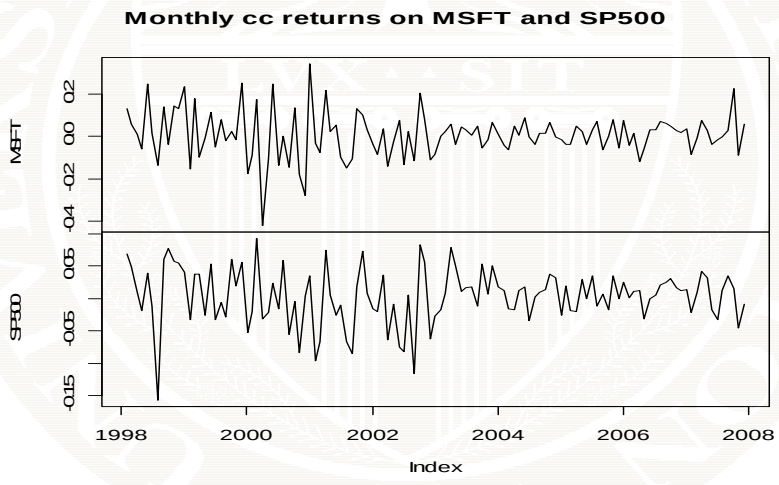
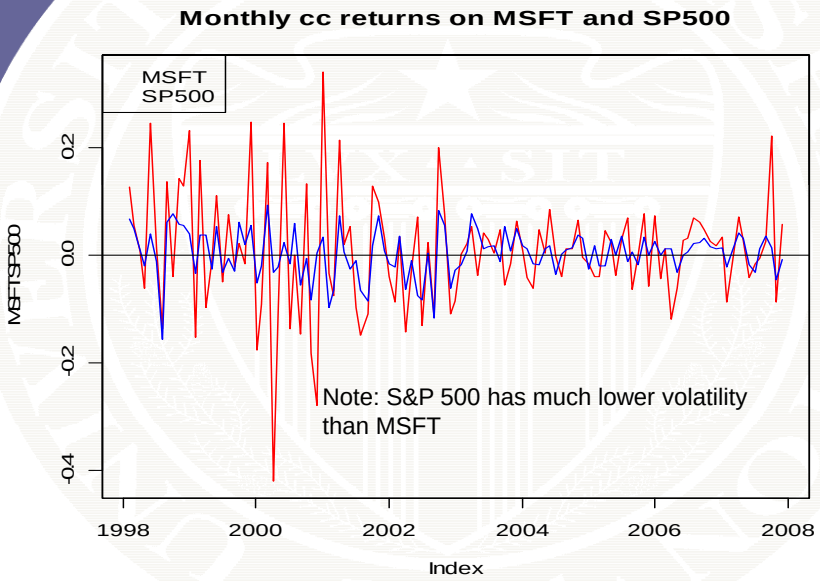
Descriptive Statistics

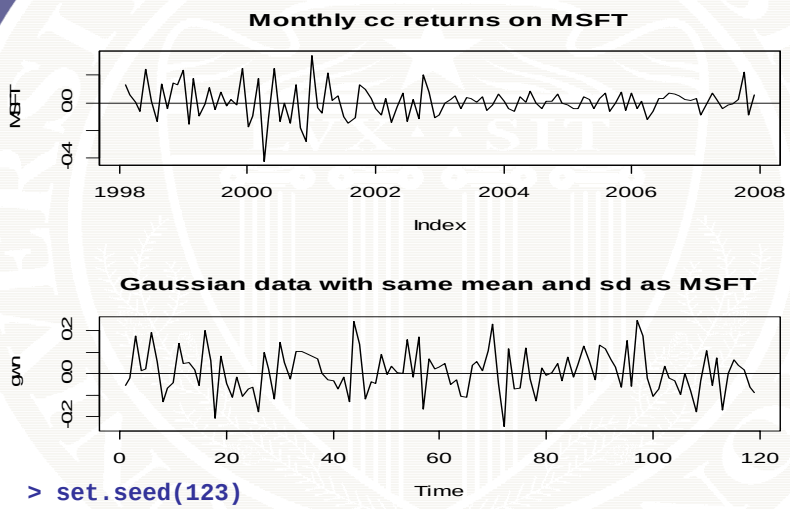
Econ 424
Fall 2008
Eric Zivot
Update: October 12, 2008

