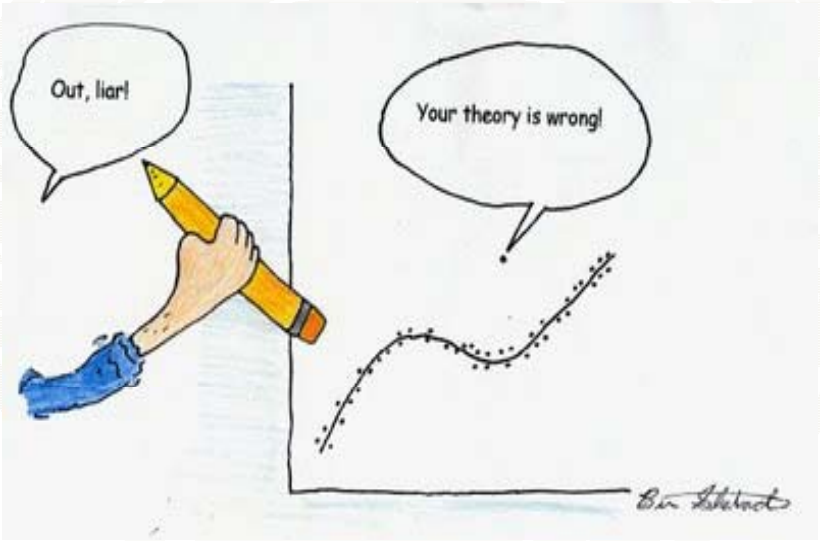
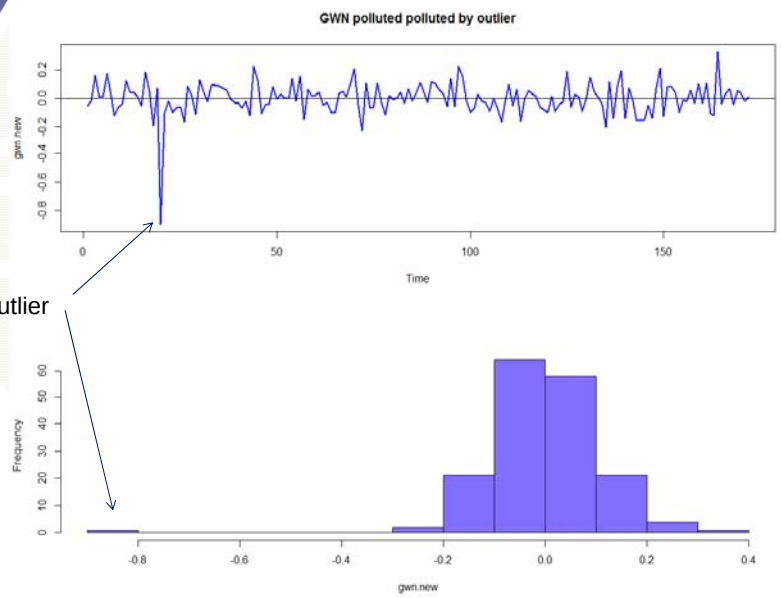


Fig. 4.9. *As one moves from (a) to (d), the curves are convex, concave, convex-concave, and concave-convex. Normal plots with these patterns indicate left skewness, right skewness, heavier tails than a normal distribution, and lighter tails than a normal distribution, respectively, assuming that the data are on the x-axis and the normal quantiles on the y-axis, as will always be the case in this textbook.*



