

Certificate in Quantitative Finance

Asset Returns: Empirical Stylized Facts

Lecture notes provided by:

Stephen Taylor

Emeritus Professor, Lancaster University Management School, UK

s.taylor@lancaster.ac.uk

25 July 2023

Relevant chapters from Stephen Taylor's book:

Asset Price Dynamics, Volatility, and Prediction

This lecture

Volatility clustering	2
Properties of daily returns	4
Properties of high-frequency returns	12

Next lecture

Volatility concepts	8
GARCH models	9 & 10

Background material

Random variables and stochastic processes	3
---	---

After today's class you will:

- Know about the most important empirical properties of asset returns,
- Know that changes in volatility explain many empirical effects,
- Have seen several examples of time series,
- Have seen several examples of autocorrelations.

↳ *statistical properties of data, std & correlations*

Getting started

Most of the information that we consider is provided by time series of asset prices.

The integer variable t represents *time* and it counts trading periods.

Very often, one trading period corresponds to one day of trading.

p_t denotes a *price* for period t . The price is usually measured at the end of the period.

A set of observations, in time order, defines a *time series*.

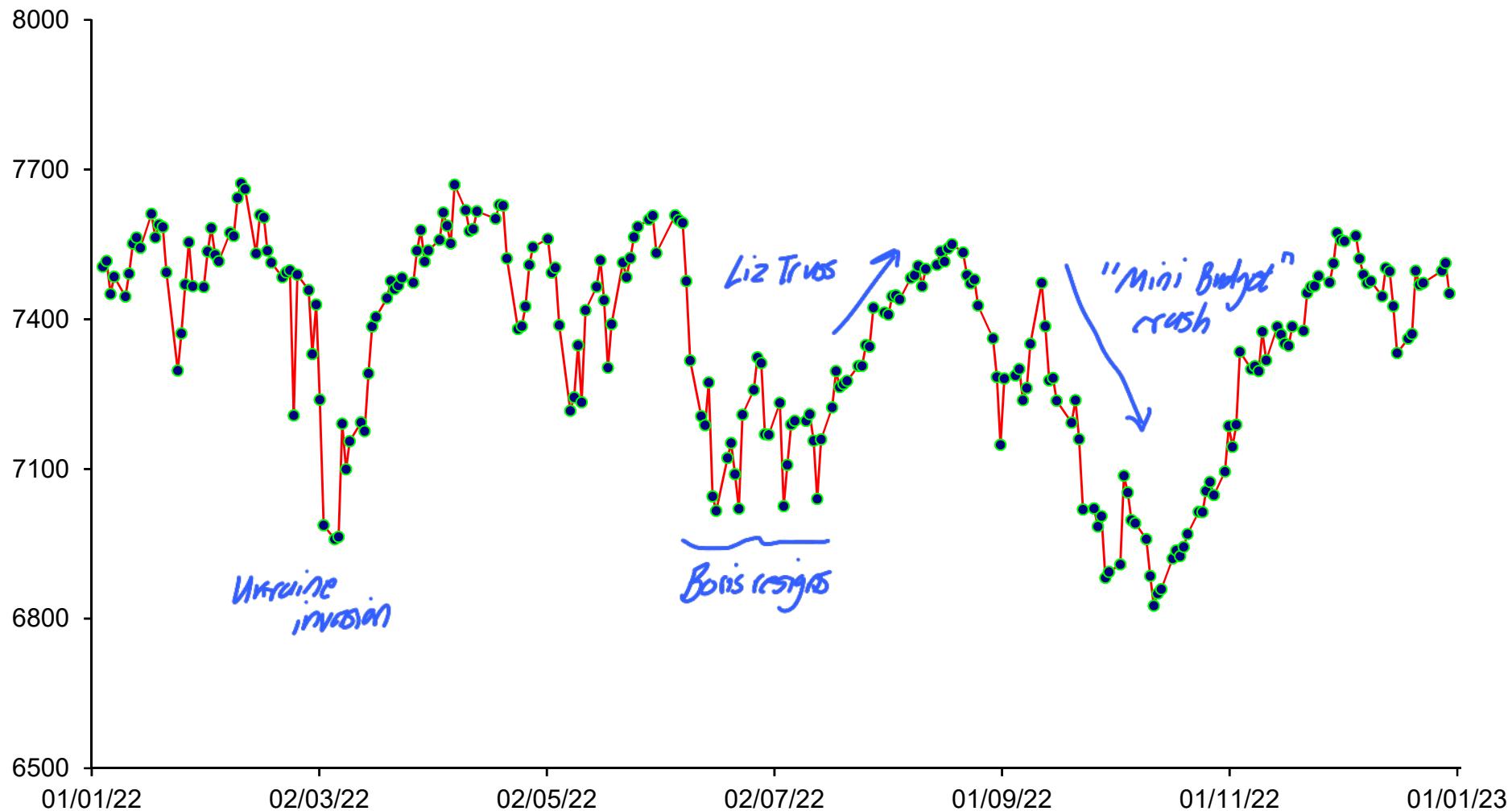
A recent example

The graph on the next page shows the time series of daily closing levels for the FTSE 100-share index from January 2022 to December 2022.

The subsequent examples are taken from the textbook and are consequently more dated.

252 trading days → green dots → closing prices

One year of FTSE 100 index levels, January 2022 to December 2022



We will encounter time series for:

- Prices,
- Returns,
- Squared returns, and other functions of returns,
- Realized variance,
- Historical and [redacted].

for option prices

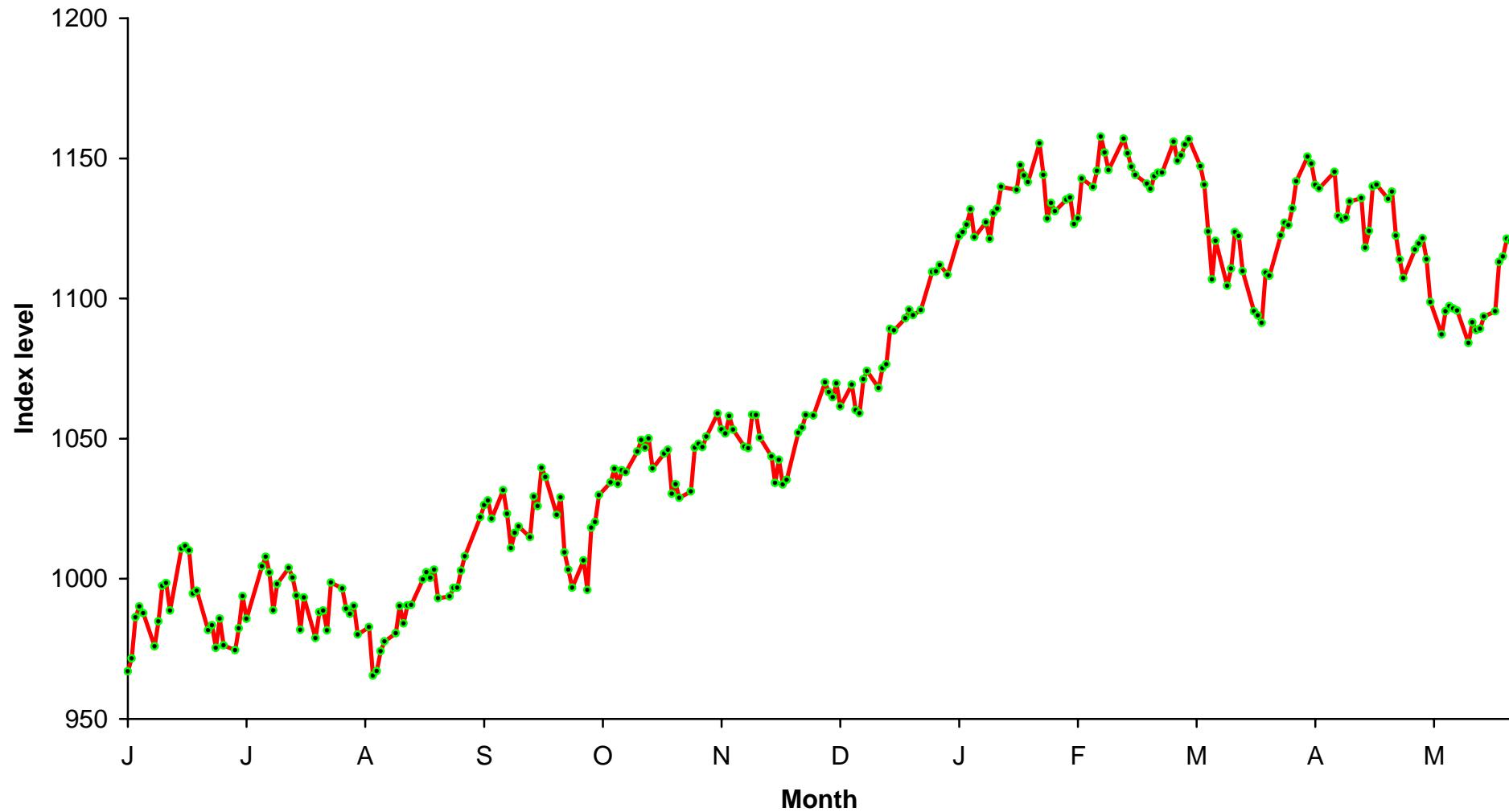
Textbook examples

- Figure 1.1 shows the time series of daily closing levels for the Standard and Poor's (S & P) 500-share index from June 2003 until May 2004.
- Figure 1.2 shows a time series of daily observations of σ

→ implied volatility

Recommended reading: Chapter 1 of the textbook, Asset Price Dynamics, Volatility, and Prediction

Figure 1.1 A year of S & P 500 index levels

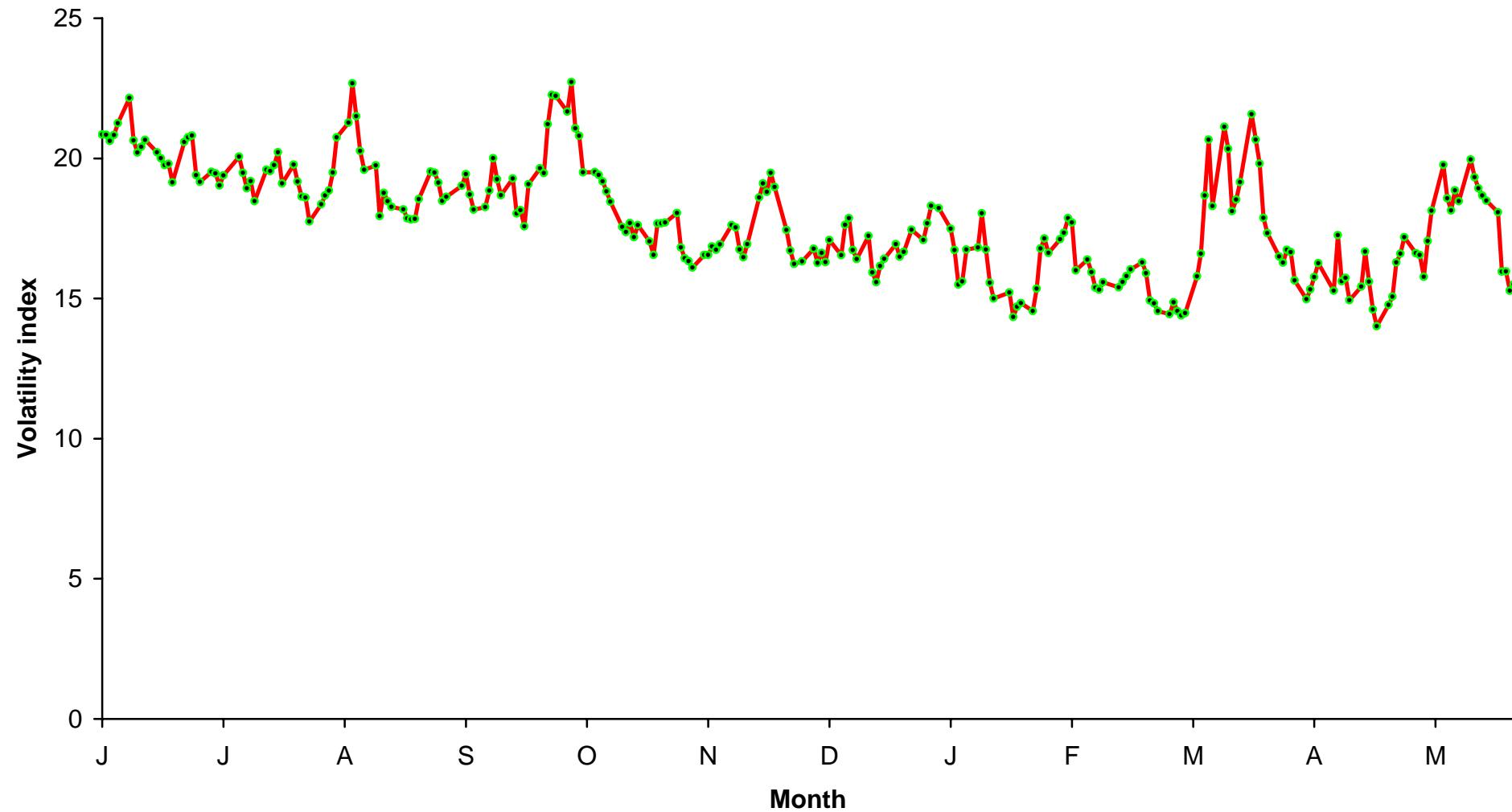


Prices: Random walk

Vol: Stationary or I1

↳ mean reversion

Figure 1.2 A year of VIX observations



Prices and returns

The word *price* may refer to a stock price, a stock index level, an exchange rate, a commodity price, a futures price, etc.

Let p_t be a representative *price* at the end of period t . Suppose we:

- buy an asset at time t ,
- receive a payoff at time t , and
- sell the asset at time t .

The *simple* return on our investment is then

$$r'_t = \frac{p_t + d_t - p_{t-1}}{p_{t-1}}, \quad = \quad \frac{\Delta \text{price} + \text{dividend}}{\text{original price}}$$

ignoring any transaction costs.

The *continuously compounded* return r_t is related to the simple return r'_t by

$$e^{r_t} = 1 + r'_t$$

and it equals the change in the logarithms of prices, adjusted for any dividend payments:

$$r_t = \log(p_t + d_t) - \log(p_{t-1}).$$

The return measures r_t and r'_t are very similar numbers, whenever returns are close to zero, because

$$1 + r'_t = 1 + r_t + \frac{1}{2} r_t^2 + \dots$$

It would be surprising if an important conclusion depended on the choice between the definitions r_t and r'_t .

We do so because the “ has the advantage that

a multi-period return is exactly the sum of single-period returns.

↳ not the case with simple discrete returns

Dividend payments are often:

- zero in most periods,
- ignored when returns are calculated, so then

$$\boxed{r_t = \log(p_t) - \log(p_{t-1})} |$$

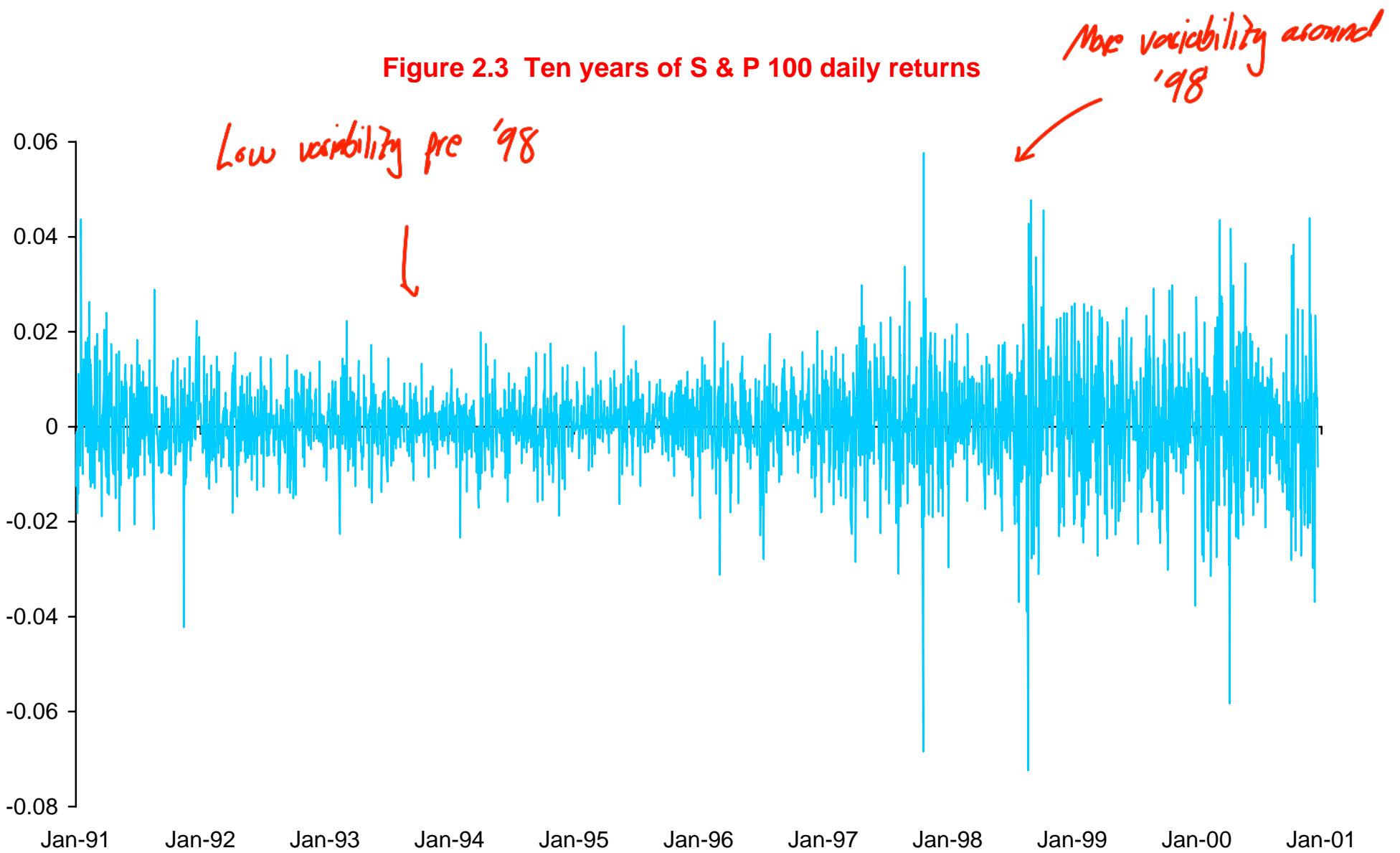
↳ therefore take logs of the returns

Example

The textbook contains several results for the S & P 100-share index,
from 1991 to 2000:

- The time series contains daily index closing levels,
- No dividend payments are included in the index, or in the return calculations,
- See Figure 2.3 for a plot of the daily returns.