Monthly Continuously Compounded Returns

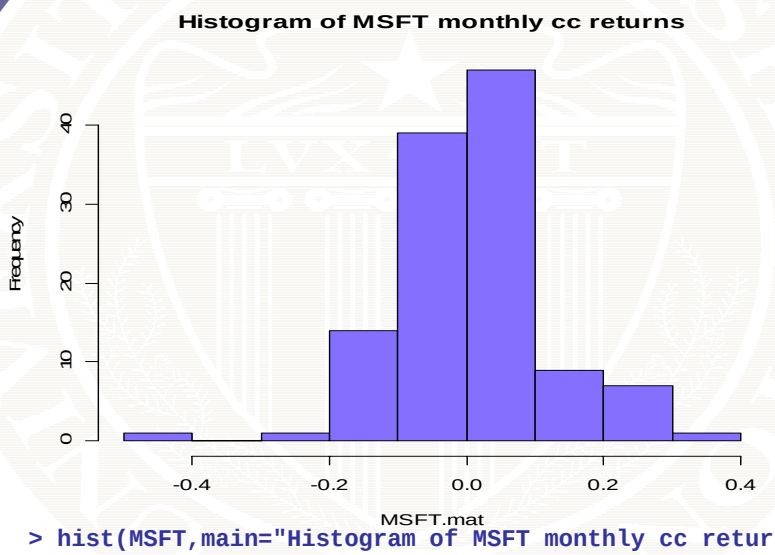




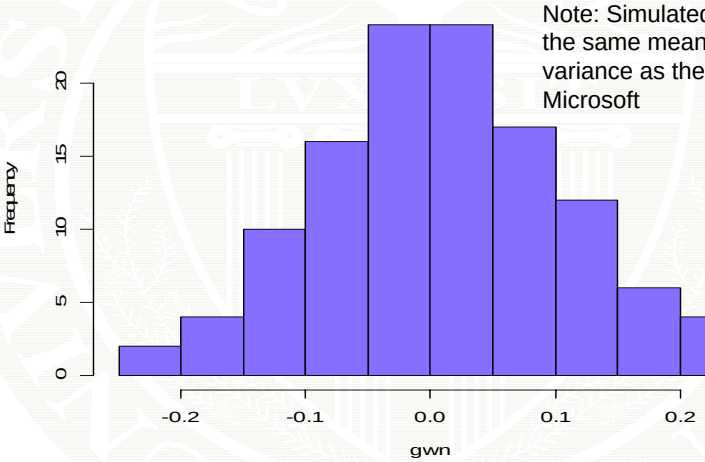
Are Monthly Returns Normally Distributed?



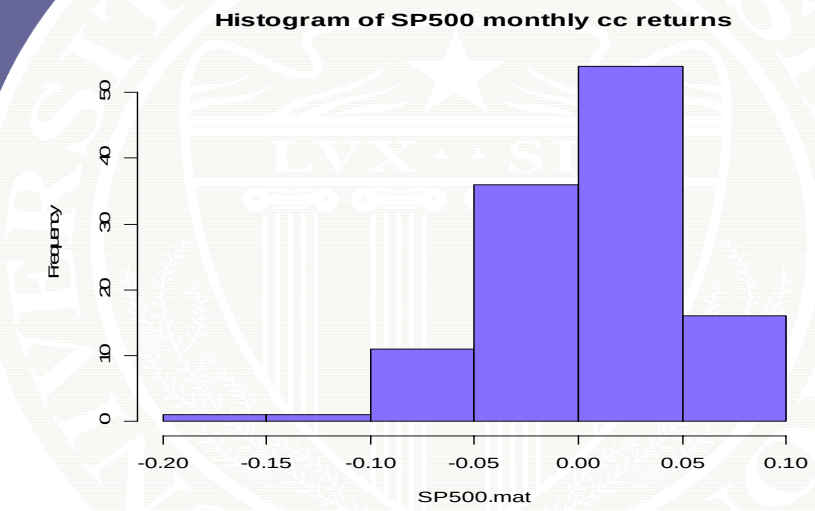

```
> set.seed(123)
> gwn = rnorm(length(MSFT),mean=mean(MSFT),sd=std(MSFT))
```




