9. Some Professional Trading Strategies

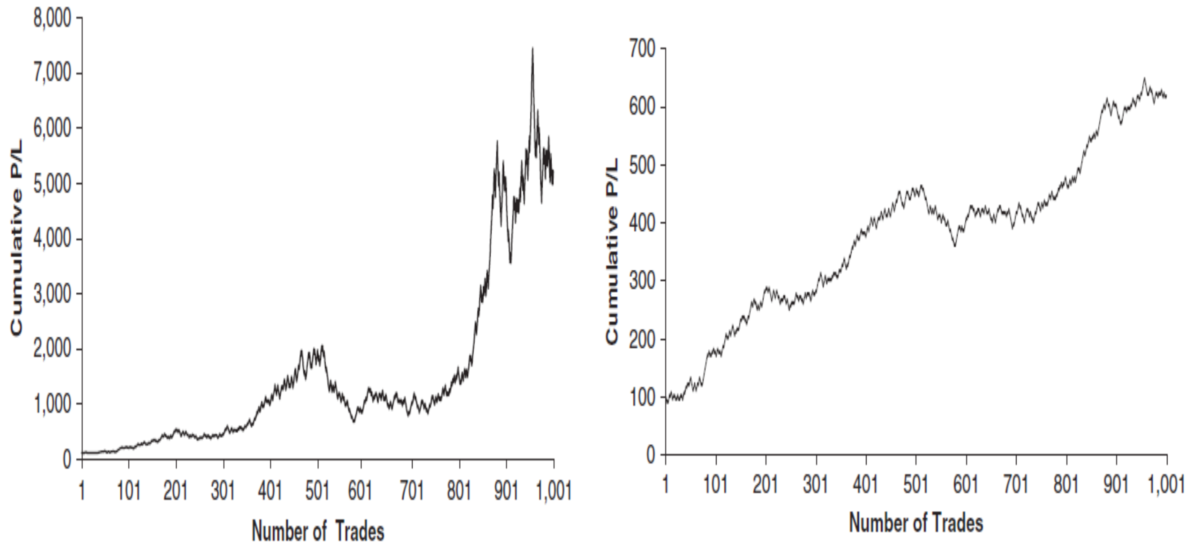
9.1. Kelly criterion in money management

9.1.1. Introduction

The presentation in this section is taken from Sinclair (2013), chapter 8.

We begin with an example. Let's say you have \$100, and you play a game where you flip a coin with probability of 0.5 for heads and 0.5 for tails. For heads, you win 2 dollars for each dollar bet, i.e. you end up with a total of \$3 after the bet. For tails, you lose your dollar bet. How much do you want to bet each time in order to optimize your earnings? Typically people use one of three possible methods:

- Trade according to "feel", i.e. change the size each time depending the trader's view.
- Trade a fixed amount every time, e.g. \$5.
- Trade a fixed fraction of the total portfolio every time.



The results can be very different. The above diagrams show some simulations about possible bankrolls with different trading strategies (from Sinclair (2013), p.132). The left diagram represents the strategy with a constant bet of 5% of the bankroll, with a final wealth of \$5207. The right diagram is a bet of \$5 for each trade, with a final wealth of \$620. The winning percentages are 550 bets out of 1000 in each case.

Intuitively, if you bet all your money on each game, you will easily lose a game and be bankrupt. However, if you bet too little, you will not make too much money. A controversial method known as the **Kelly criterion** has been derived and named after John Kelly, a physicist in Bell Labs, in 1956. Originally the method was developed in the context of information theory. Subsequently it has been popularized for blackjack gambling in casinos, as well as being used for trading in hedge funds.

9.1.2. Derivation of the Kelly criterion

We assume that the initial wealth level = W_0. In a game, when we win we will earn a%, and when we lose, the loss is b%. Furthermore, assume that each bet is a fraction f of the total current wealth. In this set up, after n wins and m losses, we will have:

Total wealth $W = W_0 \times G(f)$ where gain factor $G(f)$ is defined as
$G(f) = (1+fa)^n(1-fb)^m$,
or on a per trade basis
$G(f)^{[1/(n+m)]} = g(f) = (1+fa)^p(1-fb)^q$
$p = n/(n+m);\ q = m/(n+m)$

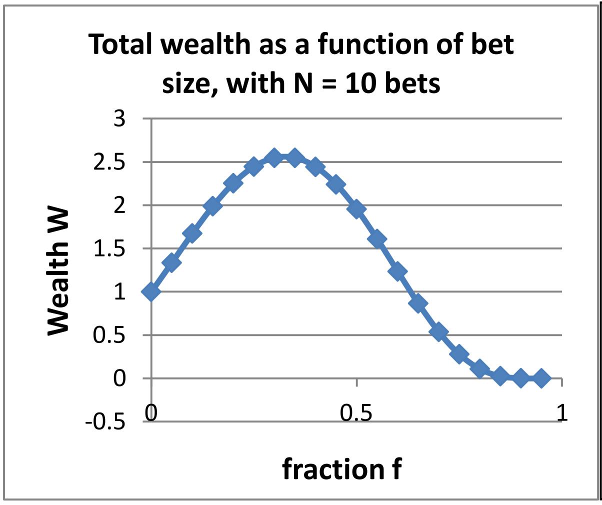
The next question is: what should be the objective function in the optimization process? It is obvious that not all objective functions are desirable. For example, we can examine the strategy for *maximizing the expected return*. Using the numbers from the earlier game, suppose that your bankroll is 1 dollar and in order to maximize the expected (mean) return, the correct strategy is for you to bet everything ($f = 1$). Let's say that the game goes on for 10 rounds. After 10 rounds, there is one chance in 1024 ($=2^{10}$) that you will have 59,049 dollars ($=3^{10}$, assuming that you win every time and triple the bankroll) and 1023 chances out of 1024 bets that you will have 0 dollars. Thus your expected wealth is 59049/1024 = 57.67 dollars (which is the arithmetic mean). We can see that the median wealth is 0 dollars. While the expected wealth is the highest, the risk is also very high as there is a very big possibility that you would end up with nothing.

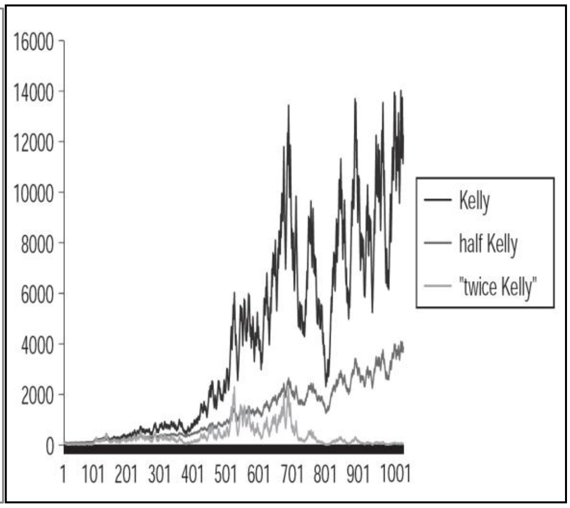
Instead of maximizing expected wealth, one can maximize a utility function, say a log utility (c.f. section 2.3.4.2). We can write: $\mathrm{Ln}[g(f)] = p\ln(1+fa)+q\ln(1-fb)$. The maximum of this function can be found by differentiating w.r.t. f and set the derivative to 0:

$$pa/(1+fa) - qb/(1-fb) = 0$$
$$f = (pa-qb)/ab$$

This fraction f is the Kelly criterion. The expected wealth after N bets is:

$$W = W_0[1 + p\ln(1+fa)+q\ln(1-fb)]^N$$





The diagram on the left shows the result when $p = 0.55$, $q = 0.45$, $a = 2$, $b = 1$, $N = 10$; Kelly ratio f is 0.325. In general, if we bet more than the Kelly ratio, the result is worse. In this example, the wealth will be lower than the initial value when $f > 0.632$. A simulation example is taken from Sinclair (2013), p.137. We can see that betting with the Kelly ratio gives the best result. Otherwise the wealth will be higher if each bet is based on "half Kelly" rather than "twice Kelly," and the difference can be quite substantial.

The case for stocks

In the stock market setting, the outcome is not binary, but will consist of a continuous range of values. The profit / loss of a trade depends on how much the stock price has changed compared to the starting level. Therefore the profit/loss profile would also follow some kind of distribution.

In order to formulate this problem, we can start with modeling the stock price by a certain probability distribution, e.g. the Geometric Brownian motion as used in the derivation of option formulas. Using the previous notation, the gain factor can now be written as:

$$G(f) = \prod_{i=1}^{N} [1 + fh(X_i)]^{p_i}$$

where X_i is a random variable, p_i is the probability of having a payoff, and $h(X_i)$ is the payoff. The expected value of the log of the gain factor is:

$$E[\ln(G(f))] = \sum_{i=1}^{N} \left[p_i \ln(1 + fh(X_i)) \right]$$

We can make use of an approximation to simplify this expression. If z is small, the logarithmic function can be approximated by:

$$\ln(1+z) = z - z^2/2 + z^3/3 - z^4/4 + \cdots$$

Using the first two terms only, and finding the maximum by differentiating w.r.t. f and setting the result to zero, we have

$$f = (\sum [p_i h(X_i)]) / (\sum [p_i (h^2(X_i)]) = r / \sigma^2$$

where r is the mean of the distribution of $h(X_i)$, and σ^2 is close to the variance of $h(X_i)$ – it should be $E[h^2(X_i)]$, which is close to the variance if the mean is small. Thus the Kelly ratio can be applied in the case of stock trading.

9.1.3. Some Properties of the Kelly Criterion

We have seen above that maximizing the expected wealth may not be a good idea for an investor. In deriving the Kelly fraction, we are actually maximizing the geometric mean of wealth, or the arithmetic mean of the log of wealth (represented by the gain factor). To summarize the properties of the Kelly criterion, in the long term (i.e. with an infinite sequence), the following propositions would definitely occur (with probability = 1):

- Maximizes the final value of the wealth (compared with any other strategy).
- Maximizes the median of the wealth.
- Minimizes the expected time required to reach a specified goal for the wealth.

However, from the diagrams above we can also see that the path of outcomes can be very volatile. We can state the probability that we reach a certain wealth level $B \times W > W_0$ before we dip to $A \times W < W_0$. This is given by (c.f. Sinclair (2013), p.139)

$$P(A,B) = \frac{1 - A^{1-2/f}}{B^{1-2/f} - A^{1-2/f}}$$

For example, if A = 0.5 and B = 2, we get $P(A,B)$ = 2/3, which means that there is 1/3 chance of having the wealth level to become half before it doubles – the risk is very high. Therefore it is common to trade at less than Kelly ratio to reduce the risk.

9.1.4. Comments about the Kelly criterion

The advantages of adopting the Kelly criterion include the following:

- It is a proportional strategy and will never lose more than the initial wealth.
- On average, this strategy outperforms others.
- Any new trade is a function of the current wealth and is not dependent on previous trades.
- One can control risk by trading at less than Kelly ratio.

Some disadvantages of the method include:

- Maximizing the log utility of wealth may not always be the best option.
- Trade size can be big if the probability of winning is high, hence the trading risk increases.
- The strategy is aggressive and the outcome can be volatile, which is worse for a risk-averse investor.
- A long time is needed before it dominates other strategies.

