

Financial markets, prices and risk

The focus of this chapter is on the statistical techniques used for analyzing prices and returns in financial markets. The concept of a stock market index is defined followed by a discussion of prices, returns and volatilities. Volatility clusters, the fat-tailed property of financial returns and observed sharp increases in correlations between assets during periods of financial turmoil (i.e., nonlinear dependence) will also be explored.

Various statistical techniques are introduced and used in this chapter for the analysis of financial returns. While readers may have seen these techniques before, Appendix A contains an introduction to basic statistics and time series methods for financial applications. The most common statistical methods presented in this chapter are implemented in the two programming languages discussed in this book: R and Matlab. These languages are discussed in more detail in Appendix B for R and Appendix C for Matlab.

We illustrate the application of statistical methods by using observed stock market data, the S&P 500 for univariate methods and a portfolio of US stocks for multivariate methods. The data can be downloaded from sources such as `finance.yahoo.com` directly within R and Matlab, as demonstrated by the source code in this chapter.

A key conclusion from this chapter is that we are likely to measure risk incorrectly by using volatility because of the presence of volatility clusters, fat tails and nonlinear dependence. This impacts on many financial applications, such as portfolio management, asset allocation, derivatives pricing, risk management, economic capital and financial stability.

The specific notation used in this chapter is:

T	Sample size
$t = 1, \dots, T$	A particular observation period (e.g., a day)
P_t	Price at time t
$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$	Simple return
$Y_t = \log \frac{P_t}{P_{t-1}}$	Continuously compounded return
y_t	A sample realization of Y_t
σ	Unconditional volatility
σ_t	Conditional volatility
K	Number of assets
ν	Degrees of freedom of the Student- t
t	Tail index

1.1 PRICES, RETURNS AND STOCK INDICES

1.1.1 Stock indices

A stock market index shows how a specified portfolio of share prices changes over time, giving an indication of market trends. If an index goes up by 1%, that means the total value of the securities which make up the index has also increased by 1% in value.

Usually, the index value is described in terms of “points”—we frequently hear statements like “the Dow dropped 500 points today”. The points by themselves do not tell us much that is interesting; the correct way to interpret the value of an index is to compare it with a previous value. One key reason so much attention is paid to indices today is that they are widely used as benchmarks to evaluate the performance of professionally managed portfolios such as mutual funds.

There are two main ways to calculate an index. A *price-weighted index* is an index where the constituent stocks are weighted based on their price. For example, a stock trading at \$100 will make up 10 times more of the total index than a stock trading at \$10. However, such an index will not accurately reflect the evolution of underlying market values because the \$100 stock might be that of a small company and the \$10 stock that of a large company. A change in the price quote of the small company will thus drive the price-weighted index while combined market values will remain relatively constant without changes in the price of the large company. The Dow Jones Industrial Average (DJIA) and the Nikkei 225 are examples of price-weighted stock market indices.

By contrast, the components of a *value-weighted index* are weighted according to the total market value of their outstanding shares. The impact of a component’s price change is therefore proportional to the issue’s overall market value, which is the product of the share price and the number of shares outstanding. The weight of each stock constantly shifts with changes in a stock’s price and the number of shares outstanding, implying such indices are more informative than price-weighted indices.

Perhaps the most widely used index in the world is the Standard & Poor 500 (S&P 500) which captures the top-500 traded companies in the United States, representing about 75% of US market capitalization. No asset called S&P 500 is traded on financial markets, but it is possible to buy derivatives on the index and its volatility VIX. For the Japanese market the most widely used value-weighted index is the TOPIX, while in the UK it is the FTSE.

1.1.2 Prices and returns

We denote asset prices by P_t , where the t usually refers to a day, but can indicate any frequency (e.g., yearly, weekly, hourly). If there are many assets, each asset is indicated by $P_{t,k} = P_{\text{time,asset}}$, and when referring to portfolios we use the subscript “port”. Normally however, we are more interested in the *return* we make on an investment—not the price itself.

Definition 1.1 (Returns) *The relative change in the price of a financial asset over a given time interval, often expressed as a percentage.*

Returns also have more attractive statistical properties than prices, such as stationarity and ergodicity. There are two types of returns: simple and compound. We ignore the dividend component for simplicity.

Definition 1.2 (Simple returns) *A simple return is the percentage change in prices, indicated by R :*

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Often, we need to convert daily returns to monthly or annual returns, or vice versa. A multiperiod (n -period) return is given by:

$$\begin{aligned} R_t(n) &= (1 + R_t)(1 + R_{t-1})(1 + R_{t-2}) \cdots (1 + R_{t-n+1}) - 1 \\ &= \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-n+1}}{P_{t-n}} - 1 = \frac{P_t}{P_{t-n}} - 1 \end{aligned}$$

where $R_t(n)$ is the return over the most recent n -periods from date $t - n$ to t .

A convenient advantage of simple returns is that the return on a portfolio, $R_{t,\text{port}}$, is simply the weighted sum of the returns of individual assets:

$$R_{t,\text{port}} = \sum_{k=1}^K w_k R_{t,k}$$

where K is the number of assets; and w_k is the portfolio weight of asset i . An alternative return measure is continuously compounded returns.

Definition 1.3 (Continuously compounded returns) *The logarithm of gross return, indicated by Y_t :*

$$Y_t = \log(1 + R_t) = \log\left(\frac{P_t}{P_{t-1}}\right) = \log P_t - \log P_{t-1}.$$

The advantages of compound returns become clear when considering multiperiod returns:

$$\begin{aligned} Y_t(n) &= \log(1 + R_t(n)) = \log((1 + R_t)(1 + R_{t-1})(1 + R_{t-2}) \cdots (1 + R_{t-n+1})) \\ &= \log(1 + R_t) + \log(1 + R_{t-1}) + \cdots + \log(1 + R_{t-n+1}) \\ &= Y_t + Y_{t-1} + \cdots + Y_{t-n+1}. \end{aligned}$$

Continuously compounded multiperiod returns are the sum of continuously compounded single-period returns. In contrast to simple returns, it is much easier to derive the time series properties of sums than of products.

The situation is different for portfolio returns since the log of a sum does not equal the sum of logs:

$$Y_{t,\text{port}} = \log\left(\frac{P_{t,\text{port}}}{P_{t-1,\text{port}}}\right) \neq \sum_{k=1}^K w_k \log\left(\frac{P_{t,k}}{P_{t-1,k}}\right).$$

where $P_{t,\text{port}}$ is the portfolio value on day t ; and $Y_{t,\text{port}}$ is the corresponding return. The difference between compound and simple returns may not be very significant for small returns (e.g., daily),

$$Y_{\text{port}} \approx \sum_{k=1}^K w_k R_k$$

and as the time between observations goes to zero, so does the difference between the two return measures:

$$\lim_{\Delta t \rightarrow 0} Y_{\text{port}} = R_{\text{port}}.$$

It will not usually matter much which measure we choose to use. For example, suppose $P_t = 1,000$ and $P_{t-1} = 950$ then:

$$R_t = \frac{1,000}{950} - 1 = 0.0526$$

$$Y_t = \log\left(\frac{1,000}{950}\right) = 0.0513.$$

The discrepancy between them becomes significant only when percent changes are high—for example, if $P_t = 1,000$ and $P_{t-1} = 700$, then:

$$R_t = \frac{1,000}{700} - 1 = 0.429$$

$$Y_t = \log\left(\frac{1,000}{700}\right) = 0.357.$$

In some situations, such as accounting, simple returns need to be used.

Another common type of returns is excess returns (i.e., returns in excess of some reference rate, often the risk free rate).

We should think of simple returns and compound returns as two different definitions of returns. They are also known as arithmetic and logarithmic returns, respectively. Simple returns are of course *correct*; investors are primarily interested in simple returns. But there are reasons for continuously compounded returns being preferable.

A key advantage is that they are symmetric, while simple returns are not. This means an investment of \$100 that yields a simple return of 50% followed by a simple return of -50% will result in \$75, while an investment of \$100 that yields a continuously compounded return of 50% followed by a continuously compounded return of -50% will remain at \$100.

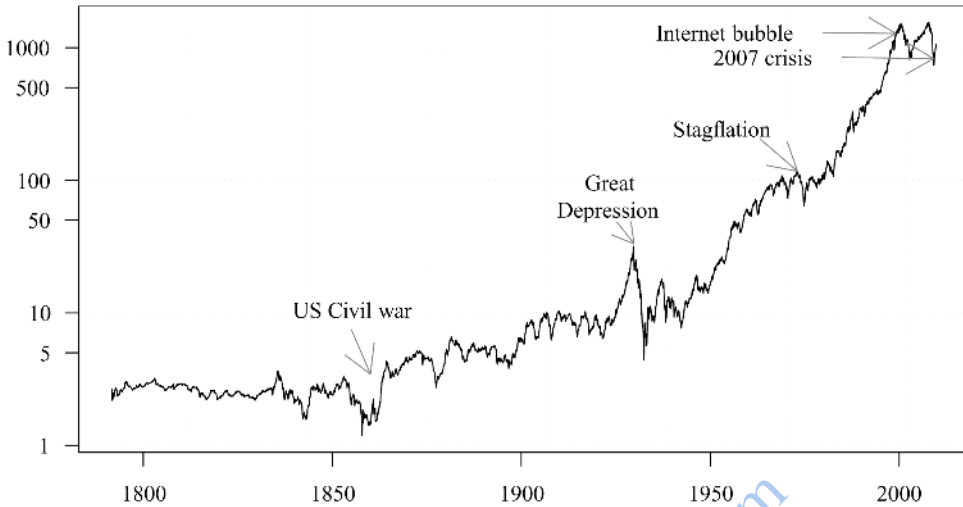


Figure 1.1. S&P 500 index August 1791 to December 2009, log scale.
Data source: Global Financial Data.

