

# Efficient Markets and predicting returns

Susan Thomas

4 September, 2017

# Information processing by markets

- ▶ Old economist notion of efficient capital markets: aggregate supply and demand in a *competitive* market, with rational traders who maximise profits.
- ▶ Market notion of efficient capital markets: Speculators trade in information: huge profits to be made if they are right, and huge losses if they are wrong.
- ▶ Any new information is rapidly assimilated into prices through such a profit-maximisation.
- ▶ The resulting equilibrium price from an “efficient market” does a pretty good job of embedding forecasts of future  $d_t$  and a sensible risk premium  $\Delta$ .

# Information processing by markets

- ▶ Old economist notion of efficient capital markets: aggregate supply and demand in a *competitive* market, with rational traders who maximise profits.
- ▶ Market notion of efficient capital markets: Speculators trade in information: huge profits to be made if they are right, and huge losses if they are wrong.
- ▶ Any new information is rapidly assimilated into prices through such a profit-maximisation.
- ▶ The resulting equilibrium price from an “efficient market” does a pretty good job of embedding forecasts of future  $d_t$  and a sensible risk premium  $\Delta$ .

# Consequences of efficient markets: price behaviour

- ▶ If traders process information efficiently and immediately into prices, the current price reflects all relevant information perfectly.
- ▶  $\implies$  prices move only in response to news and information.
- ▶  $\implies$  if you know price today, you need to know news tomorrow to predict price tomorrow.
- ▶ But news is unforecastable  $\implies$  prices are unforecastable.
- ▶ Differing views between economists and traders about price efficiency.

# Consequences of efficient markets: price behaviour

- ▶ If traders process information efficiently and immediately into prices, the current price reflects all relevant information perfectly.
- ▶  $\implies$  prices move only in response to news and information.
- ▶  $\implies$  if you know price today, you need to know news tomorrow to predict price tomorrow.
- ▶ But news is unforecastable  $\implies$  prices are unforecastable.
- ▶ Differing views between economists and traders about price efficiency.

# Recap on market efficiency

- ▶ In an efficient market, all speculators know the historical prices.  
     $\implies$  Competition between them will eliminate opportunities for earning money “for free”. (Zero-profit condition under perfect competition.)
- ▶ In the limit, when millions of smart speculators are in play, prices should become non-forecastable.
- ▶ Simplest model:  $E(P) = P_t$  where  $P_t$  is the last observed price  
     $\implies E(P - P_t) = E(r) = 0$ .
- ▶ Simplest model of EMH: returns are homoscedastic normal, with a mean  $\mu_r = 0$ .
- ▶ Also, “returns are homoscedastic normal” is a *conveniently testable statement*.
- ▶ Reality doesn't have to oblige.

# Market prices and returns

# Definition of returns

- ▶ Investment decisions are based on expected returns,  $E(r)$ .
- ▶ The market produces a time-series  $P_1, P_2, \dots$
- ▶ We focus on the percentage change in prices as the “returns”,  $r_t$ .

$$r_t = 100 \times \log P_t/P_{t-1} = 100 * (\ln(P_t) - \ln(P_{t-1}))$$

Here  $P_t$  is the “Adjusted Closing Price” (ACP in Prowess).

- ▶ One way to get  $E(r)$  is from the distribution of  $(\bar{r})$ .  
(Assumption?)



## Numerical example using Britannia prices, source: NSE website

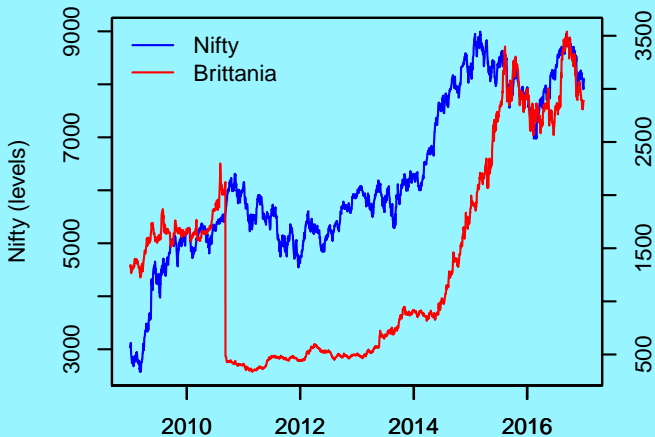
```
> brit <- read.csv(file="britannia.csv", skip=1,
  sep=",", col.names=c("Symbol", "Series", "Date",
    "PrevClose", "Open", "High", "Low", "LTP", "Closing",
    "AvPr", "Quantity", "Turnover", "NumTrades"))
> tail(brit)
```

```
> Symbol Series          Date PrevClose   Open   High
1976 BRITANNIA      EQ 26-Dec-2016   2831.20 2829.90 2831.40 27
1977 BRITANNIA      EQ 27-Dec-2016   2807.30 2810.25 2856.85 27
1978 BRITANNIA      EQ 28-Dec-2016   2826.85 2830.00 2888.00 28
1979 BRITANNIA      EQ 29-Dec-2016   2839.60 2835.00 2898.00 28
1980 BRITANNIA      EQ 30-Dec-2016   2883.15 2905.00 2922.00 28
1981 BRITANNIA      EQ 02-Jan-2017   2886.30 2901.10 2901.10 28
```