Figure 2.3 Ten years of S & P 100 daily returns



Volatility has many definitions, but they all relate to the variability of returns, usually quantified by a standard deviation.

Fundamental key concept

- There are .
- M .

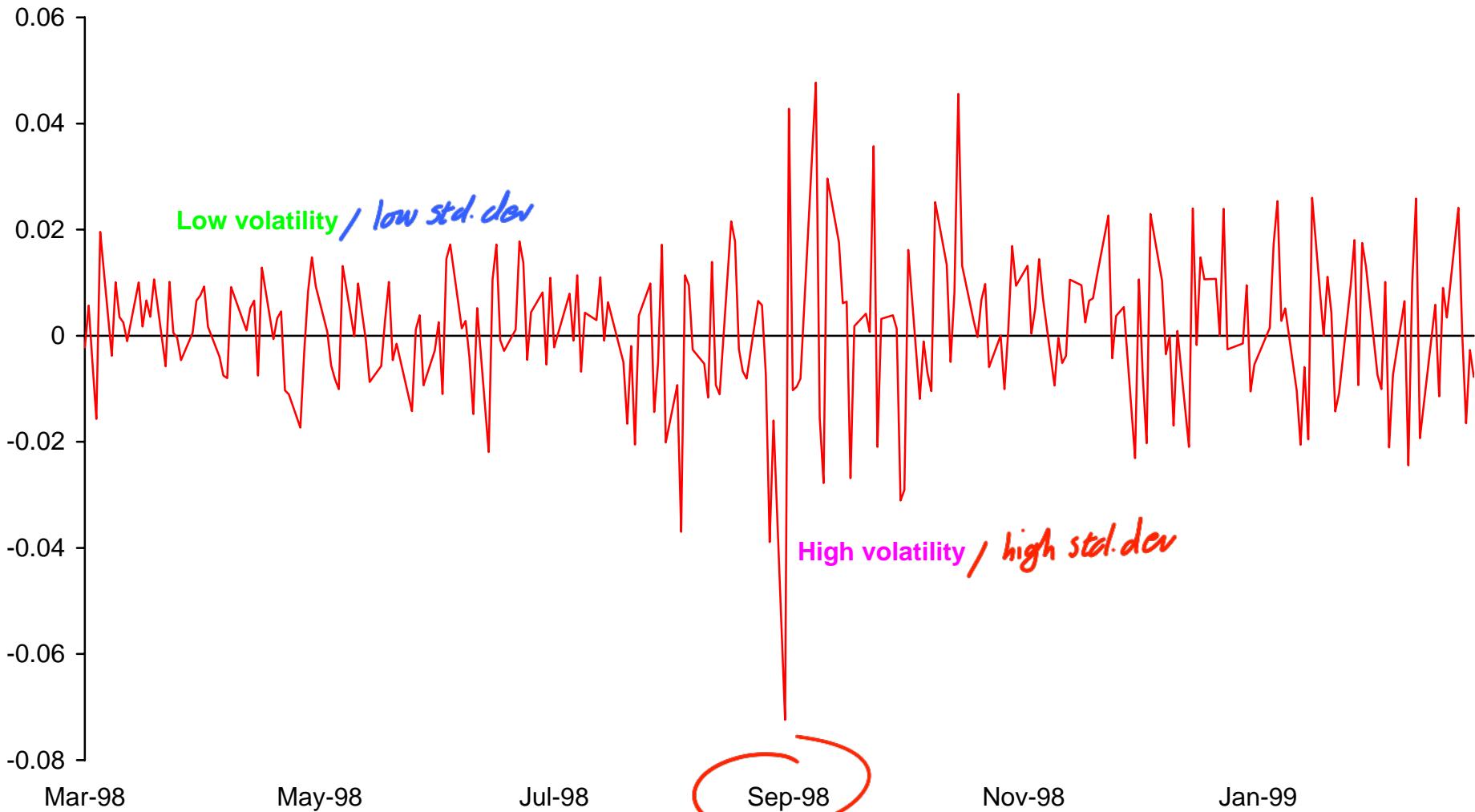
we don't know which way it will go but this property holds true

It may be difficult to see *volatility clustering* on Figure 2.3.

The effect is clearly seen during 1998, on Figure 2.5:

- High volatility cluster, late-August to mid-October.
- Low volatility before August, relative to volatility after October.

Figure 2.5 One year of S & P 100 daily returns



'98: Hedge Fund called Long Term Capital Management collapsed overnight
↳ caused turmoil in the markets, also Russian debt crisis around similar time

A recent example of prices, returns and volatility clustering

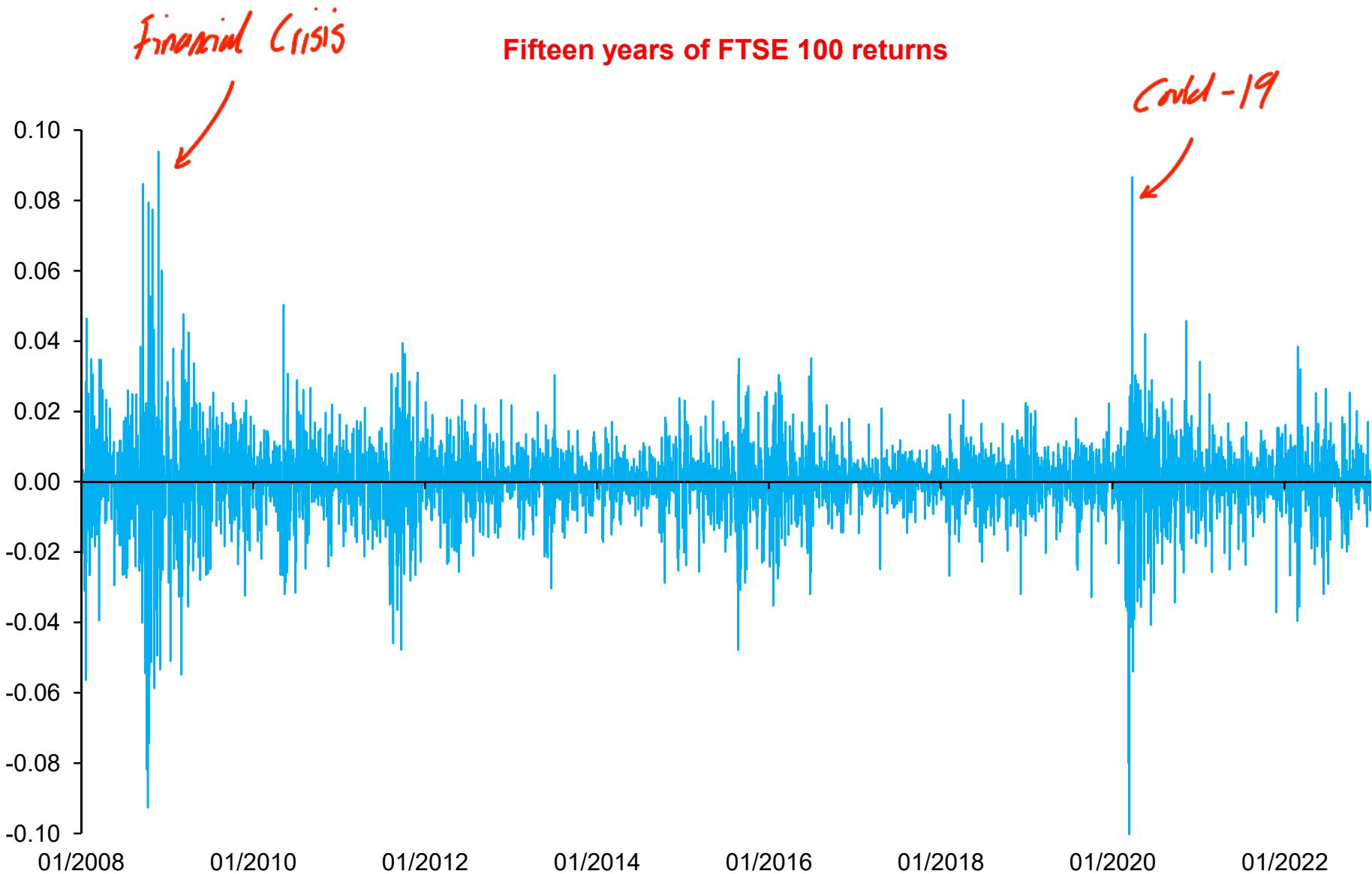
Fifteen years of FTSE 100 index levels:

- Daily closing levels, ending in December 2022.
- Volatility is generally high for 4 years, after which it is generally lower for 9 of the next 11 years.
- Volatility is exceptional from October 2008 until December 2008, and likewise during March and April 2020.

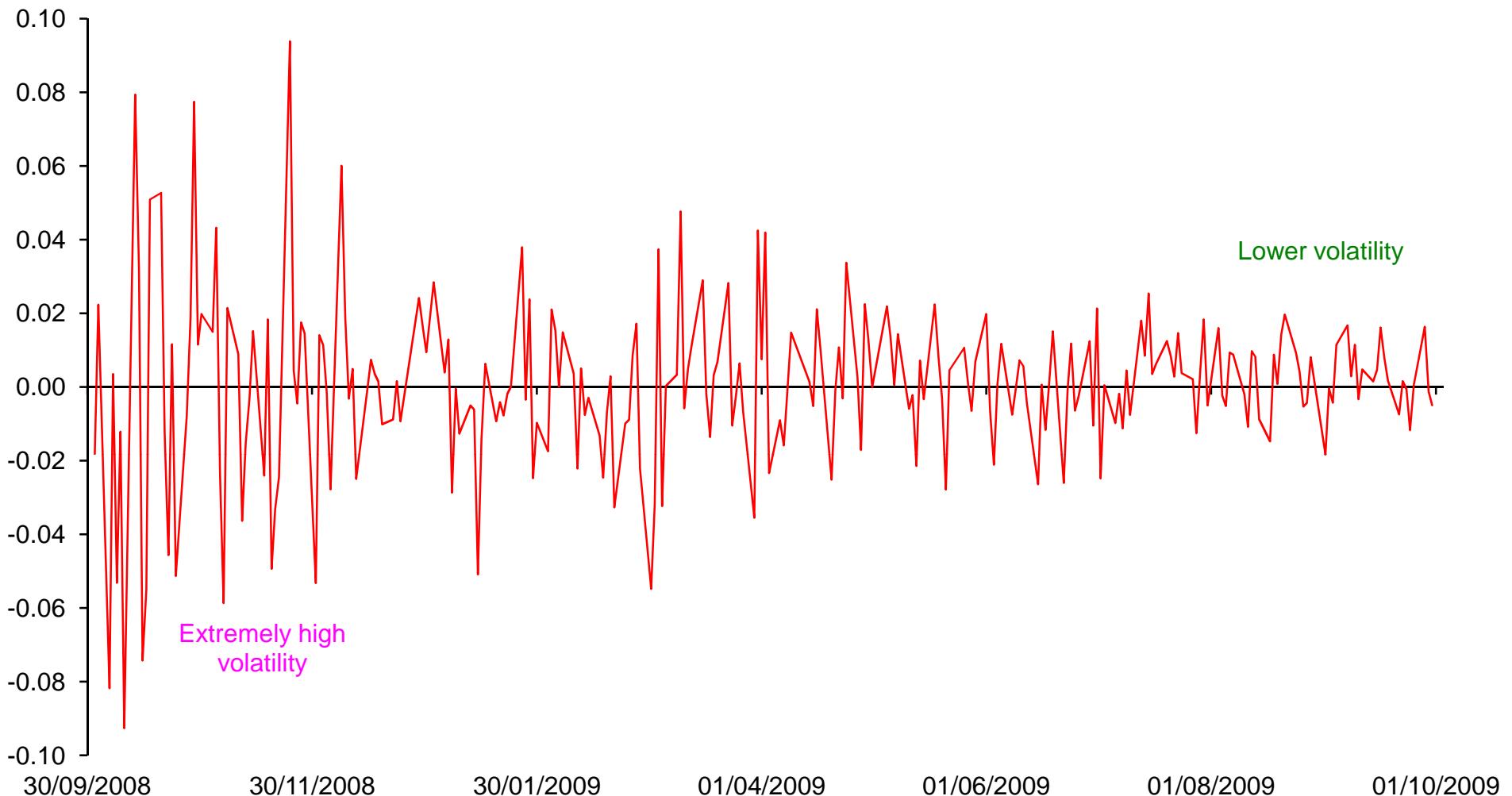
Recommended reading: Chapter 2, excluding Section 2.6.