Effect of Outliers on Descriptive Statistics

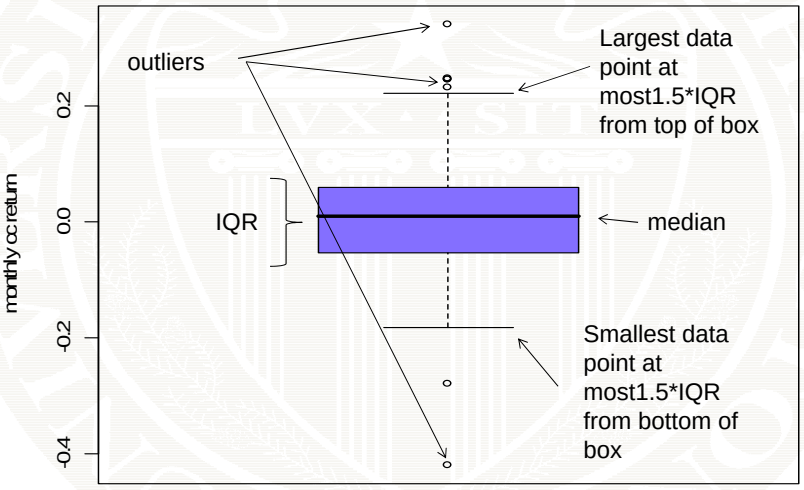


Summary statistics of polluted data

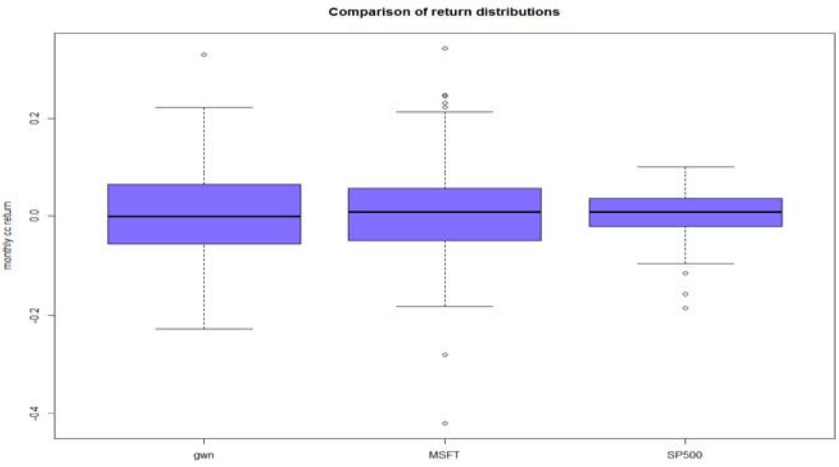
```
> tmp = cbind(gwn, gwn.new)
> apply(tmp, 2, mean)
       gwn    gwn.new
 0.0043420 -0.0006391

> apply(tmp, 2, sd)
    gwn gwn.new
0.09515 0.11746

> apply(tmp, 2, skewness)
    gwn gwn.new
 0.2842 -2.3751

> apply(tmp, 2, kurtosis)
    gwn gwn.new
 0.1557 18.3707
```

Notice how sample statistics are influenced by the single outlier

```
# outlier robust measures
> apply(tmp, 2, median)
       gwn    gwn.new
-0.0009163 -0.0009163

> apply(tmp, 2, IQR)
    gwn gwn.new
 0.1200  0.1219
```



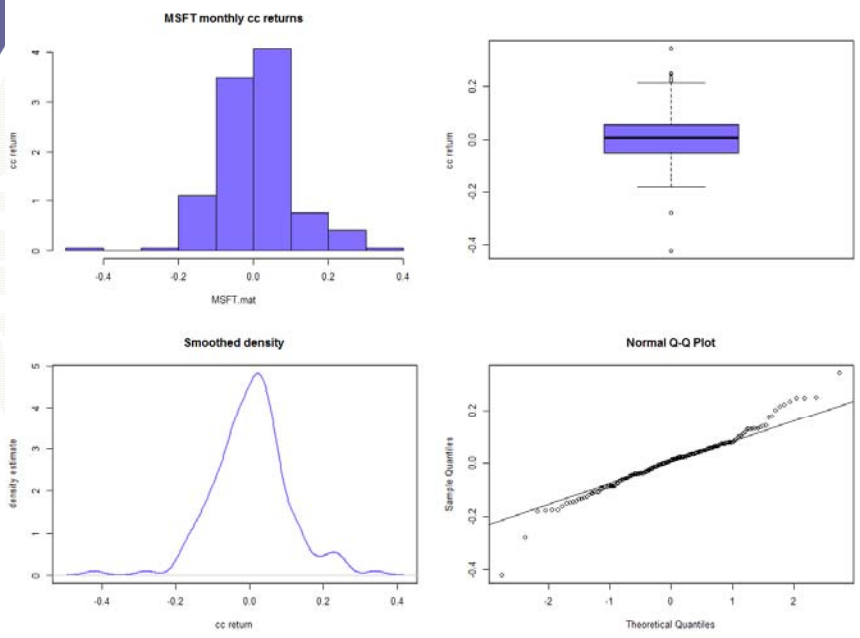

```
> boxplot(MSFT,outchar=T,main="Boxplot of monthly cc
+ returns on Microsoft",ylab="monthly cc return")
```