9.2. Trading strategies with high leverage

9.2.1. Introduction

Leveraging strategies are often employed in order to achieve a high return with limited capital. To demonstrate the use of leverage, assume that you have $10000 to buy shares. If share price goes up by 10%, your return is 10%. However, if you can borrow money to invest, the potential return is much higher. For example, you have $10000, but someone lends you an additional $40000, so that you bought $50000 worth of shares. If share price goes up by 10%, your profit is $5000, and this represents 50% of your original investment

In the financial markets, there are many ways to achieve the leverage effects. Some of these methods are:

- Margin accounts (stocks, FX). Instead of paying the full price to enter a position, only a certain percentage is required. However, there is a need to maintain a minimum margin level, otherwise there would be a forced termination of the position.
- Use of futures and options. For example, for HSI futures, we need to pay the equivalent of 1365 index points in margin (as of Jan 17, 2017) to control an asset which is roughly equal to 23000 points.
- Long/short trades (pairs trading). Very high leverage is attainable because it can require very little initial investment. Further details are given in the following sections.

We introduce a term known as *market neutral strategy*. These strategies do not make assumption as to whether the market will rise or fall, and are based on how two assets will move relative to each other. One typical example is called pairs trading, which involves entering positions in two assets. For example, we can go long 1000 shares of stock A at $10 per share, and short stock 2000 shares of stock B at $5 per share on trade date, so that no initial investment is required (other than margins). If share price of A goes up by 10% and share price of B goes up 5%, a total profit of \$1000 – \$500 = \$500 is made. In fact, as long as the relative performance of A is better than B, a profit is earned (e.g. both stocks can go down, but still a profit is possible). The strategy would incur a loss if B outperforms A, irrespective whether there is an upward or downward move in the stock prices.

9.2.2. Pairs trading

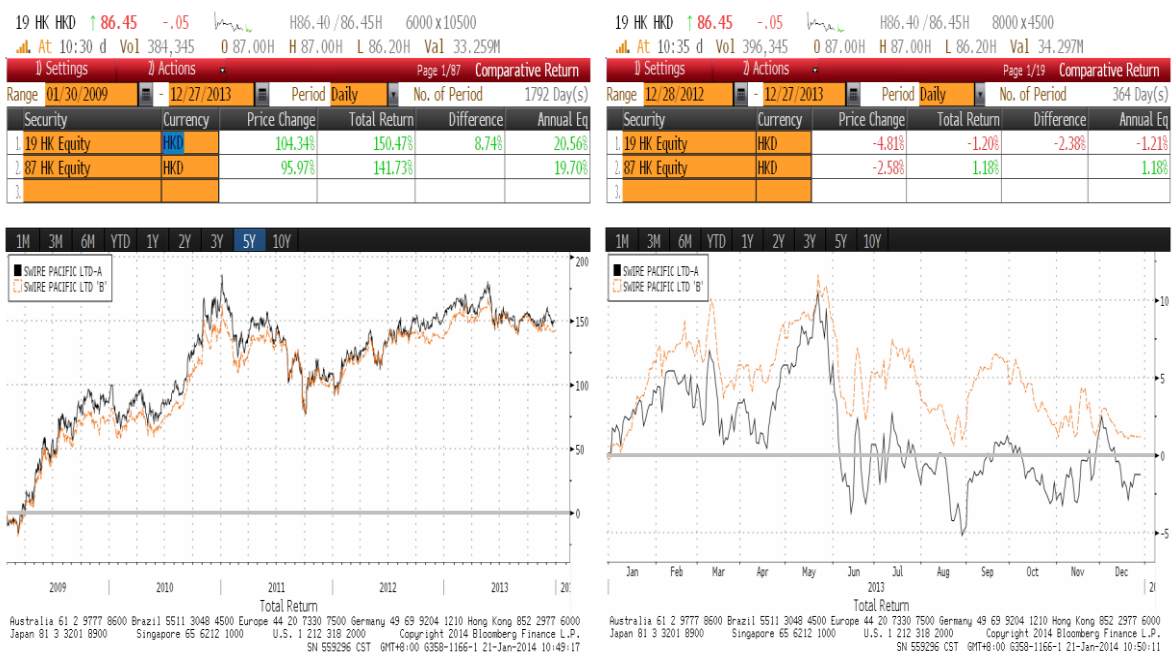
Paris trading is sometimes known as relative-value trading, where the methodology is heavily based on whether two securities will revert back to an underlying long-term relationship between them (a **"mean-reverting"** behaviour). The idea is straightforward: enter a long position in one asset and a short position in another asset. In practice, one needs to consider the cost and possibility of short selling (such as the availability of stock borrowing and regulatory constraints concerning short positions).

From a trading point of view, the key concern is to establish rules to pick the stock pairs. There are two main groups of strategies. The first type is based on fundamental company characteristics. Some of the typical strategies include:

- Different classes of security for the same company listing in the same market.
- Different listing, such as the local shares against the American Depository Receipt (ADR).
- A company listing simultaneously in different markets, e.g. A-shares vs H-shares.
- Ratio trading based on cross holdings between the companies.
- Outperformance trades based on company characteristics, e.g. industrial sectors.

The second type of strategy will not pay attention to the company's data, but the selection criterion is purely based on past performance. A popular strategy is known as statistical arbitrage, where the time series properties of historical prices would be analyzed. This will be discussed in the next section.

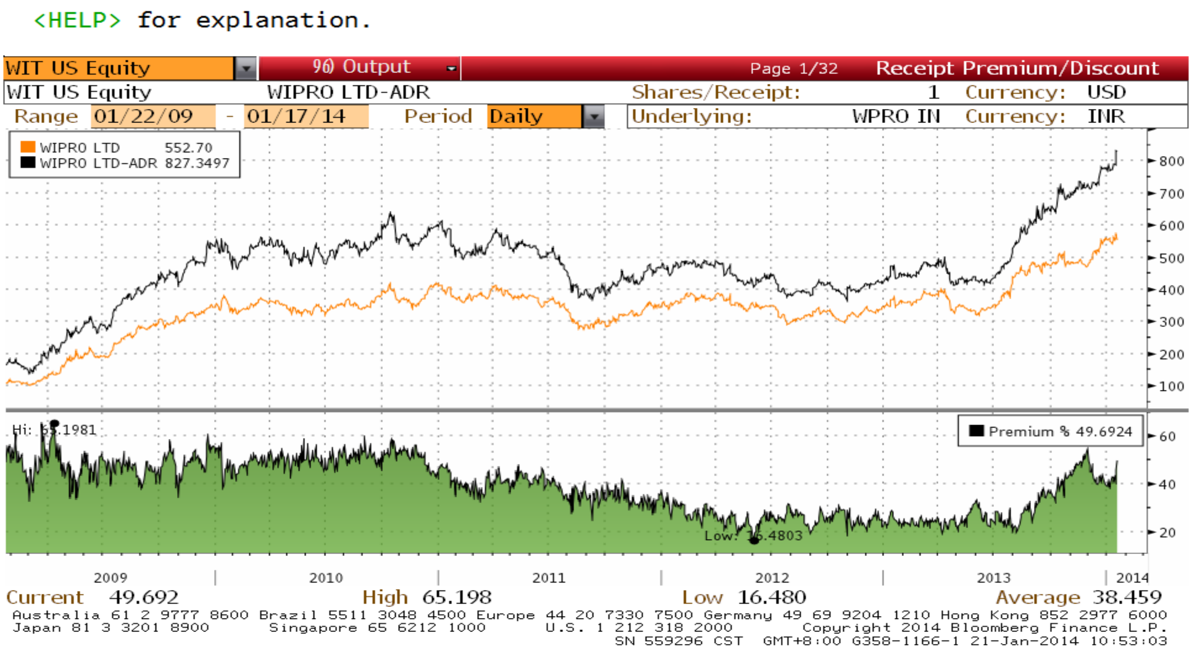
Similar class



Source: Bloomberg LP

The relative performance of two stocks, Swire Pacific ‘A’ (0019) and Swire Pacific ‘B’ (0087), are shown in the diagrams above. Both these stocks are based on the same company, although the shares differ in voting rights and dividend policy. It is noticed that the liquidity of Swire Pacific ‘B’ is much lower since it is largely owned by the majority shareholder, and the dividend yield of Swire Pacific ‘B’ is often higher than Swire Pacific ‘A’. Theoretically, there should be no difference between the performances of these shares, although short term discrepancy can be seen, mainly due to trading factors. It is thus possible to take advantage of the short term deviation from the fair value and wait for the stock prices to converge. In practice, one side of the trade may not be feasible, as Swire Pacific ‘B’ is no longer eligible for short selling according to the rules of the Hong Kong Exchange.

Long/short example: premium extraction trade



Source: Bloomberg LP

Wipro ADR (listed in New York) is usually trading at a premium to the local Indian stock due to foreign ownership rules for the local stock in India. However, the premium level keeps changing – within the five year period between 1/2009 and 1/2014, we can see that the premium has a range roughly between 20% and 60% (as seen in the lower graph). A possible trade is as follows: if it is anticipated that the premium will trade higher, one strategy is to (1) borrow the local stock and sell it; (2) convert the Indian Rupee thus received to US Dollar; (3) use the proceeds to buy the ADR. As long as this view of the market is correct, whether the company share price is going up or down is irrelevant.

Long/short example: related companies

Source: Bloomberg LP

Hutchison (0013) owns 76.4% of Cheung Kong Infrastructure (1038). As of January 2014, Hutchison's price/earning (P/E) ratio is 17 times, with a market capitalization of HKD 450 billion. Cheung Kong Infrastructure (CKI) has a P/E ratio of 12.8 times, with a market capitalization of HKD 117 billion. While the P/E ratios of companies from different sectors are not comparable, a possible trade idea is to have a long position in CKI and short position in Hutchison. From the fundamentals given above, a possible ratio is to hold 2.5 share of CKI and short 1 share of Hutchison.

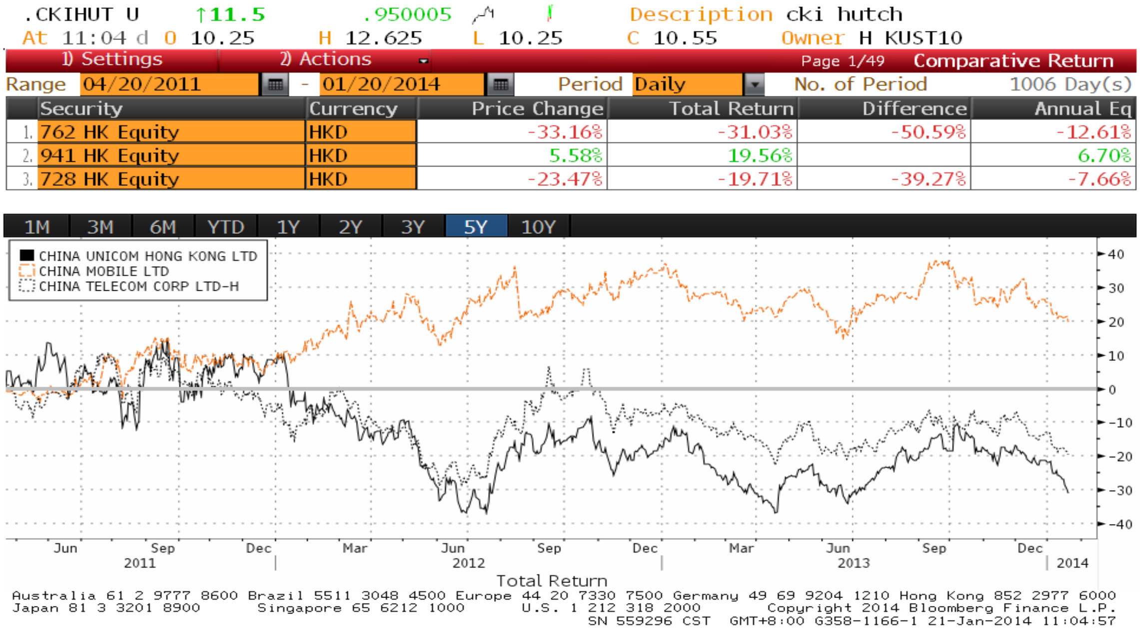
Let's say you put on the trade on May 31, 2011, when the prices of Hutchison and CKI were HK$ 90.00 and HK$ 36.75. The cost of trading the strategy = HK$ 2.50 x 36.75 – 90 = $1.875. Other transaction costs should include the borrowing fee for Hutchison, stamp duty, funding cost etc. If transaction costs are ignored and if CKI's price moves up by 1% (0.3675) while Hutchison's price remains unchanged, the portfolio return is 0.3675/1.875 = 19.6%, i.e. the portfolio has a leverage of 19.6x. This trade has been very successful in 2011. On Sep 30, 2011, the spread was worth $55.80. As of Nov 4, 2014, the spread was $41.775 (Hutchison's price was $97.10, CKI's price was $55.55). Note that this strategy is often used in short term trading instead of for long term investments.