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

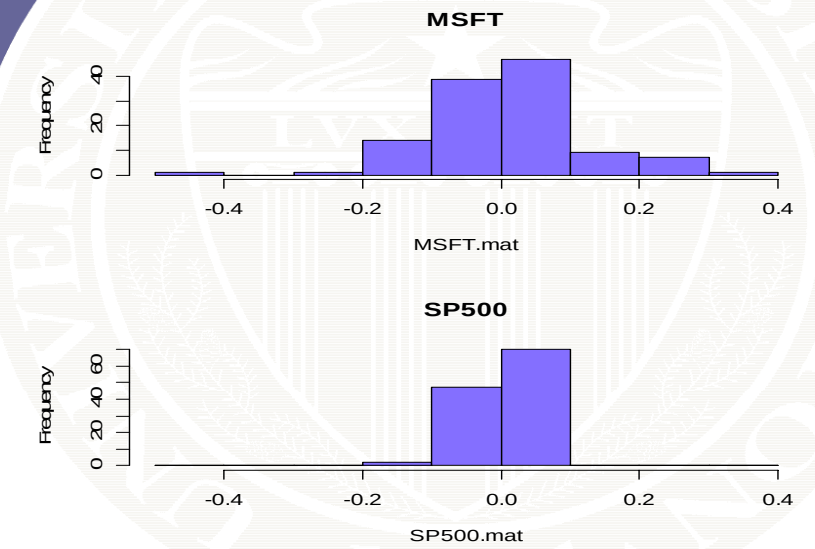



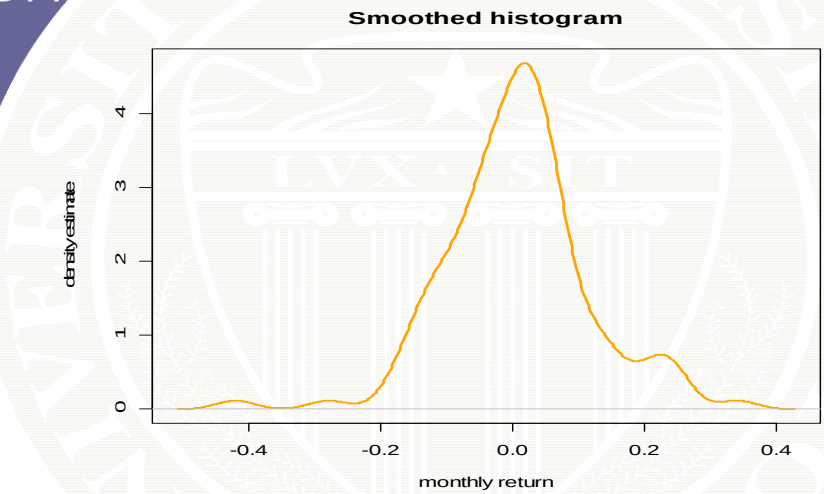

```
> hist(gwn,main="Histogram of simulated Gaussian data“,
+      col=“slateblue1”)
```



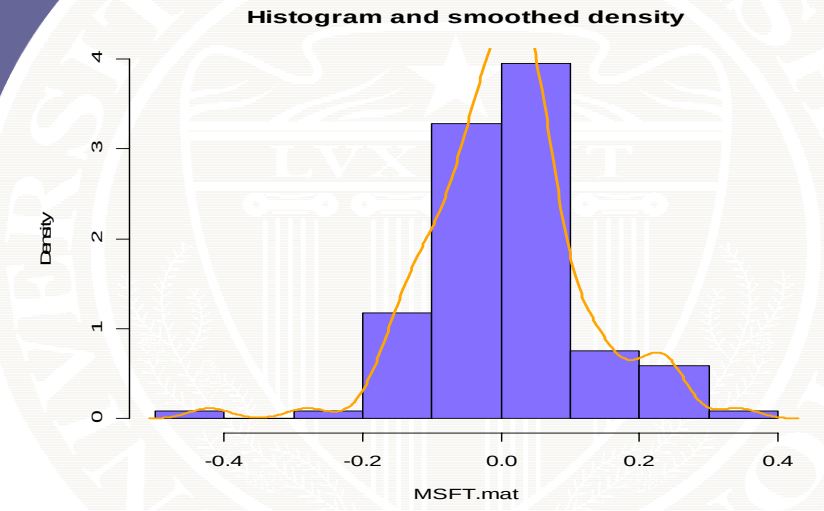

```
> hist(SP500,main="Histogram of SP500 monthly cc returns",
+       col="slateblue1")
```


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")
> points(MSFT.density, type="l", col="orange", lwd=2)
```

Computing quantiles

```
> quantile(MSFT.mat)
          0%          25%          50%          75%         100%