	Closing	AvPr	Quantity	Turnover	NumTrades
1976	2807.30	2805.73	68000	1907.90	8314
1977	2826.85	2824.84	78676	2222.47	11720
1978	2839.60	2860.05	123554	3533.70	17006
1979	2883.15	2866.23	90143	2583.71	8816
1980	2886.30	2901.90	110334	3201.78	13373
1981	2886.45	2876.68	52245	1502.92	3552

```
>
```

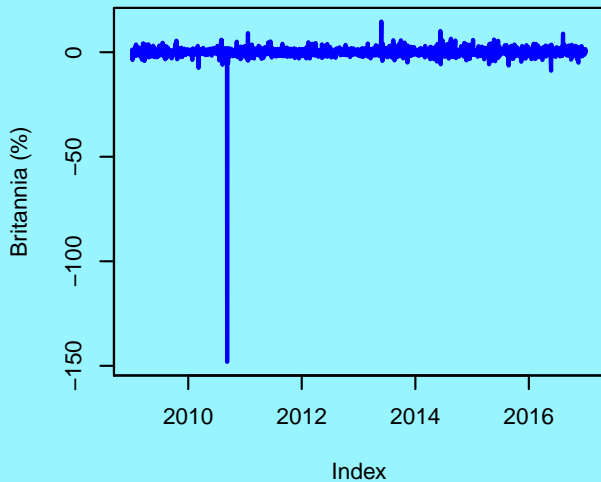
# Example: Britannia and Nifty, 2009 to 2016



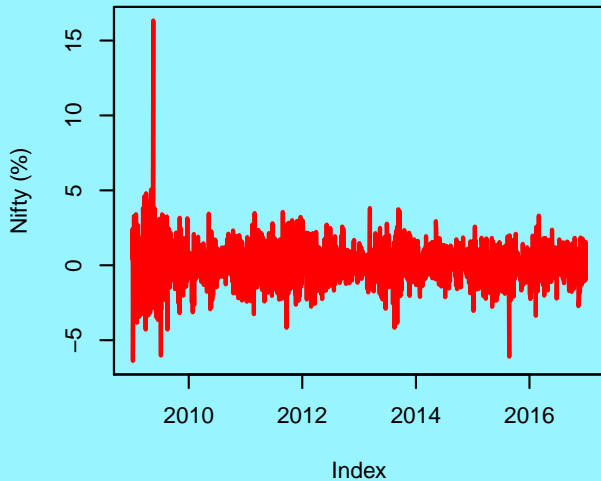
## Numerical example calculating Britannia returns

```
> library(zoo)
> mydates <- as.Date(brit$Date, "%d-%b-%Y")
> pricesb <- zoo(brit$Closing, order.by=mydates)
> returnsb <- 100*diff(log(pricesb))
> tail(returnsb)
  2016-12-26  2016-12-27  2016-12-28  2016-12-29  2016-12-30
-0.84774827  0.69398502  0.45001797  1.52202493  0.00000000
> summary(returnsb)
      Index              returnsb
Min.   :2009-01-05   Min.   : -148.12043
1st Qu.:2011-01-05   1st Qu.:  -0.84901
Median :2012-12-31   Median :    0.00891
Mean   :2013-01-02   Mean   :    0.03928
3rd Qu.:2015-01-01   3rd Qu.:    0.99999
Max.   :2017-01-02   Max.   :   14.67419
> sd(returnsb)
[1] 3.728062
>
```