Source: Bloomberg LP

There is no particular reason why a ratio of 2.5 to 1 must be used. If we use a different ratio, such as long 1.33 shares of CKI and short 1 share of Hutchison, the results are shown in the above diagram. We can see that the profit is less.

Sometimes it is even possible to construct a zero cost strategy (other than the transaction costs), so that an extremely high leverage could be attained. However, setting the trading ratio is more of an art than a science.

Other pairs trading examples



Source: Bloomberg LP

Another common strategy is to look for outperformance of one company against another within the same sector. For example, we can compare the performance of the telecom companies in China. Three such stocks listing in Hong Kong are China Unicom (0762), China Mobile (0941) and China Telecom (0728). From the diagram above, we can see that China Mobile has outperformed the other two in the last few years. One can then take a position as to whether this trend will continue. Another possible strategy is to compare the forecasts of companies in the same sector from different countries, such as Citibank vs HSBC. [1]

9.2.3. Statistical arbitrage

Another kind of pairs trading strategy does not rely on company fundamentals. Statistical arbitrage uses time series techniques to determine trading signals. Again it is based on the concept of mean reversion, which suggests that the relationship between two securities fluctuates around a constant mean. Whereas the relative behavior between the two securities may deviate from the mean temporarily, it will return to normal after a certain time. This relationship would be quantified, and the typical methods include:

- Measuring the spread or ratio between two securities.

[1] See Ganapathy Vidyamurthy, *Pairs Trading: Quantitative Methods and Analysis*, Wiley Finance (2004) for additional details.

- Regression-based methods.
- Establishing a co-integrated time series.

An arbitrage is supposed to mean that a riskless strategy is conducted by simultaneously trading different assets. On the other hand, there is no guarantee that the relationship must revert back to normal levels. As such, statistical arbitrage cannot be considered as a real "arbitrage"! However, these strategies are very popular in recent years. A detailed discussion is outside the scope of this course.[2] There has been an increasingly large literature based on high frequency trading algorithms and techniques, which can also be considered to belong to this class of strategies.

A simple example

Define the log spread as:

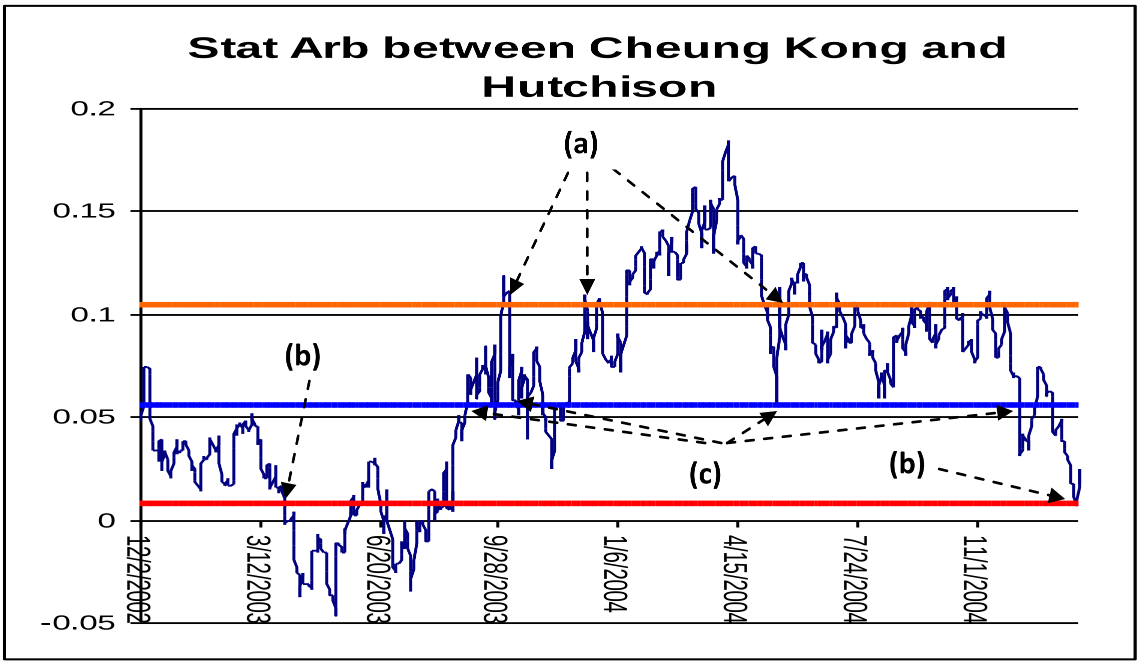
$$s_t = \ln(y_t) - \ln(x_t)$$

where x_t and y_t are the prices of the two stocks at time t. The mean and variance are defined by:

$$\bar{s} = \frac{1}{T}\sum_{t=1}^{T} s_t,\ \sigma_s^2 = \frac{1}{T-1}\sum_{t=1}^{T}(s_t - \bar{s})^2$$

Simple trading rules can be set up, for example:

- If $s_t > \bar{s} + k\sigma_s$, go short in stock y and long in stock x
- If $s_t < \bar{s} - k\sigma_s$, go long in stock y and short in stock x
- when s_t returns to $\bar{s}$ close out of position



We use historical data for Cheung Kong and Hutchison to illustrate this trading idea. k has been chosen to be 1.8. (a) represents the signal to sell Cheung Kong and buy Hutchison; (b)

[2] Refer to the appendix at the end of this chapter for an introduction to time series techniques.

is the signal to buy Cheung Kong and sell Hutchison; and when (c) is breached the existing position is closed out in order to realize the profit. The trading strategy will have a profit of HKD 18,900 for trading size of 1000 shares each (before transaction cost) over the monitoring period. This is a naïve example which is unrealistic, and real life statistical arbitrage strategies would aim to trade much more frequently while hoping to make a small profit for each trade. Hence it is very important to make use of historical data to train the model in order to identify optimal signals for profitable trades.

Co-integrated time series

A more sophisticated model is to employ the idea of co-integration. Stock prices are often modelled as a random walk process, which is non-stationary. In 1987 Clive Ranger and Robert Engle proposed that for two non-stationary time series x_t and y_t , there may be some kind of relationship between the two: if a specific linear combination becomes a stationary series (say $y_t - \gamma x_t$), the two original series are known as co-integrated. (This idea is crucial in earning the Nobel prize in 2003). More formally, if ε_{xt} and ε_{yt} are white noise processes,

$$y_t = \alpha_y (y_{t-1} - \gamma x_{t-1}) + \varepsilon_{yt}$$
$$x_t = \alpha_x (y_{t-1} - \gamma x_{t-1}) + \varepsilon_{xt}$$

An alternative representation is the *common trends model* from Stock and Watson:

$$y_t = n_{yt} + \varepsilon_{yt}$$
$$x_t = n_{xt} + \varepsilon_{xt}$$

where n_{xt} and n_{yt} are the non-stationary components of the time series and ε_{xt} and ε_{yt} are the stationary components. We can write the co-integrating combination as:

$$y_t - \gamma x_t = (n_{yt} - \gamma n_{xt}) + (\varepsilon_{yt} - \gamma \varepsilon_{xt})$$

Since this combination is stationary, it means that the non-stationary component must be 0, thus:

$$n_{yt} - \gamma n_{xt} = 0 \text{ or } n_{yt} = \gamma n_{xt}$$

Issues in strategy design

The first step in designing an appropriate trading strategy is to identify stock pairs that could potentially be co-integrated. From finance theories such as the APT model, stock prices can be related to common factors, so many stocks can potentially satisfy this property. The next step is to conduct a test of co-integration: determine the co-integration coefficient γ and examine the spread series to ensure it is stationary and mean-reverting. Once a satisfactory pair of stocks has been identified, one needs to set up trading rules by defining the entry and exit levels, taking into account of trading costs and anticipated slippage due to market factors.

A successful implementation of the strategy would only be possible based on extensive back-testing of models using historical data. Therefore it is very important to pay attention to data quality. Closing prices may need to be adjusted because of various technical factors such as corporate actions. Furthermore, if a strategy involves comparing stock prices from two different countries, especially for countries at different time zones, there may be a need to make adjustments due to different market hours and holidays.

Not all companies can engage in pairs trading strategies. In particular, there may be different regulatory concerns. In Hong Kong, short selling is not allowed for every stock.[3] Some types of companies such as mutual funds or retirement funds are governed by ordinances and leveraged trading is not allowed. Banks must fulfilled capital adequacy ratios (CARs), with 8% of Tier 1 capital being the global standard, and thus there is a limit to the level of leverage that can be attained. Of course, un-regulated hedge funds are not subject to these restrictions.

9.3. Index arbitrage
9.3.1. Introduction

A straightforward application of the arbitrage relationship is the index arbitrage strategy, which is also known as cash-and-carry arbitrage. Recall from section 6.2.2 that the fair forward value of the index is given by: $F=(I-D)\text{ x}(1+rt)$, where I=current index, D=dividends, r = interest rate, and t = time to maturity. If the futures contract is trading above or below F, arbitrage opportunity exists. Unlike the previously described strategy such as statistical arbitrage, index arbitrage is a real arbitrage, i.e. if the unwinding of the trades at maturity can be conducted without slippage, a guaranteed profit can be made.

The strategy can be one of two types. Three trades are needed in each operation:

- If futures is trading above F (i.e. futures trading "at a premium"), the strategy is known as *regular arbitrage*: sell futures, borrow money at rate *r* for a fixed term, and buy stock basket (index).
- On the other hand, if futures is trading below F (i.e. futures trading "at a discount"), the strategy is known as *reverse arbitrage*: buy futures, borrow stock basket and short sell the stocks, and lend the money received at rate *r* for a fixed term.

Index arbitrage example

The following market rates are available: spot index level = 23900, interest rate = 0.10%, expected dividend = 100 points, time to maturity = 0.0694 yrs, index futures trading at 23890 points. The fair value of futures is given by:

$$F = (23900 - 100)\text{ x }(1+rt) = 23801.65$$

Since the calculated fair value is less than the market price of the futures, the correct strategy is regular arbitrage, i.e. borrow money, buy stock basket, and sell futures at 23890. The theoretical expected profit at maturity = 23890 – 23801.65 = 88.35 points per basket (before cost).

Why would arbitrage opportunity exist?