Continuously compounded returns also play an important role in the background of many financial calculations. They are a discrete form of *continuous time* Brownian motion,¹ which is the foundation for derivatives pricing and is used to model the changes in stock prices in the Black–Scholes model.

1.2 S&P 500 RETURNS

The S&P 500 index has been published since 1957 but Global Financial Data, a commercial vendor, go back as far as 1791. The log of the monthly close of the S&P 500 from 1791 until 2009 can be seen in Figure 1.1. One needs to be careful when looking at a long time series of prices as it is easy to reach misleading conclusions.

The first observation is on 1791/08/31 when the index had a value of \$2.67, while the value on the last day of the sample, 2009/12/31, was \$1,115.1. This implies that the index has risen in value by 41,660%, or 2% per year. This analysis, however, overlooks depreciation in the value of the dollar (i.e., inflation). We can calculate how much one dollar has increased in value from 1791 to 2009 using the five different techniques shown in Table 1.1.

Using the CPI, the real increase in the value of the index has actually been a measly 1.4% per year. This does not, however, represent the total returns of an investor as it ignores dividend yield.

We show the compound returns in Figure 1.2. There is high volatility during the American Civil War in the 1860s, the Great Depression in the 1930s, the stagflation of the 1970s and the Asian crisis in 1997, among others. Prolonged periods of high volatility are generally associated with great uncertainty in the real economy.

¹ Brownian motion, also called Wiener process, is a centered, zero mean Gaussian process $\{W_t; t\}$.

Table 1.1. Increase in value of one dollar from 1791 to 2009 using five different techniques.

\$23.66	Consumer price index
\$22.73	GDP deflator
\$397.91	Unskilled wage
\$942.48	Nominal GDP per capita
\$70,902.74	Relative share of GDP

Calculated from <http://www.measuringworth.com/uscompare>

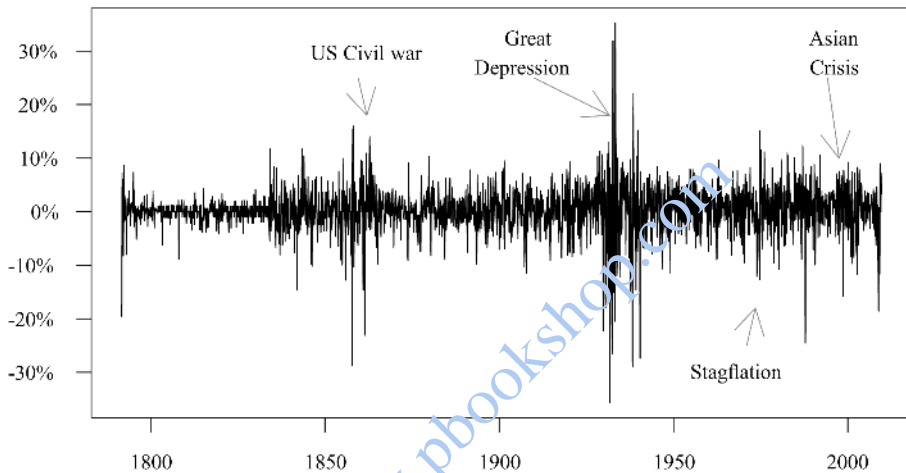


Figure 1.2. Returns on the monthly S&P 500 index from August 1791 to December 2009.

1.2.1 S&P 500 statistics

A selection of summary statistics for daily S&P 500 returns from 1928 to 2009 is presented in Table 1.2. The daily mean is very small at 0.019% while daily volatility is around 1.2%. The fact that the daily mean is only one-fiftieth of daily volatility will simplify the construction of risk measures as we can effectively assume it is zero, without

Table 1.2. S&P 500 daily return summary statistics, 1928–2009

Mean	0.019%
Standard deviation	1.16%
Min	-22.9%
Max	15.4%
Skewness	-0.4
Kurtosis	22.5
Autocorrelation (one lag) of returns	3.4%
Autocorrelation (one lag) of squared returns	21.7%
Jarque–Bera (p -value)	0.0%
Ljung–Box 20 lags (p -value)	0.0%
Ljung–Box squared returns 20 lags (p -value)	0.0%

loss of generality. Furthermore, the mean grows at a linear rate while volatility grows approximately at a square root rate, so over time the mean dominates volatility.

The lowest daily return of -23% corresponds to the stock market crash of 1987, while the best day in the index, 15% , was at the end of the Great Depression. The returns have a small negative skewness and, more importantly, quite high kurtosis.

Finally, the returns have a daily autocorrelation of about 3% while squared returns have an autocorrelation of 22% . Squared returns are a proxy for volatility. The 22% autocorrelation of squared returns provides very strong evidence of the predictability of volatility and volatility clusters.

The table also shows a test for normality, the Jarque–Bera (JB) test, first-order autocorrelations of returns and returns squared, and finally a test for the presence of an autocorrelation up to 20 lags, a Ljung–Box (LB) test.

1.2.2 S&P 500 statistics in R and Matlab

The results in Table 1.2 can be easily generated using R or Matlab. It is possible to directly download stock prices into R or Matlab from several websites, such as `finance.yahoo.com`. In some of the examples in this chapter we use data going back to the 1700s; data that old were obtained from Global Financial Data.

The following two R and Matlab code listings demonstrate how S&P 500 daily prices from 2000 until 2009 can be downloaded from `finance.yahoo.com`, where the stock market symbol for the S&P 500 is `^gspc`. An active internet connection is required for this code to work, but it is straightforward to save the returns after downloading them. One issue that comes up is which data field from `finance.yahoo.com` to use. One might think it best to use closing prices, but that is usually not correct, because over time we observe actions that change the prices of equities such as stock splits and stock buybacks, without affecting the value of the firm. We therefore need to use the *adjusted closing prices* which automatically take this into account. For the S&P 500 this makes no difference, but for most stock prices it does. Therefore, it is good practice to use adjusted closing prices by default.

We use the R function `get.hist.quote()` from the `tseries` library. We then convert the prices into returns, and plot the returns. By default, `get.hist.quote()` returns a four-column matrix with open and closing prices, as well as the high and low of prices. To get adjusted closing prices in R we need to include `quote="AdjClose"` in the `get.hist.quote()` statement. Note that prices and returns in R are represented as a time series object while in Matlab they are simply vectors. The function `{\tt coredata}` is discussed on p. 94.

Listing 1.1. Download S&P 500 data in R

```
library("tseries")           # load the tseries library
price = get.hist.quote(instrument = "^gspc", start = "
  2000-01-01", quote="AdjClose") # download the prices,
  from January 1, 2000 until today
y=diff(log(price))           # convert the prices into returns
plot(y)                      # plot the returns
y=coredata(y)                # strip date information from returns
```

In Matlab it is equally straightforward to download prices. It is possible to use the GUI function, `FTSTool` from the financial and data feed toolboxes; however, it may be easier to use the Matlab function `urlread()` which can directly read web pages, such as `finance.yahoo.com`. Several free user-contributed functions are available to ease the process, such as `hist_stock_data()`.² `finance.yahoo.com` returns the data sorted from the newest date to the oldest date, so that the first observation is the newest. We want it sorted from the oldest to newest, and the R procedure does it automatically; unfortunately, the Matlab procedure does not, so we have to do it manually by using a sequence like `end:-1:1`. Of course, it would be most expedient to just modify the `hist_stock_data()` function.

Listing 1.2. Download S&P 500 data in Matlab

```
price = hist_stock_data('01012000','31122000','^gspc');
                                     % download the prices, from
                                     % January 1, 2000 until
                                     % December 31, 2009
y=diff(log(price.Close(end:-1:1)))    % convert the prices into
                                     % returns
plot(y)                               % plot the returns
```

After having obtained the returns `y`, we can calculate some sample statistics; they are given in Listing 1.3.

Listing 1.3. Sample statistics in R

```
library(moments)
sd(y)
min(y)
max(y)
skewness(y)
kurtosis(y)
acf(y,1)
acf(y^2,1)
jarque.bera.test(y)
Box.test(y, lag = 20, type = c("Ljung-Box"))
Box.test(y^2, lag = 20, type = c("Ljung-Box"))
```

² It can be obtained directly from the webpage of the Matlab vendor <http://www.mathworks.com/matlabcentral/fileexchange/18458-historical-stock-data-downloader>.

