Bootstrap and Bayesian Methods in Finance

Lecture 2: Random Walk Hypothesis & Distribution of Returns

Matthew Pollard

matthewcpollard.googlepages.com

August 29, 2008

1 Introduction

- Bootstrap: testing the random walk hypothesis
  - Resampling under null
  - Problem of spurious auto correlation
  - Calculating power of test
  - Demonstration in R
- Bayesian: testing the distribution of daily returns.
  - Models: Normal, Student-t and Double Exponential
  - Comparison with GMM (*Tom's thesis*)

2 Random Walk Hypothesis

The random walk hypothesis says that returns r_t are *independent* from previous returns,

$$r_t \perp r_{t-k}, \ \ \forall k.$$

Not the same as $r_t \sim$ i.i.d, so not strictly "random walk". A testable is uncorrelated returns:

$$E[(r_t - \overline{r})(r_{t-k} - \overline{r})] = 0 \tag{1}$$

Problems:

- Returns non-normal, changing volatility $\Longrightarrow$ test statistics have complicated, unknown distribution.
- Spurious auto-correlation: bid-ask bounce, stale prices.

3 Why use Bootstrap?

- Ideal where distribution of test statistic is unknown / non-normal. Doesn't require large samples or asymptotic normality.

- Pick any test statistic, choice is irrelevant. Want to pick most powerful.

- Want distribution of test statistic under null of independence. Bootstrap automatically chops up dependency structure in resampling

- *Very easy* to implement:

  Given time series $(r_1, r_2, r_3, ..., r_T)$,
  for (i in 1 to B)
  pick T random numbers from 1 to T with replacement. $\{t_1, t_2, t_3, ...\}$
  Calculate $\hat{S}_b = S(r_{t_1}, r_{t_2}, r_{t_3}, ...)$

4 R Code for Bootstrap

```
boot<-function(x,theta=mean,R=600){
    data<-matrix(sample(x,size=length(x)*R,replace=T),nrow=R)
    return(apply(data,1,theta))
    }

boxpierce<-function(x){length(x)*sum((acf(x,plot=F)$acf[-1])^2)}

S<-boxpierce(data)

Sstar<-boot(data,theta=boxpierce,R=600).

p.value<-mean(S>Sstar)
```

5 Spurious Auto-Correlation

Example: Bid-ask bounce. General Motors intraday prices, tick = 1/8.