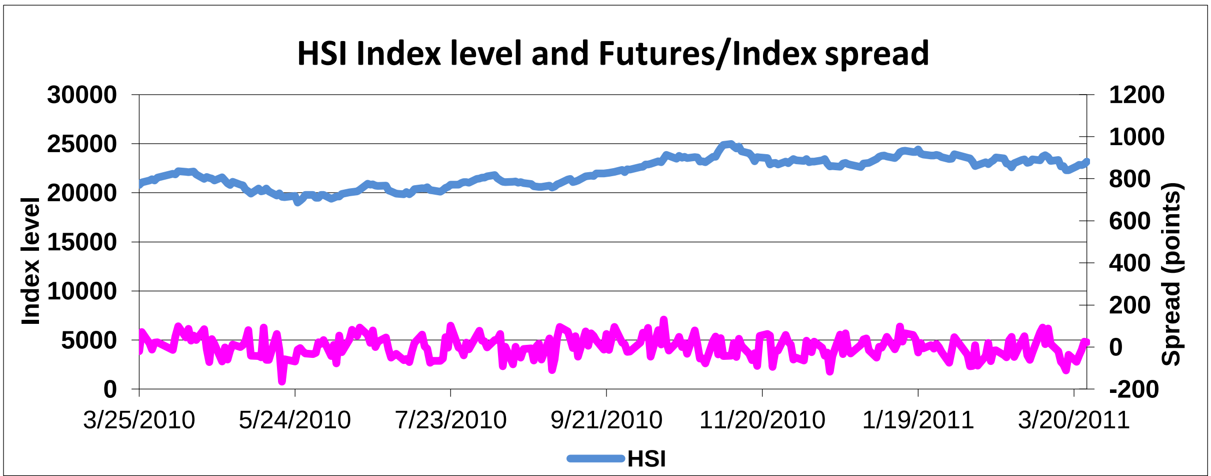
While finance theory usually assumes that arbitrage opportunities are rare, in reality they can frequently be seen, although profits can be small when various costs are taken into account. Reasons for the existence of these opportunities include:

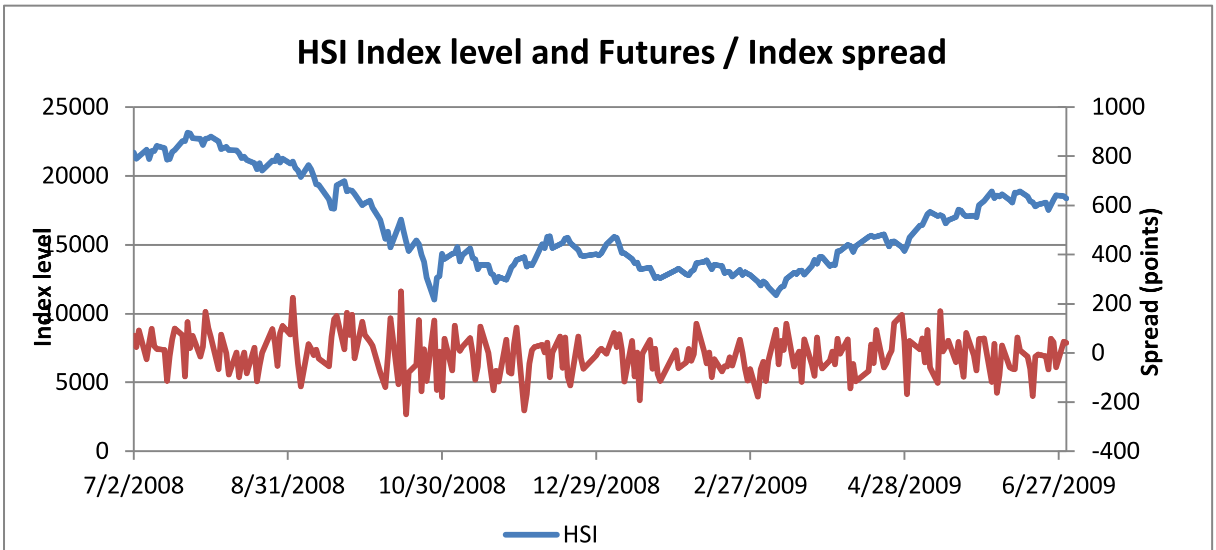
[3] Check the "Designated securities eligible for Short Selling" in http://www.hkex.com.hk under "Market Operations / Securities Trading Information".

- Convenience of futures trading as a speculative tool, and thus there can often be a temporary discrepancy between the cash and futures markets if sensitive news come out.
- Some stocks within the index may be very illiquid which would affect index calculation.
- Regulatory issues – it is not possible to short sell in certain Asian markets, therefore reverse arbitrage could not occur, and futures could trade well below its fair value. Companies which have enough inventories of stocks may be able to take advantage of this situation.
- Technical issues – limits in stock price movements exist in some markets, therefore futures could move more than the cash market when extreme moves are seen. For example, stocks may be limited to move only 5%, whereas theoretically the futures price can move more.

9.3.2. Index arbitrage in practice

In a well-behaved market, the premium and discount of the futures prices around the cash index level fluctuate from one side to another. This is also evident during the volatile periods in the most recent financial crisis in late 2008 / early 2009.



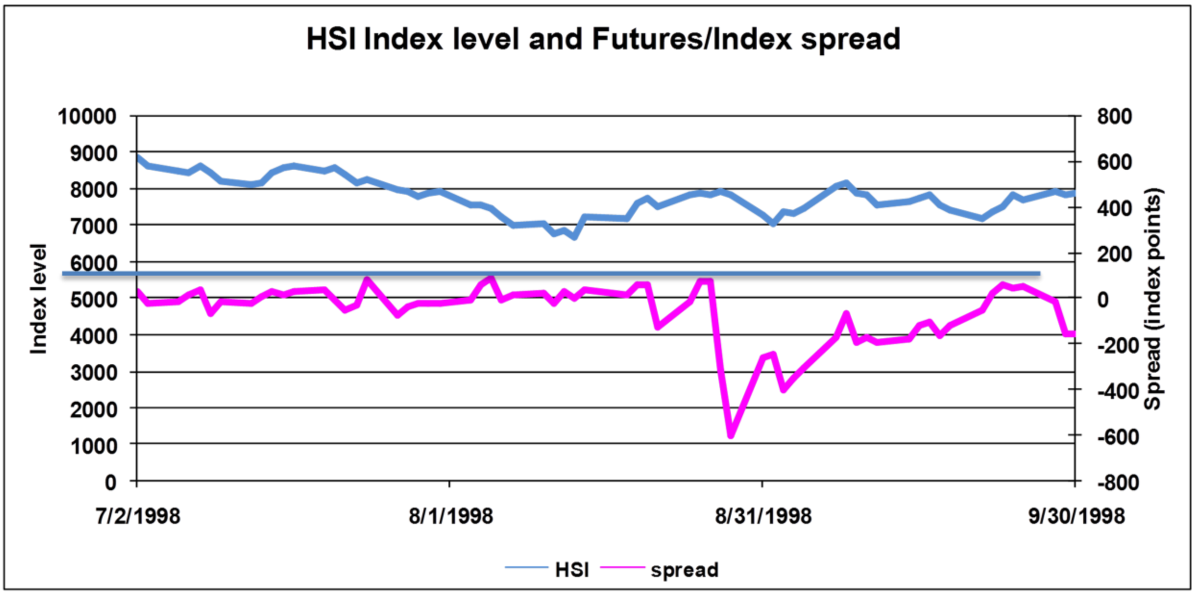


In engaging the strategy, a number of practical issues need to be taken care for:

- Most indices are only updated in regular intervals. For example, Hang Seng Index is published every 2 seconds using the last traded prices of each stock, and the Shanghai

Composite Index gets updated every 5 seconds. Instead of relying on the published index, one must implement the exact formula so that real time calculation of index value is available.

- In order to calculate the actual cost of the replication strategy, we need to use the correct side of the stock bid/ask price. For a higher chance of executing the trades, we can use all bid prices if we are selling stocks, and all ask prices if we are buying stocks.
- Other market parameters also need to be accurate, including the spread between borrow and deposit rates, and the best estimation of future dividends (if not announced already).
- Electronic trading facilitates the simultaneous execution of multiple stock orders, but one must know the market well and check the liquidity of stocks, especially for some of the more minor components in the index. For example, is it possible to complete the buy/sell orders when we need to trade stock baskets worth 50 futures contracts? Sometimes we may have to make use of a smaller "tracking" basket to represent the index (c.f. S&P 500 has 500 constituent stocks). However, this would certainly increase the slippage (i.e. tracking error). Another cause of slippage is the rounding errors due to board lots. If the trade size is too small, the resulting basket would deviate from the theoretical value because of board lot constraints.
- Transaction costs: as of Jan 2017, in Hong Kong the stamp duty and transaction fee together is 0.1077% of transaction amount (roughly equal to 25 index points when HSI is at 23000). Therefore, the discrepancy should at least be twice this level in order for the strategy to become profitable, even assuming no further slippage in profit realization (see the last section).



A reverse arbitrage strategy can be more risky, because one needs to borrow stock and short sell, whereas the stock borrowing may not always become available. One extreme example was the incident in Hong Kong in August 1998.[4] From the diagram above,

[4] c.f. Paul Draper and Joseph K.W. Fung, "Discretionary Government Intervention and the Mis-pricing of Index Futures," *Journal of Futures Markets*, Special Issue 23(12):1159-89, 2003.

we can see that shortly before the maturity of the August contract, futures was trading at a discount of almost 7% to the cash index level. However, stocks could not be borrowed. Those companies that engaged the reverse arbitrage strategy earlier in the month (when the discount was at 3% to 4%, which in itself was almost a record level) were forced to close out the position before maturity, and suffered a great loss as a result.

stock code	name	shares outstanding (millions)	10-Feb close	board lot	shares required for 25 futures	rounded
5	HSBC	13,413	79.2	400	52,225	52,400
700	Tencent	1,117	533.5	100	4,350	4,400
941	China Mobile	6,034	73.55	500	23,493	23,500
939	CCB	84,146	5.18	1,000	327,632	328,000
1299	AIA Group	12,044	35.65	200	46,895	46,800
1398	ICBC	69,435	4.64	1,000	270,353	270,000
3988	Bank of China	79,441	3.18	1,000	309,313	309,000
883	CNOOC	17,859	12.1	1,000	69,536	70,000
13	Hutchison	2,132	96.45	1,000	8,300	8,000
27	Galaxy Entertainment	2,325	69.6	1,000	9,053	9,000
857	PetroChina	21,099	7.67	2,000	82,151	82,000
1	Cheung Kong	1,390	112.7	1,000	5,411	5,000
2628	China Life Insurance	7,441	20.55	1,000	28,973	29,000
386	Sinopec	25,513	5.78	2,000	99,339	100,000
1928	Sands China	2,419	58.2	400	9,418	9,600
388	Hong Kong Exchanges	1,103	121.1	100	4,296	4,300
16	Sun Hung Kai Prop	1,352	95.15	1,000	5,265	5,000
2318	Ping An Insurance	2,034	60.55	500	7,921	8,000
2	CLP	1,895	58.05	500	7,378	7,500
11	Hang Seng Bank	765	119	100	2,978	3,000

The above shows a partial list of stocks required for 25 HSI futures as of Feb 10, 2014 (see section 5.1.1.). On that day, HSI closed at 21579.26 points. For the equivalent of 25 futures contracts, the size of the basket is: 25 x 21579.26 x 50 = $26974075 ($50 being the value per index point). Using the proportions given in the “shares outstanding” column, the size of one basket is the summation of (shares outstanding x stock price), which is $6927788.53 in this example. Hence 25 futures contracts is equivalent to 26974075/6927788.53 = 3.8936 baskets. This is the multiplier used to work out the exact number of shares for each stock. For example, the theoretical number of shares of Tencent to be traded = 1117.27 x 3.8936 = 4350 shares. Once we obtain the figure for each stock, we can apply the board lot constraints to work out the sizes of the orders.