-0.420894211 -0.052845790  0.009839652  0.059997232  0.341619651

> qnorm(p=c(0.01,0.05), mean=mean(MSFT.mat), sd=sd(MSFT.mat))
[1] -0.2467863 -0.1725060

> quantile(MSFT.mat,probs=c(0.01,0.05))
        1%         5%
-0.2620067 -0.1464874

> qnorm(p=c(0.01,0.05), mean=mean(SP500.mat), sd=sd(SP500.mat))
[1] -0.09669998 -0.06737741

> quantile(SP500.mat,probs=c(0.01,0.05))
        1%         5%
-0.1130098 -0.0759229
```

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

Summary Statistics

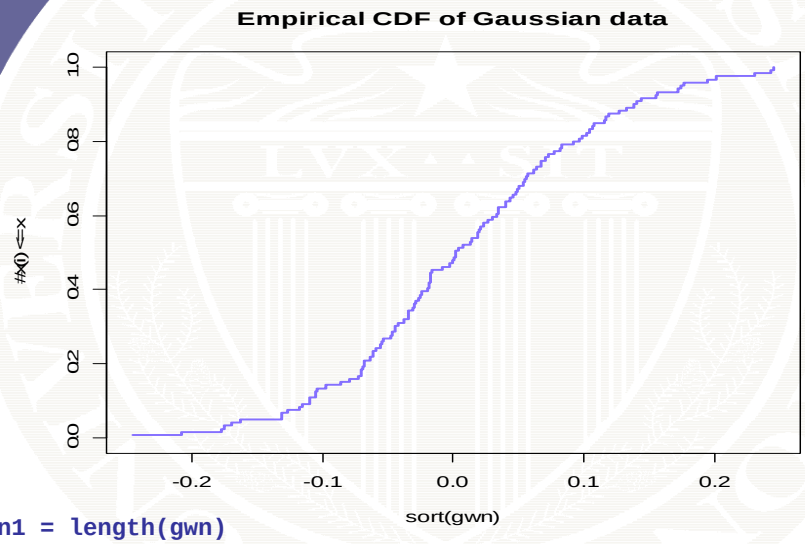
```
> mean(MSFT.mat)
[1] 0.006777013