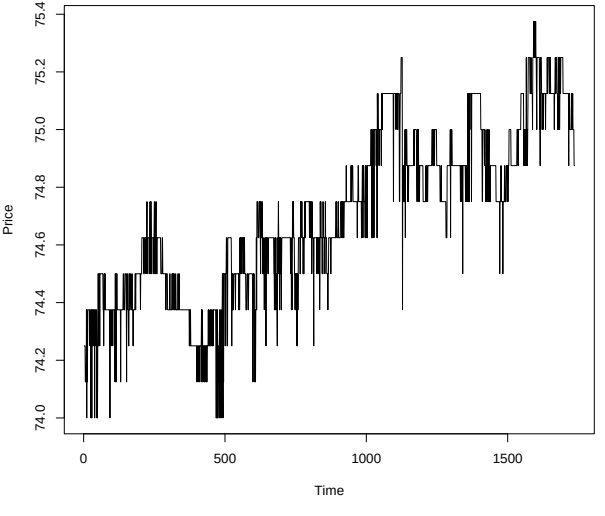


Figure 1: *Tick Data in GE Stock, 9 September to 11 September 1992.*

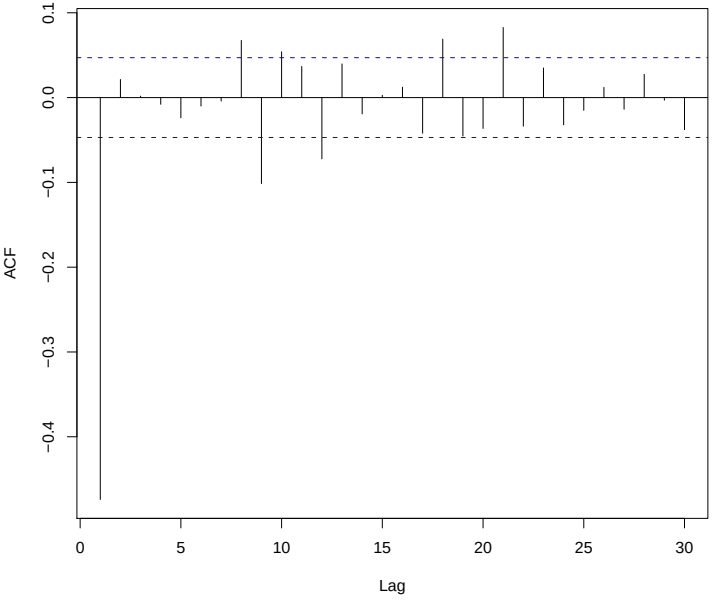


Figure 2: *Sample auto-correlation function* $\gamma(h)$. *First lag* $\gamma(1)$ *is -0.47.*

Roll's Model for Bid-Ask bounce:

$$P_t = P_t^* + I_t \cdot \frac{S}{2}$$

$$\begin{aligned} P_t^* &= \text{fundamental value independent of tick} \\ T &= \text{tick size} \\ I_t &= \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases} \quad \text{and } i.i.d \end{aligned}$$

Result: $Corr(\Delta P_t, \Delta P_{t-1}) = \frac{-T^2/4}{\sigma^2 + T^2/2} < 0$

With GM stock, $T = \frac{1}{8}$, $\sigma^2 = 2.25\text{e} - 06$,

$$\begin{aligned} \text{Predicted } Corr(\Delta P_t, \Delta P_{t-1}) &= -0.49 \\ \text{Estimated } Corr(\Delta P_t, \Delta P_{t-1}) &= -0.47 \end{aligned}$$

6 Test Statistic

Let $\rho(k) = E[(r_t - \mu)(r_{t-k} - \mu)]$. Want to test

$$H_0(\text{random walk}) = \{\rho(1) = \rho(2) = ... = 0\}$$

With bootstrap can arbitrary pick any test statistic.
Our test statistic is the Box-Pierce:

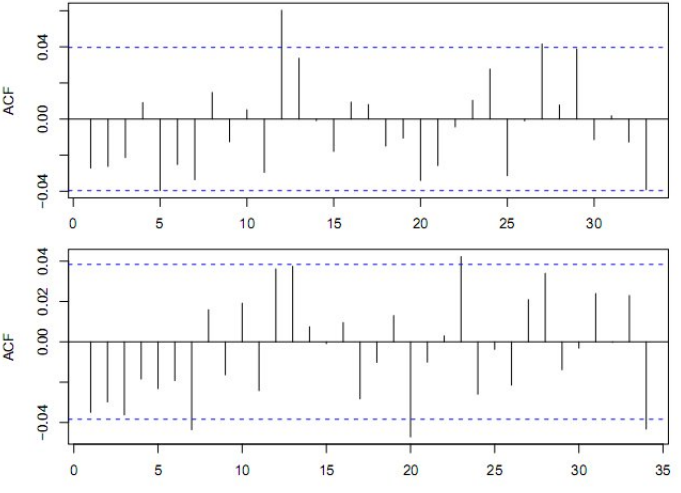
$$\hat{S} = T \sum_{t=1}^{K} \hat{\rho}(k)^2$$

This has an χ^2_K asymptotic distribution for large T, but finite small sample properties is non-Chi.

We use the bootstrap to find distribution of $\hat{S}|H_0$.

7 Data

- S&P 500 Index daily returns, 1998 to 2007.
  - Possible non-synchronous trading effects via stale prices.
- S&P 500 futures index returns, 1998 to 2007.
  - Contracts traded continuously in high volume, so should circumvent non-synchronous effects.
- Sample auto-correlations look small:


ACF
0.04
0.00
−0.04
0
5
10
15
20
25
30
ACF
0.04
0.02
0.00
−0.04
0
5
10
15
20
25
30
35

8 Results

Raw data = r_t, volatility removed "pre-whitens" data by resampling $r_t/\hat{\sigma}_t$,