Listing 1.4. Sample statistics in Matlab

```

% JPL and MFE toolboxes
mean(y)
std(y)
min(y)
max(y)
skewness(y)
kurtosis(y)
sacf(y,1,[],0)
sacf(y.^2,1,[],0)
jarquebera(y)
[q, pval]=ljungbox(y,20)
[q, pval]=ljungbox(y.^2,20)

```

1.3 THE STYLIZED FACTS OF FINANCIAL RETURNS

Extensive research on the properties of financial returns has demonstrated that returns exhibit three statistical properties that are present in most, if not all, financial returns. These are often called the three *stylized facts* of financial returns:

Volatility clusters

Fat tails

Nonlinear dependence

The first property, volatility clusters, relates to the observation that the magnitudes of the volatilities of financial returns tend to cluster together, so that we observe many days of high volatility, followed by many days of low volatility.

The second property, fat tails, points to the fact that financial returns occasionally have very large positive or negative returns, which are very unlikely to be observed, if returns were normally distributed.

Finally, nonlinear dependence (NLD) addresses how multivariate returns relate to each other. If returns are linearly dependent, the correlation coefficient describes how they move together. If they are nonlinearly dependent, the correlation between different returns depends on the magnitudes of outcomes. For example, it is often observed that correlations are lower in bull markets than in bear markets, while in a financial crisis they tend to reach 100%.

Each of those stylized facts is discussed in turn in the following sections.

1.4 VOLATILITY

The most common measure of market uncertainty is volatility.

Definition 1.4 (Volatility) *The standard deviation of returns.*

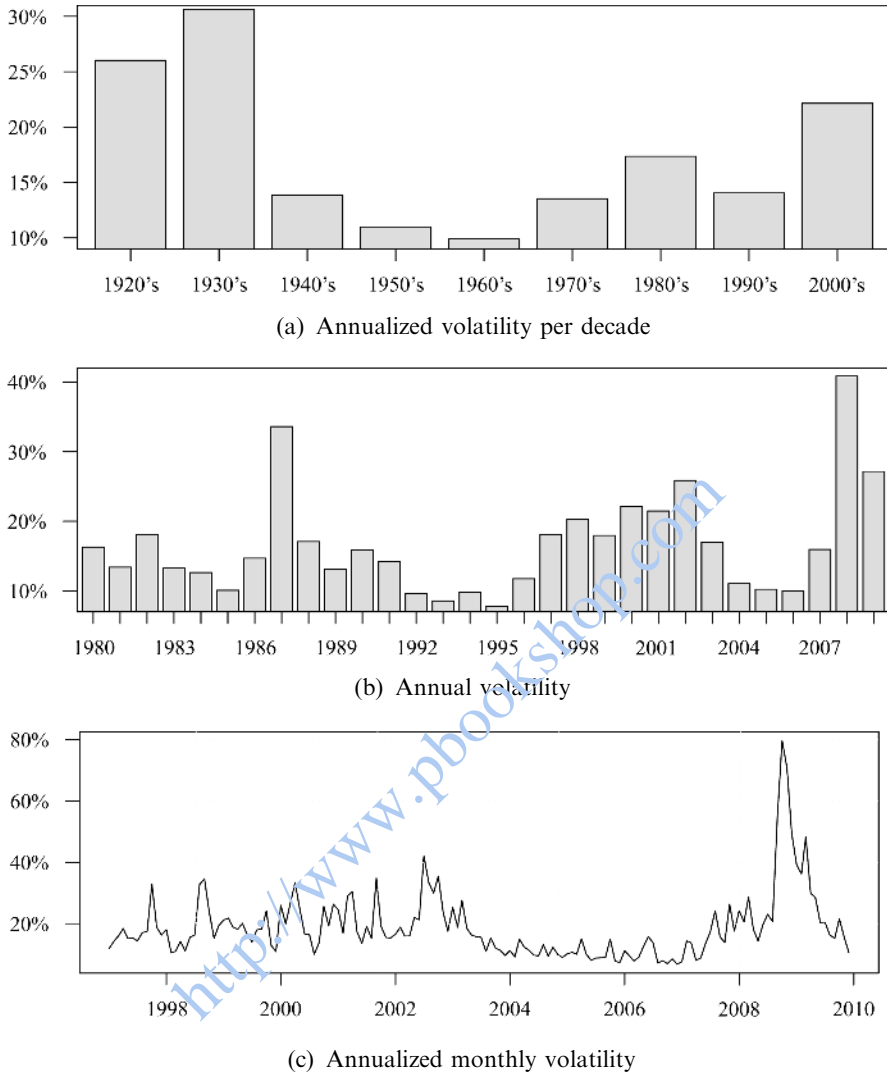


Figure 1.3. Volatility cycles.

We further explore the nature of volatility in the S&P 500 index by calculating volatility in subperiods of the data. This calculation is repeated for daily returns over decades, years and months (see Figure 1.3).

Panel (a) of Figure 1.1 shows volatility per decade from 1928 to 2009; we can see clear evidence of cyclical patterns in volatility from one decade to the next. Volatility is lowest in the 1960s and highest during the Great Depression in the 1930s. Note that 1920s' values only contain a part of 1929 and that the Great Depression started in 1929.

Focusing on more recent events, panel (b) shows volatility per year from 1980. The most volatile year is 2008, during the 2007–2009 crisis, followed by the stock market

crash year of 1987. The calmest year is 1995, right before the Asian crisis; 2004–2006 are also quite relaxed.

However, the fact that volatility was very low in 1995 and 2005 does *not* imply that risk in financial markets was low in those years, since volatility can be low while the tails are fat. In other words, it is possible for a variable with a low volatility to have much more extreme outcomes than another variable with a higher volatility. This is why volatility is a misleading measure of risk.

Finally, panel (c) shows average daily volatility per month from 1995. Again, it is clear that volatility has been trending downwards, and has been very low from 2004. This is changing as a result of the 2007–2009 crisis.

Taken together, the figures provide substantial evidence that there are both long-run cycles in volatility spanning decades, and short cycles spanning weeks or months. In this case, we are observing cycles within cycles within cycles. However, given we have many fewer observations at lower frequencies—such as monthly—there is much more statistical uncertainty in that case, and hence the plots are much more jagged.

The crude methods employed here to calculate volatility (i.e., sampling standard errors) are generally considered unreliable, especially at the highest frequencies; more sophisticated methods will be introduced in the next chapter.

1.4.1 Volatility clusters

We use two concepts of volatility: *unconditional* and *conditional*. While these concepts are made precise later, for our immediate discussion unconditional volatility is defined as volatility over an entire time period, while conditional volatility is defined as volatility in a given time period, conditional on what happened before. Unconditional volatility is denoted by σ and conditional volatility by σ_t .

Looking at volatility in Figure 1.3, it is evident that it changes over time. Furthermore, given the apparent cycles, volatility is partially predictable. These phenomena are known as *volatility clusters*.

We illustrate volatility clusters by simulations in Figure 1.4, which shows exaggerated simulated volatility clusters. Panel (a) shows returns and panel (b) shows volatility. In the beginning, volatility increases and we are in a high-volatility cluster, then around day 180 volatility decreases only to increase again after a while and so on.

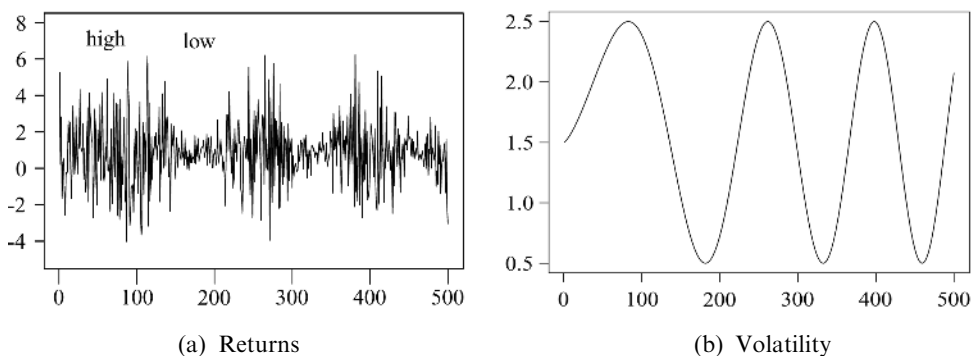


Figure 1.4. Exaggerated simulated volatility clusters.

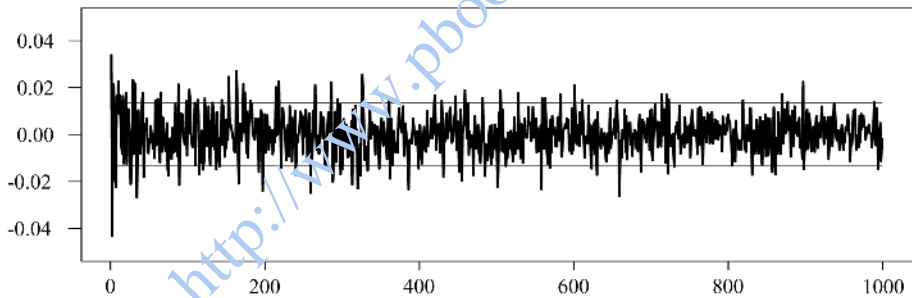
Almost all financial returns exhibit volatility clusters (i.e., the market goes through periods when volatility is high and other periods when volatility is low). For example, in the mid-1990s volatility was low, while at the beginning and end of the decade it was much higher. This feature of financial time series gained widespread recognition with the publication of Engle (1982) and is now one of the accepted stylized facts about asset returns. If we can capture predictability in volatility, it may be possible to improve portfolio decisions, risk management and option pricing, among other applications.