> var(MSFT.mat)
             MSFT
MSFT 0.01188019

> sd(MSFT.mat)
      MSFT
0.1089963

> skewness(MSFT.mat)
[1] -0.0800148

> kurtosis(MSFT.mat)
[1] 1.970088
```

Skewness() function is in package TSA

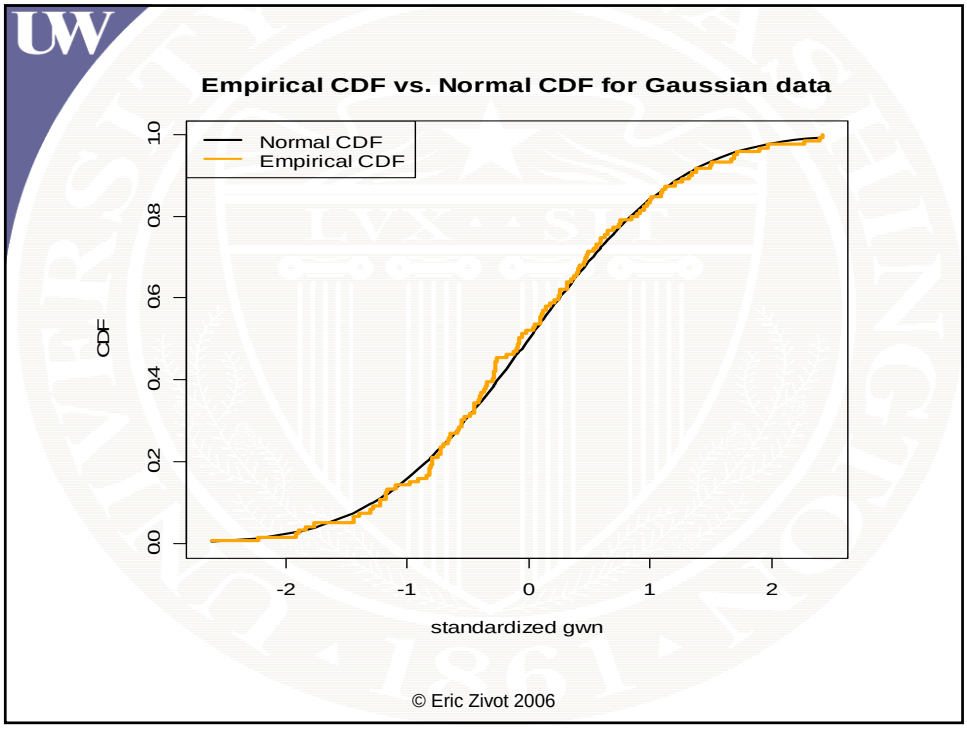
kurtosis() function is in package TSA and computes excess kurtosis

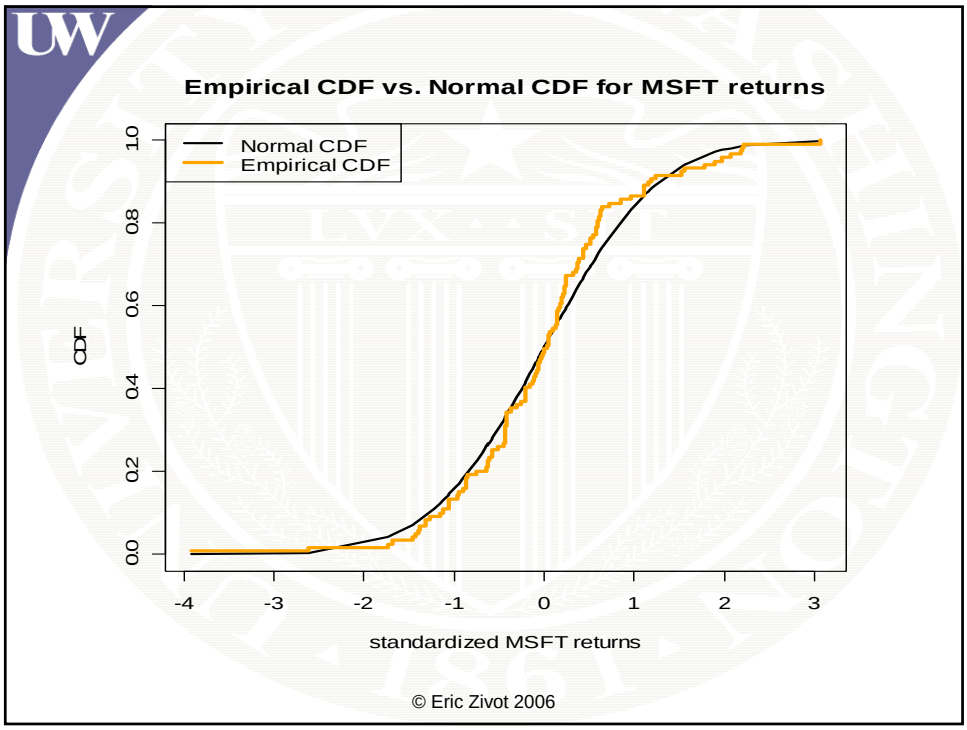
Summary Statistics

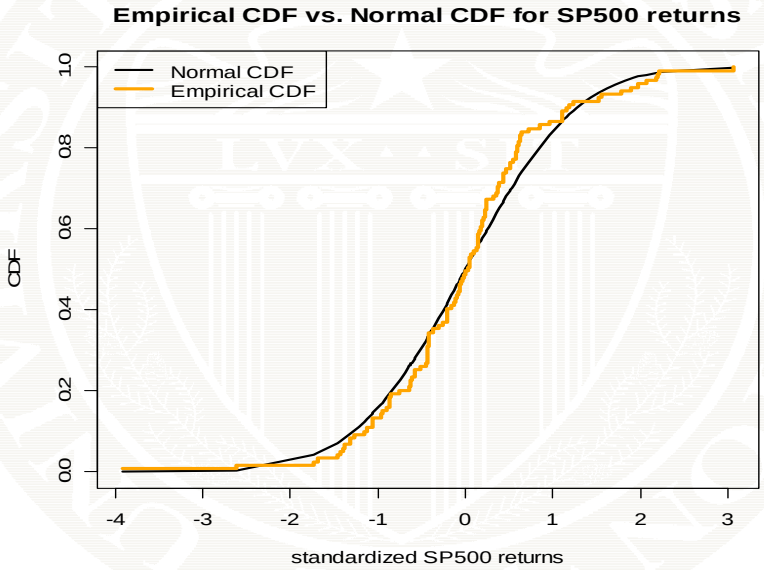
```
> apply(MSFTSP500.mat, 2, mean)
      MSFT      SP500
0.006777013 0.003395489