9.3.3. Profit realization

Once an arbitrage trade has been carried out, the theoretical arbitrage profit is “locked in.” However, unless the position is closed successfully, the profit could not be realized. Given that the arbitrage condition is only satisfied at maturity, the profit can be realized only if the stocks are unwound at exactly at the price level of the futures final settlement price, because it is only at the maturity date when the futures price and the underlying index level are guaranteed to converge.

There are a few different ways to realize the profit. The straightforward way is to unwind the whole trade at maturity. Depending on the market, there may be difficulties in execution. For example, one needs to know exactly how futures settlement price is determined. In Hong Kong, the Expected Average Settlement Price (EAS) is based on calculating the index level at every five minute interval and taking the average throughout the day. Thus the correct unwind strategy is to liquidate the stock basket in equal proportions at every five minute interval. This is impossible to achieve in practice, and thus execution slippage would occur. On the other hand, in a market like Tokyo, the opening print is used in deciding the settlement price of the futures contract. Therefore it is guaranteed that the whole stock basket order can be completed other than the transaction costs to be paid.

Another way to realize the profit is to look for market opportunity to reverse the initial trades. For example, if the original trade was carried out when futures was trading at a premium, i.e. a regular arbitrage, reversing the trades mean that the positions would be unwound when an opportunity for a reverse arbitrage exists. This method of unwinding is potentially the most profitable, but sometimes there may not be such an opportunity before the current futures contract expires. In this case, one could roll over to the next futures contract (maybe at a gain or loss) in order to wait for a good trading opportunity.

9.4. Leveraged exchanged traded fund (ETF)[5]

9.4.1. Introduction

Exchange Traded Funds (ETFs) have been very popular in recent years. These instruments are seen to be transparent investment vehicles as their performances would track the underlying indices closely. Another product which appears to be similar to the vanilla ETF, known as the leveraged ETF, has been listed in the US markets since 2006.

There had been great mis-understanding about the risk and behavior of this product among many investors, including professional traders and investment advisors. Originally these products are supposed to be designed for short term traders and hedge funds only, but many investors treat these to be leveraged bets of the underlying indices. It turned out that in the US, most of these ETFs underperformed the underlying indices, and some of these funds suffered big (unexpected) losses during the financial crisis in 2008 and early 2009. In Asia, South Korea, Japan, Taiwan have listed leveraged ETFs, and some of these funds performed extremely well in 2010 when the market rebound. Hong Kong has finally introduced a limited set of these funds in 2016.

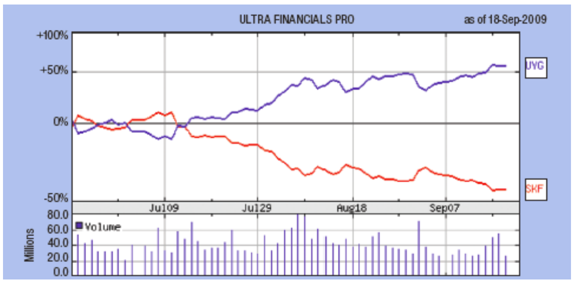
Leveraged ETFs are available for sectors including equities, commodities, fixed income, and foreign exchange. Basically, the daily return of the ETF is $R = x \times I$, where x is the leveraged factor and I is the daily index return. Typically x = 2 or 3 (positive leveraged) or x = –1, –2, –3 (inverse leveraged). For example,

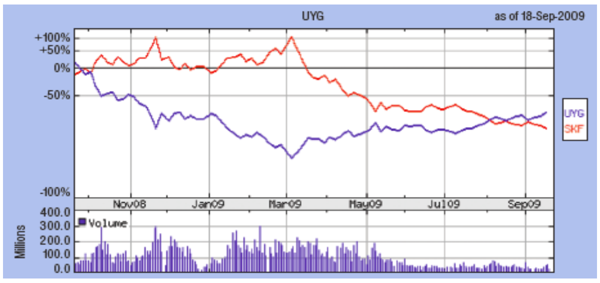
- Proshare UltraPro Dow30 (UDOW): 3x leveraged on Dow Jones Index
- Proshare UltraPro Short Dow30 (SDOW): –3x leveraged on Dow Jones Index

In the ETF prospectuses, it is stated clearly that the return horizon is only 1-day. However, the misunderstanding is that the fund can track the long term performance of the index. An explanation is given in section 9.4.4.

[5] Good references include Cheng and Madhavan (2009) and Avellaneda and Zhang (2009).

9.4.2. Performance of some funds





The above diagrams show the performances of two ETFs in 2008/9.[6] These two funds track the Dow Jones Financial index offered by Proshares (UYG and SKF); one is 2x, the other is –2x. The left diagram shows the fund prices between June and September 2009. The index went up during this period, and the performances of the two funds seem to be a mirror image of each other, which is what one might expect to see. However, if we look at a longer term picture, from Sep 2008 to Sep 2009, we see that both funds lost money during the period. This may seem surprising, because if one fund returns 2x the index while the other returns –2x, how is it possible that both can lose money within one year? To be fair, this is not due to the bad management of the fund manager as the performance is not due to tracking error.

We show two more examples of extreme performances

	Naïve return expectation	Actual return	Difference
Russell 2000 Index (RTY)	–15.8%	–15.8%	
Russel 2000 Ultra Long 2x (UWM)	–15.8% x 2 = –31.6%	–52.7%	–21.1%
Russell 2000 Short (RWM)	+15.8%	–16.5%	–32.3%
Russell 2000 Ultra Short 2x (TWM)	+15.8% x 2 = +31.6%	–51.3%	–82.9%

During the period between January 2008 and December 2009, the Russell 2000 index dropped 15.8%. The fund which is supposed to return 2x the performance dropped 52.7% in this period. Even the investors who hold the two funds which return –1x and –2x lost substantially.

[6] Source: Avellaneda and Zhang (2009).

Date	FXI	FXI return	FXP	FXP return
10/27/2008	19.29		183.01	
10/28/2008	23.15	20.0%	118.70	-35.1%
10/29/2008	22.34	-3.50%	115.48	-2.71%
10/30/2008	25.48	14.06%	87.03	-24.64%
10/31/2008	24.99	-1.92%	89.00	2.26%
11/3/2008	25.28	1.16%	86.00	-3.37%
11/4/2008	27.02	6.88%	75.20	-12.56%
11/5/2008	24.48	-9.40%	87.00	15.69%
11/6/2008	22.54	-7.92%	101.99	17.23%
11/7/2008	25.43	12.82%	76.32	-25.17%
11/10/2008	26.43	3.93%	69.50	-8.94%
11/11/2008	24.88	-5.86%	76.66	10.30%
11/12/2008	23.92	-3.86%	82.58	7.72%
11/13/2008	27.47	14.84%	60.40	-26.86%
11/14/2008	24.98	-9.06%	69.68	15.36%
11/17/2008	24.93	-0.20%	69.96	0.40%
11/18/2008	24.11	-3.29%	74.61	6.65%
11/19/2008	22.13	-8.21%	86.23	15.57%
11/20/2008	21.02	-5.02%	93.00	7.85%

In this example, FXI tracks the FTSE/Xinhua China 25 Index, FXP is –2x this index. While we can see that the daily movements do not match the claimed performance exactly, the difference after one month is surprising. While FXI total return is about 9% throughout this period (from 19.29 to 21.02), FXP's return is –49.2% (from 183.01 to 93.00).
(Source: Euan Sinclair, Volatility Trading, 2nd ed. (2013), p.232)

9.4.3. Daily hedging demand

It may be against intuition that the hedging that needs to be carried out would be in the same direction no matter whether it is leveraged or inverse leveraged. This is analogous to the case for options, where both in the case of short call or short put, the position is short gamma.

Assume that the NAV of a fund at time t_n is $A(t_n)$ with leverage x (which means that original exposure $L(t_n)$ is $xA(t_n)$), and the return for the period between t_n and t_{n+1} is r_n. Before re-hedging (at time t_{n+1}), the exposure is:

$$L_{BR}(t_{n+1}) = L(t_n)\,[1 + r_n] = xA(t_n)[1 + r_n]$$

whereas the NAV of the fund is

$$A(t_{n+1}) = A(t_n)[1 + xr_n]$$

which would require a new exposure of

$$L(t_{n+1}) = xA(t_{n+1}) = A(t_n)[1 + xr_n]$$

Therefore the re-hedging that needs to be carried out

$$\begin{aligned}\Delta(t_n) &= L(t_{n+1}) - L_{BR}(t_{n+1}) \\ &= xA(t_n)[1 + xr_n] - xA(t_n)[1 + r_n] = (x^2 - x)A(t_n)r_n\end{aligned}$$

Note that, in this expression, the term $(x^2 - x)$ is always positive for any integer x if x is not equal to 1. Therefore, no matter whether the fund has positive leverage or inverse leverage, the hedging action is the same, i.e. increase exposure on a day when the index goes up, and decrease exposure on a day when the index goes down.

Case 1	$S(t_n)$	$A(t_n)$	$L(t_n)$	$L_{BR}(t_n)$	$\Delta(t_n)$
n=0	100	100	200		
n=1	90	80	160	180	–20
n=2	99	96	192	176	+16
Case 2					
n=0	100	100	-100		
n=1	90	110	-110	-90	–20
n=2	99	99	-99	-121	+22
Case 3					
n=0	100	100	-200		
n=1	90	120	-240	-180	–60
n=2	99	96	-192	-264	+72

The hedging action is shown in the examples above. The parameters are:

- Case 1: $x = 2$; Case 2: $x = -1$; Case 3: $x = -2$
- index returns on day 1 and 2 are –10% and +10%

We can see that the directions of the trades are identical in every case.

9.4.4. Special features of the Leveraged ETF

We would explain why theoretically the ETF's performance can be so different from the long term index performance. Assume we have an ETF that pays the return of the index with an initial level of 100. Further assume that the return of the index is down 10% on day 1, and up 10% on day 2. Thus the index levels on days 0 to 2 are: 100, 90, 99. We can immediately see that with two equal % moves, the ETF loses 1% over 2 days. In fact, if there is an ETF that has a leverage of +2, it would have lost 4% (100, 80, 96). This is simply because $(1-x)(1+x) = 1-x^2 < 1$ where x is the daily return. Therefore if the underlying index is volatile, an ETF tracking its daily percentage movement would show a drag in the performance even though the final index level may not move much.

Another characteristic of the leveraged ETF is that its performance is path dependent and is not related to the total return in the monitoring period. A simple two period example is shown below:

Path 1	$S(t_n)$	$A(t_n)$	$L(t_n)$	$L_{BR}(t_n)$	$\Delta(t_n)$
n=0	100	100	200		
n=1	90	80	160	180	–20
n=2	99	96	192	176	+16
Path 2					
n=0	100	100	200		
n=1	100	100	200	200	0
n=2	99	98	196	198	–2

In this example, the leverage of the fund is $x = 2$. In both these paths, the index level ends at 99, whereas the NAVs of the fund are different (96 and 98).

If the index follows a geometric Brownian motion, it can be shown that over a period of time t_n, the return of the fund's NAV is given by:[7]

$$\frac{A(t_n)}{A(t_0)} = C\left[\frac{S(t_n)}{S(t_0)}\right]^x, \; C = \exp\left[\frac{(x-x^2)\sigma^2 t_n}{2}\right]$$

Note that the scaling factor C is less than one as $(x-x^2) < 0$, but will never be negative, i.e. the fund's NAV will not drop below 0.