# Britannia returns



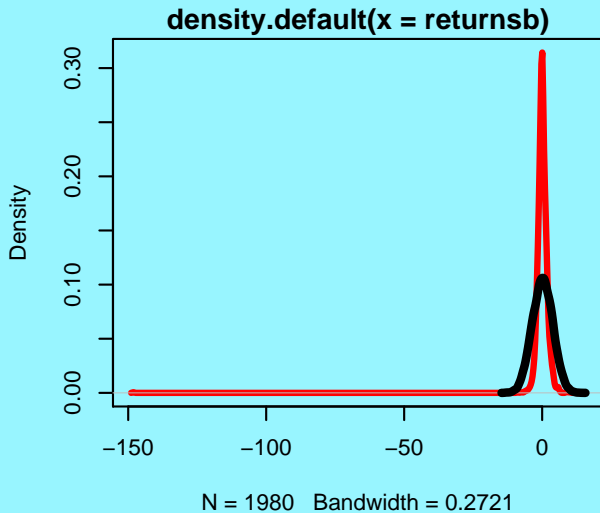
# Nifty returns



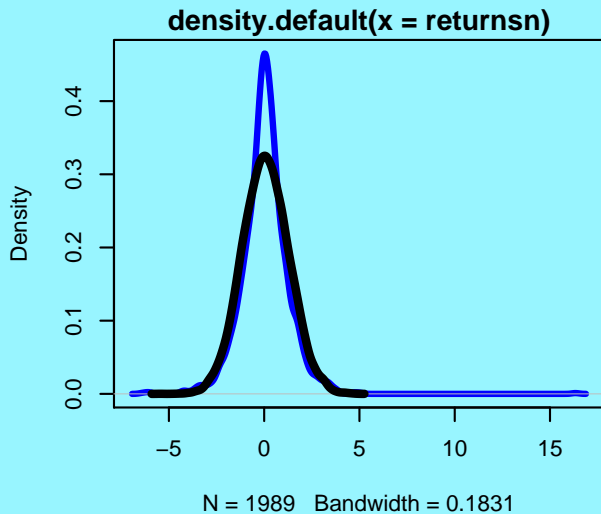
## Numerical example of the histogram of Britannia returns vs. a gaussian distribution

```
> simulb <- rnorm(10*length(returnsb), mean(returnsb)
                 sd(returnsb))
> pdf("2016to2009histb.pdf", width=4.0, height=2.8,
      bg="cadetblue1", pointsize=8)
> par(mai=c(.8, .8, .2, .8))
> plot(density(returnsb), col="red", lwd=3)
> lines(density(simulb), col="black", lwd=4)
> dev.off()
```

# Britannia vs. a gaussian distribution



# Nifty vs. a gaussian distribution





# **EMH and implications for the data generating process of price**

# Setting up a mathematical model

- ▶ Rational expectations about prices which contain all information at all times implies that:

$$\begin{aligned}P_{t+1} &= EP_{t+1} + \epsilon_{t+1} \\ E_t(P_{t+1} - EP_{t+1}) &= E_t\epsilon_{t+1} = 0\end{aligned}$$

- ▶ The orthogonality property:  $\epsilon_{t+1}$  must be independent of all information at  $t$ .

For example,  $\epsilon_{t+1} = \rho\epsilon_t + \nu_{t+1}$  violates EMH.

- ▶ Think of  $\epsilon_{t+1}$  as *unexpected* profit/loss.

$\implies$  the only way to make expected profits is to be informed: know  $\epsilon_{t+1}$  before everyone else.

(Note: The information has to eventually become public.)

# Setting up a mathematical model

- ▶ Rational expectations about prices which contain all information at all times implies that:

$$\begin{aligned}P_{t+1} &= EP_{t+1} + \epsilon_{t+1} \\ E_t(P_{t+1} - EP_{t+1}) &= E_t\epsilon_{t+1} = 0\end{aligned}$$

- ▶ The orthogonality property:  $\epsilon_{t+1}$  must be independent of all information at  $t$ .

For example,  $\epsilon_{t+1} = \rho\epsilon_t + \nu_{t+1}$  violates EMH.

- ▶ Think of  $\epsilon_{t+1}$  as *unexpected* profit/loss.

⇒ the only way to make expected profits is to be informed: know  $\epsilon_{t+1}$  before everyone else.

(Note: The information has to eventually become public.)

# Prices as a random walk

- ▶  $P_{t+1} = P_t + \epsilon_{t+1}$  and  
 $\epsilon \sim iid(0, \sigma^2)$ .

- ▶ Then

$$E_t P_{t+1} = P_t + E_t \epsilon_{t+1} = P_t$$

- ▶ Implications:

- ▶ Innovations to the DGP are permanent.

$$\log P_{t+1} = \log P_t + \epsilon_t$$

$$\text{And, } \log P_{t+1} = \log P_{t-k} + \sum_{i=0}^k \epsilon_{t-i}$$

- ▶ The best estimate of the forecasted price  $P_{t+1}$  is  $P_t$ .
- ▶ This is true for forecasts at all horizons,  $h$ , in the future.

$$E(P_{t+h}) = P_t$$

- ▶ These are also properties of a time series with a unit root.

# A random walk is non-forecastable

- ▶  $P_{t+1} = P_t + \epsilon_{t+1}$
- ▶ Forecastability is focussed on any new information/pattern,  $\epsilon_{t+1}$  over  $P_t$ . This is a problem because:
  1.  $\epsilon_{t+1}$  tends to be a small change over  $P_t$ .
  2.  $\epsilon_{t+1}$  is a random number.

$\epsilon_t$  tends to be white noise.

- ▶ Speculators focus on picking patterns in the data, either in the short run or the long run.

But random draws also have some **non-zero** probability of (a) runs and (b) temporal serial correlation.

# The market efficiency debate, part I

- ▶ A strong statement: **zero** forecastability of returns.
- ▶ Some people get excited when a  $t$  stat of 2.5 turns up, they have “rejected the  $H_0$  of market efficiency”.
- ▶ There is a lot of talk about “inefficient markets” based on such rejections.
- ▶ But no forecasting equation has substantial power.
- ▶  $H_0$  can be rejected, but with a tiny  $R^2$ , the process is mostly white noise!
- ▶ This is a statistical problem, not an economic one.

# The market efficiency debate, part I

- ▶ A strong statement: **zero** forecastability of returns.
- ▶ Some people get excited when a  $t$  stat of 2.5 turns up, they have “rejected the  $H_0$  of market efficiency”.
- ▶ There is a lot of talk about “inefficient markets” based on such rejections.
- ▶ But no forecasting equation has substantial power.
- ▶  $H_0$  can be rejected, but with a tiny  $R^2$ , the process is mostly white noise!
- ▶ This is a statistical problem, not an economic one.