15 years of FTSE 100 index levels, Jan. 2008 to Dec. 2022

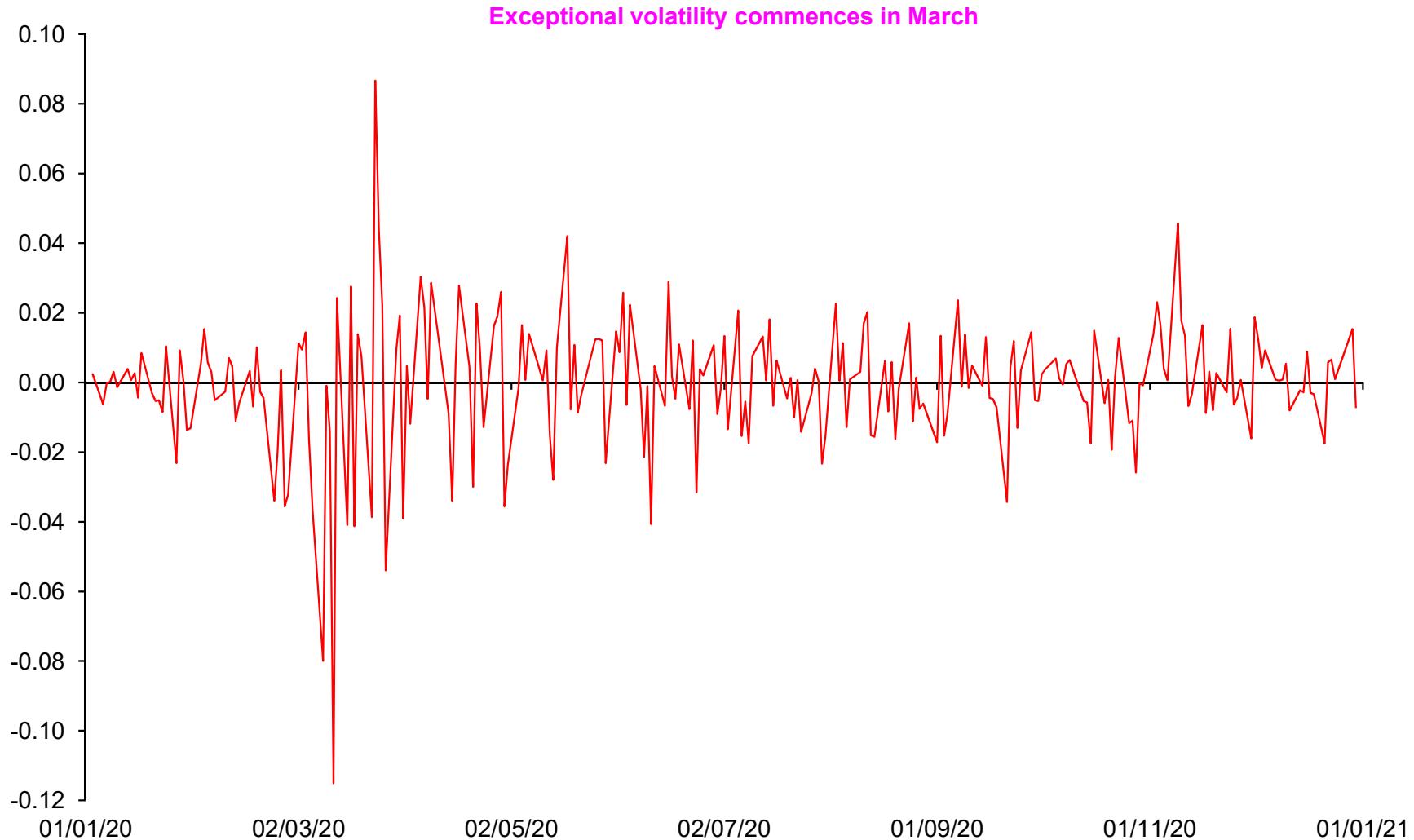




One year of FTSE 100 returns, Oct. 2008 to Sep. 2009

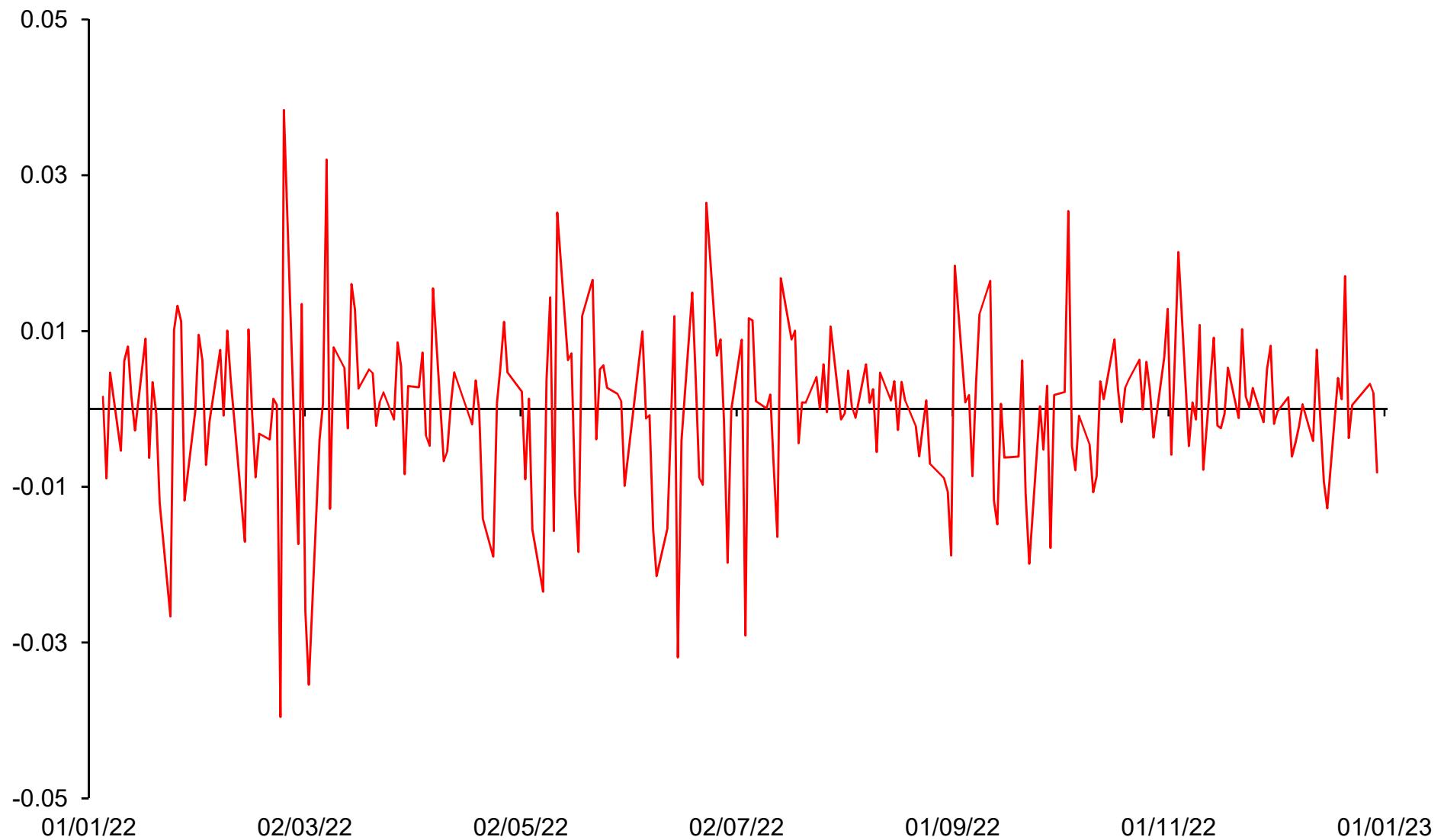


One year of FTSE 100 returns, January 2020 to December 2020



Volatility Clustering is everywhere and not to be ignored

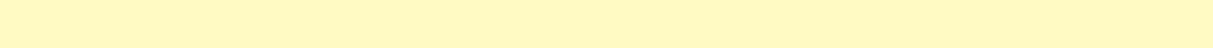
One year of FTSE 100 returns, January 2022 to December 2022



of daily asset returns

1. Means

The average return is small measured over one day, so consider averages over one year. These are:

-  
- Typically estimated from 20th century data as 

- 

Many calendar anomalies have been reported for average returns, sometimes based on 100 years of U.S. stock index returns.

Some anomalies have disappeared in recent years and we generally **ignore them**.

U.S. anomalies include:

Day-of-the-week – [REDACTED] than for other days of the week.

Day-of-the-month – Means are much higher in the first half than the second half of the month.

Month-of-the-year – [REDACTED]
especially for small firms.

Holidays – Returns on [REDACTED] than for other days.

Evidence of these largely gone and have been traded out

2. Standard deviations

Estimates vary substantially for short time series because of volatility clustering. Long time series provide the following approximate ranges for standard deviations:

The diagram illustrates the formula σ/\sqrt{N} where $N = 252 \text{ days}$. A red bracket groups the first two columns under the heading "One day return". An arrow points from this bracket to the third column under the heading "One year return".

One day return		σ/\sqrt{N} $N = 252 \text{ days}$
Currencies	0.6% to 0.9%	10% to 14%
Stock indices	0.7% to 1.3%	11% to 21%
Large US firms	1.2% to 2.0%	19% to 32%
Commodities	1.0% to 2.0%	16% to 32%

Generally the [REDACTED] ;

[REDACTED]

There is evidence for calendar effects in standard deviations.

The s.d. is a volatility measure and volatility can be expected to partially depend on

news. [REDACTED],

Thus the s.d. may be higher:

1. On Mondays than on other days,
2. On trading days that follow holidays.

Minor effects have been documented but they indicate that a relatively small amount of relevant information is announced when markets are closed.

Relevant reading: Sections 4.3 and 4.4.

Calendar effects are covered at length in Section 4.5, but they can be ignored.

3. The distribution of daily returns

A *stylized fact* is an empirical property that is observed for:

- (almost) all markets,
- (almost) all large datasets,
- (almost) all time periods.

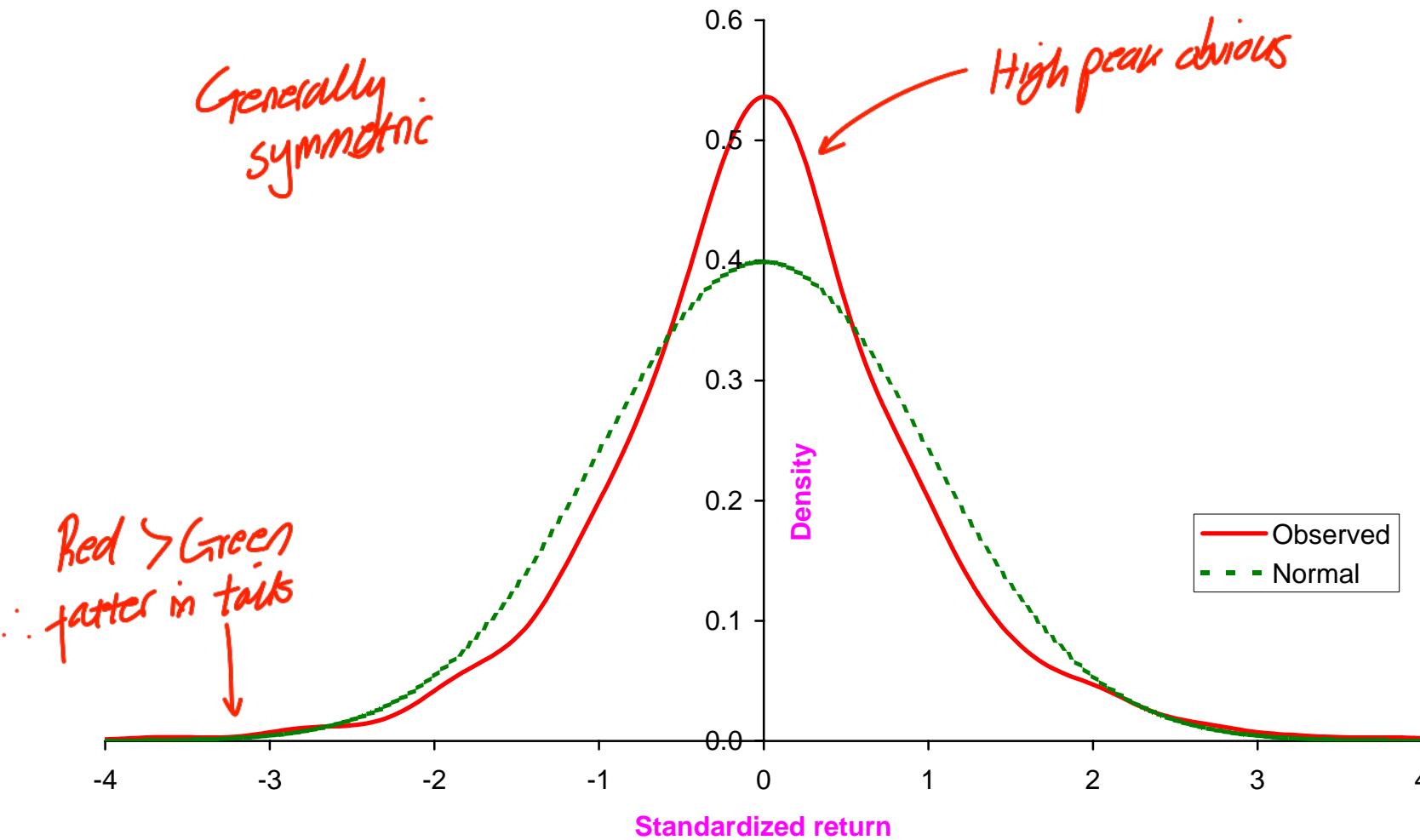
Instead, we can say of the distribution:

- It is approximately symmetric.
- It has *fat tails*.
- It has a *high peak*.

Figure 4.1 illustrates the density for S & P-500 returns.

We all see the high peak, but can we see the fat tails?

Figure 4.1
S&P-500 returns distribution



$$\text{Standardized return} = \frac{\text{Return} - \mu}{\sigma}$$

Compared with a normal distribution, estimates of the probability density of daily returns have:

- [REDACTED]
- [REDACTED]
- Less overall probability elsewhere.

Observations ... from the mean are often called ... and sometimes called *extreme values*.

They occur on average, very approximately, 3 to 4 times a year for daily returns.

↳ when looking over a very long time period

How big is a typical outlier?

- If the unconditional daily s.d. is 1.2%, from 15 recent years of FTSE-100 returns,
- And the FTSE-100 index is at 8000,
- Then 3 s.d. away from the mean is the same as a close-to-close rise or fall equal to approximately 288 points.

↳ shouldn't be surprised when we see this, mathematical expectation / probability

The frequency of daily returns more than 4 s.d. from the mean is around one or two a year; it is one every sixty years for a normal distribution.

4th moment scaled by
 $(\text{2nd moment}/\sigma)^2$

defined for n returns by :

$$k = \frac{\frac{1}{n-1} \sum_{t=1}^n (r_t - \bar{r})^4}{\left(\frac{1}{n-1} \sum_{t=1}^n (r_t - \bar{r})^2 \right)^2}.$$

Daily Returns → more outliers
Monthly / Annual Returns → closer to normal dist.

The null hypothesis that returns are both i.i.d. and come from a normal distribution is sometimes tested by evaluating

$$z = \frac{k - 3}{\sqrt{24/n}}$$

and comparing z with the standard Normal distribution. Usually large values of z are obtained, which reject the null hypothesis.

Non-normality has important implications for:

-
-
- - ↳ Black Scholes assumes all returns are normally distributed
 - ↳ other models that take into account non-normality

- Some returns in a sample come from higher volatility periods when the standard deviation of returns is relatively high. → *Assume not 'normal dist'*
- Other returns come from lower volatility periods. → *Assume normal dist*
- The complete sample is then obtained from a mixture distribution.
- Whenever returns come from a mixture of normal distributions, that have different variances, the mixture will have theoretical kurtosis > 3 .

For a proof, see Section 8.4.

We can **not** say which distribution best describes returns.



Some people have argued for a distribution which has infinite variance but the empirical evidence is solidly against this idea.

Recommended reading: Sections 4.6 and 4.7.

Section 4.8 includes a discussion of specific distributions for returns.

4. Correlation between returns

A scatter diagram for returns on consecutive days, r_t and r_{t+1} , will show that there is very little correlation between consecutive returns.

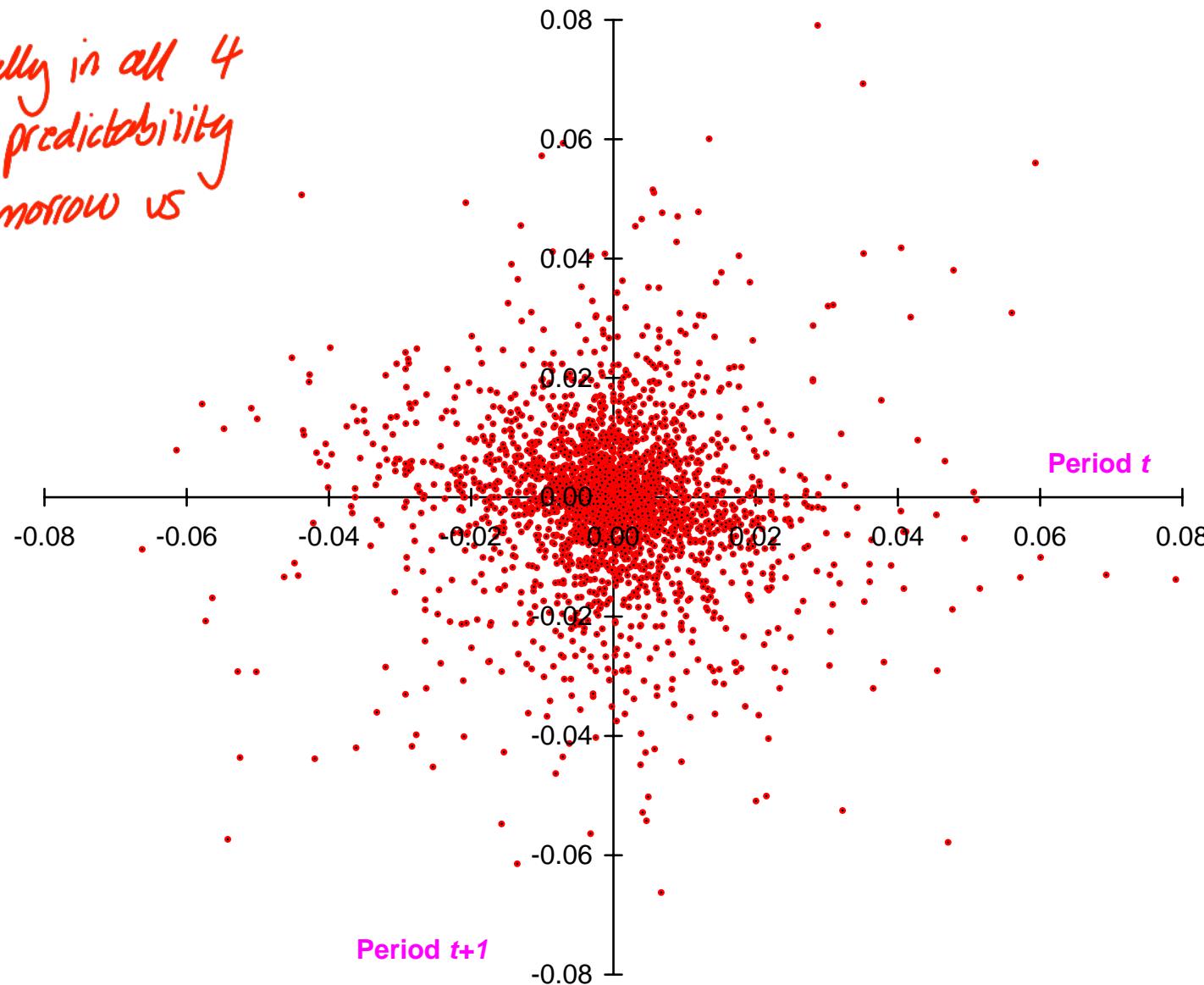
Example: Figure 4.5 is a scatter diagram for daily gold returns, 1981 to 1990.

τ is a time interval between returns

• Most red dots clustered around 0,0

• Dots fall equally in all 4 quadrants, no predictability
of returns tomorrow vs today

Figure 4.5
Gold returns in consecutive periods



Scatter diagrams show that there is



little or no correlation

The general lack of much correlation, regardless of the time that separates returns,
gives:

Statements like this can be found as far back as research by Working in 1934 and
Kendall in 1953.

auto : meaning correlation between itself at different time, period $t = \tau$

$\hat{\rho}$: hat meaning estimate from historic data

$$\hat{\rho}_\tau = \frac{\sum_{t=1}^{n-\tau} (r_t - \bar{r})(r_{t+\tau} - \bar{r})}{\sum_{t=1}^n (r_t - \bar{r})^2}$$

for a set of positive lags τ .