> apply(MSFTSP500.mat, 2, sd)
      MSFT      SP500
0.10899630 0.04302687

> apply(MSFTSP500.mat, 2, skewness)
      MSFT      SP500
-0.0800148 -0.6922284

> apply(MSFTSP500.mat, 2, kurtosis)
    MSFT    SP500
1.970088 1.126041
```

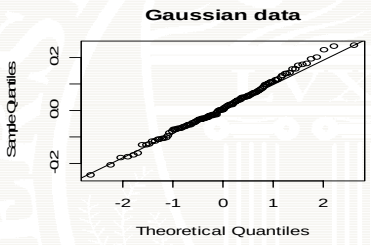

```
> n1 = length(gwn)
> plot(sort(gwn),(1:n1)/n1,type="s",ylim=c(0,1), col="slateblue1"
+ main="Empirical CDF of Gaussian data", ylab="#x(i) <= x")
```

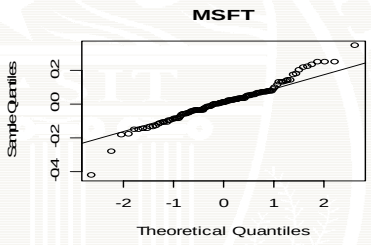


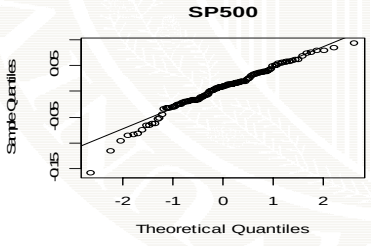




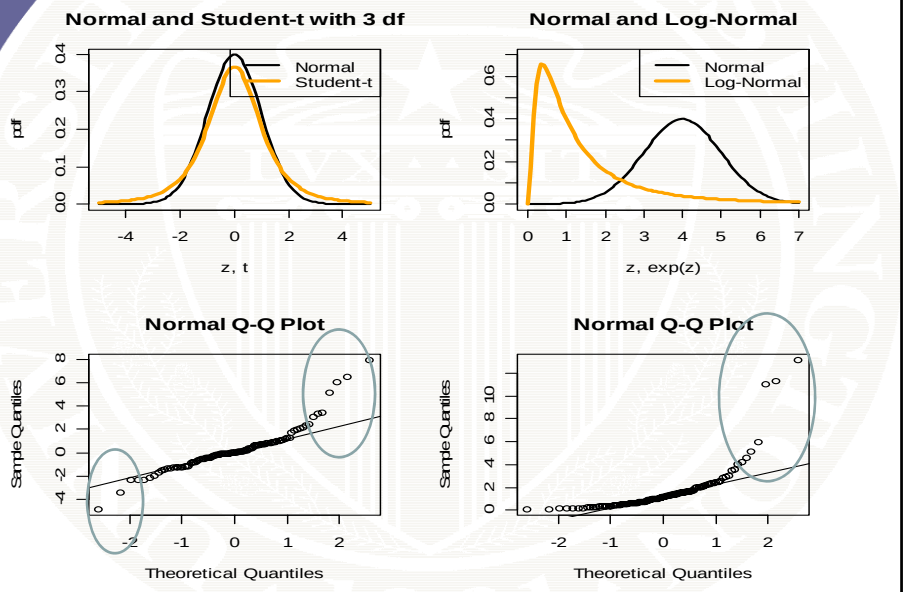
QQ-plots against Normal Distribution

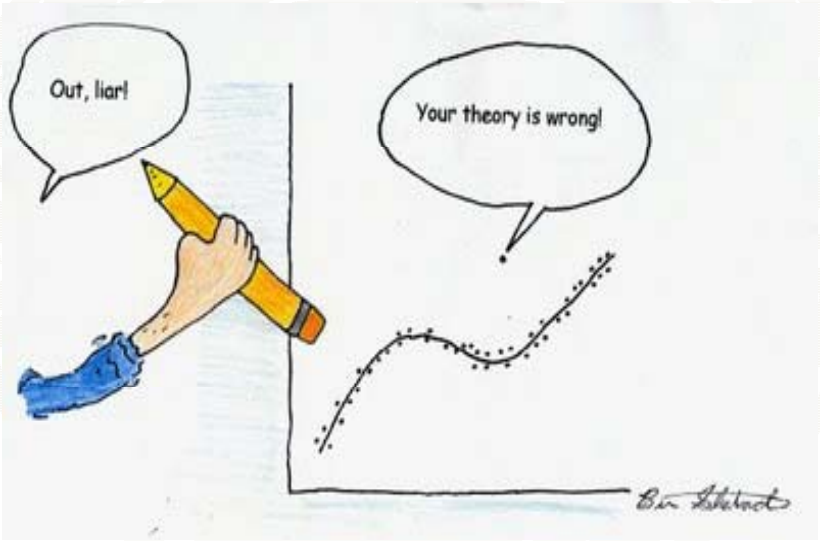




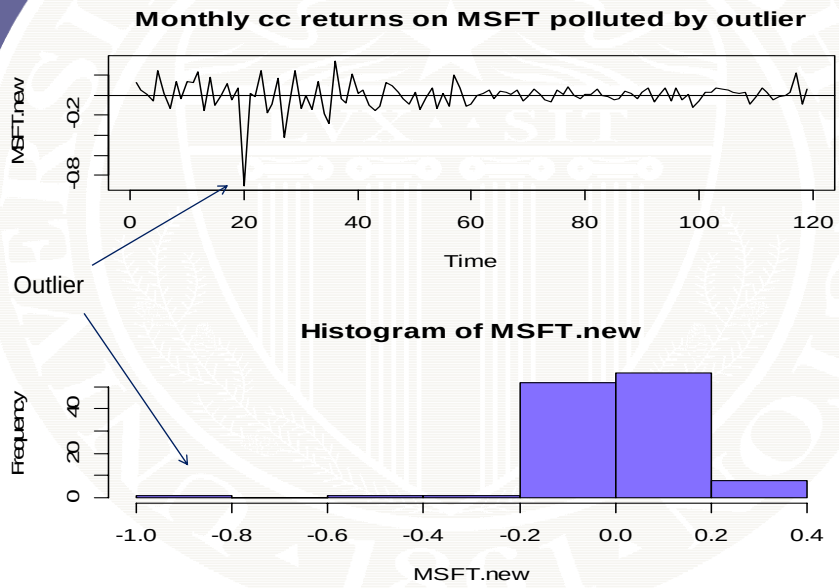