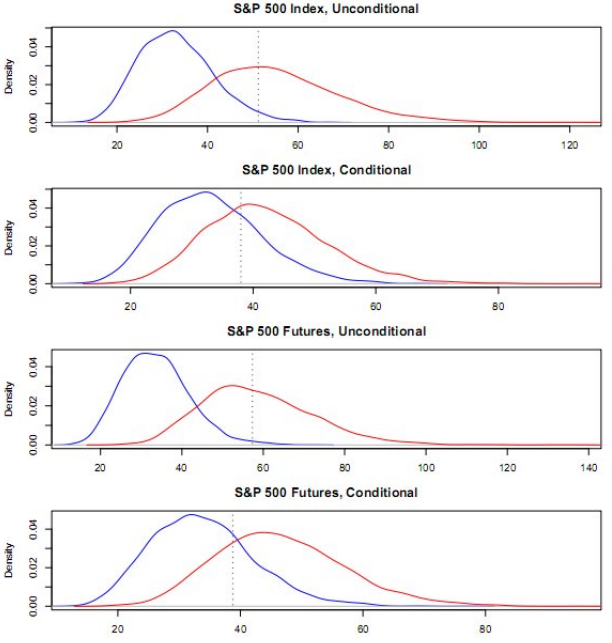
$$\hat{\sigma}_t = \hat{\sigma}_{t-1}\lambda + (1-\lambda)\varepsilon_t^2,$$

Exponential smoother estimate. Using $\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I(\hat{S} > \hat{S}_b)$,

Data	p-value
Index, Raw	0.024
Futures, Raw	0.010
Index, Volatility Removed	0.26
Futures, Volatility Removed	0.25

At α =5%, we:

- reject H0 for the raw index and futures returns
- accept H0 for the volatility removed index and futures returns.


S&P 500 Index, Unconditional
Density
0.00
0.02
0.04
20
40
60
80
100
120
S&P 500 Index, Conditional
Density
0.00
0.02
0.04
20
40
60
80
S&P 500 Futures, Unconditional
Density
0.00
0.02
0.04
20
40
60
80
100
120
140
S&P 500 Futures, Conditional
Density
0.00
0.02
0.04
20
40
60
80

9 Distribution of Stock Returns

- Bayesian: Bayes Factor vs GMM: goodness-of-fit J statistic.

- Many classical statistics tests for normality: Jarque-Bera, Kolmogrov-Smirnov, Pearson's Chi-square.
  Test Format:

$$H_0: \;\; r_t \sim \mathcal{N}(\mu, \sigma^2) \;\; \text{vs} \;\; H_A: \;\; r_t \nsim \mathcal{N}(\mu, \sigma^2)$$

  Not clear what the alternative hypothesis is; under defined.

- Such issues to no not arise in Bayesian hypothesis testing. The comparison is explicit:

$$H_0: \;\; \mathcal{M}_A \;\; \text{vs} \;\; H_B: \;\; \mathcal{M}_B$$

10 Bayes Factor and Model Probabilities

Compare the likelihood of two models: use the Bayes Factor,

$$BF(\mathcal{M}_0, \mathcal{M}_A) = \frac{p(Y|\mathcal{M}_A)}{P(Y|\mathcal{M}_0)}$$

When $BF \geq 1$, evidence for $\mathcal{M}_A$ over $\mathcal{M}_0$. Rules of thumb: $BF > 10$ "substantial evidence" for $\mathcal{M}_A$, BF>20, "decisive evidence".

If several model comparisons, easier to use posterior model probability

$$p(\mathcal{M}_1|Y) = \frac{p(Y|\mathcal{M}_1)}{\sum_k p(Y|\mathcal{M}_k)}$$

where the likelihood function $p(Y|\mathcal{M}_k) = \int p(Y|\Theta_k, \mathcal{M}_k)p(\Theta_k|\mathcal{M}_k)d\Theta_k$. Pick $\mathcal{M}_1$ over $\mathcal{M}_2$ if more probable. Usually, one model clearly dominates.

11 GMM test

Smith (1990) shows how to use GMM to test for a given distribution.