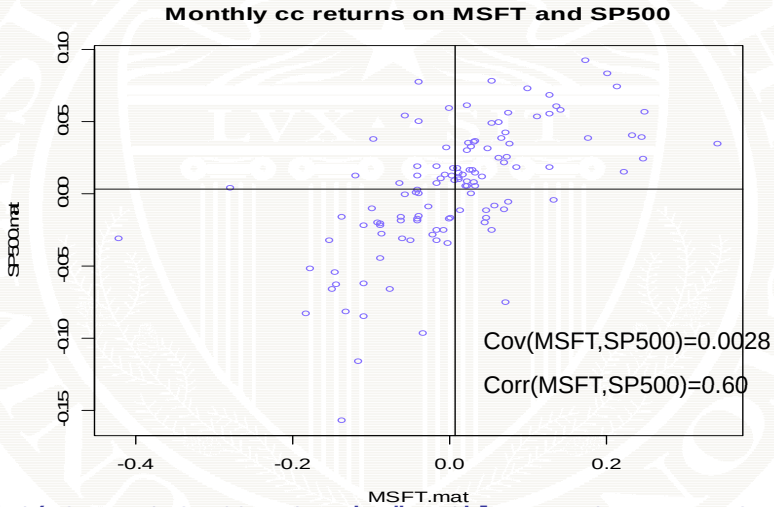

```
> boxplot(gwn,MSFT,SP500,names=c("gwn","MSFT","SP500"),outchar=T,
+ main="Comparison of return distributions", ylab="monthly cc
+ return")
```

Four Graph Summary

```
par(mfrow=c(2,2))
# plot 1
    hist(MSFT.mat,main="MSFT monthly cc returns",
         probability=T, ylab="cc return",
         col="slateblue1")
# plot 2
    boxplot(MSFT.mat,outchar=T, ylab="cc return",
            col="slateblue1")
# plot 3
    plot(MSFT.density,type="l",xlab="cc return",
         col="slateblue1", lwd=2,
         ylab="density estimate", main="Smoothed
         density")
# plot 4
    qqnorm(MSFT.mat)
    qqline(MSFT.mat)
par(mfrow=c(1,1))
```

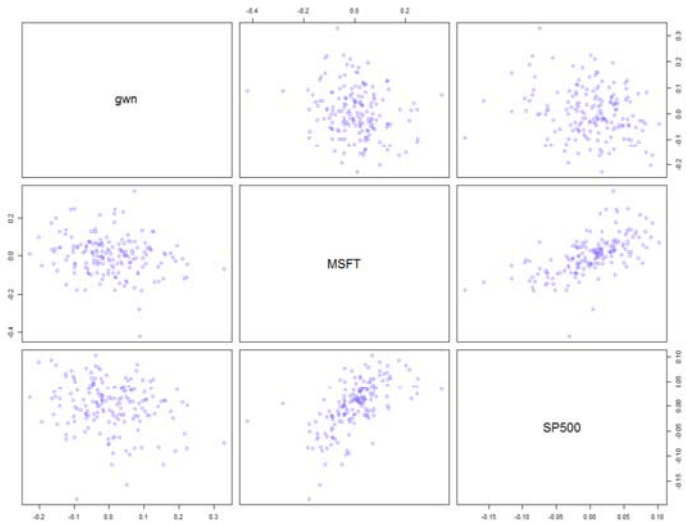



Scatterplot




```
> plot(MSFT.mat,SP500.mat,main="Monthly cc returns on MSFT
+      and SP500")
> abline(h=mean(SP500))    # horizontal line at SP500 mean
> abline(v=mean(MSFT))     # vertical line at MSFT mean
```