The above formula provides some clues for the long term erosion of return of the leveraged ETF. The leveraged ETF will outperform the static ETF only when volatility is low and the underlying asset (or index) moves in a constant trend. For example, when the index goes up, the re-balancing of the ETF effectively means that the exposure would be rising in a rising market, hence it will outperform the ETF which is not re-balanced. This is seen in the term $[S(t_n)/S(t_0)]^x$ which tends to dominate (in the short run), and this is a convex relationship.

However, as shown in the examples above, the leveraged ETF cannot track index returns for periods longer than one day, and the leveraged ETF also certainly underperforms the index / asset in the long run. Even ignoring tracking errors and transaction costs, the return depends on the index volatility and the time horizon. For example, if $x = -2$, $\sigma = 30\%$, $t_n = 5$, index return is -10%, $C = 0.259$. The fund value is only 32% of its original value even though the index has only drop 10%.

9.4.5. Options of leveraged ETFs

Given the popularity of leveraged ETFs, options on certain leveraged ETFs are also available in the market. For an option based on a leveraged ETF on the S&P 500 index, clearly the implied volatility should somehow be linked to the volatility of a non-leveraged ETF on the same index. However, empirically the connection is not very obvious, as the market does not seem to be efficient. This is an area which is still undergoing development. More information can be found in Ahn, Haugh and Jain (2014) and Leung and Sircar (2014).

Additional references

Ahn, Andrew, Martin Haugh, and Ashish Jain (2014). "Consistent Pricing of Options on Leveraged ETFs." Working paper.

Avellaneda, Marco and Stanley Zhang (2009). "Path-dependence of Leveraged ETF returns." Working paper.

Cheng, Minder and Anath Madhavan (2009). "The Dynamics of Leveraged and Inverse Exchange-Traded Funds." Working paper.

Leung, Tim and Ronnie Sircar (2014). "Implied Volatility of Leveraged ETF Options." Working paper.

Sinclair, Euan. *Volatility Trading*, 2nd edition, Wiley (2013), chapter 8.

Tsay, Ruey. *Analysis of Financial Time Series*, 3rd edition, Wiley (2010).

Vidyamurthy, Ganapathy. *Pairs Trading: Quantitative Methods and Analysis*, Wiley Finance (2004).

Wilmott, Paul. *Paul Wilmott Introduces Quantitative Finance*, 2nd Edition, Wiley (2007), 1:306-309.

[7] This result is given in both Cheng and Madhavan (2009) and Avellaneda and Zhang (2009).

Appendix: A brief introduction to financial time series analysis

A.1. What is a financial time series?

When we make observations of financial variables across different time periods, we can obtain a data series. The observations can be made with different frequency. For example, we can monitor the Hang Seng Index everyday and record the closing level. However, for economic data such as inflation, we may only have one data point every quarter or every year.

The simplest modeling case is univariate time series analysis, which refers to the modeling of a single sequence of data. There is an element of uncertainty involved in financial time series analysis. Most of the time we can only provide estimates and give a statistical description of its underlying behavior. This is achieved by the use of summary statistics (such as mean/variance) and/or graphical methods (such as a time plot). We can also try to identify a suitable statistical model to describe the data generating process, which can then be applied to forecasting, i.e. provide some kind of estimate of the future values of the series.

A.2. Fundamental concepts

Asset returns

Most financial analysis involves returns instead of prices because returns have more attractive statistical properties. Let P_t be the price of an asset at time t. One period simple return R_t is given by :$P_t = P_{t-1}(1+ R_t)$ or $R_t = (P_t - P_{t-1})/ P_{t-1}$. Multi-period simple return: holding an asset for k periods between times $t-k$ and t gives:

$$1+R_t[k] = (1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1}) = \prod_{j=0}^{k-1}(1+R_{t-j})$$

Annualized return is the geometric mean:

$$R_{at}[k] = \left[\prod_{j=0}^{k-1}(1+R_{t-j})\right]^{1/k} - 1 \approx \frac{1}{k}\sum_{j=0}^{k-1}R_{t-j} \text{ for small } R$$

Commonly used summary statistics

Let $\{r_1, r_2, \ldots, r_T\}$ be a set of T data points. We have the following definitions:

- Mean (or expectation) of returns $$\mu_r = E(r_t) = \frac{1}{T}\sum_{t=1}^{T} r_t$$
- Variance of returns $$Var(r_t) = E[(r_t - \mu_r)^2] = \frac{1}{T}\sum_{t=1}^{T}(r_t - \mu_r)^2$$
- Sample variance (unbiased estimate of the population variance) $$\sigma_r^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \mu_r)^2$$

Correlation functions

We want to answer the question: is there a relationship between the value of a variable now and the value observed *k* time steps in the past? A useful benchmark is to look for co-variance or correlation. We can make use of the following parameters:

- Lag-k auto-covariance

$$\gamma_k = \mathrm{cov}(r_t, r_{t-k}) = E[(r_t - \mu_r)(r_{t-k} - \mu_r)]$$

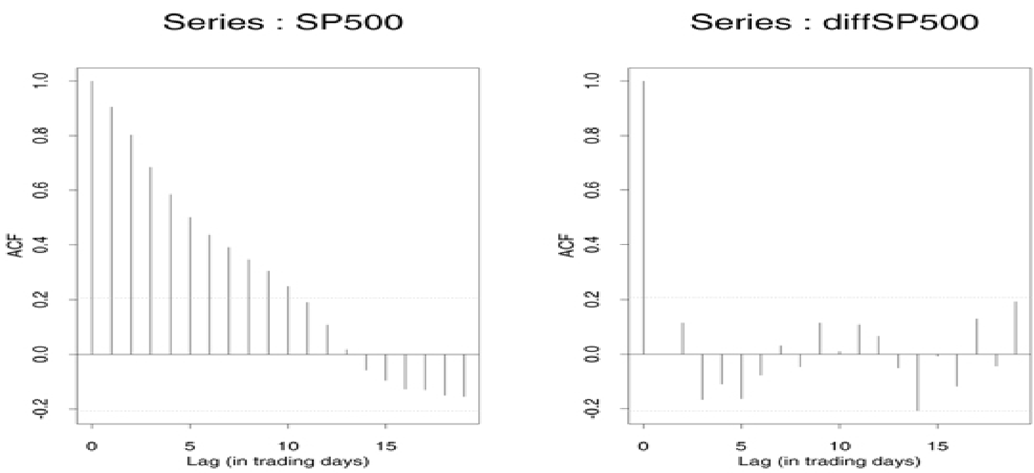
- Serial (or auto-) correlation

$$\rho_k = \frac{\mathrm{cov}(r_t, r_{t-k})}{\mathrm{var}(r_t)} = \frac{\gamma_k}{\gamma_0}$$

- Sample autocorrelation function (ACF)

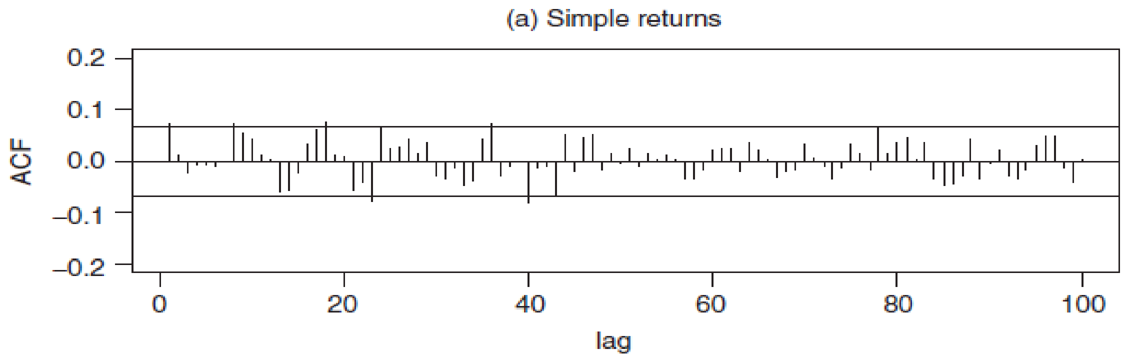
$$\rho_k = \frac{\sum_{t=1}^{T-k} (r_t - \mu_r)(r_{t+k} - \mu_r)}{\sum_{t=1}^{T} (r_t - \mu_r)^2}$$

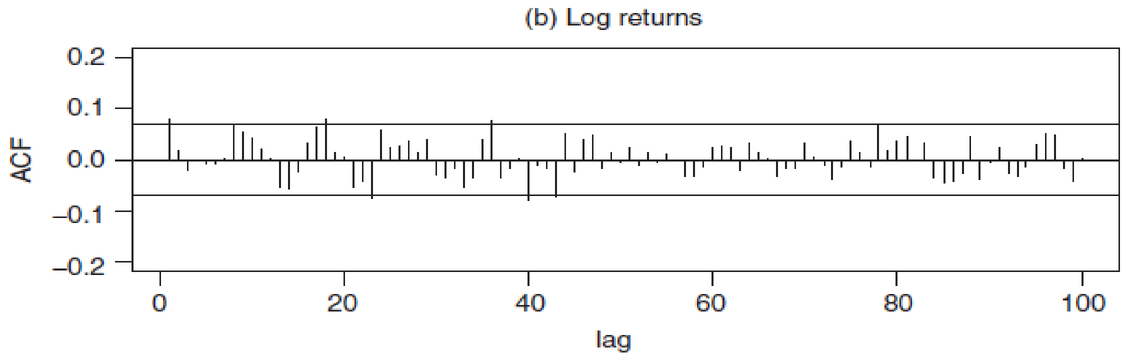
The plot of the ACF against time lag k is called the *correlogram*.



The above example comes from Chatfield (2000). The left diagram shows the ACF at varying time lag *k* using 90 trading days of data from S&P 500 index starting from March 16, 1999. It is seen that when the time lag increases, the autocorrelation decreases almost linearly. However, this may simply be caused by the overlapping data. If the same series is pre-processed by a technique known as differencing, the result is very different, as shown in the right diagram. There is no noticeable auto-correlation even with a short time lag. The differencing technique will be discussed in section A.3.4.

Another example shows the sample ACF of monthly simple returns and log returns of IBM stock from Jan 1926 to Dec 1997 (from Tsay (2005), p.28). Again no noticeable correlation is found at different time lags.





Statistical distribution of returns

Mean and variance are known as the first and second moments of a distribution. We can also define the third and fourth moments as follows. Skewness (third moment) can be used to measure the asymmetry of the distribution:

$$S_r = \frac{1}{(T-1)\sigma_r^3}\sum_{t=1}^{T}(r_t - \mu_r)^3$$

Kurtosis (fourth moment) is used to measure the tail behavior:

$$K_r = \frac{1}{(T-1)\sigma_r^4}\sum_{t=1}^{T}(r_t - \mu_r)^4$$

In the above, the sample statistics are used to estimate the population statistics (i.e. the average is being calculated with a denominator of T–1 instead of T).

Tests for normal distribution

If the number of samples is large, we can test whether the series follows a normal distribution via the following:

Test for symmetry: $S_r \sim N(0, \frac{6}{T})$, $N()$ is a normal distribution

Test for tail thickness: $K_r - 3 \sim N(0, \frac{24}{T})$, $N()$ is a normal distribution

Jaruq-Bera test (a joint test): $JB = K_r^2 + S_r^2 \sim \chi_2^2$

where χ_2^2 denotes a chi-squared distribution with two degrees of freedom. The null hypothesis is rejected if the test statistics falls outside of the range predicted by these distributions.