→ period

The hat symbol (^) indicates that

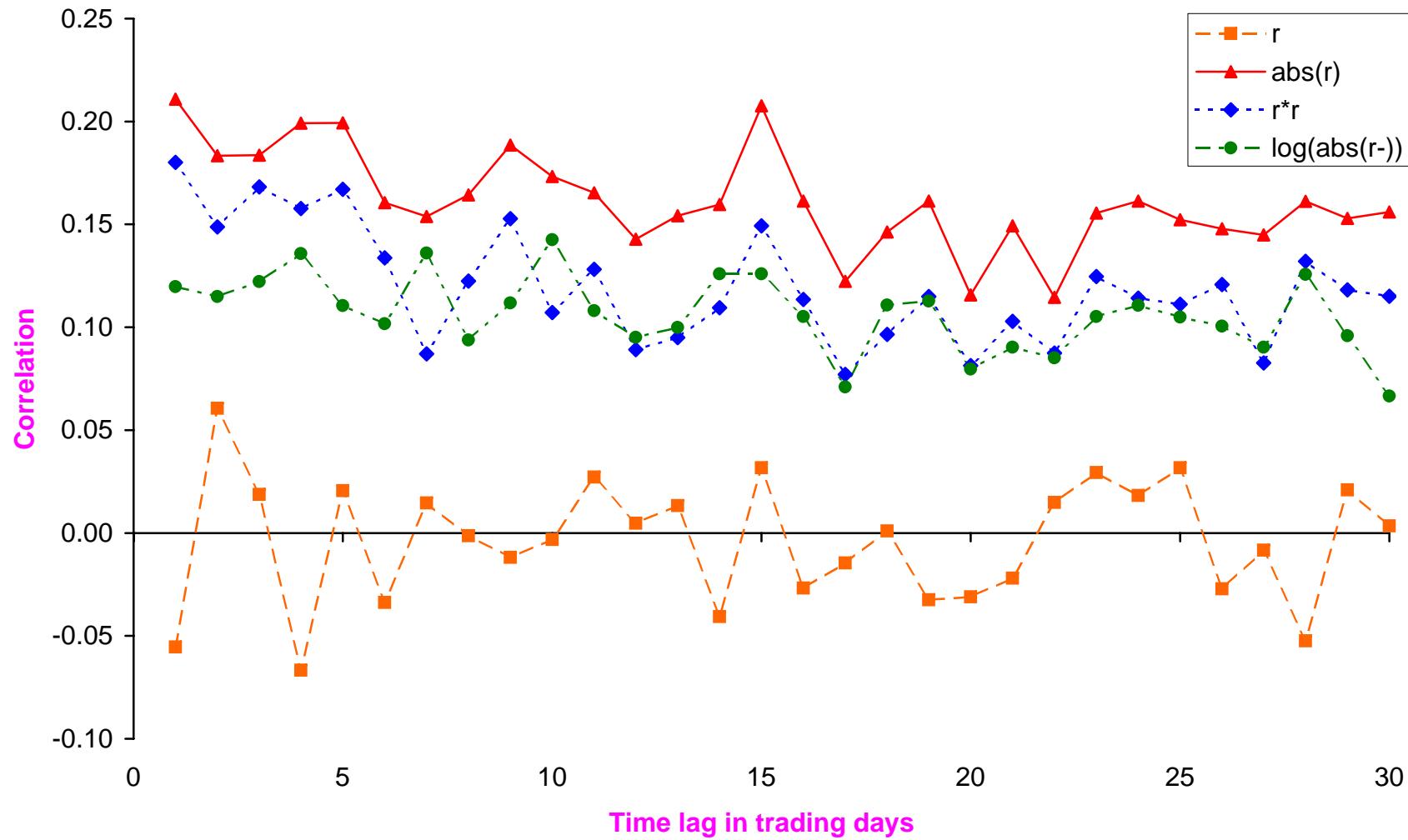
$\{R_t\}$ generates the observed returns, $\{r_t, 1 \leq t \leq n\}$.

There are other formulae that provide almost identical results, for example the Excel function CORREL can be used to find one autocorrelation.

Example: Figure 4.8 - The autocorrelations of the gold returns, at lags 1 to 30, are nearly all within the range from -0.05 to 0.05 for a series of 2522 returns.

The sample autocorrelation $\hat{\rho}_\tau$ can be considered to be an estimate of a population parameter ρ_τ . There are many tests of the null hypothesis of **no correlation**, $H_0 : \rho_\tau = 0$, but we will not discuss these. *If curious then see Chapters 5 and 6.*

Figure 4.8
Gold autocorrelations



The more specific hypothesis that r_t and $r_{t+\tau}$ are observations from 

i.i.d
↑

 can be tested more easily, as follows:

- Null hypothesis: the stochastic process is i.i.d.
- Sampling theory: $\hat{\rho}_\tau \sim N(0, 1/n)$, approximately, for an i.i.d. process.
- Test statistic: calculate $z_\tau = \sqrt{n} \hat{\rho}_\tau$ from n observations.
- Null distribution: for large n , can use $z_\tau \sim N(0, 1)$.
- Test result: for a 5% significance level, reject the null hypothesis if either $z_\tau < -1.96$ or $z_\tau > 1.96$.
- Here and elsewhere $N(\mu, \sigma^2)$ denotes the Normal distribution having mean equal to μ and variance equal to σ^2 .

The autocorrelation function can be combined into the Q statistic

which is routinely calculated by time series software:

$$Q_k = n \sum_{\tau=1}^k \hat{p}_\tau^2 = \sum_{\tau=1}^k z_\tau^2.$$

*n = number of observations
 \hat{p}_τ = autocorrelations*

The null hypothesis is then rejected if Q_k exceeds a critical point given by the chi-squared distribution with k degrees-of-freedom, χ_k^2 .

This test applies the result that the sample autocorrelations for lags 1 to k are almost independent random variables for an i.i.d. process.

For the gold series, whose autocorrelations are shown on Figure 4.8, $n \approx 2500$,

$\hat{\rho}_\tau \sim N(0, 0.02^2)$ and the null hypothesis is rejected at lags 1, 2 and 4 but not at the other lags (significance level 5%).

Also, $Q_{30} = 67.9$ and this exceeds the 5% critical point which is 43.8.

Recommended reading: Section 4.9.

5. Correlation between functions of returns

A scatter diagram for

Figure 4.6 is

an example for absolute gold returns.

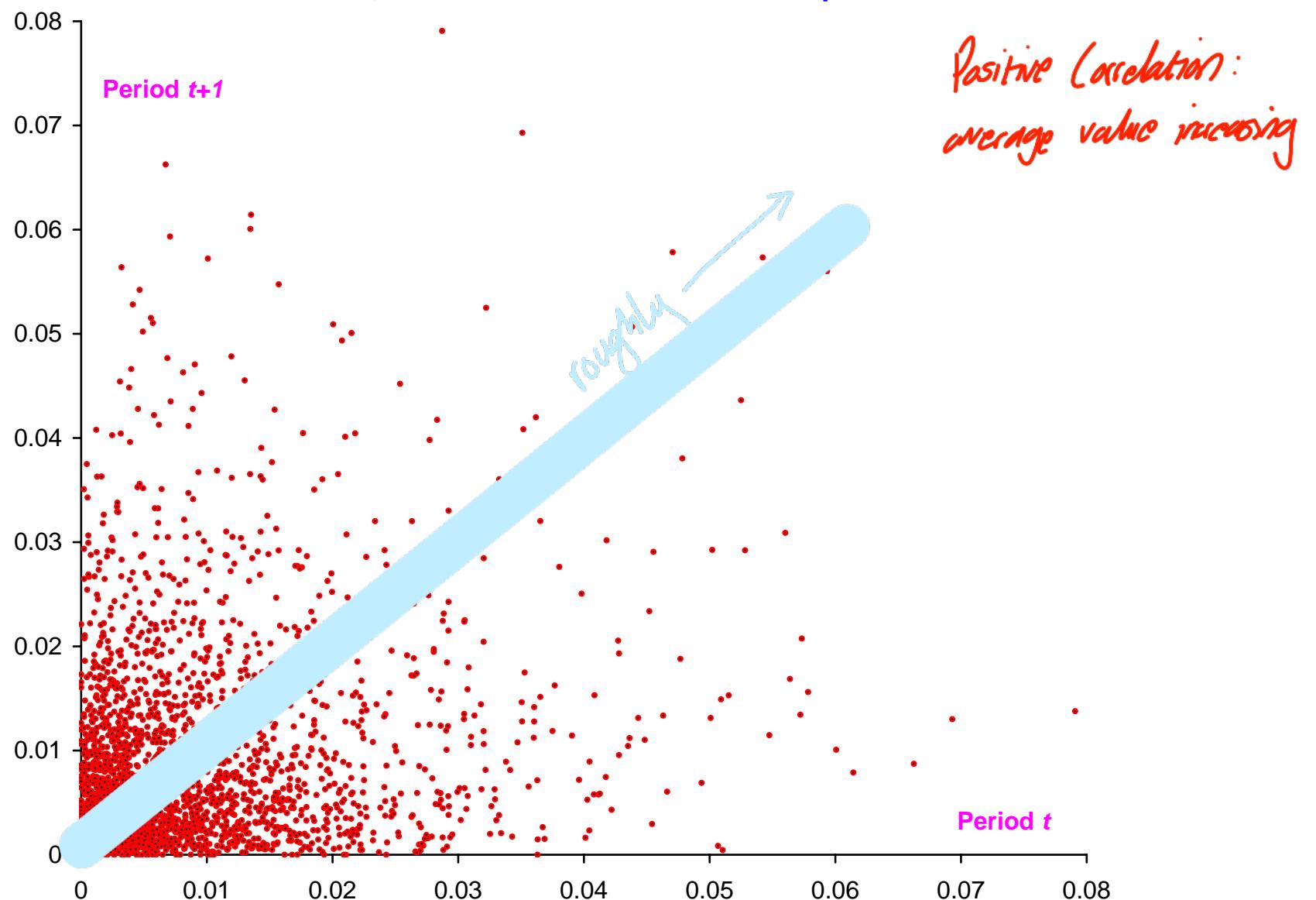
Sample autocorrelations can be calculated for several functions of returns, such as:

- absolute returns $|r_t|$,

we'll look at this next

- squared returns r_t^2 ,
- logarithms of mean-adjusted absolute returns $\log(|r_t - \bar{r}|)$.

Figure 4.6
Gold returns, absolute values in consecutive periods



These calculations produce positive sample autocorrelations which support:

the random walk hypothesis

→ doesn't work for returns themselves

See Figure 4.8 again for the gold series. All the autocorrelations for functions of absolute returns exceed the maximum autocorrelation for returns.

Large positive correlations for absolute returns:

- reject the hypothesis that absolute returns come from an i.i.d. process,
- consequently (and easily) show that **the returns process is not i.i.d.**

Gold example - Each sample autocorrelation for any of $|r_t|$, r_t^2 , $\log(|r_t - \bar{r}|)$, with lag between 1 and 30, rejects the i.i.d. hypothesis at the 5% level. Most reject at much lower levels. The values of the Q_{30} statistic are 1999, 1139 and 885 for these three functions of returns, compared with 68 for returns.

in SF3?

We can say the positive dependence in functions of returns persists for many lags.

Figure 4.12 shows averages of sample autocorrelations for 20 series.

is visible for hence there is

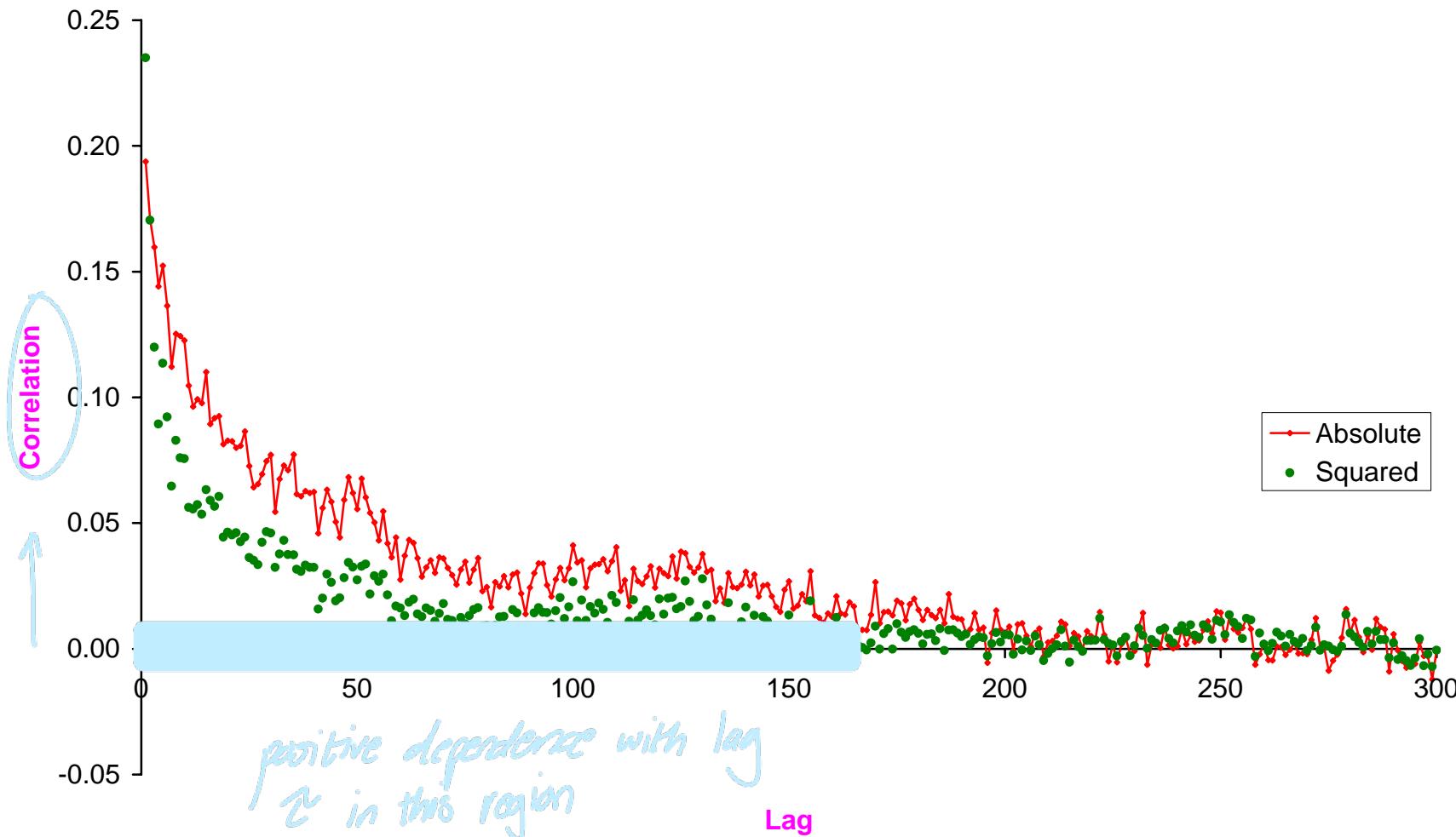
→ explained by vol. clustering

$0 \leq \tau \leq 6$ months:
positive dependence

Ding, Granger and Engle consider the autocorrelations of $|r_t|^\lambda$ for very many lags and

all positive λ . There is some evidence that the most linear dependence occurs when λ is near to 1.

Figure 4.12
Autocorrelations of absolute and squared returns, averages across 20 series



[REDACTED]

[REDACTED]

[REDACTED] on consecutive days ensures a relatively [REDACTED]

[REDACTED]

- Conversely, if [REDACTED] then it is [REDACTED] that both

[REDACTED]

- Then, [REDACTED]

[REDACTED]

For a proof, again see Section 8.4.

The [redacted] are the same as
those of [redacted]

They include:

- Volatility changes and can be predicted fairly accurately.
doesn't move fast through time
- In continuous time, prices do not follow geometric Brownian motion.
as GBM assumes constant vol.
- Option traders should apply improvements on the Black-Scholes formula.
to capture "not normal" returns distribution

Recommended reading: Section 4.10.

Properties of high-frequency asset returns

Prices can be analysed using observations:

- For every trade and/or quotation. This is often complicated, because :
 - Buy and sell quotes differ, with trades at bid, ask or some intermediate level.
 - Times between price observations are variable.
 - Often have thousands of observations a day, sometimes there are millions.

• At some intermediate time interval or one

every time interval or one every time interval. The time interval

Textbook example: FTSE-100 futures prices, March 2000 contract on 22 December 1999. The market was then electronic, order-driven and open from 08:00 to 17:30.

Figure 12.1 shows *transaction* (= *trade*) prices.