```
> par(mfrow=c(2,2))
> qqnorm(gwn)
> qqline(gwn)
> qqnorm(MSFT)
> qqline(MSFT)
> qqnorm(SP500)
> qqline(SP500)
> par(mfrow=c(1,1))
```





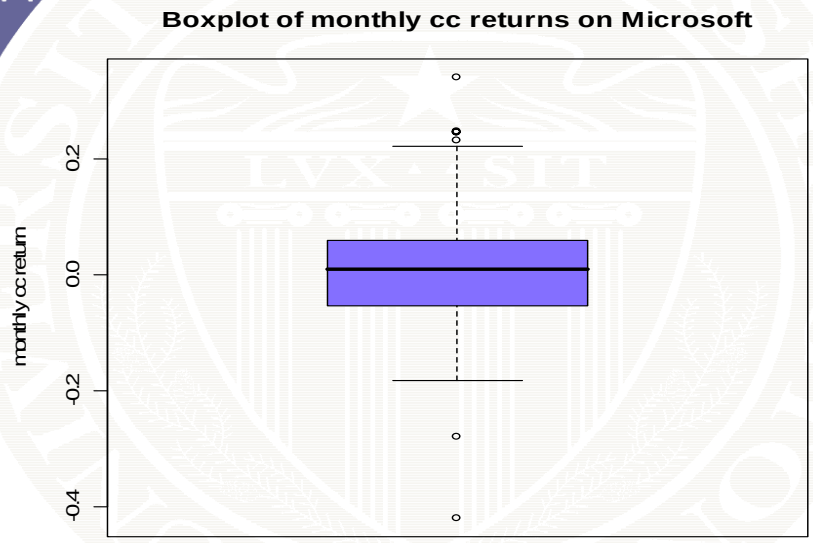
Effect of Outliers on Descriptive Statistics



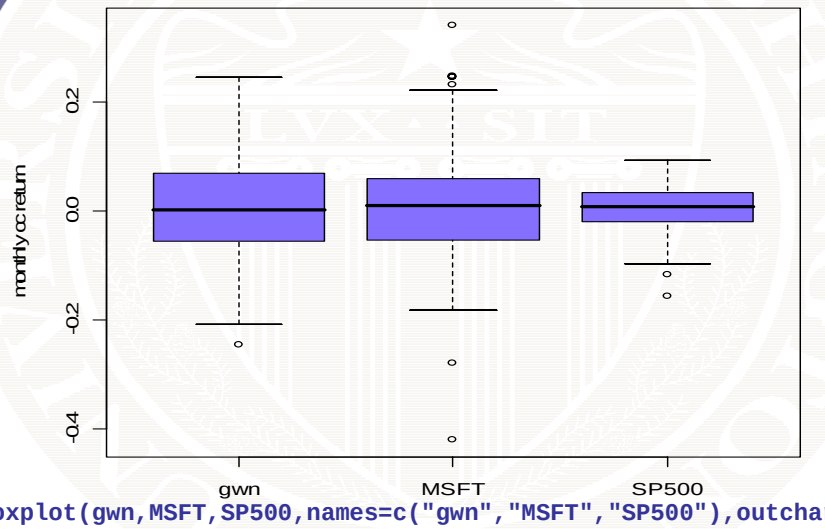
Summary statistics of polluted data

```
> tmp = cbind(MSFT.mat, MSFT.new)

> apply(tmp, 2, mean)
       MSFT      MSFT.new
 0.0067770126 -0.0006029446

> apply(tmp, 2, sd)
     MSFT  MSFT.new
0.1089963 0.1370640

> apply(tmp, 2, skewness)
     MSFT   MSFT.new
-0.0800148 -2.3420254

> apply(tmp, 2, kurtosis)
     MSFT  MSFT.new
 1.970088 14.835069
```

Notice how sample statistics are influenced by the single outlier



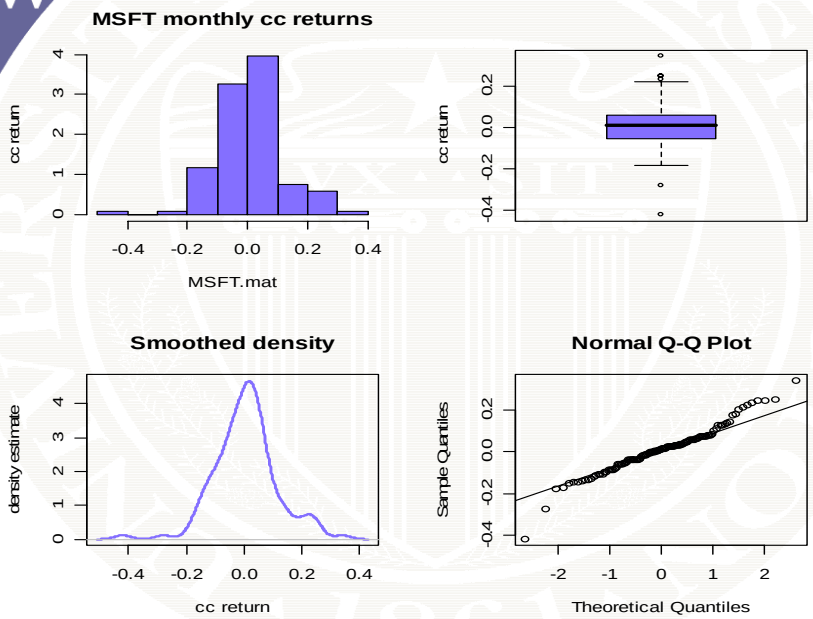