## The market efficiency debate, part II

- ▶ All tests of EMH are conditional on whether we have the correct model of expected returns.
- ▶ Example: the random walk assumes zero expected returns. But a zero expected return on an investment of any horizon is not economically sensible!
- ▶ Alternative model:
  1. No dividend.
  2. But investors hold as long as expected capital gains are constant.

$$R_{t+1} = k + \epsilon_{t+1}$$

where  $\epsilon_{t+1}$  is iid and independent of  $\Omega_t$ .

- ▶ Then,  $\log P_{t+1} = k + \log P_t + \epsilon_{t+1}$   
Random walk with a drift term, where  $k$  can be the sum of the risk-free rate and an equity premium.
- ▶ Tests of randomness that ignore this infer that EMH is violated.



# Alternative model: prices as martingale processes

- ▶ The martingale process: expectations are conditional on information at  $t$ .

1. If  $x$  is random discrete variable, then  $E(X) = \sum_{i=1}^J \pi_i X_i$ .

2. If  $x$  is uniform discrete variable, then

$$E(X) = \int_{-\infty}^{\infty} X f(X) dX.$$

3. If  $x_t$  is a martingale variable, then  $E_t(x_{t+1} | \Omega_t) = x_t$

- ▶ Since prices are log-normally distributed, then prices being martingale processes means:

$$\log p_t = \log p_{t-1} + \epsilon_t$$

where  $\epsilon_t$  is iid as  $N(0, \sigma^2)$ .

- ▶ Implication:  $E_t(R_{t+1} | \Omega_t) = 0$ .

This is the property of a *fair game*.

# Alternative model: prices as martingale processes

- ▶ The martingale process: expectations are conditional on information at  $t$ .
  1. If  $x$  is random discrete variable, then  $E(X) = \sum_{i=1}^J \pi_i X_i$ .
  2. If  $x$  is uniform discrete variable, then  $E(X) = \int_{-\infty}^{\infty} X f(X) dX$ .
  3. If  $x_t$  is a martingale variable, then  $E_t(x_{t+1} | \Omega_t) = x_t$
- ▶ Since prices are log-normally distributed, then prices being martingale processes means:

$$\log p_t = \log p_{t-1} + \epsilon_t$$

where  $\epsilon_t$  is iid as  $N(0, \sigma^2)$ .

- ▶ Implication:  $E_t(R_{t+1} | \Omega_t) = 0$ .  
This is the property of a *fair game*.

# Alternative model: prices as martingale processes

- ▶ The martingale process: expectations are conditional on information at  $t$ .
  1. If  $x$  is random discrete variable, then  $E(X) = \sum_{i=1}^J \pi_i X_i$ .
  2. If  $x$  is uniform discrete variable, then  $E(X) = \int_{-\infty}^{\infty} X f(X) dX$ .
  3. If  $x_t$  is a martingale variable, then  $E_t(x_{t+1} | \Omega_t) = x_t$
- ▶ Since prices are log-normally distributed, then prices being martingale processes means:

$$\log p_t = \log p_{t-1} + \epsilon_t$$

where  $\epsilon_t$  is iid as  $N(0, \sigma^2)$ .

- ▶ Implication:  $E_t(R_{t+1} | \Omega_t) = 0$ .  
This is the property of a *fair game*.

# Alternative model: prices as martingale processes

- ▶ The martingale process: expectations are conditional on information at  $t$ .
  1. If  $x$  is random discrete variable, then  $E(X) = \sum_{i=1}^J \pi_i X_i$ .
  2. If  $x$  is uniform discrete variable, then  $E(X) = \int_{-\infty}^{\infty} X f(X) dX$ .
  3. If  $x_t$  is a martingale variable, then  $E_t(x_{t+1} | \Omega_t) = x_t$
- ▶ Since prices are log-normally distributed, then prices being martingale processes means:

$$\log p_t = \log p_{t-1} + \epsilon_t$$

where  $\epsilon_t$  is iid as  $N(0, \sigma^2)$ .

- ▶ Implication:  $E_t(R_{t+1} | \Omega_t) = 0$ .  
This is the property of a *fair game*.

# Implication of EMH tests with martingale prices

- ▶ EMH is defined as having the fair game property for unexpected stock returns.
  - ▶ "Asset prices fully and instantaneously rationally reflect all available relevant information." (Fama 1969,1971)
  - ▶ "Asset prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed the marginal costs."
- ▶ On average, the *abnormal/unexpected* return is zero.

**Reference:** John Y Campbell, Andrew W. Lo, Craig A. MacKinlay, 1995, "The econometrics of financial markets", published by Princeton University Press.

# Implication of EMH tests with martingale prices

- ▶ EMH is defined as having the fair game property for unexpected stock returns.
  - ▶ "Asset prices fully and instantaneously rationally reflect all available relevant information." (Fama 1969,1971)
  - ▶ "Asset prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed the marginal costs."
- ▶ On average, the *abnormal/unexpected* return is zero.