- One trade, on average, every 20 seconds.
- ... but, median inter-trade time 5 seconds.
- More frequent trading after markets open in the U.K. (197 from 08:00 to 08:30) and the U.S. (323 from 14:30 to 15:30).
- ... and

Figure 12.1 FTSE 100, March futures trades on 22 Dec 1999

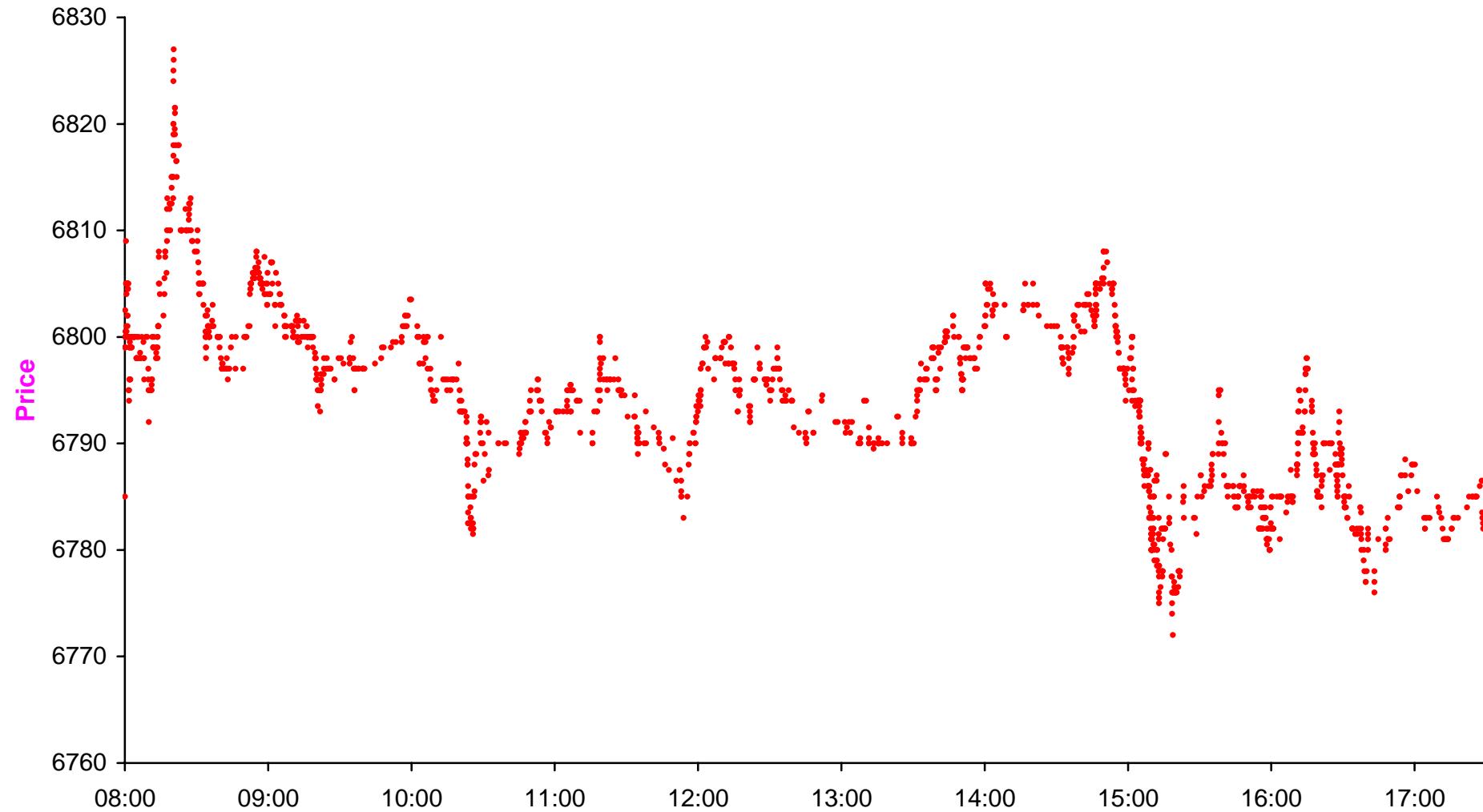


Figure 12.2 covers the 15 minutes from 15:45 to 16:00.

- [red] are shown by [blue] connected by lines.
- [red] are shown by [blue].
- Similar numbers of bid, ask and trade prices are shown (but can't be sure everything is recorded).
- Trades occur at either the most recent bid or the most recent ask.

Figure 12.3 shows the *spread* (= ask minus bid), for contemporaneous quotes.

- Minimum 0.5, maximum 12, mode 1, median 2, average 2.2.

Recommended reading: Sections 12.2 and 12.3.

Figure 12.2 Bids, asks and trades for fifteen minutes

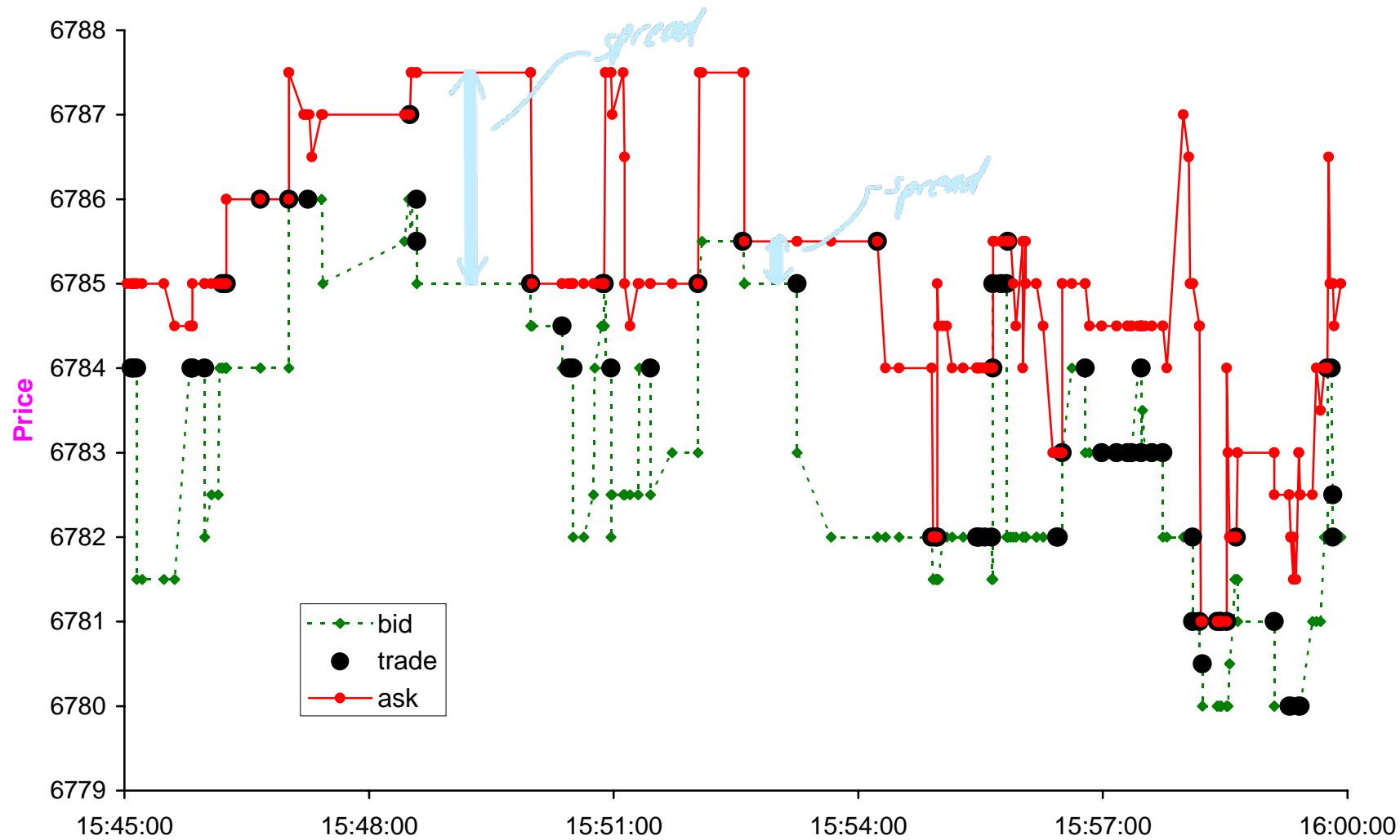
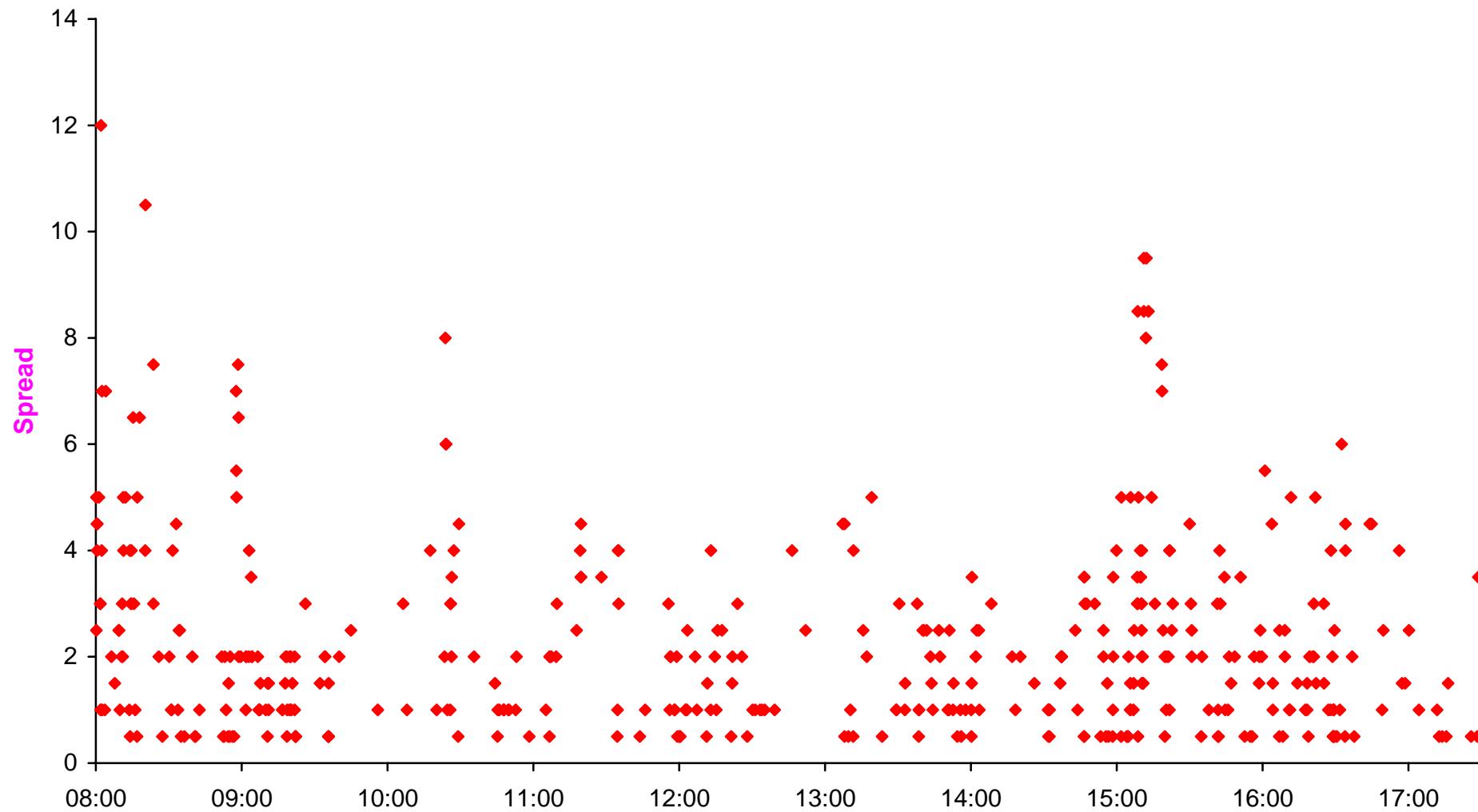


Figure 12.3 Bid/ask spreads



Another example: Prices for the S&P 500 ETF (SPY) on 3 January 2012, extracted from the TAQ database.

SPY trades on several U.S. exchanges.

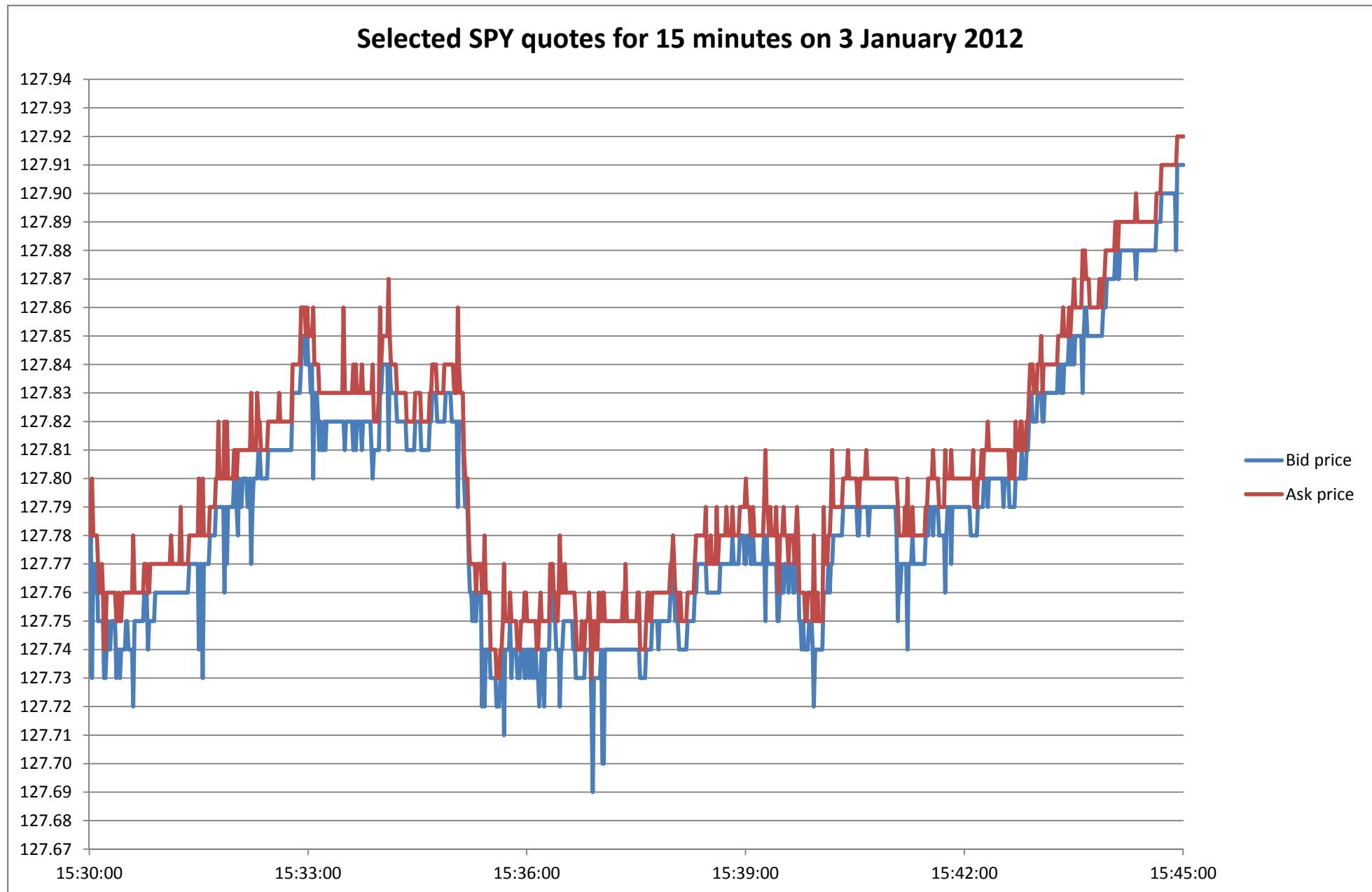
From 15:30 to 15:45,

15 minutes

- There are
- The chart shows the first pair of quotes for each second.
- These quotes are not always the most competitive.

huge quantity of data

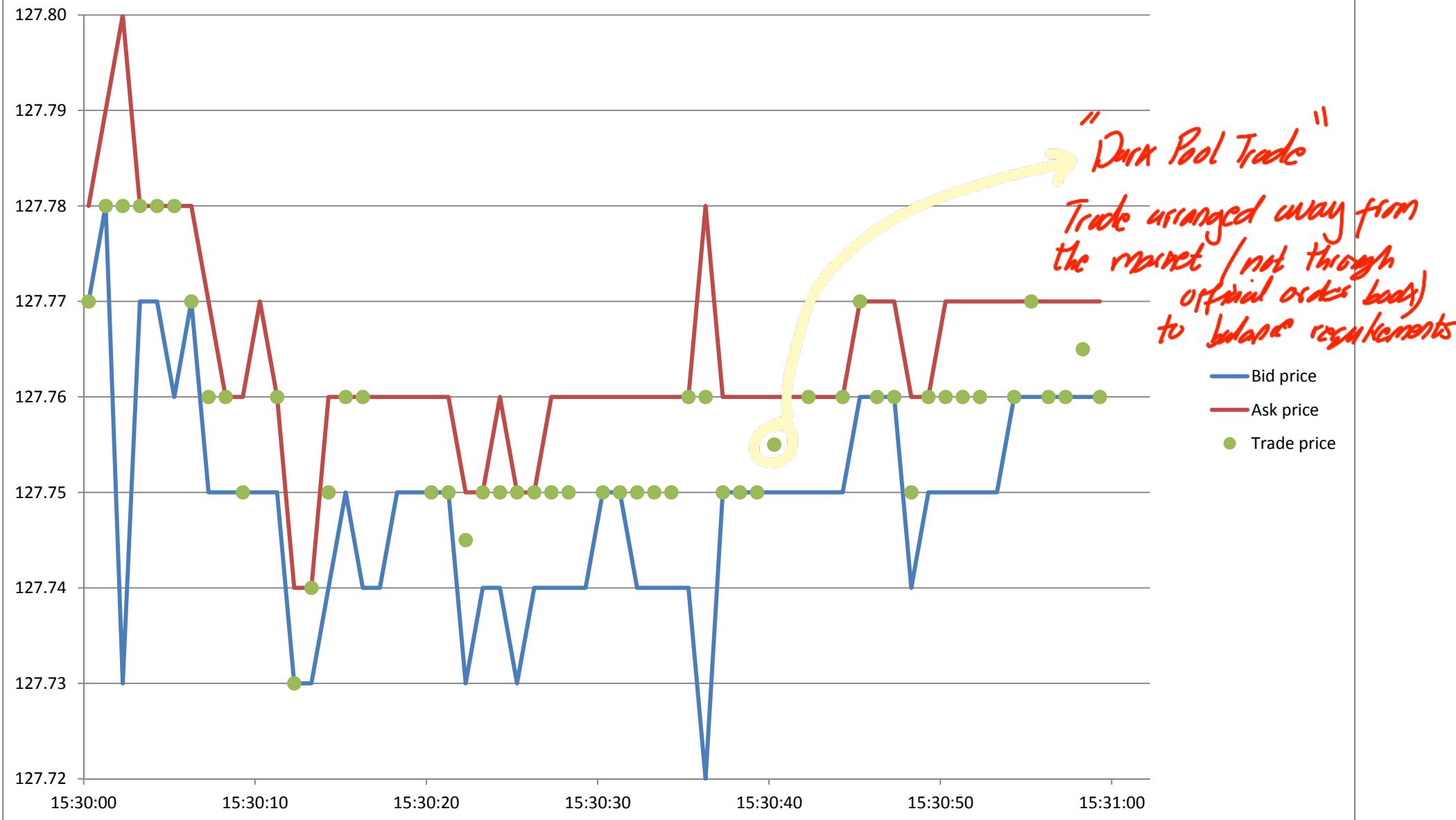
Spread usually one tick as this is most competitive ask - bid



From 15:30:00 to 15:30:59,

- The chart shows the first trade and the first pair of quotes for each second.
- A few seconds contain no trades.
- Trades times are *not* exactly equal to quote times.
- Some recorded trade prices include fractions of a cent; these are ‘dark pool’ prices.