```
> boxplot(MSFT,outchar=T,main="Boxplot of monthly cc
+ returns on Microsoft",ylab="monthly cc return")
```

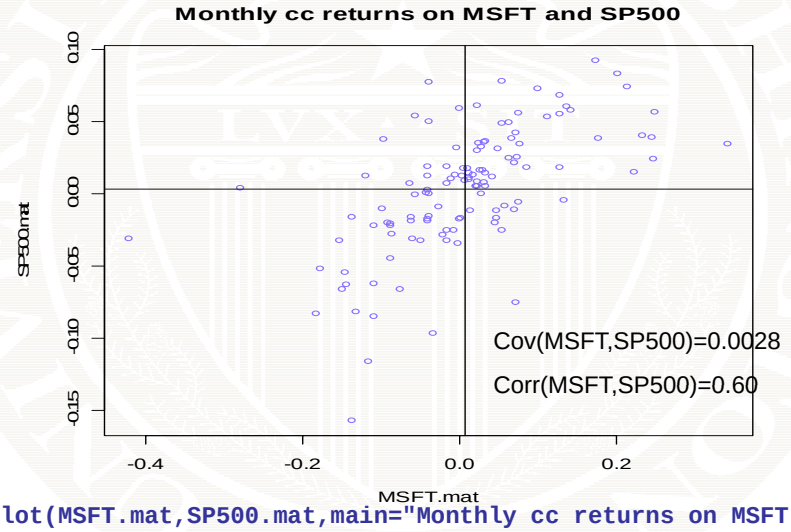

```
> boxplot(gwn,MSFT,SP500,names=c("gwn","MSFT","SP500"),outchar=T,
+ main="Comparison of return distributions", ylab="monthly cc
+ return")
```

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))
```

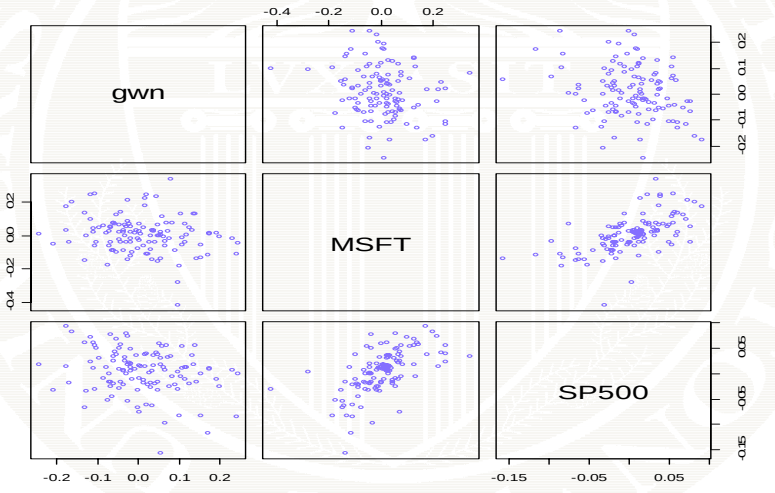



Scatterplot




```
> 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

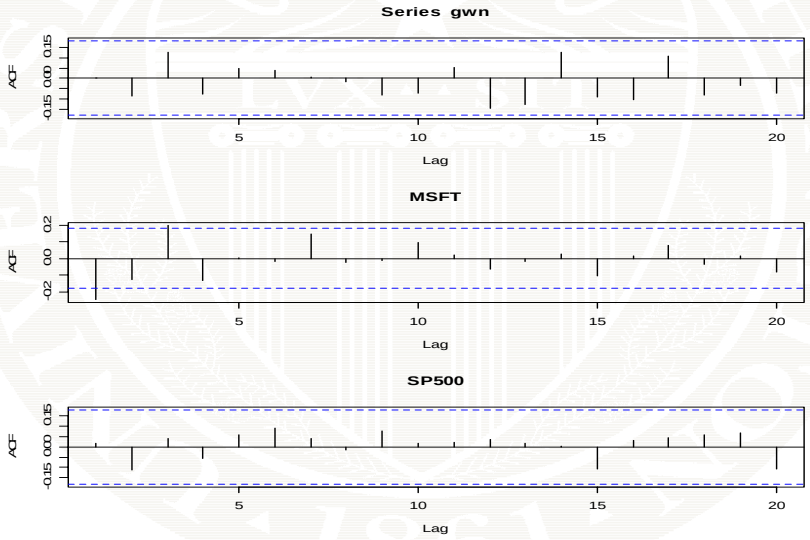



```
> pairs(cbind(gwn,MSFT,SP500))
```

Sample Covariances and Correlations

```
> var(cbind(gwn,MSFT.mat,SP500.mat))
              gwn         MSFT        SP500
gwn    0.009474997 -0.002035362 -0.001033916
MSFT  -0.002035362  0.011880194  0.002819044
SP500 -0.001033916  0.002819044  0.001851312

> cor(cbind(gwn,MSFT.mat,SP500.mat))
             gwn       MSFT       SP500
gwn    1.0000000 -0.1918405 -0.2468629
MSFT  -0.1918405  1.0000000  0.6011050
SP500 -0.2468629  0.6011050  1.0000000
```


Sample Autocorrelations



Summary of Stylized Facts for Monthly CC Returns

- Returns appear to be approximately normally distributed
- Many assets are contemporaneously correlated
- Assets are approximately uncorrelated over time (no serial correlation)