**Reference:** John Y Campbell, Andrew W. Lo, Craig A. MacKinlay, 1995, "The econometrics of financial markets", published by Princeton University Press.

# EMH: Implications

- ▶ If price is the correct discounted value of future cashflows, then:
  1. There are no arbitrage opportunities: you only get extra returns if you take on additional risk.
  2.  $E(r)$  of any asset is a function only of the risk premium on equity.
    - ⇒ E(excess returns) across any pair of assets ought not to differ persistently
- ▶ Both hold given a fixed information set.
- ▶ Research question: Does no-arbitrage actually hold in an efficient market?

# Tests of EMH



# Categories

- ▶ Tests of EMH are categorised depending upon the information captured by market prices.
- ▶ The test categories are:
  1. Weak form: tests based on publicly observed information.
  2. Semi-strong form: based on information that is originally observed by a few, and then becomes publicly disclosed.
  3. Strong form: based on information that only a small set of investors could be privy to.
- ▶ For example, testing for autocorrelation in a price series is a **weak form** test of EMH.

The tests are based on prices, which are publicly observed.
- ▶ Most tests of EMH are semi-strong form.

# Tests of EMH

- ▶ Weak form: ACF, Variance Ratio analysis (Nelson and Plosser 1985, Summers 1988).  
Effects studied: serial correlation, seasonal effects (such as day of week, budget day, end of year effects).
- ▶ Semi-strong form: Event–study analysis (Brown and Warner 1980, 1985).  
Effects studied: corporate action (such as dividend announcements, bonus issues, rights issues, debt issues, defaults, etc), institutional changes (such as introduction of derivatives markets, changes in laws to shareholders/creditors, etc).
- ▶ Strong form:  
Effects studied: mutual fund/institutional fund performance wrt stock market index.

# Statistical vs. economic tests of EMH

- ▶ Statistical tests of EMH are joint tests of the market efficiency in the context of a given asset pricing model. Example, all the first tests of EMH were based on the null of the random walk model of prices.
- ▶ Modification: Tests incorporate empirical deviations from normality such as skewness and heteroscedasticity.
- ▶ Economic tests focus on whether the market has arbitrage opportunities or not:
  - ▶ Are excess returns are independent of information sets?
  - ▶ Is there persistently abnormal returns to trading strategies?
  - ▶ Is the market price equal to the fundamental value?

## EMH: some early references

- ▶ *The Variation of Certain Speculative Prices*, Benoit Mandelbrot, Journal of Business, Vol. 36, No. 4 (Oct., 1963), pp. 394-419
- ▶ *Proof that properly anticipated prices fluctuate randomly*, Paul Samuelson, Industrial Management Review, 41-49, 1965.
  - ▶ Informationally efficient market,
  - ▶ prices are unforecastable if they are “properly anticipated”.
- ▶ *Random walks in stock market prices*, Eugene Fama, Financial Analysts Journal, pages 55-59, 1965.
  - ▶ Measuring the statistical properties of prices.
  - ▶ Resolving the difference between technical (time series of prices) and fundamental (accounting and financial data) analysis.

# HW

1. Obtain the full ACP time-series from Prowess for your company. Calculate returns for the series.
  - 1.1 Break the time series up into halves - Period I and Period II.
  - 1.2 Do the runs test for the full period, and for periods I and II separately.
  - 1.3 Graph the ACF and estimate an AR model for the full period, and for the half periods separately.
2. Get the time series for the NSE-50 (Nifty) index and repeat the above tasks for this time series as well.

What are any qualitative and quantitative differences between the time series aspects of the company vs. Nifty.

# The remaining slides

- ▶ A few standard tests of weak-form efficiency:
  - ▶ Runs tests
  - ▶ AutoCorrelation Functions
  - ▶ Variance Ratios

# Statistical tests of EMH: tests of randomness

# Runs test

- ▶ A returns sequence as follows  $- +, +, + -$  is
  1. a positive run and
  2. a run of length 3.
- ▶ Runs can have different directions  $(+, -, 0)$  and different lengths.
- ▶ Randomness of returns implies certain properties of runs.

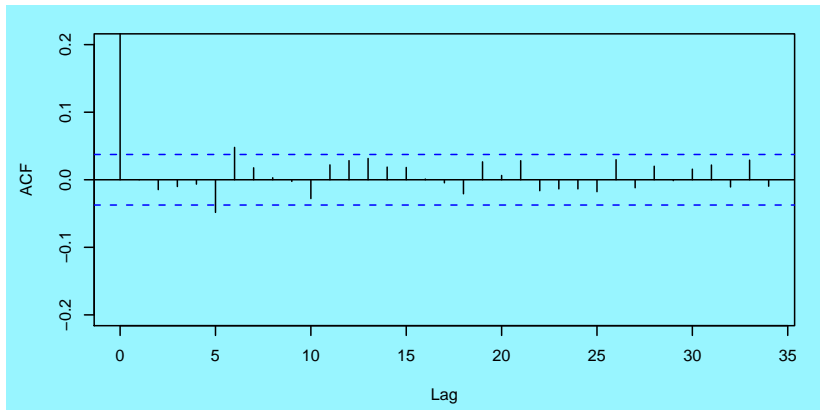


# Autocorrelation coefficients

- ▶ If a series of data is “random”, then it will have no significant autocorrelation coefficients.
- ▶  $H_0 : \rho = 0$
- ▶ The standard deviation for the autocorrelation coefficient approximated by