Pairwise Scatterplots

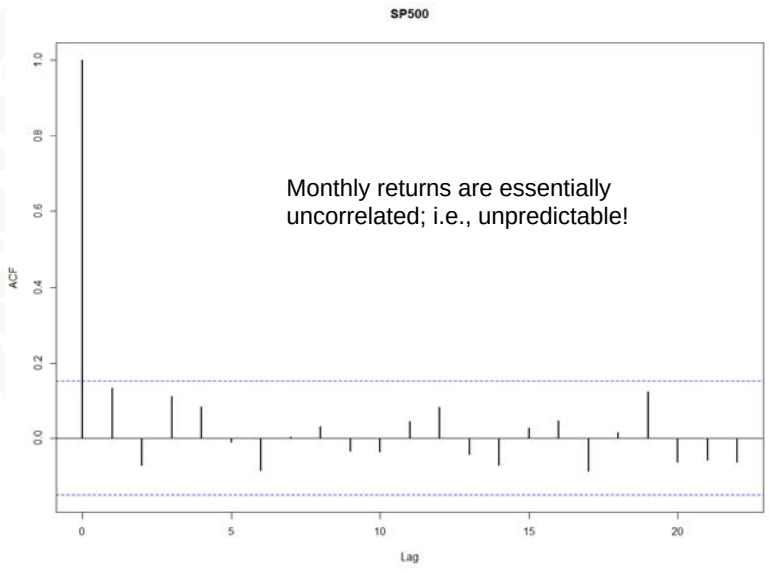



```
> pairs(cbind(gwn,MSFT,SP500))
```

Sample Covariances and Correlations

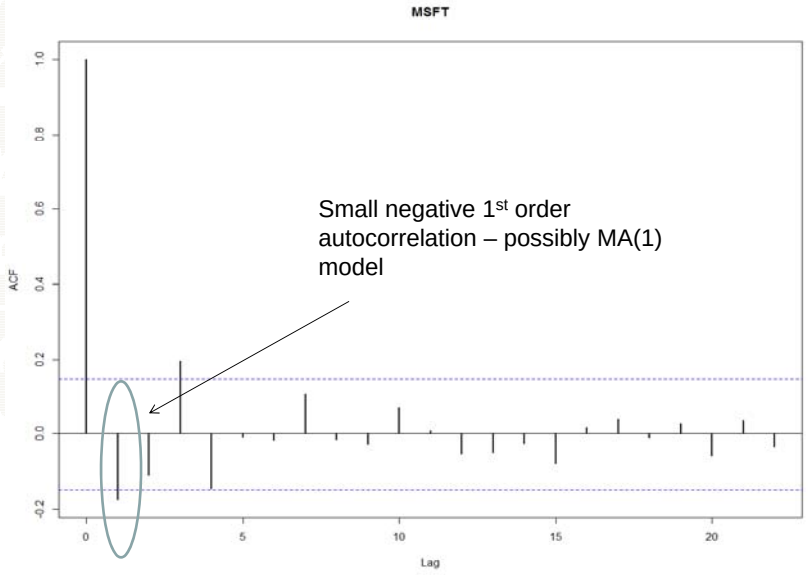
```
> var(cbind(gwn,MSFT.mat,SP500.mat))
              gwn      MSFT      SP500
gwn     0.0090534 -0.001856 -0.0009685
MSFT   -0.0018563  0.010044  0.0029993
SP500  -0.0009685  0.002999  0.0023494

> cor(cbind(gwn,MSFT.mat,SP500.mat))
            gwn    MSFT   SP500
gwn     1.0000 -0.1947 -0.2100
MSFT   -0.1947  1.0000  0.6174
SP500  -0.2100  0.6174  1.0000
```


Sample ACF for S&P 500



Monthly returns are essentially uncorrelated; i.e., unpredictable!

Sample ACF for MSFT



Stylized Facts for Monthly CC Returns

- Returns appear to be approximately normally distributed
  - Some noticeable negative skewness and excess kurtosis
- Individual asset returns have higher SD than diversified portfolios
- Many assets are contemporaneously correlated
- Assets are approximately uncorrelated over time (no serial correlation)