- Given test distribution $x \sim f(\Theta)$, calculate theoretical 1 to M moments, construct residual vector $g_T(\Theta) = E_T[h(x_t)]$
- Calculate weighting matrix S in $g^T(\Theta)S^{-1}g(\Theta)$ if tractable, numerically estimate it by 2 stage process if not.
- Set $\hat{\Theta} = \arg\min[g^T(\Theta)S^{-1}g(\Theta)]$
- Calculate $J_T = Tg^T(\hat{\Theta})S^{-1}g(\hat{\Theta})$, under null $\sim \chi^2_{M-\#\Theta}$.

12 Example: Test for Normality

GMM test of normal tests if sample skewness is 0 and kurtosis is $3\sigma^4$.

$$g_T(\mu, \sigma^2) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} x_t - \mu \\ (x_t - \mu)^2 - \sigma^2 \\ (x_t - \mu)^3 \\ (x_t - \mu)^4 - 3\sigma^4 \end{pmatrix}$$

Analytically tractable S matrix, tractable J-statistic.

$$S = \begin{pmatrix} \sigma^2 & 0 & 3\sigma^4 & 0 \\ 0 & 2\sigma^4 & 0 & 12\sigma^6 \\ 3\sigma^4 & 0 & 15\sigma^6 & 0 \\ 0 & 12\sigma^6 & 0 & 96\sigma^8 \end{pmatrix}, \quad TJ_T = \frac{1}{6}\left(\hat{\gamma} + \frac{(\hat{\kappa} - 3)^2}{4}\right)$$

Reduces exactly to the Jarque-Bera test.

13 Models Considered

Models for the distribution of daily logarithmic returns, $r_t := \log(\frac{P_t}{P_{t-1}})$

$$\begin{aligned} \mathcal{M}_{Normal}: \quad r_t &\sim \mathcal{N}(\mu, \sigma^2) \\ \mathcal{M}_{Student-t}: \quad r_t &\sim \mu + \sigma\mathcal{T}(k) \\ \mathcal{M}_{D-exponential}: \quad r_t &\sim \mathcal{DE}(\mu, \sigma^2) \end{aligned}$$

Laplace distribution defined by

$$f(x|\mu, \sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right)$$

basically a "double-exponential," a mirrored exponential spliced with another.

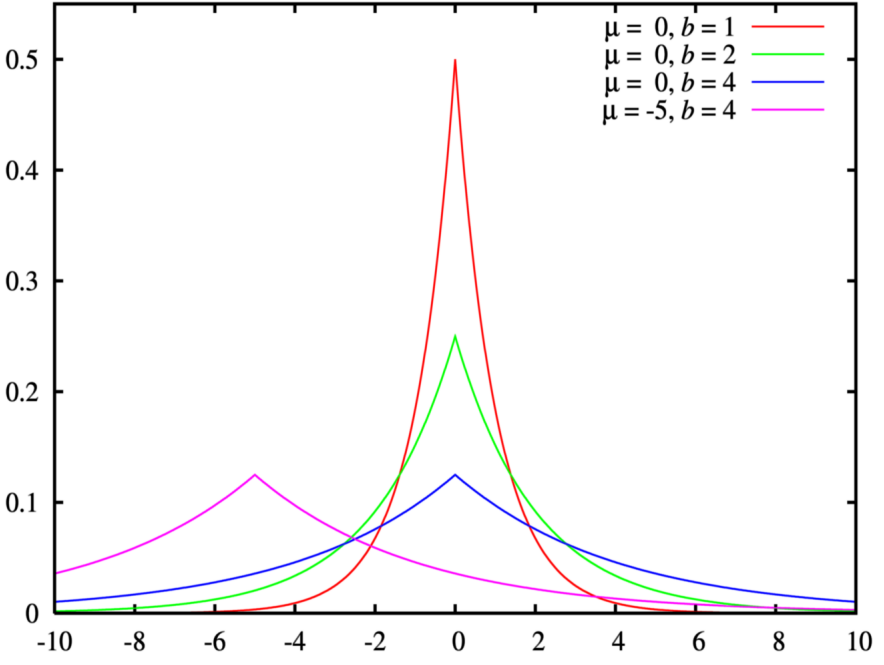


Figure 3: Laplace distribution with different scale and location parameters.

14 Results

Data: daily S&P 500 index returns, 1998 to 2007. Diffusive prior distributions.