Stationary time series

A useful concept that is often applied in financial time series analysis is stationarity. A time series $\{r_t\}$ is said to be strictly stationary if the joint distribution of $(r_{t1}, r_{t2}, \ldots, r_{tk})$ is identical to that of $(r_{t1+t}, r_{t2+t}, \ldots, r_{tk+t})$ for all t, where k is an arbitrary positive integer and $(t_1, t_2, \ldots, t_k)$ is a collection of k positive integers. In order words, the distribution is invariant under time shift. This condition is seldom satisfied in practice. A less stringent condition, known as *weakly stationary* time series, has more applications. A time series $\{r_t\}$ is weakly stationary if:

(1) $E(r_t) = \mu$, which is a constant, when different segments of data are taken;
(2) $\mathrm{Cov}(r_t, r_{t-k}) = \gamma_k$, which only depends on k;

i.e. only the first two moments (mean and covariance) of the series are time-invariant. Given that a normal distribution can be defined with mean and variance alone, it is evident that if the returns follow a normal distribution, a weakly stationary time series is also strictly stationary. Furthermore, we note that: (1) $\gamma_0 = \mathrm{Var}(r_t)$; and (2) $\gamma_{-k} = \gamma_k$.

Most time series analysis are based on weakly stationary series, and this is the basis for using models to make meaningful inferences in the future.

White noise process

In a white noise process, the sequence of data r_t consists of independent and identically distributed random variables with finite mean and variance. If r_t is normally distributed with mean 0 and variance σ^2, it is called a Gaussian white noise. In this series, all ACFs are zero (because all variables are independent with each other by construction). Furthermore, a Gaussian white noise process is stationary because its mean and variance are constant.

Linear time series models

A time series r_t is linear if $r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$

where μ is the mean of r_t, $\{a_t\}$ is a white noise process.
If r_t is weakly stationary,

$$E(r_t) = \mu; \quad \mathrm{var}(r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2$$

Lag-m auto-covariance of r_t is

$$\gamma_m = \mathrm{cov}(r_t, r_{t-m}) = E\left[\left(\sum_{i=0}^{\infty} \psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j a_{t-m-j}\right)\right] = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+m}$$

A.3. Some useful models

The basic stochastic processes that are applicable in modeling financial time series include:
- Moving average (MA) process
- Autoregressive (AR) process
- Autoregressive moving-average (ARMA) process
- Autoregressive integrated moving-average (ARIMA) process

While these processes may not be too useful for derivative trading, more sophisticated volatility models (e.g. the ARCH family) are built from these processes.

A.3.1. AR process

One of the most basic schemes is the AR process, where the regression is constructed with lagged variables. An example is the growth rate of quarterly US real GNP from 1947 to 1991:

$$r_t = 0.005 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \hat{\sigma}_a = 0.01$$

In this formulation, the growth rate r_t depends on the growth rates of the past three quarter, and a_t is a random error term. This is quite similar to a factor model, only that the "factors" come from past values of r_t rather than on separate predictor variables.

AR(1) model

The simplest formulation of the AR process is sometimes known as the Markov process:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t$$

a_t is from a random process with mean 0 and variance σ_a^2, and r_t depends on only one past variable. Taking the expectation and under stationarity conditions, we have $E(r_t) = E(r_{t-1}) = \mu$, hence

$$\mu = \phi_0 + \phi_1\mu \quad \text{or } E(r_t) = \mu = \phi_0 / (1-\phi_1)$$

Writing $\phi_0 = (1 - \phi_1)\mu$, the model can be re-written as

$$x_t = r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t \quad \text{or} \quad x_t = \phi_1 x_{t-1} + a_t$$

Repeated substitutions then give

$$x_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots = \sum_{i=0}^{\infty} \phi_1^i a_{t-i}$$

Since the above summation involves terms with powers of ϕ_1, the series would only converge when $-1 < \phi_1 < 1$. This is also a necessary and sufficient condition for which the AR(1) model is to be weakly stationary.

We note that since $E[(r_{t-1} - \mu)\, a_t] = 0$ (given that r_{t-1} is a function of a_{t-i} for i >=0 and the a's are independent), we can deduce that

$$\mathrm{var}(r_t) = \phi_1^2\ \mathrm{var}(r_{t-1}) + \sigma_a^2$$

Under the stationary assumption, $\mathrm{var}(r_t) = \mathrm{var}(r_{t-1})$, therefore

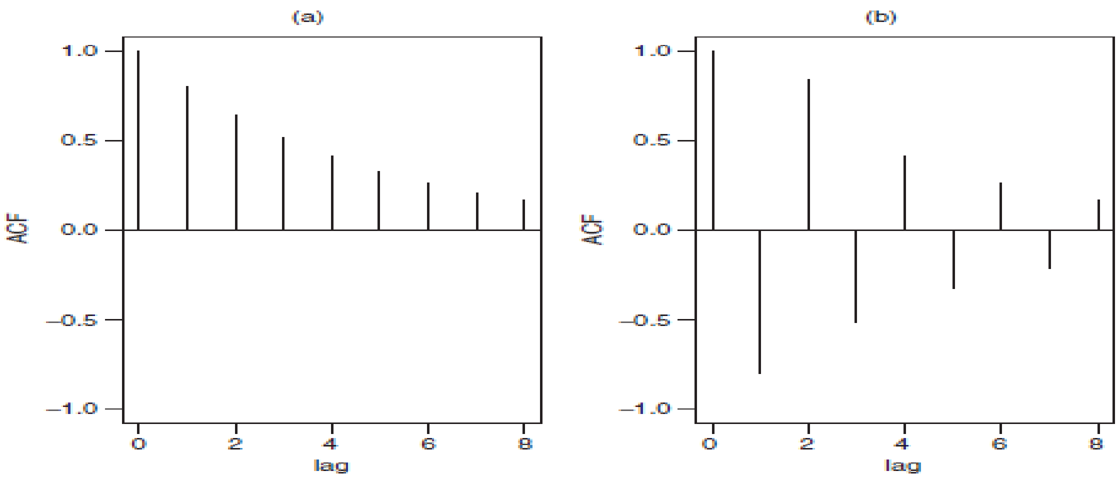
$$\mathrm{var}(r_t) = \frac{\sigma_a^2}{1-\phi_1^2}$$

We can also derive the correlation and ACF of the AR(1) model, assuming weak stationarity:

$$\begin{aligned} r_t - \mu &= \phi_1(r_{t-1} - \mu) + a_t \\ E[a_t(r_t - \mu)] &= \phi_1 E[a_t(r_{t-1} - \mu)] + E[a_t^2] = \sigma_a^2 \\ \gamma_m &= E[(r_{t-m} - \mu)(r_t - \mu)] \\ &= \phi_1 E[(r_{t-m} - \mu)(r_{t-1} - \mu)] + E[a_t(r_{t-m} - \mu)] \\ &= \begin{cases} \phi_1\gamma_1 + \sigma_a^2 & \text{if } m = 0 \\ \phi_1\gamma_{m-1} & \text{if } m > 0 \end{cases} \end{aligned}$$

The ACF of r_t satisfies $\rho_m = \phi_1\rho_{m-1}$ for $m \geq 0$ or $\rho_m = \phi_1^m$, which means that the ACF of an AR(1) series decays exponentially with rate ϕ_1 starting at ρ_0=1.

Examples of autocorrelation functions of the AR(1) model are show below.



The respective constants in the models are: (a) $\phi_1 = 0.8$; (b) $\phi_1 = -0.8$. Note that the ACF drops off gradually, as there will always be some correlation between the values since they come from linear combinations of past realizations from a white noise series which are common to all samples.

Properties of an AR(2) model

An AR(2) model has the form $r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t$
Using the same argument as in the AR(1) case, the expected value is:

$$E(r_t) = \mu = \frac{\phi_0}{1-\phi_1-\phi_2}$$

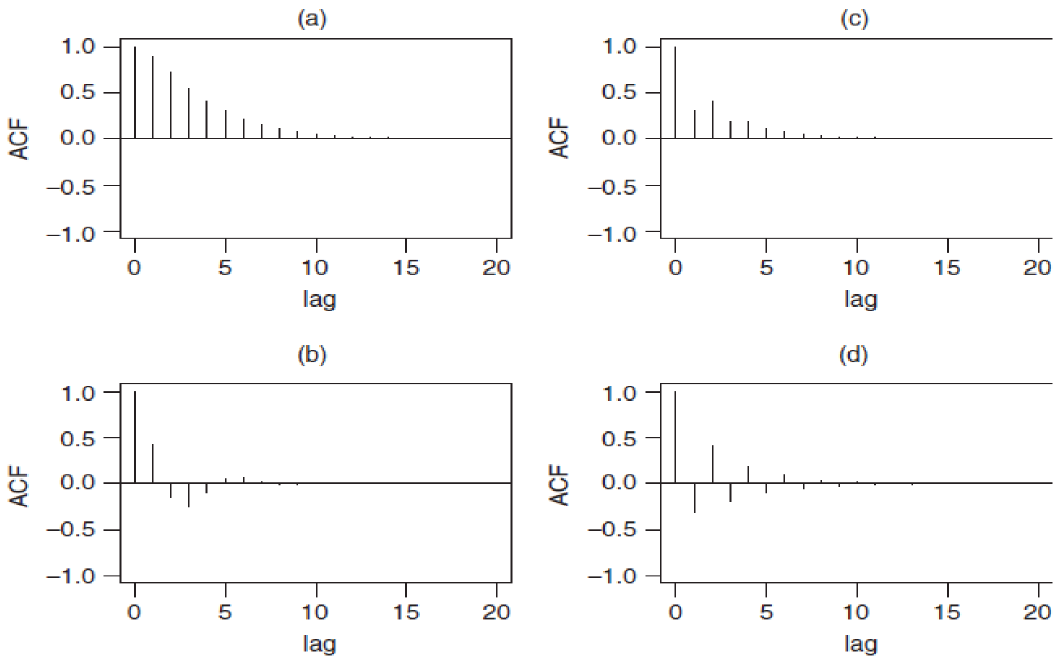
By re-writing the model as: $r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t$
We can multiply this equation by ($r_{t-m} - \mu$), take expectation, and using the result $E[(r_{t-m} - \mu)a_t] = 0$ for $m>0$, we have

$$\gamma_m = \phi_1 \rho_{m-1} + \phi_2 \rho_{m-2} \quad \text{for } m > 0$$

The ACF functions are thus: $\rho_0 = 1, \ \rho_1 = \phi_1 / (1 - \phi_2)$

$$\rho_m = \phi_1 \rho_{m-1} + \phi_2 \rho_{m-2} \quad \text{for } m \geq 2$$

Some example autocorrelation functions of AR(2) model are shown below:



(a): $\phi_1 = 1.2$ and $\phi_2 = -0.35$; (b): $\phi_1 = 0.6$ and $\phi_2 = -0.4$;
(c): $\phi_1 = 0.2$ and $\phi_2 = 0.35$; (d): $\phi_1 = -0.2$ and $\phi_2 = 0.35$.