$$\sigma_\rho = 1/\sqrt{N}$$

# ACF for M&M daily returns



# AR model for daily returns

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.001966	0.019062	0.103	0.9178
ar2	-0.014461	0.019038	-0.760	0.4475
ar3	-0.010136	0.019040	-0.532	0.5945
ar4	-0.005683	0.019041	-0.298	0.7653
ar5	-0.048529	0.019039	-2.549	0.0108 *
ar6	0.047550	0.019061	2.495	0.0126 *
intercept	0.082461	0.056337	1.464	0.1433

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1

Fit:

sigma<sup>2</sup> estimated as 8.68

# To watch out for: small chinks in randomness!

- ▶ Daily  $\sigma$  of returns was 2.95;  
This implies variance of 8.7025.
- ▶ The AR(6) residual has variance of 8.68.  
This comes to a  $R^2$  of 0.0025.
- ▶ Even if there appears to be *statistically significant* correlation, we know:

smart speculators are attacking the data every day, discovering patterns, making money on them, and thus eliminating these correlations,

if they find it *economically efficient* to do so.

- ▶ EMH is a story about speculators who learn.

# To watch out for: small chinks in randomness!

- ▶ Daily  $\sigma$  of returns was 2.95;  
This implies variance of 8.7025.
- ▶ The AR(6) residual has variance of 8.68.  
This comes to a  $R^2$  of 0.0025.
- ▶ Even if there appears to be *statistically significant* correlation, we know:  
smart speculators are attacking the data every day, discovering patterns, making money on them, and thus eliminating these correlations,  
if they find it *economically efficient* to do so.
- ▶ EMH is a story about speculators who learn.

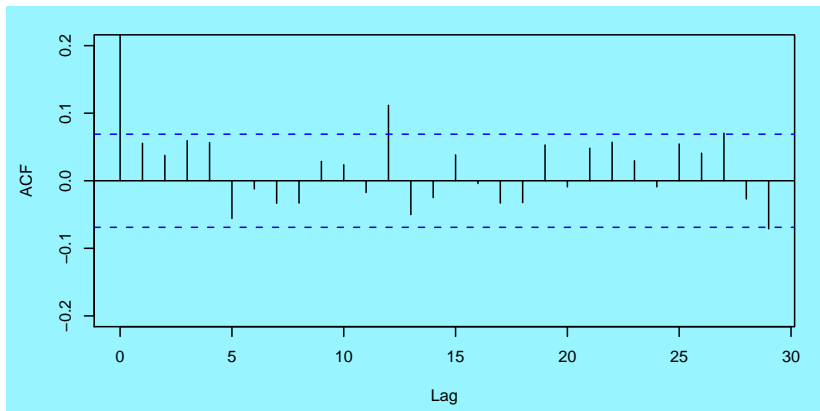
# To watch out for: small chinks in randomness!

- ▶ Daily  $\sigma$  of returns was 2.95;  
This implies variance of 8.7025.
- ▶ The AR(6) residual has variance of 8.68.  
This comes to a  $R^2$  of 0.0025.
- ▶ Even if there appears to be *statistically significant* correlation, we know:  
smart speculators are attacking the data every day, discovering patterns, making money on them, and thus eliminating these correlations,  
if they find it *economically efficient* to do so.
- ▶ EMH is a story about speculators who learn.

## Summary statistics about weekly returns

```
> rw <- prices2returns(p)
> summary(r)
      Index                r
Min.   :1990-01-12   Min.   : -37.9039
1st Qu.:1994-02-25   1st Qu.:  -3.4892
Median :1998-01-09   Median :   0.1102
Mean   :1997-12-25   Mean    :   0.3082
3rd Qu.:2001-11-23   3rd Qu.:   3.9093
Max.   :2005-10-07   Max.    :  30.3682
> sd(r)
[1] 6.551645
>
```

# ACF for M&M weekly returns





## AR model for weekly returns

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.04835	0.03516	1.375	0.1691
ar2	0.03169	0.03514	0.902	0.3671
ar3	0.05501	0.03513	1.566	0.1174
ar4	0.05286	0.03510	1.506	0.1320
ar5	-0.06485	0.03510	-1.848	0.0646 .
ar6	-0.01281	0.03511	-0.365	0.7152
intercept	0.30832	0.23007	1.340	0.1802