Selected SPY trades and quotes for one minute on 3 January 2012



1. Stylized facts for intraday returns → 5 of these facts instead of 3 seen before

1. • [redacted]

→ Kurtosis ↑ as $\tau \downarrow$

2. • [redacted] Any important

dependence is usually negative and between consecutive returns. → see bid ask bounce feature

3. • There is [redacted]. It occurs at many low lags. [redacted]

as inter day absolute returns we saw before

↳ distinctive patterns between intra-day absolute returns when looking at a clockwork pattern

- Figure 12.4, FX pattern repeats after 48 thirty-minute periods (one calendar day).
- Figure 12.5, equity pattern repeats after 77 five-minute periods (one trading day).

4. • The with a
- *volatility clustering (daily)*
→ *AND inter-day patterns of volatility*
5. • There are short that follow

Recommended reading: Section 12.4.

Deutschmark / \$
30 min
absolute
returns

Figure 12.4 Autocorrelations for DM/\$ 30-minute absolute returns

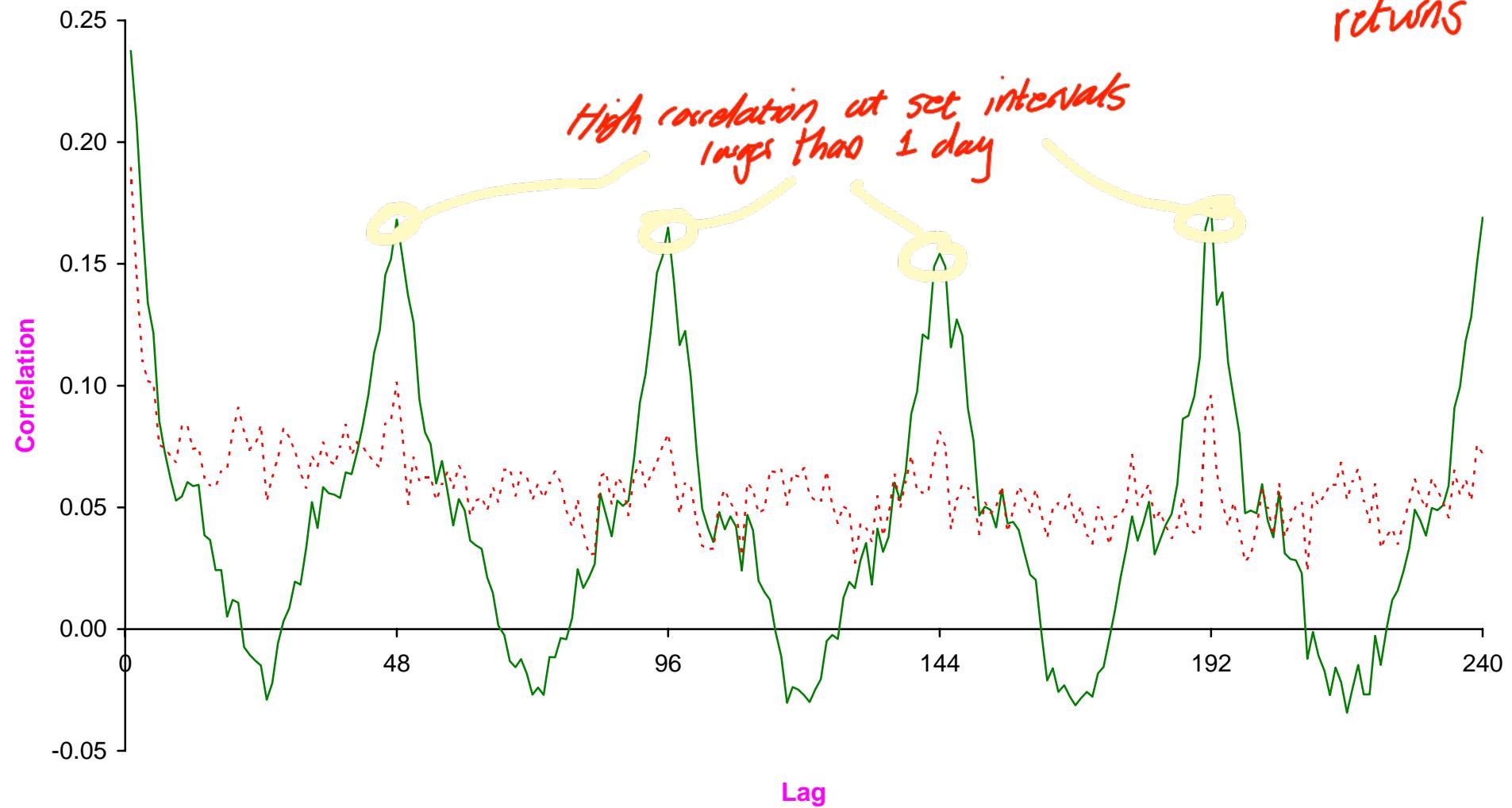
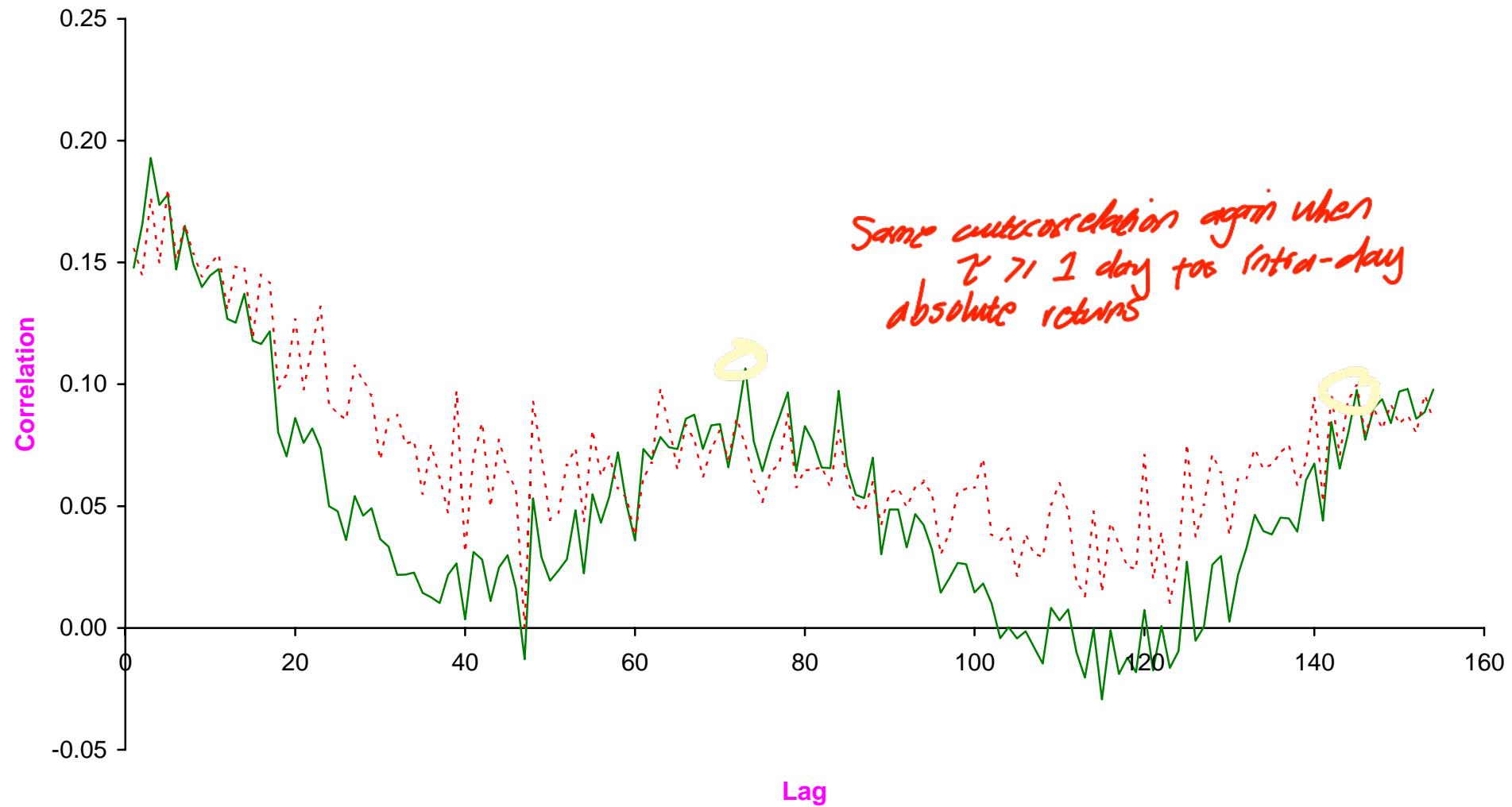


Figure 12.5 Autocorrelations for intraday absolute S & P returns



2. Intraday volatility patterns

→ What causes these vol. patterns?

There are distinctive volatility patterns that depend on:

- The [redacted]
- The [redacted]
- [redacted] about unemployment, trade balances, GNP, inflation, money supply, etc.

→ Important: fundamentals remain but patterns change

These [redacted], e.g. U.S. macro news around 14:00 EST has become more important than macro news at 10:00 EST in recent years.

t,

, from Nov. 1993 to July 1998.

It shows variance *proportions* (factors) for 91 five-minute intervals when the market was open - from 08:35 to 16:10 local time. These,

- Sum to 1 across the day.
- Are proportional to the sample variances of five-minute returns for the intervals.
- Are highest when the market opens at 08:35.
- Increase slightly when U.K. macro. news is announced at 09:30 local time.
- Decline from 09:35 to 13:30.

FIGURE 12.7

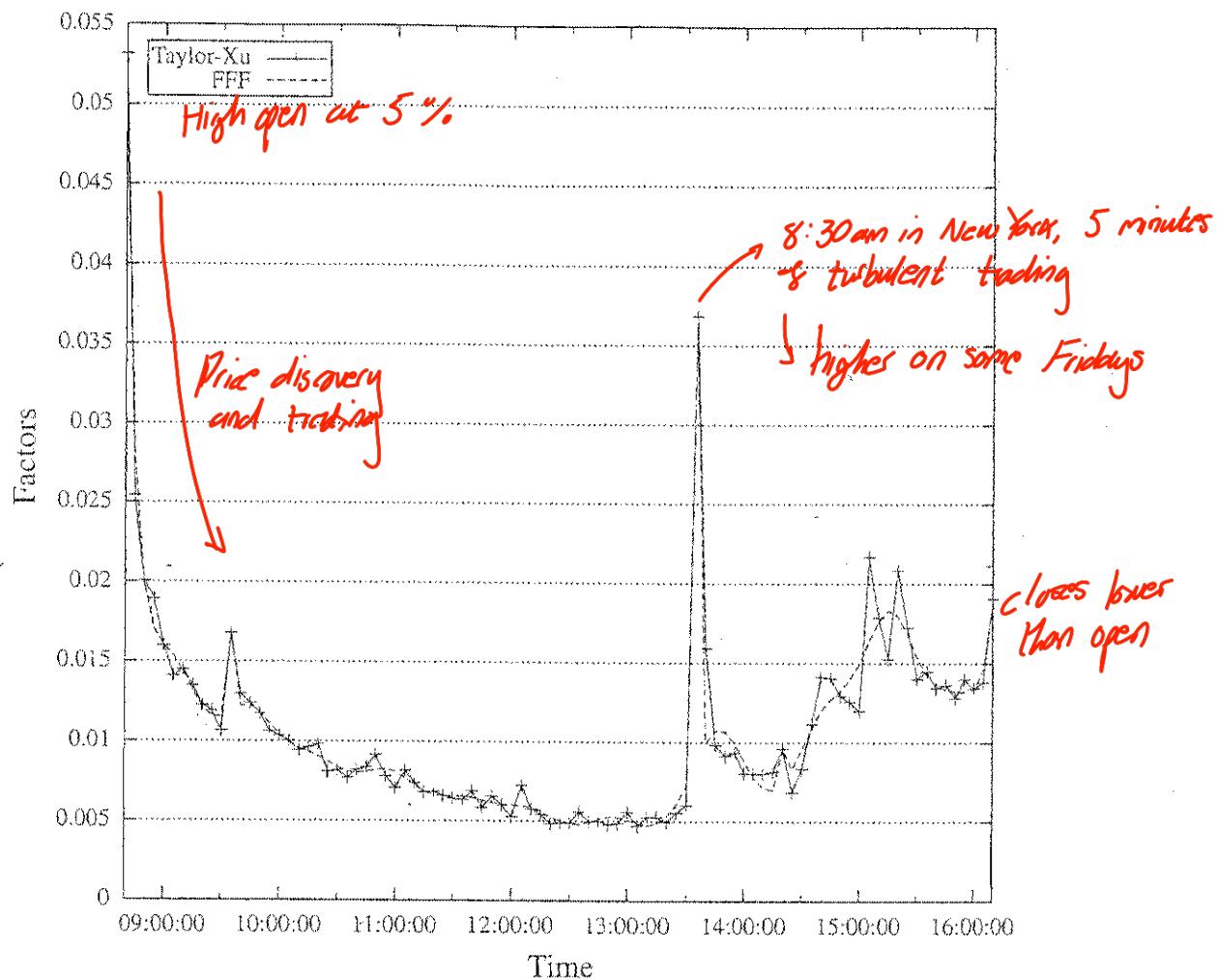


FIGURE 2

Five-minute fitted open-market variance proportions for the FTSE-100 futures index, using all days of the week, for the period from 18/11/1993 to 17/07/1998.

- **Market Volatility** is higher on **some** days, particularly some **Fridays**.
 - Are **market movements** random?
 - Only refer to the open-market period. About 30% of all price variation occurs when the market is closed.

Figure 12.8 shows estimates of the London pattern **by day** of the week.

- Monday has the highest volatility peak at the open, presumably reflecting the longer closed market period.
- Friday has the highest mid-day peak, because the most important U.S. macro. news is released then.
- Otherwise, the pattern is similar for all five days.

However, U.S. macro. news is only released on a half of all Fridays. Ederington and Lee show that the volatility spike only occurs on the announcement days.