1.4.2 Volatility clusters and the ACF

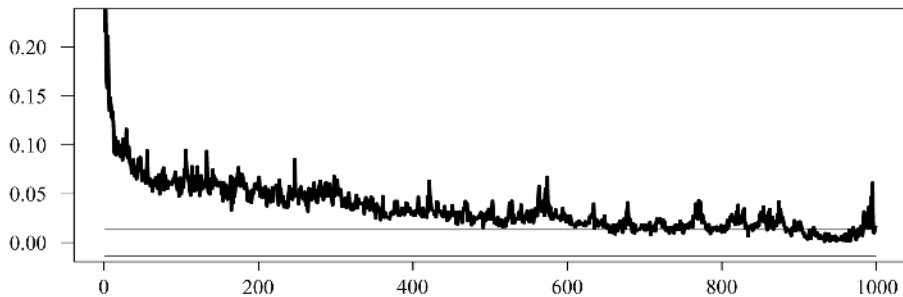
A standard graphical method for exploring predictability in statistical data is the *autocorrelation function* (ACF). The ACF measures how returns on one day are correlated with returns on previous days. If such correlations are statistically significant, we have strong evidence for predictability.

Panel (a) of Figure 1.5 shows the ACF of S&P 500 returns along with a 95% confidence interval, where most autocorrelations lie within the interval. Contrast this with the ACF of squared returns in panel (b) where it is significant even at long lags, providing strong evidence for the predictability of volatility.

We can test for the joint significance of autocorrelation coefficients over several lags by using the Ljung–Box (LB) test. We do the LB test using 21 lags of daily S&P 500



(a) Returns



(b) Squared returns

Figure 1.5. Autocorrelation plots of daily S&P 500 returns, 1929–2009, along with a $\pm 95\%$ confidence interval.

Table 1.3. Ljung–Box test for daily S&P 500 returns, 1929–2009

<i>T</i>	<i>Ljung–Box test, 21 lags</i>	<i>p-value</i>
21,334	147.7	$<2.2e^{-16}$
2,500	110.0	$4.763e^{-14}$
100	14.5	0.846

Table 1.4. Ljung–Box test for squared S&P 500 returns, 1929–2009

<i>T</i>	<i>Ljung–Box test, 21 lags</i>	<i>p-value</i>
21,334	9,158.1	$<2.2e^{-16}$
2,500	3,812.1	$<2.2e^{-16}$
100	33.2	0.0438

returns (i.e., approximately the number of trading days in a calendar month). The test is performed using the full sample size, as well as the most recent 2,500 and 100 observations; the results are given in Table 1.3. We can also use the Engle LM test to test for volatility clusters.

Table 1.3 shows there is significant return predictability for the full sample, but not for the most recent observations. This does not mean a violation of market efficiency, since we would need to adjust the returns for risk, the risk-free rate and transaction costs.

The same procedure is repeated for squared returns; the results are shown in Table 1.4. The reason for focusing on squared returns is that they are proxies for volatilities; most forecast procedures for volatilities, like those in the next chapter, use squared returns as their main input. The p -value for the smallest sample size of squared returns is much lower than the corresponding value for returns. Tables 1.3 and 1.4 demonstrate that it is easier to predict volatility than the mean.

The code necessary to carry out ACF plots and the Ljung–Box test in R and Matlab is given in the following listings.

Listing 1.5. ACF plots and the Ljung–Box test in R

```

library(MASS,stats)      # load stats and MASS package
q = acf(y,20)
plot(q[2:20])
q = acf(y^2,20)
plot(q[2:20])
b = Box.test(y,lag=21,type="Ljung-Box")

```

Listing 1.6. ACF plots and the Ljung–Box test in Matlab

```

sacf (y, 20)
sacf (y.^2, 20)
ljungbox (y, 20)

```

1.5 NONNORMALITY AND FAT TAILS

Many applications assume that S&P 500 index returns are normally distributed. Table 1.5 shows some return outcomes and probabilities based on this assumption (e.g., where the probability of a return less than -2% is 3.5%).

Table 1.2 shows that the biggest one-day drop in the index was 23% . If S&P 500 index returns were indeed normally distributed, then the probability of that one-day crash would be 2.23×10^{-97} according to Table 1.5. In other words, the crash is supposed to happen once every 10^{95} years (accounting for weekends and holidays). To put this into context, scientists generally assume that the earth is about 10^7 years old and the universe 10^{13} years old. Assuming normality this equates to believing that the crash of 1987 only happens in one out of every 12 universes. We are doing slightly better on the best day of the index which only has a probability of occurrence once every 10^{41} years under normality.

Table 1.5. Outcomes and probabilities of daily S&P 500 returns assuming normality, 1929–2009

Returns above or below	Probability
1%	0.865
2%	0.035
3%	0.00393
5%	2.74×10^{-06}
15%	2.70×10^{-43}
23%	2.23×10^{-97}

However, it can argued that the crash of 1987 was an anomaly, so assuming normality for all the other days would be relatively innocuous. But is this really the case? Figures 1.6(a, b) show the most extreme daily returns per decade and year, respectively. It is clear that there are still many more extremes than Table 1.5 predicts.

An alternative way of analyzing the distribution of the S&P 500 index is shown in Figure 1.7. Panel (a) plots the histogram of the returns and superimposes the normal distribution with the same mean and variance. Panel (b) shows both the normal distribution and the empirical distribution of the returns, while panel (c) blows up the left tail of the distributions. We can observe from these three figures that

1. The peak of the return distribution is much higher than for the normal distribution.
2. The sides of the return distribution are lower than for the normal distribution.
3. The tails of the return distribution are much thicker (fatter) than for the normal distribution.

In other words, there are more days when very little happens in the market than predicted by the normal and more days when market prices change considerably.

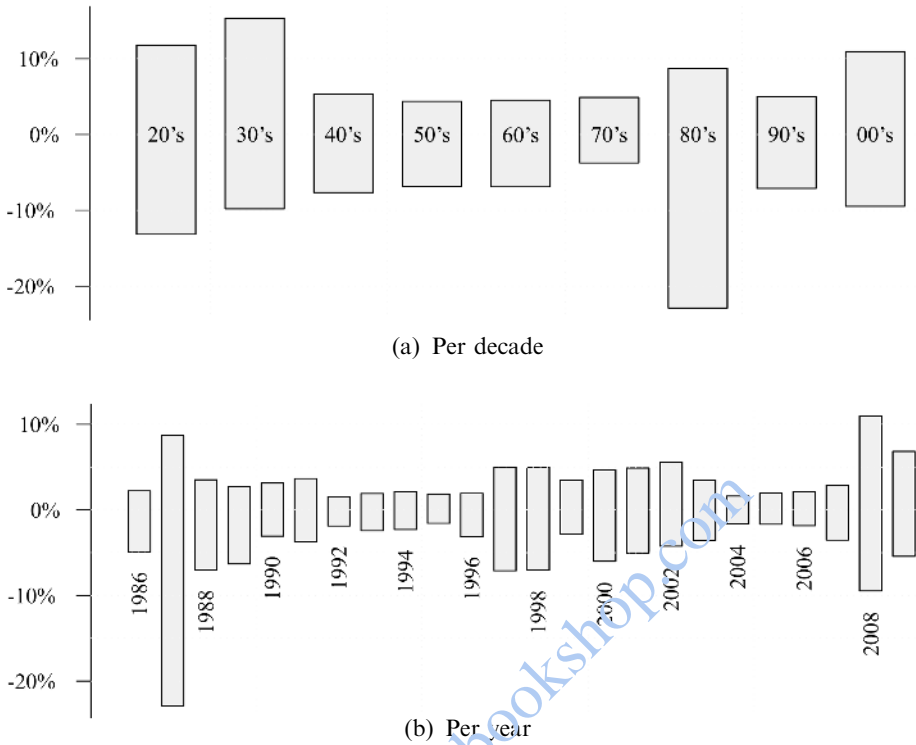


Figure 1.6. Maximum and minimum daily S&P 500 returns.