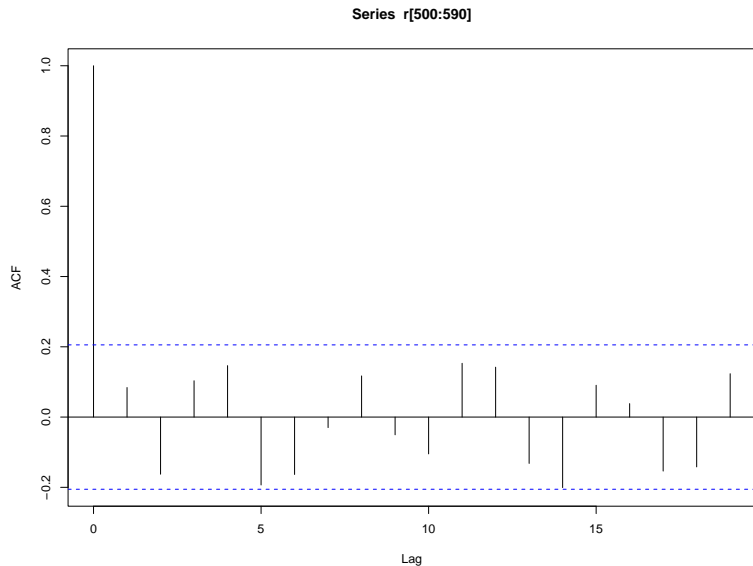
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1

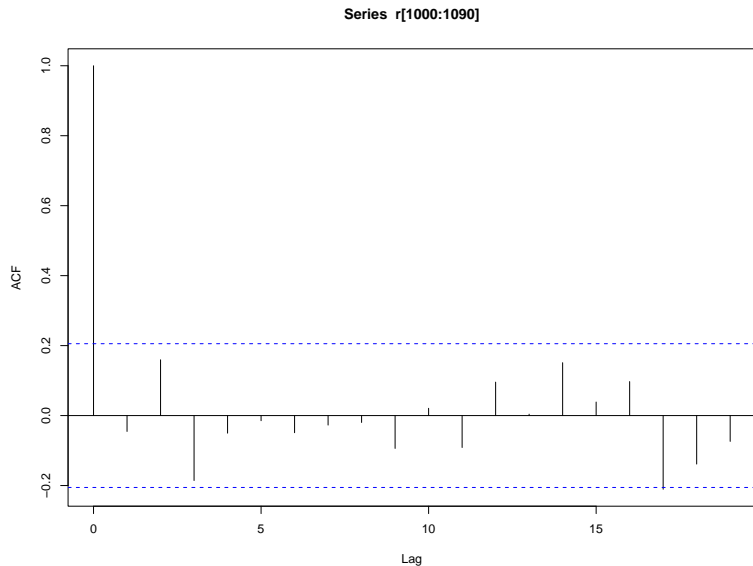
Fit:

$\sigma^2$  estimated as 42.37

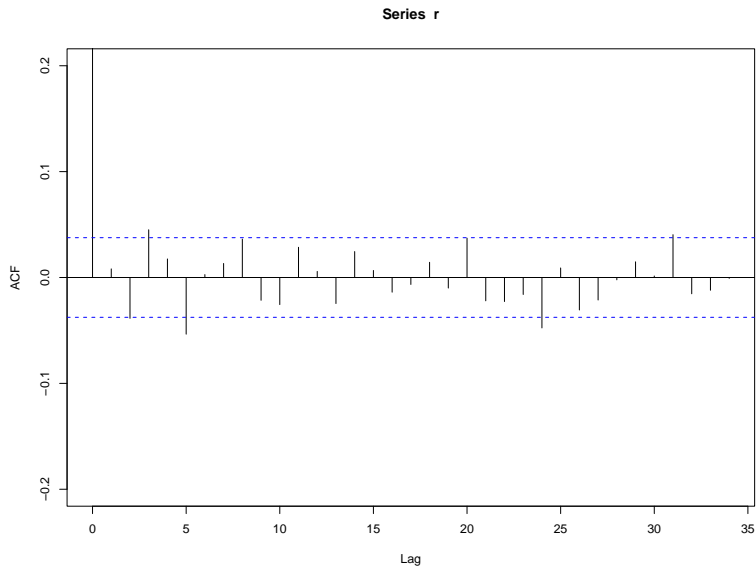
# Example: Nifty, 90 days – segment 1



# Example: Nifty, 90 days – segment 2



# Example: Nifty, 2000 days



## Variance ratio: definition

- ▶ If innovations are independent, and the distribution has constant variance, then  $\sigma_K^2$ , the variance of returns over  $k$  periods is  $K\sigma_1^2$ .
- ▶ The Variance Ratio at lag  $K$  is defined as  $VR(K)$  where

$$VR(K) = \frac{V(K)}{V(1)} \frac{1}{K}$$

- ▶ Under the null of iid returns,  $VR(K) = 1$  for any  $K$ .

# From the idea of $\sqrt{T}$ scaling to a test

- ▶ Okay, so we believe that in a fairly efficient, homoscedastic market, we will get  $\sqrt{T}$  scaling of volatility.
- ▶ But how can we look at data from the realworld and reject the null?
- ▶ This need a test.

## Calculating $VR(k)$

- ▶ In order to calculate  $V(k)$ , daily returns are aggregated over  $k$  periods.
- ▶ Cochrane 1988 formulation:

$$VR(k) \sim 1 + 2 \sum_{j=1}^{k-1} \frac{k-j}{k} \hat{\rho}_j$$

$\hat{\rho}_j$  is the estimated autocorrelation coefficient at lag  $j$ .