FIGURE 12.8

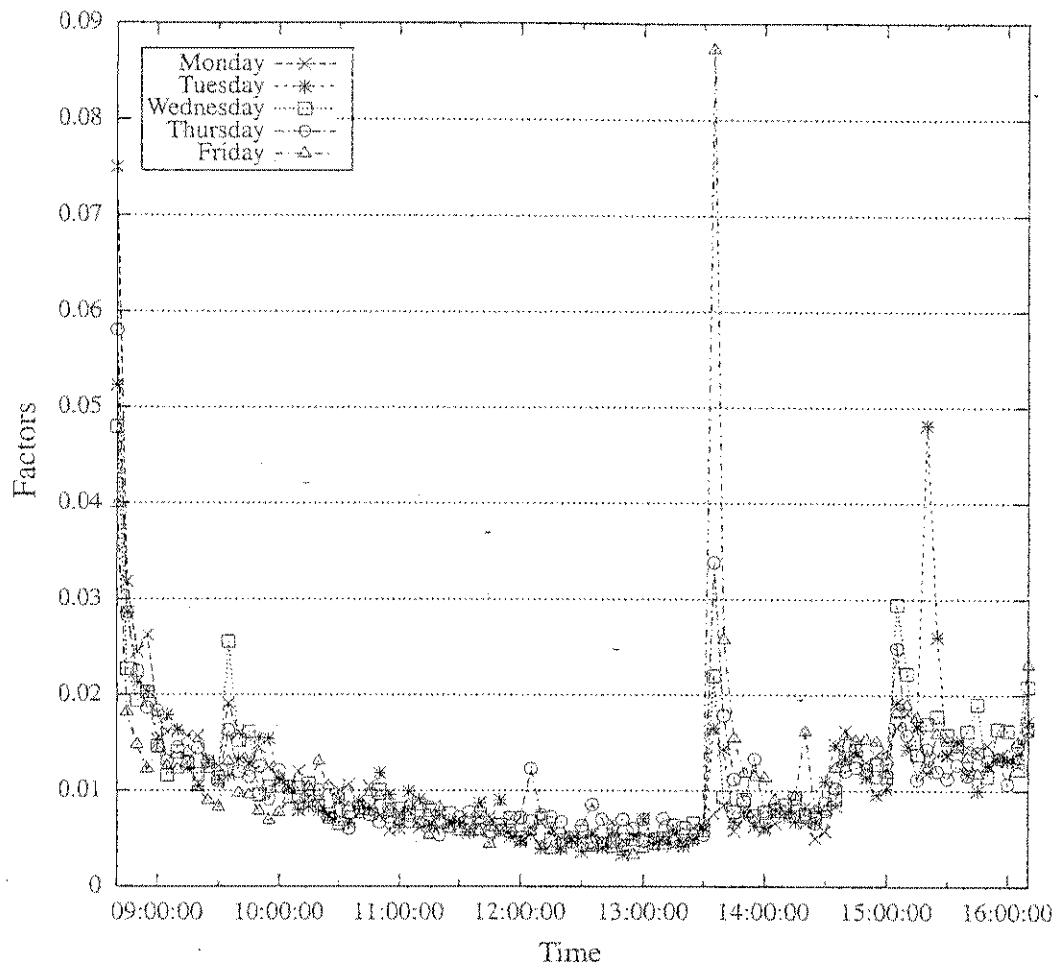


FIGURE 1

Five-minute open-market variance proportions for the FTSE-100 futures index, by day of the week, for the period from 18/11/1993 to 17/07/1998.

standard for Forex

The multipliers are scaled standard deviations and are:

- Relatively high from 07:30 GMT to 19:30 GMT in London (intervals 13 to 24).
- Higher when both U.S. and European dealers are active, from 12:30 to 18:30 GMT in London (07:30 to 13:30 EST in New York).
- Very high in the hour that contains the U.S. macro. news releases at 08:30 EST.
- Similar across the five business days of the week, except in the macro. news period.

Recommended reading: Section 12.5.

FIGURE 12.9

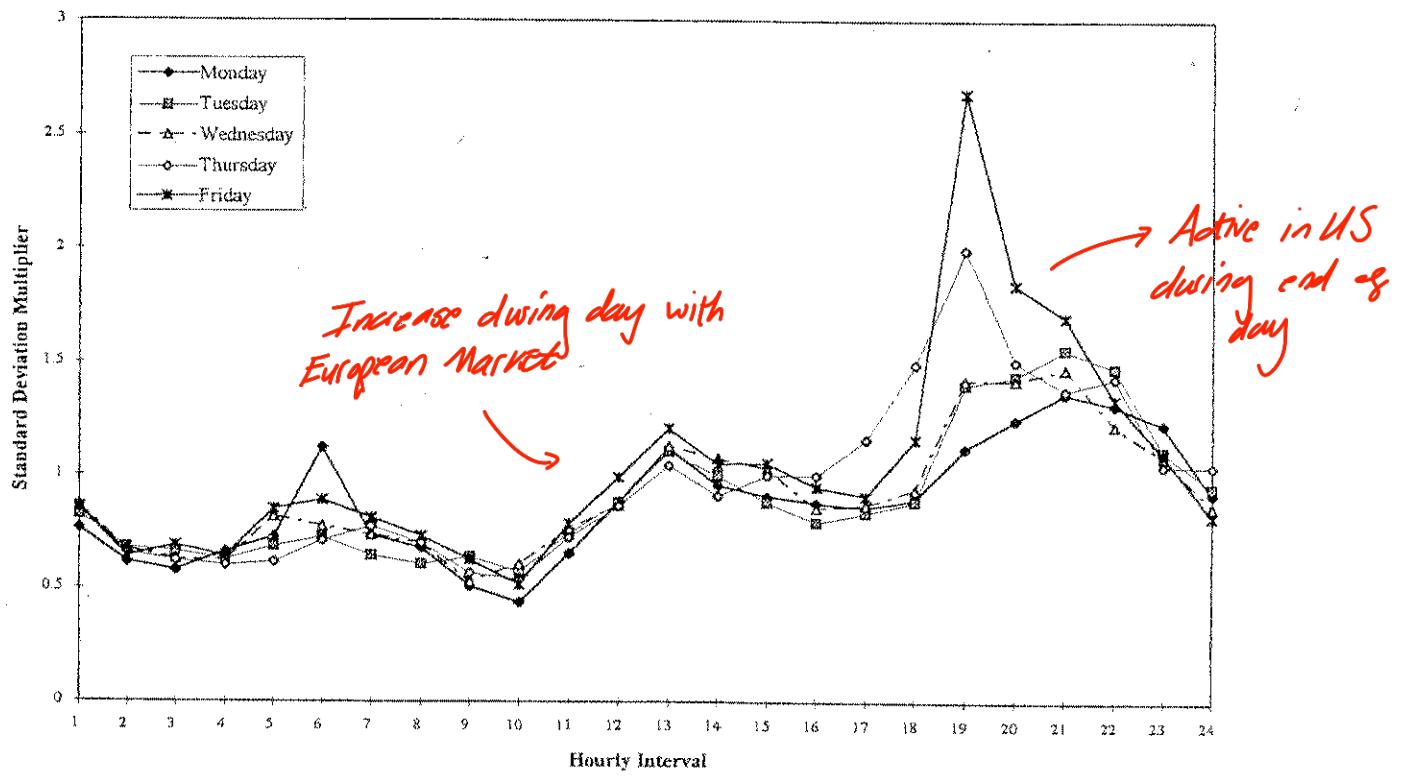


Fig. 4. DM/\$ intra-day standard deviation multipliers.

3. Realized volatility

Let $r_{t,j}$, $j = 1, 2, \dots, N$ represent the high frequency (e.g. five-minute) returns on day t .

Then the *realized volatility*, defined by

$$\sigma_t = \sqrt{\sum_{j=1}^N r_{t,j}^2},$$

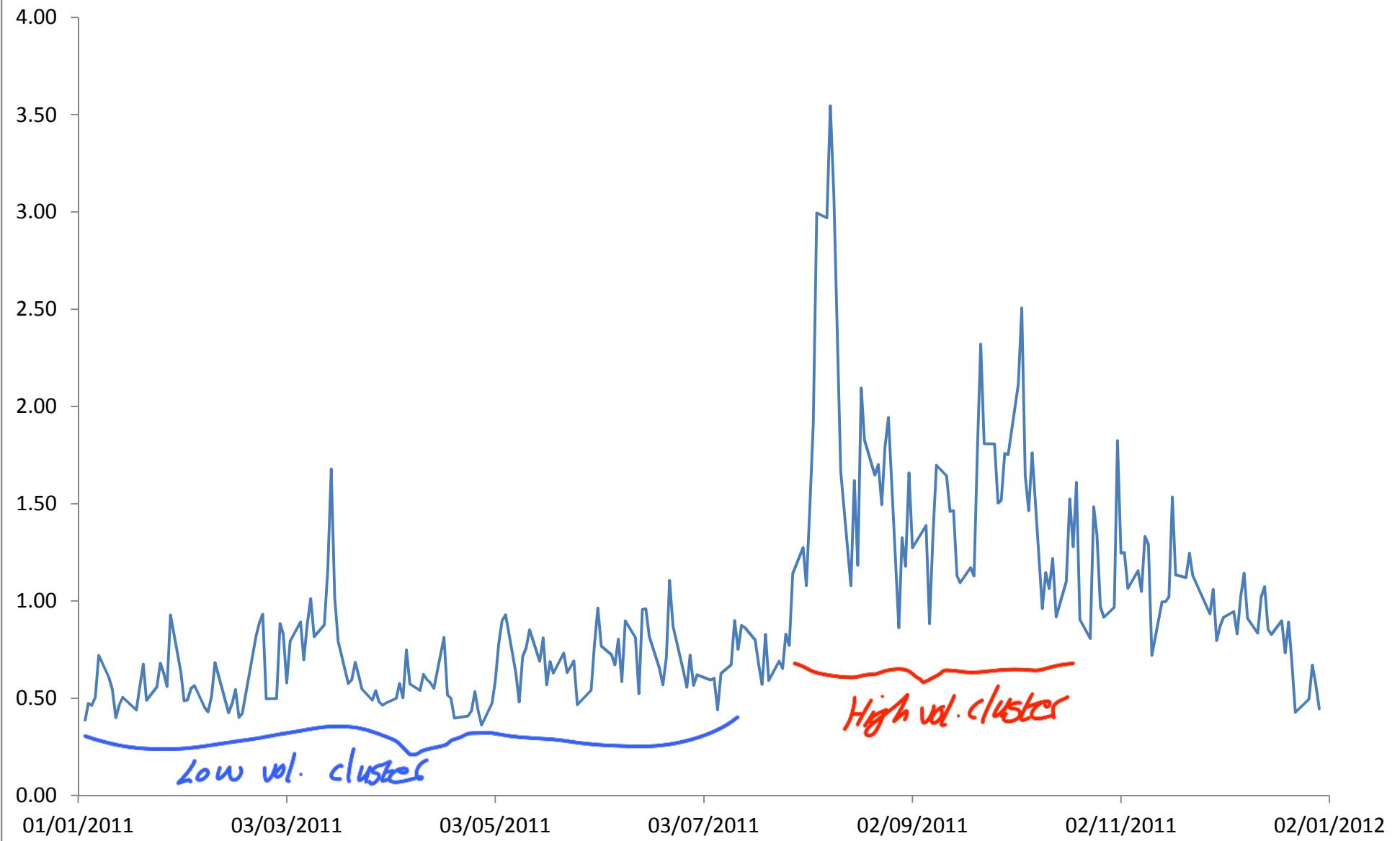
r_t = return on day t
 t - day
 j = intra-day period

is a fairly accurate measure of the volatility during trading hours on day t .

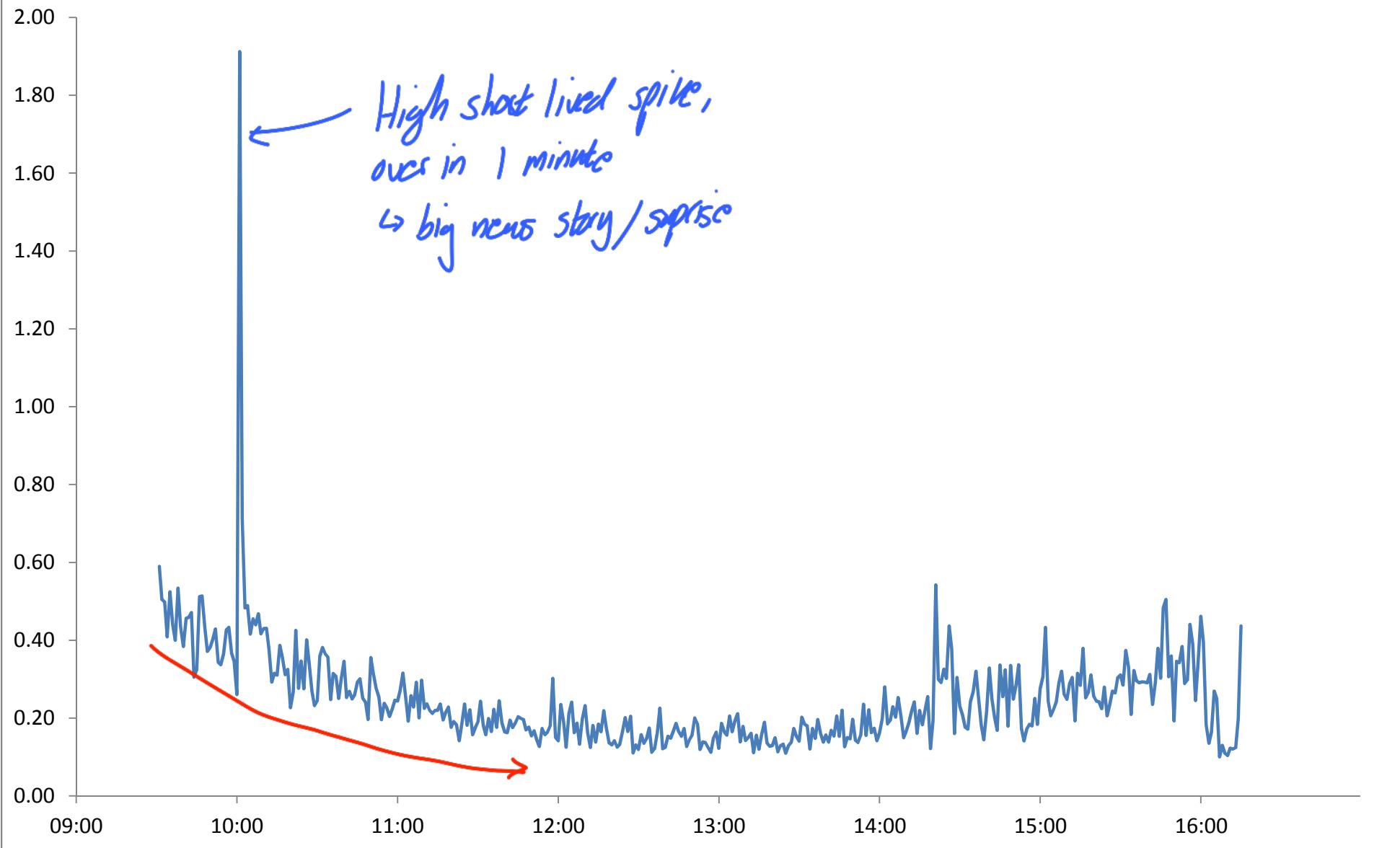
File **SPY_2011_by_minute for Fitch students.xlsb**:

- Contains SPY trade prices for each full trading day in 2011.
- There are 406 prices for each day, one a minute from 09:30 to 16:15 EST inclusive.
- Realized variance shoots up in August.
- Is visibly higher during Aug/Sep/Oct than during Apr/May/Jun.
- A realized s.d. of 1% is the same as an annualized open-market volatility of 16%.
- Percentages of daily variance sum to 100 across 405 one-minute intervals.
- Clear peak at 10:00 EST, explained by macroeconomic news.
- More volatility around open and close, than mid-day.

Realized standard deviation, percent units



Percentage of daily variance



Characteristic U shape towards midday → rises slowly to close after this

For many markets it is found that:

- The distribution of $\ln(r_t)$, unlike the distribution of returns. See Figure 12.12:

the dotted curve (normal) is near the solid curve (empirical data).

- The distribution, through time, of σ_t is **approximately lognormal**.

See Figure 12.13: again the two curves are similar.

- The distribution of $\ln(\sigma_t)$. See Figure 12.14.

Further reading: Sections 12.8 and 12.9. Both sections contain technical material, particularly 12.8.

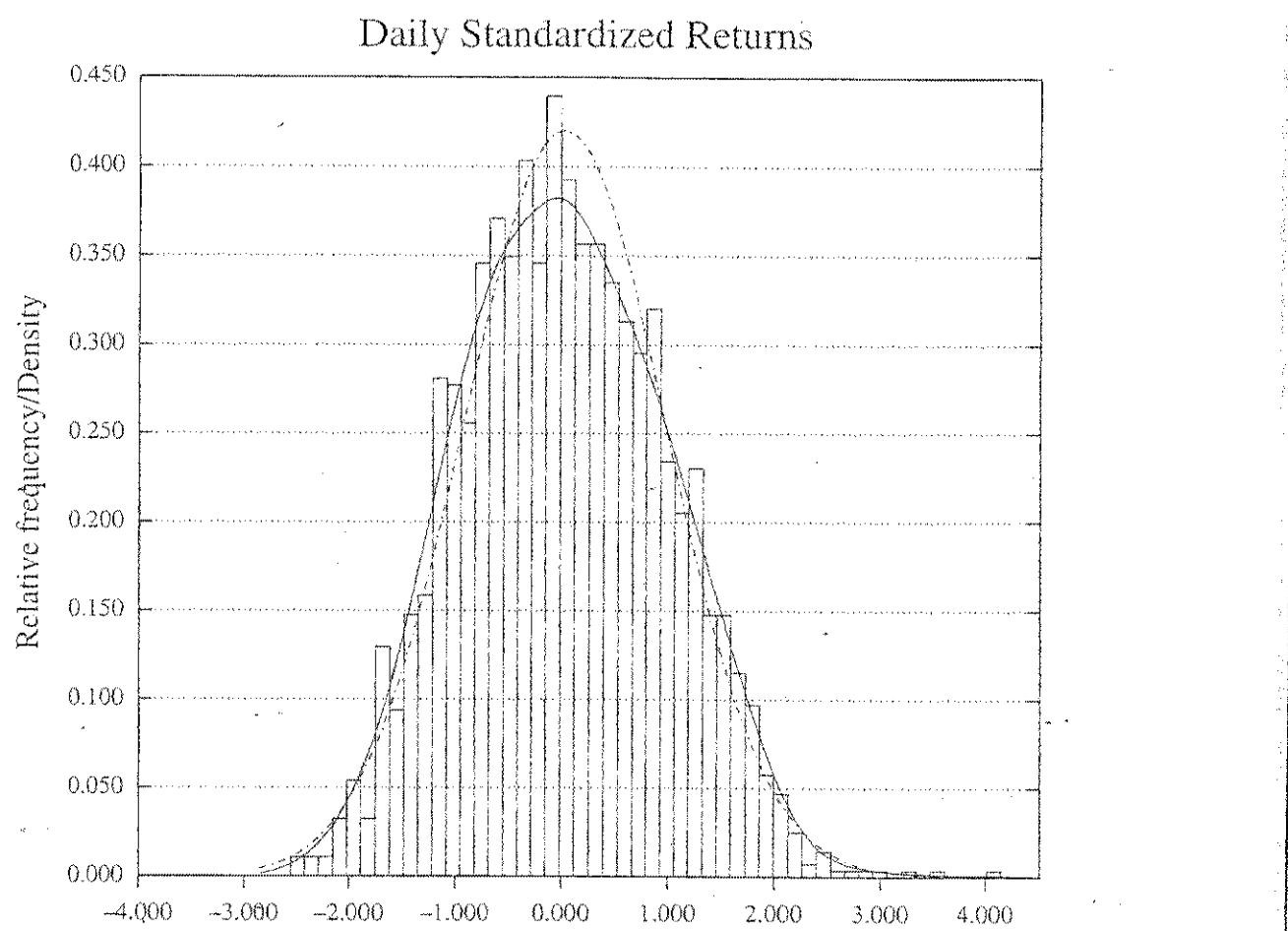


FIGURE 10
Frequency distribution of FTSE-100 index futures
from March 1990 to July 1998.