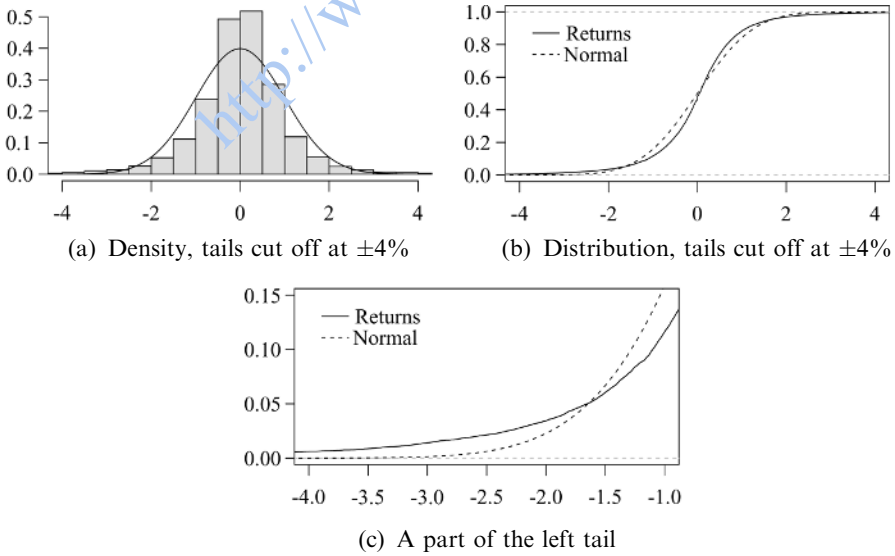


Figure 1.7. Empirical density and distribution of S&P 500 index returns for 2000–2009 compared with the normal distribution.

1.6 IDENTIFICATION OF FAT TAILS

There are two main approaches for identifying and analyzing the tails of financial returns: statistical tests and graphical methods. Statistical tests compare observed returns with some base distribution, typically but not always the normal. Graphical methods relate observed returns with values predicted from some distribution, often the normal.

1.6.1 Statistical tests for fat tails

From the above we can see that one important feature of financial returns is that they exhibit what is known as *fat tails*. We give an informal definition of fat tails below, while the formal definition can be found in Definition 9.1.

Definition 1.5 (Fat tails) *A random variable is said to have fat tails if it exhibits more extreme outcomes than a normally distributed random variable with the same mean and variance.*

This implies that the market has more relatively large and small outcomes than one would expect under the normal distribution, and conversely fewer returns of an intermediate magnitude. In particular, the probability of large outcomes is much higher than the normal would predict. The fat-tailed property of returns has been known since Mandelbrot (1963) and Fama (1963, 1965).

A basic property of normally distributed observations is that they are completely described statistically by the mean and the variance (i.e., the first and second moments). This means that both skewness and kurtosis are the same for all normally distributed variables (i.e., 0 and 3, respectively). Skewness is a measure of the asymmetry of the probability distribution of a random variable and kurtosis measures the degree of peakedness of a distribution relative to the tails. High kurtosis generally means that more of the variance is due to infrequent extreme deviations than predicted by the normal, and is a strong, but not perfect, signal that a return series has fat tails. Excess kurtosis is defined as kurtosis over and above 3.

This suggests that a quick and dirty (makeshift) test for fat tails is to see if kurtosis exceeds 3. Recall that in Table 1.2 we found excess kurtosis to be 20, which is pretty strong evidence against normality.

Consequently, one can test for normality by seeing if skewness and excess kurtosis are significantly different from zero. A well-known test in this category is the Jarque–Bera (JB) test. Another common test for normality is the Kolmogorov–Smirnov test (KS) which is based on minimum distance estimation comparing a sample with a reference probability distribution (e.g., the normal distribution).

The KS test has the advantage of making no assumptions about the data distribution except the continuity of both distribution functions (i.e., technically speaking it is nonparametric and distribution free). It is sometimes claimed that the KS test is more powerful than the JB test because it considers the entire distribution. The KS test is sensitive to differences in both the location and shape of the cumulative distribution function, and a relatively large number of observations are required to reject the null in

practice. However, in most cases the KS and JB tests coincide. The KS test is done in R using the function `ks.test()`, while in Matlab one can use `kstest()` from the statistics toolbox.

1.6.2 Graphical methods for fat tail analysis

A number of graphical methods exist to detect the presence of fat tails. While such graphical methods cannot provide a precise statistical description of data, they can indicate if tails are fat or thin and can reveal information about the nature of how data deviate from normality. We have seen an example of this already in Figure 1.6(a) but better techniques exist.

QQ plots

Perhaps the most commonly used graphical method for analyzing the tails of distributions is the QQ plot (quantile–quantile plot). It is similar to the comparison of distributions in Figure 1.6(a), but is more accurate. QQ plots are used to assess whether a set of observations have a particular distribution, or whether two datasets have the same distribution. The QQ plot compares the quantiles of the sample data against the quantiles of a reference distribution. The code to draw QQ plots in R and Matlab is given in the following listings.

Listing 1.7. QQ plots in R

```
library(car)
qq.plot(y, envelope=F)           # normal
qq.plot(y, distribution="t", df=5, envelope=F) # Student-t
```

Listing 1.8. QQ plots in Matlab

```
% Statistics toolbox
qqplot(y) % only normal available
```

The QQ plot for the S&P 500 against the normal is shown in Figure 1.8(a).

The x -axis shows the standard normal while the y -axis measures outcomes from the data. The straight line is the normal prediction. We see that many observations seem to deviate from normality, both on the downside and on the upside, as the QQ plot has a clear S shape. The returns seem to have fatter tails than the normal but can we discover how fat the tails are?

Some idea of tail fatness can be obtained by comparing the data with a fat-tailed distribution. For example, the Student- t has fat tails, where the degrees of freedom indicate how fat the tails actually are.

In Figure 1.8(b) the Student- t with 5 degrees of freedom—that is, $t(5)$ —is chosen as the reference distribution. The returns clearly seem to be fat relative to the $t(5)$ both on

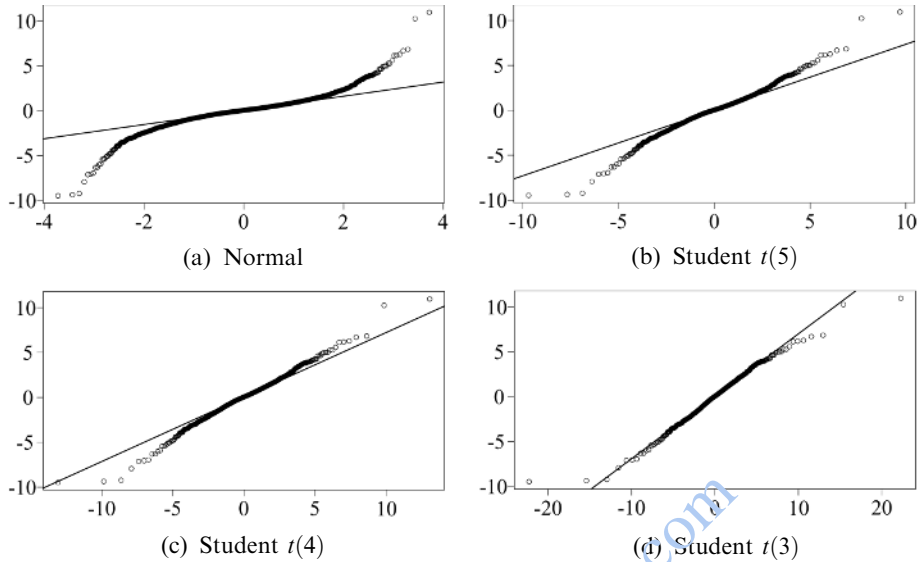


Figure 1.8. QQ plots for daily S&P 500 returns, 1989–2009.

the downside and the upside. Next we try the $t(4)$ in panel (c)—the data seem to be getting closer, the upside approximately matches the $t(4)$ while the downside is still quite fat. However, in looking at the $t(3)$ in panel (d), the data appear thin relative to the $t(3)$ both on the upside and downside. The conclusion is that S&P 500 returns have tails that are approximately equal to a $t(4)$, where the lower tail seems to be a tad thicker than the upper tail. We discuss later how we can estimate the degrees of freedom.

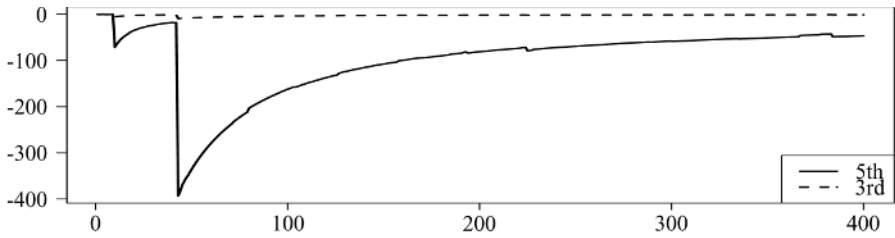
Sequential moments

An alternative graphical technique for detecting fat tails is a *sequential moments plot*. It is based on the formal definition of fat tails discussed in Chapter 9, which focuses on extreme value theory. There, the thickness of the tail of a distribution is measured by the tail index, indicated by ν . The lower the tail index the thicker the tails. In the special case of the Student- t distribution, the tail index corresponds to the degrees of freedom.