- ▶ Fama and French 1988, 1989:

$$r_{t,t+k} = \alpha_k + \beta_k r_{t-k,t} + \epsilon_{t,t+k}, \text{ and}$$
$$\beta_k \sim \frac{\hat{\rho}_1 + 2\hat{\rho}_2 + \dots + (k+1)\hat{\rho}_{k+1} + \dots + \hat{\rho}_{2k-1}}{k + 2[(k-1)\hat{\rho}_1 + \dots + \hat{\rho}_{k-1}]}$$

where  $\beta_k$  is distributed around 0, and negative values indicate mean reversion.

## Inference for $VR(k)$

- ▶ The test statistic has to be adjusted for the heteroskedasticity.
- ▶ Lo, Mackinlay (1988): a heteroskedasticity consistent estimator for  $VR(k)$ .

$$\sqrt{T}(VR(k) - 1) \sim N(0, \theta_k)$$

where

$$\theta_k = 4 \sum_{i=1}^{T/k-1} \left(1 - \frac{i}{k}\right)^2 \hat{\delta}_i$$
$$\hat{\delta}_i = T \sum_{j=i+1}^T \frac{\sigma_j^2 \sigma_{j-i}^2}{\sigma_j^4}$$

- ▶ Kim, Nelson, Startz (1988) use bootstrap and randomisation to infer the distribution for the VR.



# Economic interpretation of the VR observations

- ▶ When prices show positive deviations from 1 in the short term, followed by negative deviation in the longer term, it is referred to as the “mean-reversion” property of prices.
  - ▶ Prices over-react and overshoot the “mean-level” prices initially ( $VR > 1$ ).
  - ▶ Prices then “revert” to the mean over a longer period.
- ▶ The earlier literature also identified varying **magnitudes** of mean-reversion in different periods.  
For example, mean-reversion was much stronger in the pre-WWII period as compared to in the post-WWII period.

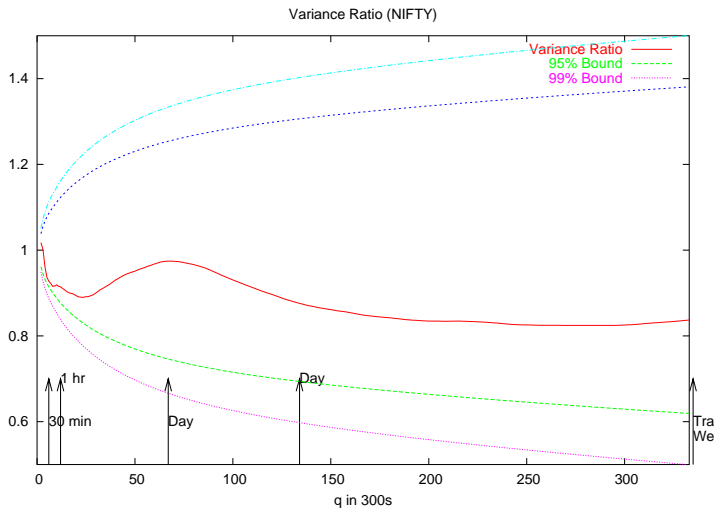
## Causes for mean-reversion

- ▶ On the short-run, bid-ask spread causes a negative serial correlation: Roll (1984).
- ▶ Across stocks of different liquidity, those with higher liquidity will have smaller serial correlation: Hasbrouck (1991).
- ▶ For a portfolio containing stocks of different liquidity, the same information will get absorbed sooner by some stocks, a little later by others.  
This ought to cause positive serial correlation in an index: Lo and Muthuswamy (1996).

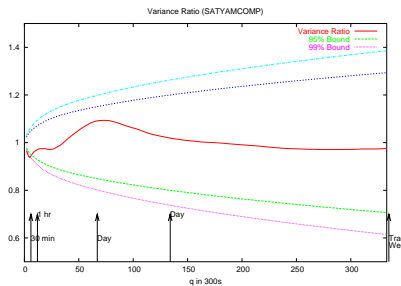
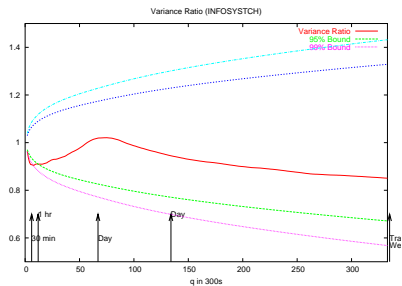
# Causes for mean-reversion

- ▶ HF Finance: These deviations are even more pronounced when the horizon reduces to within the day – to hour/minutes/seconds.
- ▶ The behaviour of the VR using extremely high frequency data becomes a story of how information transmits into prices.  
This can be studied at the level of individual stocks, pairs of stocks and the entire market.
- ▶ HF data helps trace out the path of market efficiency.

# Serial correlation in Nifty, March 1999 to February 2001



# Serial correlation in IT stocks, March 1999 to February 2001



# Serial correlation in manufacturing stocks, March 1999 to February 2001