Figure 12.13

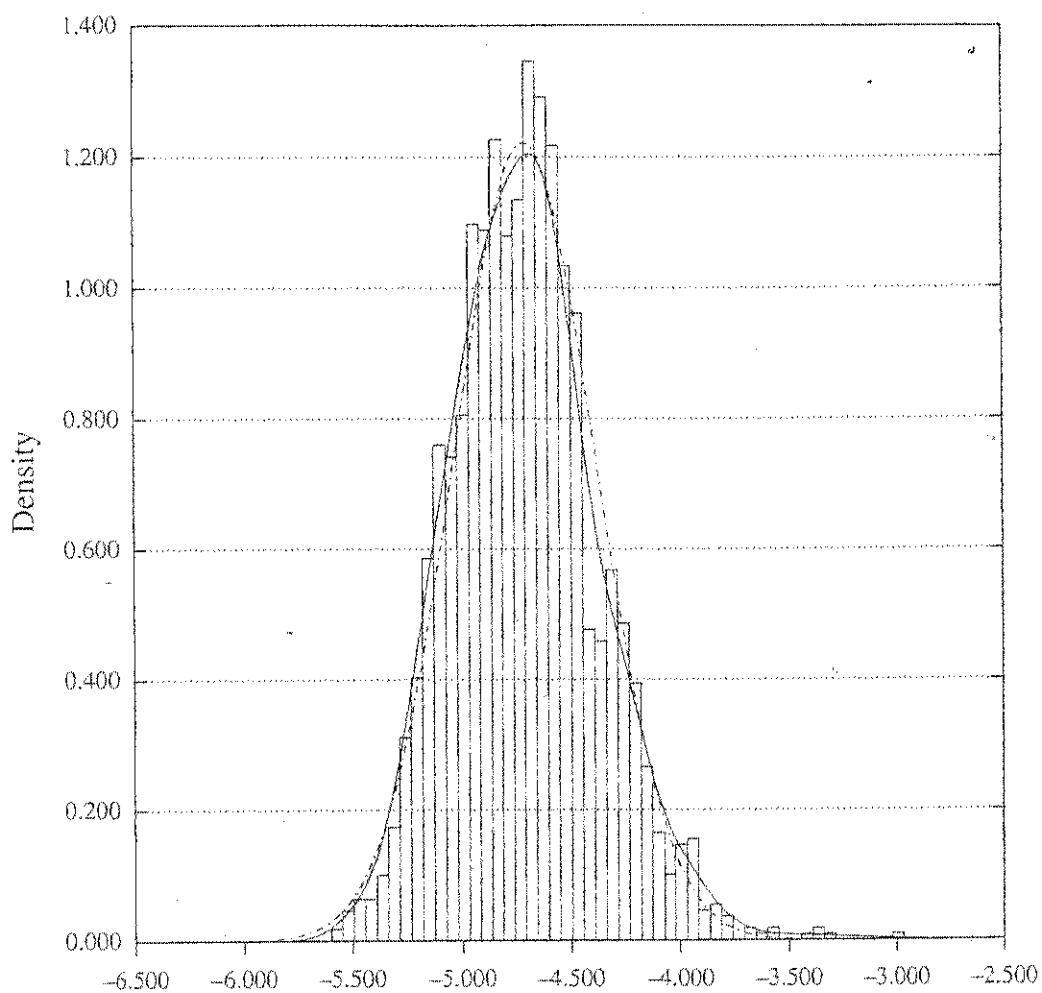


FIGURE 5

The distribution of the logarithm of realized volatility for the FTSE-100 index from 1990 to 1998.

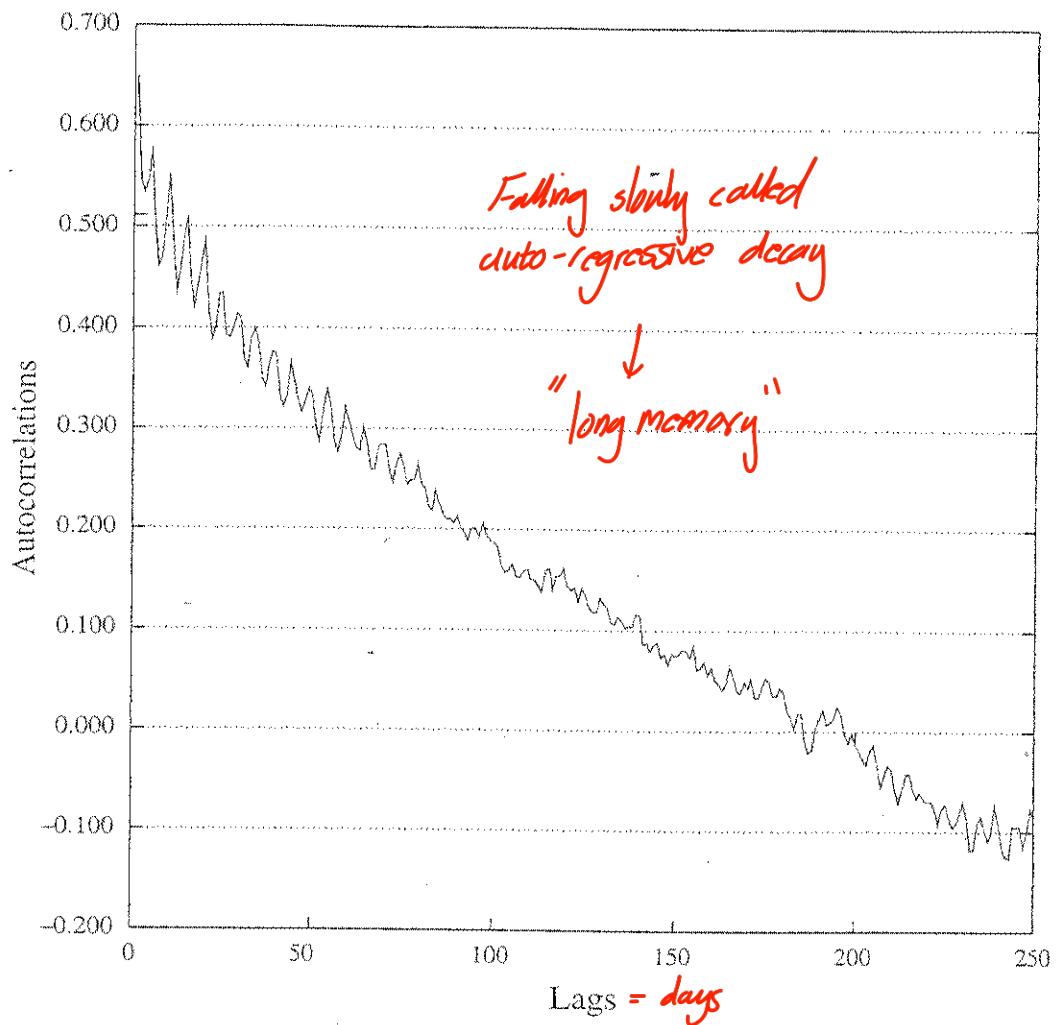


FIGURE 6

for the FTSE-100 index
from March 1990 to July 1998.

4. Further insights from high-frequency prices

Information about future volatility

- There is additional information about future volatility.
This is true for both equity and FX markets.
- There is sometimes incremental information in five-minute returns, when compared with that of implied volatilities.

News and volatility

- Specific items of news, 
 -  For example,
- the quantity of news headlines displayed by Reuters seems to be only weakly correlated with volatility.

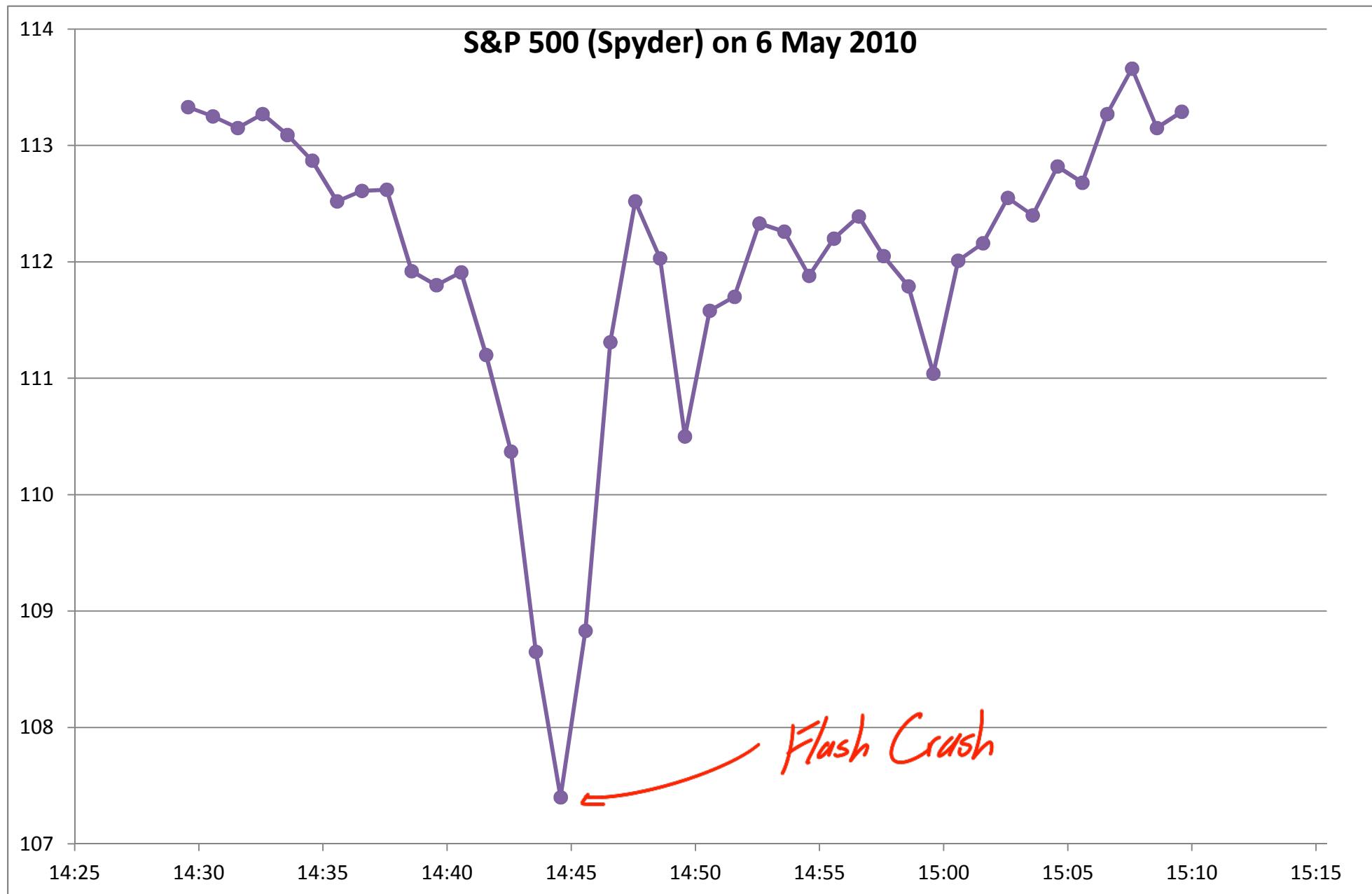
Exceptional price movements

Explanations of this event are controversial!?

→ Potentially a trades that stuffed quotes for market manipulation

U.S. stock indices collapsed around 14:45 but rapidly recovered.

The [redacted] around this
event was equivalent to an [redacted]



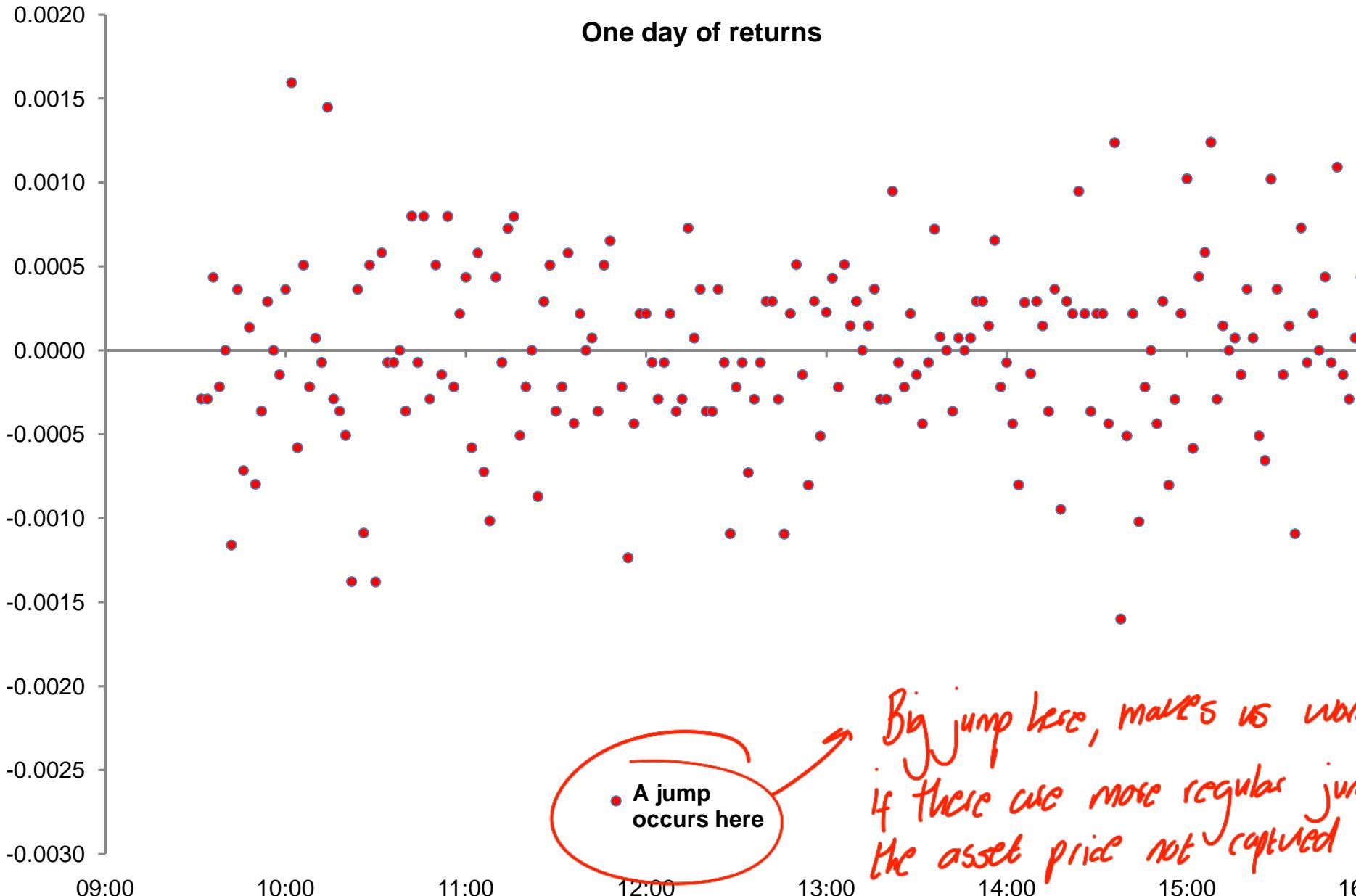
Price jumps

It would be

For example, consider SPY on 22 April 2008, with prices recorded every two minutes.

A test motivated by Andersen, Bollerslev & Dobrev (J Econometrics, 2007) finds a jump between 11:48 and 11:50 EST. See the prices and returns on the two following pages.





A jump occurs here

Big jump here, makes us wonder
if there are more regular jumps in
the asset price not captured by
diffusion processes alone
↳ Enter: Jump - Diffusion Process