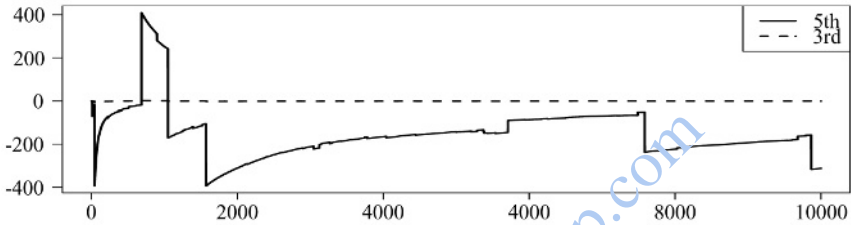
This suggests a simple graphical method of testing for tail thickness by using sample *moments* of data. The m th centered moment is given by:

$$E[(X - \mu)^m] = \int_{-\infty}^{\infty} (x - \mu)^m f(x) dx. \quad (1.1)$$

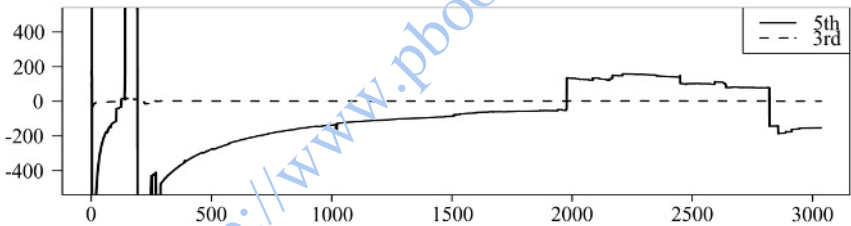
This integral does not have a finite solution for all m and all distributions. In particular, if the distribution is fat tailed, we can only compute the moments for $m < \nu$. The implication is that if the number of *bounded moments* is finite (i.e., we cannot calculate the moments in (1.1) for all $m > 0$), the data must have *fat tails*. In the case of the normal distribution we have $\nu = \infty$, so we can compute all moments in this case.



(a) Sequential moments from a $t(4)$



(b) A few more ...



(c) Oil price returns 02/01/1990–13/08/2002

Figure 1.9. Sequential moments.
Data source: Datastream.

We can therefore measure tail thickness by graphically plotting the moments of a dataset as we add more and more observations:

$$\frac{1}{t} \sum_{i=1}^t x_t^m.$$

Such sequential moment plots are shown in Figure 1.9, where panels (a) and (b) show the fifth and third sequential moments for simulated data from a Student- $t(4)$ distribution. As expected, the third moment converges but the fifth does not. Panel (c) shows the third and fifth sequential moments for returns on oil prices. Here too we find that the fifth moment does not converge while the third does, indicating that the tail index of oil returns is between 3 and 5. More formal methods for estimating the tail index are presented in Chapter 9.

1.6.3 Implications of fat tails in finance

The fact that returns are non-normal has important consequences in finance. Many methods in portfolio theory and derivative pricing assume normality of returns, and break down in the absence of normality. It is, however, in the field of risk management where the normality of returns is crucially important. An assumption of normal distribution for risk calculations leads to a gross underestimation of risk. This has been widely recognized:

“... as you well know, the biggest problems we now have with the whole evolution of risk is the fat-tail problem, which is really creating very large conceptual difficulties. Because as we all know, the assumption of normality enables us to drop off the huge amount of complexity in our equations ... Because once you start putting in non-normality assumptions, which is unfortunately what characterizes the real world, then these issues become extremely difficult.”

Alan Greenspan (1997)

Financial risk modeling is usually concerned with obtaining the probability of important but infrequent events, such as market crashes or how much money we may expect to lose in a worst case scenario. For such applications, the main concern is on the far left tail of the return distribution. Assuming normality will therefore lead to an underestimation of risk and may induce the wrong investment choice and perhaps leaving us underhedged, or overexposed.

Risk managers are in a difficult situation. If they assume normality, they are underestimating risk, often with dire consequences. However, the use of non-normal techniques is highly complicated, and unless correctly used, may lead to incorrect outcomes.

Option pricing is also crucially dependent on the underlying distribution. If the return distribution is indeed fat tailed, then using the Black–Scholes model will lead to underpricing.

The presence of fat tails has caused problems for many financial institutions. For example, in the beginning of the 2007–2009 crisis, several banks lost large amounts of money on their quantitative trading funds, such as Goldman Sachs as reported by the *Financial Times*:

“For reasons that are still unclear, shares began to move in ways that were the opposite of those predicted by computer models. These moves triggered selling by the funds as they attempted to cover their losses and meet margin calls from banks. This in turn exacerbated the share price movements.”

Financial Times (2007)

This had a strong negative impact on Goldmans. Its GEO fund was down more than 30% and its flagship Global Alpha fund, a quant fund, lost 27% by that point in 2007.

“We were seeing things that were 25-standard deviation moves, several days in a row,” said David Viniar, Goldman’s chief financial officer. “There have been issues in some of the other quantitative spaces. But nothing like what we saw last week.”

Financial Times (2007)

The *Wall Street Journal* notes the problems facing Lehmans, quoting Mr. Rothman, a University of Chicago Ph.D,

“Wednesday is the type of day people will remember in quant-land for a very long time ... Events that models only predicted would happen once in 10,000 years happened every day for three days.”

Wall Street Journal (2007)

1.7 NONLINEAR DEPENDENCE

The final stylized fact of financial returns is nonlinear dependence (i.e., the observation that the dependence between different return series changes according to market conditions). For example, most of the time, the prices of assets move relatively independently of each other, but in a crisis they all drop together.

In practice, joint extreme outcomes are more likely to occur than predicted by multivariate normality and linear correlations. For example, the probability of the joint FTSE and S&P 500 crash in 1987 is 10^{-69} if measured by historical data and assumptions of bivariate normality.

Most statistical models assume that the relationship between different returns is linear. Suppose X and Y denote returns on two assets, then—if they are *linearly dependent*—the conditional expectation $E(Y|X)$ is a linear function of X . If this is true, we can measure the strength of their *linear* dependence by using correlations, such as Pearson’s correlation coefficient ρ .

It is important to keep in mind that if $E(Y|X)$ cannot be expressed as a linear function of X , then ρ does not adequately capture the dependence structure between the two variables.

While it is tempting to conclude that the two data series are independent if the correlation coefficient is zero, in general this is not true, as illustrated by a simple example.

Example 1.1 Consider a random variable X which is symmetrically distributed about zero, such as the normal, and let Y be another random variable defined as $Y = X^2$. In this case, Y is completely dependent on X while the correlation coefficient between them is zero, because the correlation coefficient only detects linear dependencies between two variables.

Considerable recent research has shown that the assumption of linear dependence does not generally hold for asset returns, where correlations are usually lower in bull markets than in bear markets. Furthermore, if financial data were jointly normally distributed, correlations would decrease for extreme events whereas empirically we see that correlations tend to increase to one in a crisis, as demonstrated by the example in Section 1.7.1.

To capture such phenomena, models of nonlinear dependence allow the dependence structure to change according to market conditions. In this case, linear correlations overestimate dependence in non-crisis periods and underestimate correlations during crises.

Research such as Ang et al. (2001) and Patton (2002) has found that these nonlinear dependence structures command a premium in the market as investors require higher expected returns for portfolios where assets are highly correlated under bad market conditions. Aside from asset allocation, applications in risk analysis, economic capital and financial stability also focus on large outcomes. In such applications it is essential to address nonlinear dependence.

1.7.1 Sample evidence of nonlinear dependence

We illustrate nonlinear dependence with long-run correlations between the returns of three (former) investment banks, Morgan Stanley, Goldman Sachs and Bear Stearns, and one nonfinancial firm, Microsoft, for the time period May 5, 1999–September 12, 2007 (see Table 1.6). Unsurprisingly, financials are relatively highly correlated, while Microsoft is less correlated with financials at around 40%.

During the first round of the 2007–2009 crisis (August 2007), the correlations of all stocks increased dramatically. For example, Table 1.6(b) shows the correlations between Morgan Stanley and Goldman Sachs increasing from 81% to 94%. Such a high correla-

Table 1.6. Return correlations and means for Microsoft (MSFT), Morgan Stanley (MS), Goldman Sachs (GS) and Bear Stearns (BSC)

(a) Daily return correlations (May 5, 1999–September 12, 2007)				
	MSFT	MS	GS	
MS	44%			
GS	44%	81%		
BSC	38%	74%	71%	
(b) Daily return correlations (1 August 2007–15 August 2007) during the opening events of the 2007 crisis				
	MSFT	MS	GS	
MS	93%			
GS	82%	94%		
BSC	82%	92%	89%	
(c) Daily mean returns				
	MSFT	MS	GS	BSC
1999/5/5–2007/9/12	−0.007%	0.028%	0.049%	0.050%
2007/8/1–2007/8/15	−0.252%	−1.094%	−1.208%	−1.468%

Data source: CRSP.

tion indicates that the two stocks almost move in lockstep. Even Microsoft was affected by this and its correlation with financials increased substantially.

Furthermore, Table 1.6(c) shows how actual stock prices were affected by the crisis. The effect on Microsoft is relatively insignificant, but all the financials saw their mean daily return of about 0.05% fall to -1.5% . It is an empirical fact that very high correlations are usually associated with very negative returns.

It is straightforward to write an R or Matlab program to download stock prices automatically and calculate correlations. We start by modifying Listings 1.1 and 1.2. We want to calculate correlations during the period of the financial crisis (i.e., from 2007/6/1 to 2009/12/31). Bear Stearns went into bankruptcy in 2008 so we exclude it from the sample. First, we show the R code in Listing 1.9, which is then followed by the Matlab code in Listing 1.10.