Model	μ	σ^2	k	$p(\mathcal{M}_k \mid Y)$
$\mathcal{M}_{Normal}$	4.26e-4	1.12e-4		1.49e-267
$\mathcal{M}_{Student-t}$	4.96e-4	1.46e-4	5.59	1.38e-07
$\mathcal{M}_{Laplace}$	5.49e-4	1.43e-4		0.999

Very strongly reject normal and student-t distribution in favor of Laplace distribution.

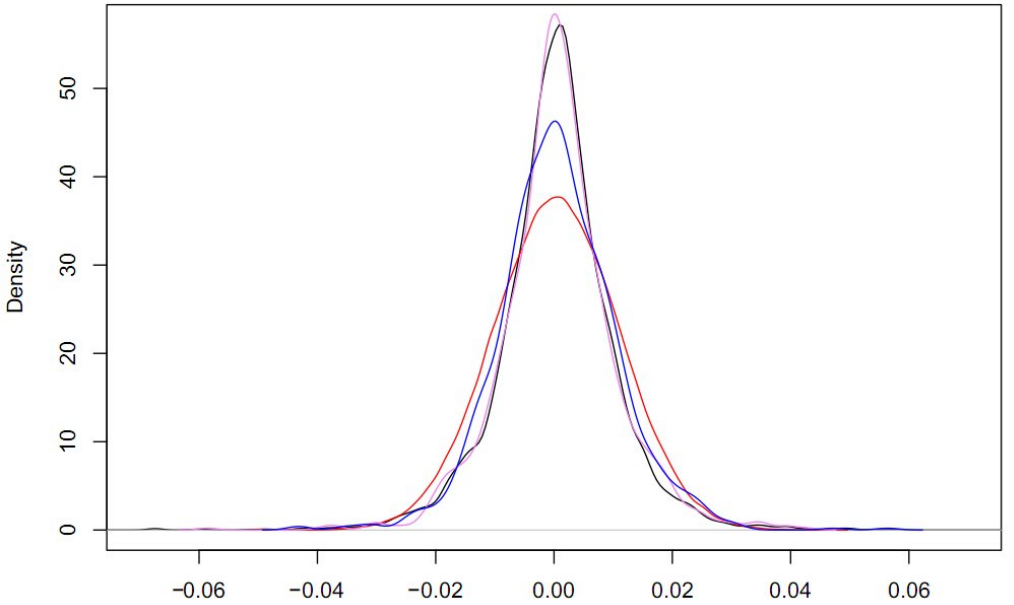


Figure 4: Three fitted models and actual data . *Black*: empirical density of returns, *pink:* Laplace distribution, *Red:* Normal distribution and *Blue:* Student-t distribution.

15 Conclusion

- Different tools for different problems,
  - Bootstrap ideal for testing random walk hypothesis due to independent resampling method.
  - Bayesian ideal for testing multiple hypotheses and formal model comparisons.
- Results:
  - Accept random walk hypothesis.
  - Reject normal and student-t distribution of returns in favour of Laplace distribution.

Thank You! Now time for Questions.

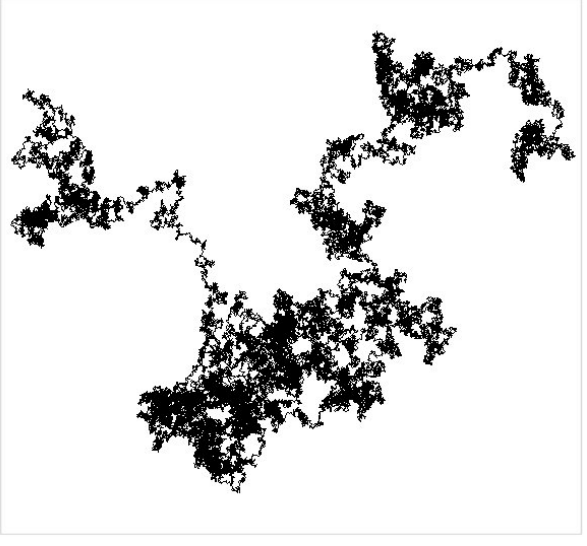


Figure 5: *A Brownian Motion.* If only financial markets behaved this nicely.