Listing 1.9. Download stock prices in R

```
price1 = get.hist.quote(instrument = "msft", start = "2007-06-01",
  end = "2009-12-31", quote="AdjClose")
price2 = get.hist.quote(instrument = "ms", start = "2007-06-01",
  end = "2009-12-31", quote="AdjClose")
price3 = get.hist.quote(instrument = "GS", start = "2007-06-01",
  end = "2009-12-31", quote="AdjClose")
p=cbind(price1,price2,price3)      # combine price vectors into a
                                # matrix
y=diff(log(p))
cor(y)                          # calculate correlation matrix
```

Listing 1.10. Download stock prices in Matlab

```
price = hist_stock_data('01062007', '31122009', 'msft', 'ms', 'gs');
price=[price(1).AdjClose(end:-1:1),price(2).AdjClose(end:-1:1),
  price(3).AdjClose(end:-1:1)]
y=diff(log(price))
corr(y)      % calculate correlation matrix
```

1.7.2 Exceedance correlations

One method for documenting the presence of nonlinear dependence is by using *exceedance correlations* as proposed by Longin and Solnik (2001) and Ang and Chen (2002).

Consider two stock returns X and Y which have been standardized (mean zero and variance one). Exceedance correlations show the correlations of the two stocks as being conditional on exceeding some threshold, that is:

$$\tilde{\rho}(p) \equiv \begin{cases} \text{Corr}[X, Y|X \leq Q_X(p) \text{ and } Y \leq Q_Y(p)], & \text{for } p \leq 0.5 \\ \text{Corr}[X, Y|X > Q_X(p) \text{ and } Y > Q_Y(p)], & \text{for } p > 0.5, \end{cases} \quad (1.2)$$

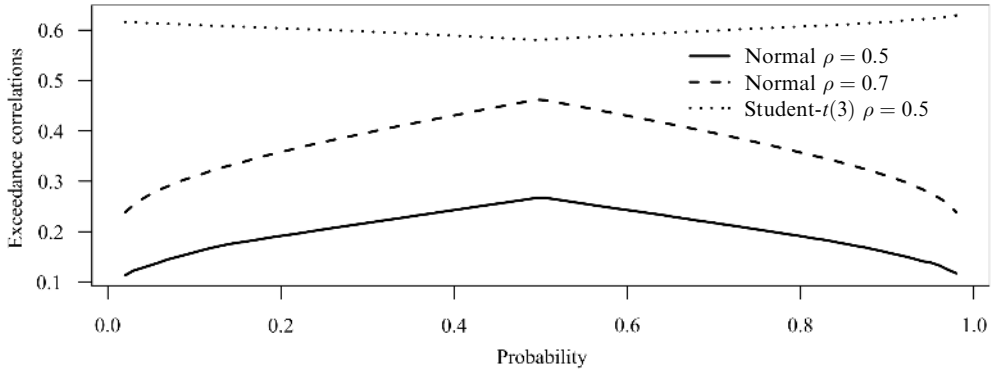


Figure 1.10. Exceedance plot for the bivariate normal and the Student- t .

where $Q_X(p)$ and $Q_Y(p)$ are the p th quantiles of X and Y , respectively, given a distributional assumption. The shape of the exceedance correlation plot depends on the underlying distribution of data.

Exceedance correlations for the standard bivariate normal and the Student- t distributions are shown in Figure 1.10. The horizontal axis shows the probability (we go from 0 to 0.5, and then from 1 to 0.5) and the vertical axis shows the correlation between the two returns given that both exceed that quantile (i.e., exceedance correlations). The plot is nonlinear in p but symmetric. Exceedance correlations decrease for the normal as we go out towards extreme quantiles, while they increase for the Student- t .

Figure 1.11 shows the empirical exceedance correlations for daily returns on Disney and IBM over 24 years, superimposed with exceedance correlations for the bivariate normal and the bivariate Student- $t(3)$ distributions with the same correlation coefficient. The exceedance correlations exhibit substantial asymmetry. The stock returns become highly correlated at the left extreme, with correlations steadily decreasing when we move to the right of the distribution. This is precisely the type of dependence structure that risk-averse investors dislike.

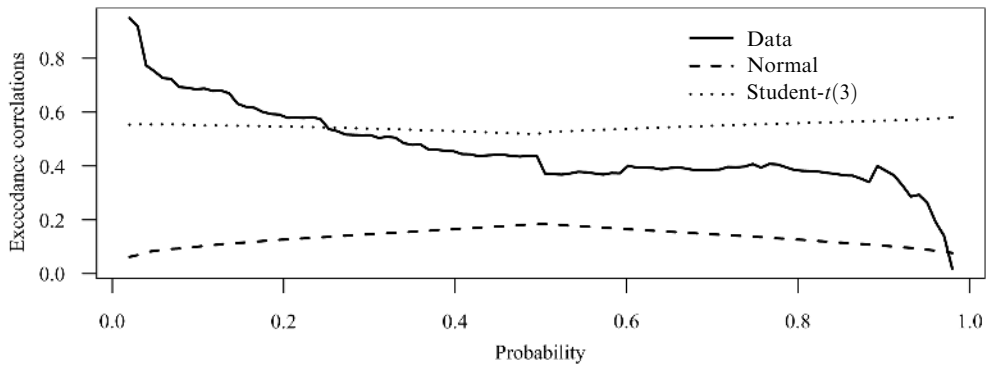


Figure 1.11. Exceedance plot for Disney and IBM returns (January 1, 1986–November 3, 2009). Data source: Yahoo Finance.

1.8 COPULAS

The correlation analysis shown in Table 1.6 along with the exceedance plots help in identifying the nature of nonlinear dependence (NLD). However, this still leaves the question of how to model NLD more formally. One approach is discussed in Chapter 3 (i.e., multivariate volatility models), but an alternative method is to use *copulas*. For more on the application of copulas in finance, see, for example, Patton (2009).

Copulas provide the means to create a multivariate distribution with a range of types of dependence. We start by taking the returns on each return series separately (called marginal returns), and transform them to the uniform distribution using the probability integral transformation discussed below.

Armed with returns transformed to a uniform random variable, we then proceed by modeling the dependence structure between these uniforms using a copula. Since the probability integral transform is invertible, the copula also describes the dependence between the original variables (returns).

In other words, we separate out the distribution of individual assets from the distribution that links them together. For example, each asset can have a normally distributed return, but taken jointly the portfolio is much more correlated in adverse market situations than in bull markets. The copula provides information about how assets behave together.

1.8.1 The Gaussian copula

Recall the normal (or Gaussian) distribution, where $\Phi(\cdot)$ is the normal distribution, and $\Phi^{-1}(\cdot)$ is the inverse normal distribution. U and V are uniform ($U, V \in [0, 1]$) random variables, and $\Phi_\rho(\cdot)$ is the bivariate normal distribution with correlation coefficient ρ . The function $C(\cdot)$ in

$$C(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

is then known as the *Gaussian copula function*, made famous by Li (2000), whose work on the Gaussian copula enabled the pricing of structured credit products (like subprime CDOs) which subsequently got blamed for the 2007–2009 crisis. The copula provides the information that links the two univariate (also known as marginal) random variables together.

For an example of a Gaussian copula see Figure 1.12(a), which shows a cross plot (scatter plot) from a bivariate normal distribution, while panel (c) shows the joint distribution and panel (e) the contours.

1.8.2 The theory of copulas

The joint distribution of multiple random variables is composed of information about each variable separately, as well as information about how the various random variables are linked together. Suppose X and Y are two random variables representing the returns of two different stocks,

$$X \sim f$$

$$Y \sim g.$$

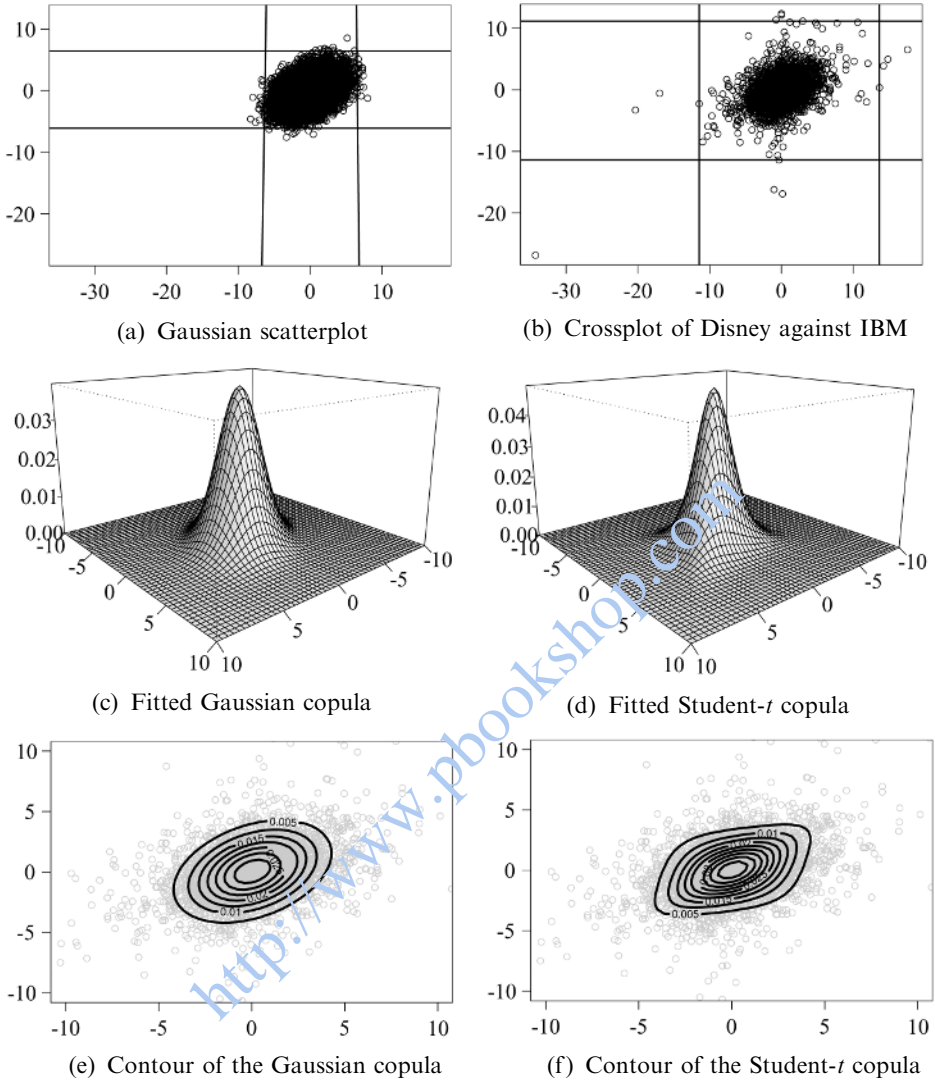


Figure 1.12. Copula plots for daily Disney and IBM returns, 1986–2009.

Together, the joint distribution and the marginal distributions are represented by the joint density:

$$(X, Y) \sim h.$$

The idea behind the copula approach is that that we focus separately on marginal distributions (F, G) and the function that combines them into the joint distribution, H . That function is the *copula*. In other words, the copula extracts information on the dependence structure from the joint distribution.

The probability integral transformation

The first step in applying copulas is to transform X and Y into random variables that are distributed *uniformly* between zero and one, which removes individual information from the bivariate density h . The *probability integral transformation* due to Fisher (1925) states that:

Theorem 1.1 *Let a random variable X have a continuous distribution F , and define a new random variable U as:*

$$U = F(X).$$

Then, regardless of the original distribution F , $U \sim \text{Uniform}(0, 1)$.

The theorem says that any continuous distribution can be transformed into a uniform variable; therefore, knowing the distribution of the uniform random variable does not imply anything about the original distribution. The probability integral transformation is a strictly increasing transformation hence it is invertible. This means that we can identify the dependence between two variables with the dependence between their transforms.

On to copulas

Apply Theorem 1.1 to the two returns, X and Y , to obtain two uniform variables:

$$\begin{aligned} U &= F(X) \\ V &= G(Y). \end{aligned}$$

A copula is a probability distribution on a unit cube for which every marginal distribution is uniform on the interval $[0, 1]$. The copula contains all the dependence information in the original bivariate density h , but none of the individual information. Sklar (1959) provides the main theory for copulas.

Theorem 1.2 *Let F be the distribution of X , G the distribution of Y and H the joint distribution of (X, Y) . Assume that F and G are continuous. Then there exists a unique copula C such that:*

$$H(X, Y) = C(F(X), G(Y)). \quad (1.3)$$

In applications we are more likely to use densities; Sklar's theorem allows us to decompose joint density by:

$$h(X, Y) = f(X) \cdot g(Y) \cdot c(F(X), G(Y)).$$

Nelsen (1999) provides a corollary to the above theorem that allows us to extract the copula from any given multivariate distribution and use it independently of the original marginal distributions. In other words, we can construct a joint distribution from *any* two marginal distributions and *any* copula, and extract the implied copula and the marginal distributions from any joint distribution.

1.8.3 An application of copulas

We illustrate the application of copulas using the same Disney and IBM data as in Figure 1.11. Recall that the data showed greater dependence in the negative quadrant.

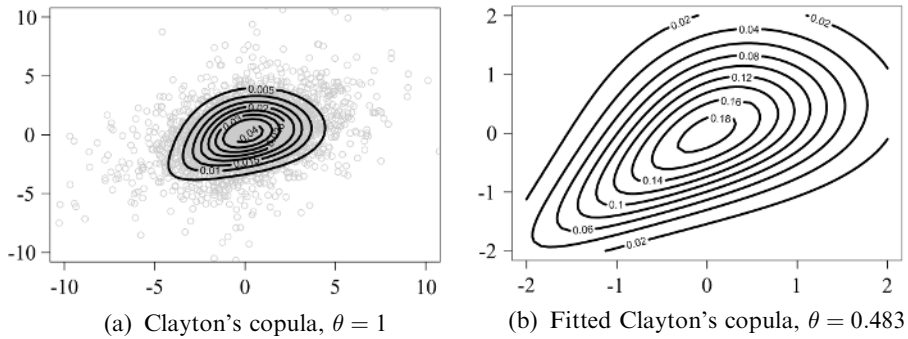


Figure 1.13. More examples of copulas, same data as in Figure 1.12.

The results are shown in Figure 1.12. Panel (b) shows a cross plot of stock returns along with their 0.05% and 99.95% quantiles. The vast majority of data points seem to be centered around zero, but there are a few outliers from both stocks that lie outside their sample quantiles.

We can analyze the dependence by comparing simulated data from the bivariate normal distribution with stock returns where both have the same covariance matrix. Panel (a) shows that simulated bivariate normal data do not have the same joint extremes as stock returns.

We estimated both a Gaussian and a Student- t copula for the data. The estimated parameter for the normal copula is the correlation coefficient, ρ , while for Student- t , the estimated parameters are the correlation coefficient and the degrees of freedom, ν . Fitted copula densities are shown in panels (c) and (d). It is hard to compare distributions from three-dimensional graphs, and the corresponding contour plots in panels (e) and (f) provide a clearer picture. The t -copula clearly has more correlated extremes, on both the upside and downside, as the plot is more stretched towards the top right and bottom left corners.

There are a number of copulas available; we present here the Clayton copula, which is often used in financial applications of return dependence (see Figure 1.13). The Clayton copula is from the Archimedean class of copulas and, unlike the normal and the Student- t , is asymmetric, exhibiting greater dependence in the negative tail than in the positive tail, with parameter θ a measure of the strength of dependence. Panel (a) shows the Clayton copula with $\theta = 1$, while panel (b) shows the estimated copula with the same Disney and IBM data as above.

1.8.4 Some challenges in using copulas

The main problem with copulas—demonstrated in the example to an extent—is that we can specify any type of dependence structure we want, where the copula need not be affected by the marginal densities we choose. In the example above, marginal densities were assumed to be the normal distribution. Currently, goodness-of-fit tests for copulas are not common and it is unclear whether a copula that has a good fit yields a good fit for the distribution of data.

One example of the possible misuse of copulas is when the rating agencies used Gaussian copulas to estimate the default correlation in a pool of mortgages, with damaging consequences, as shown in the 2007–2009 crisis.

No economic theory of copulas exists; hence, there is little guidance in choosing copulas. In turn, this means we have the freedom to choose any type of dependence structure we want.

1.9 SUMMARY

Many applications in finance are based on the assumption that returns on financial assets are IID normally distributed. This assumption has been analyzed in detail in this chapter. By using a sample of stocks of the most important stock market index in the world, the S&P 500, as well as a selection of stocks on the New York Stock Exchange, we have demonstrated that IID normality does not hold. Furthermore, it is straightforward to show similar results for most financial assets.

It is well known that financial returns follow a complicated and ever-changing probability distribution function where we can only hope to statistically model a very small part of the distribution of returns at any one time. Often, the underlying application dictates which part of the statistical distribution of returns we focus on.

The stylized facts of financial markets we have examined are

Volatility clusters

Fat tails

Nonlinear dependence

These stylized facts seem to hold for most if not all basic (i.e., non-derived) financial assets regardless of asset type, sampling frequency, observation period or market.

Other empirical results only hold some of the time. For example, return distributions are usually skewed either to the left or to the right. Returns often have a strong positive autocorrelation over long periods of time during bull markets such as the internet bubble, or negative autocorrelation during prolonged bear markets such as the Japanese stock market since 1990. At the highest frequencies, returns tend to have negative autocorrelations, but positive autocorrelations at the lowest frequencies. However, no regular first-moment or third-moment patterns about returns exist and the overwhelming conclusion is that we cannot profitably forecast prices in a systematic way using simple methods.

The focus in this chapter has been on empirically identifying these stylized facts using a range of statistical techniques. In subsequent chapters we focus on statistically modeling financial returns, primarily for the purpose of forecasting risk.

<http://www.pbookshop.com>