A.3.2. MA process

Another class of models that are useful in modeling return series in finance is the moving-average (MA) models. An MA(1) model can be seen as an infinite order AR model, but with some constraints in the parameters

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t$$

Alternatively, it can be rewritten as

$$r_t = \mu + a_t - \theta_1 a_{t-1}$$

which is the general form of an MA(1) model.
An MA(q) model has the form

$$r_t = \mu + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Since MA models are finite linear combinations of a white noise sequence where the first two moments are time-invariant, <u>these models are always weakly stationary</u>.
For the MA(1) model, with the properties of a_t, it is obvious that expected value: $E(r_t) = \mu$.
For the variance, given that a_t and a_{t-1} are uncorrelated by definition,

$$\operatorname{var}(r_t) = \sigma_a^2 + \theta_1^2 \sigma_a^2 = (1 + \theta_1^2)\sigma_a^2$$

Similarly, the variance of an MA(q) model is

$$\mathrm{var}(r_t) = (1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)\sigma_a^2$$

Autocorrelation is a property which can be used to identify the order of an MA model. Assume that $\mu = 0$, and multiply the model by r_{t-m}, we get:

$$r_{t-m}r_t = r_{t-m}a_t - \theta_1 r_{t-m}a_{t-1}$$

Take expectation, $\gamma_1 = -\theta_1\sigma_a^2, \gamma_m = 0 \text{ for } m > 1$

Therefore

$$\rho_0 = 1, \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\theta_1}{1+\theta_1^2}, \rho_m = 0 \text{ for } m > 1$$

Thus for an MA(1) model, only the lag-*1* ACF is not zero but all higher order ACFs are zero. In practice, if we suspect that the time reference can be represented as an MA process and when we plot a series of ACFs with different lags and find that only lag-1 ACF is not zero, it is very likely that the series can be characterized by an MA model, and we can assume that it is an MA(1) model (but note the discussion in section A.4.2. below). For an MA(q) series that is only linearly related to its first q lagged values, it actually has a "finite memory."

We give a simple demonstration from the MA(1) model:

$$\begin{aligned} r_t &= \mu + a_t - \theta_1 a_{t-1} \\ r_{t+1} &= \mu + a_{t+1} - \theta_1 a_t \\ r_{t+2} &= \mu + a_{t+2} - \theta_1 a_{t+1} \end{aligned}$$

Terms that are one time interval apart have one common white noise realization, thus there would be some correlation between them. For example, between r_t and r_{t+1}, a_t is the common white noise term. However, r_t and r_{t+2} have no common white noise realization, and given that a_i's are independent, theoretically the correlation between r_t and r_{t+2} is 0.

A.3.3. ARMA process

A financial series may require a high number of parameters in its specification. An advantage of the ARMA model is that it can combine the ideas of AR and MA models into a compact form, with a small number of parameters. Although this is rarely used in forecasting returns, the form is relevant to some of the more popular volatility models such as GARCH.

Properties of ARMA(1,1) model

Standard equation for ARMA(1,1) model

$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1}$$

Left hand side is the AR component and the right hand side gives the MA component. (1,1) refers to the fact that both r and a depend on only one lagged variable. The expectation in the ARMA(1,1) model is: $E(r_t) = \mu = \phi_0/(1-\phi_1)$ which is the same as the AR(1) model.

Assume ϕ_0=0, variance of r_t is given by:

$$\text{var}(r_t) = \frac{(1-2\phi_1\theta_1 + \theta_1^2)\sigma_a^2}{1-\phi_1^2}$$

The ACF of the ARMA(1,1) model is very similar to the AR(1) case, except for lag = 1. Define $\gamma_m = \text{covar}(r_t, r_{t-m})$ where m is the lag, we have

$$\gamma_1 - \phi_1\gamma_0 = -\theta_1\sigma_a^2$$

Note that for AR(1), $\gamma_1 - \phi_1\gamma_0 = 0$

For $m>1$, $\gamma_m - \phi_1\gamma_{m-1} = 0$

In terms of the ACF,

$$\rho_1 = \phi_1 - \frac{\theta_1\sigma_a^2}{\gamma_0},\ \rho_m = \phi_1\rho_{m-1}, \text{ for } m>1$$

Thus the exponential decay of the ACF of an ARMA(1,1) model starts with lag 2, and it does not cut off at any finite lag.

Representations of ARMA(1,1) model

The model can be expressed in three equivalent representations.

- *ARMA form*: the most compact, and is useful in estimation and forecasting
$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1}$$
- *AR form*: Tells how r_t depends on its past values
$$r_t = \phi_0 + a_t + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \cdots$$
- *MA form*: Tells how r_t depends on the past shocks
$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

A.3.4. Achieving stationarity

Many financial time series are non-stationary. However, Box and Jenkins have suggested a procedure to convert a non-stationary series to a stationary series, via a simple method called regular differencing. For an original series $\{x_1,\ldots, x_N\}$, form a new series $\{y_1,\ldots, y_N\}$ via the following procedure:

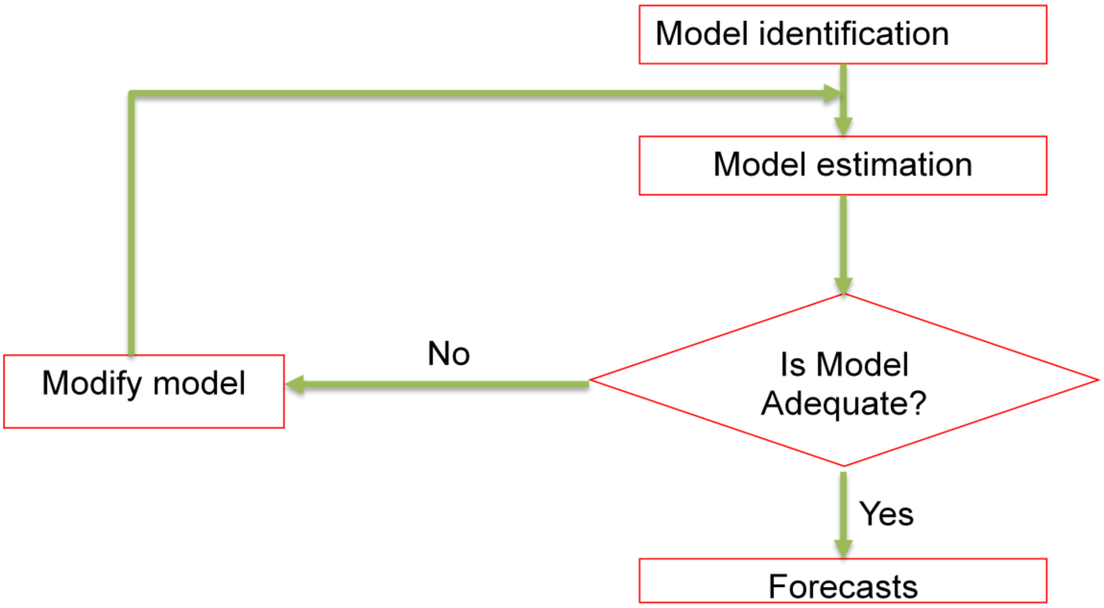
- 1st order differencing: $y_t = \Delta x_t = x_t - x_{t-1}$
- 2nd order differencing: $y^2{}_t = \Delta^2 x_t = x_t - 2x_{t-1} + x_{t-2}$

A model of the form $y_t^d = \phi_0 + \sum_{i=1}^{p}\phi_i y_{t-i}^d + a_t - \sum_{i=1}^{q}\theta_i a_{t-i}$

is known as an Integrated ARMA (or ARIMA) model of order (p,d,q); usually d is 1 (representing first order differencing).

A.4. Box-Jenkins approach to forecasting

A.4.1. Systematic process



From Box and Jenkins (1970).
The forecasting procedure can be divided into four steps:

- Model identification

Determine the order of a model, e.g. should we use AR(2) or AR(3)? This question can be answered via the help of some tools such as the calculation of autocorrelations and/or partial-autocorrelations (to be discussed below).

- Model estimation

Once a model is chosen, the objective is to minimize the sum of squares of errors in order to estimate the model parameters.

- Model validation

Certain diagnostics are used to check the validity of the model.

- Model forecasting

The estimated model is used to generate forecasts and confidence limits of the forecasts can also be provided.

A.4.2. Identifying AR models in practice

If it is suspected that the AR process should be the model of choice, we still need to decide the order or the process, i.e. how do we know whether an AR(2) or AR(3) model fits the data better? The shape of the ACF does not help much because it can be a mixture of exponential decay functions and it is not easy to isolate the order.

Consider the AR models in consecutive orders:

$$\begin{aligned}
&\mathrm{AR}(1): \ r_t = \phi_{0,1} + \phi_{1,1} r_{t-1} + a_{1t} \\
&\mathrm{AR}(2): \ r_t = \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + a_{2t} \\
&\mathrm{AR}(3): \ r_t = \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + a_{3t} \\
&\qquad\qquad \vdots \qquad\qquad\quad \vdots
\end{aligned}$$

The coefficients in these models can be estimated by the least squares method. By definition, the estimate of $\phi_{1,1}$ is called the lag-*1* sample *Partial Autocorrelation* (PACF) of r_t, estimate of $\phi_{2,2}$ is called the lag-*2* sample PACF, and so on. In terms of interpretation, the lag-*2* sample PACF shows the added contribution of r_{t-2} to r_t over the AR(1) model. Note that PACF is very different in character from ACF.

For an AR(p) model, the lag-p sample PACF should not be zero, but the estimate of $\phi_{j,j}$ should be close to zero for all $j>p$. In practice the PACF will not drop to zero, but we can use a statistical measure (e.g. at 5% significance level) to decide which orders can be rejected.

We can summarize the behavior of ACF and PACF for different types of models below:

Model	**ACF**	**PACF**
AR(p)	Tails off gradually	Cuts off after p lag
MA(q)	Cuts off after q lags	Tails off gradually
ARMA(p,q)	Tails off gradually	Tails off gradually

Table 2.1. Sample Partial Autocorrelation Function and Akaike Information Criterion for the Monthly Simple Returns of CRSP Value-Weighted Index from January 1926 to December 1997

p	1	2	3	4	5
PACF	0.11	−0.02	−0.12	0.04	0.07
AIC	−5.807	−5.805	−5.817	−5.816	−5.819
p	6	7	8	9	10
PACF	−0.06	0.02	0.06	0.06	−0.01
AIC	−5.821	−5.819	−5.820	−5.821	−5.818

In this example, using the PACF information, arguably an AR(3) or AR(5) model can be appropriate (taken from: Tsay (2005), p.41), as the PACF is getting smaller after order 4 or order 6. (But note that there is no absolute “cut-off” level which can be used to determine exactly what order is correct.)

Another tool that is sometimes useful to determine the order of a model is the *information criterion*. A model with a lot of parameters will always provide a better fit to the data, but additional parameters may not be optimal for forecasting. To decide whether adding more variables makes a better model, we can calculate the Akaike Information Criterion (AIC):

$$AIC(p) = \ln(\sigma_p^2) + 2p/T$$

where σ_p^2 is the estimated variance of the samples of an AR(p) model, T is the number of samples, and the optimum AR order p is identified as the one with minimum AIC.

Alternatively one can calculate the Bayesian Information Criterion (BIC):

$$BIC(p) = \ln(\sigma_p^2) + p\ln(T)/T$$

As shown in the example above, calculating the AIC may not always help much as the values obtained for different orders can actually be very close to each other.

Parameter estimation and model checking

Suppose we have an AR process of order p, with mean μ, given by

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \cdots + \phi_p(r_{t-p} - \mu) + a_t,\ t = p+1, \cdots, T$$

The parameters μ, ϕ_1, …, ϕ_p may be estimated by least squares by minimizing

$$S = \sum_{t=p+1}^{N} \left[r_t - \mu - \phi_1(r_{t-1} - \mu) - \cdots - \phi_p(r_{t-p} - \mu) \right]^2$$

with respect to μ, ϕ_1, …, ϕ_p. The fitted model is

$$\hat{r}_t - \mu = \hat{\phi}_1(r_{t-1} - \mu) + \cdots + \hat{\phi}_p(r_{t-p} - \mu)$$

and the associated residual is $\hat{a}_t = r_t - \hat{r}_t$. If the model is adequate, this residual series should behave as a white noise, and tests such as the Ljung-Box statistics of the residuals can be used to check its validity.

A.4.3. Forecasts in the AR(1) model

We illustrate the forecasting process with the model $x_t = \phi_1 x_{t-1} + a_t$.
At time n, the 1-step ahead forecast for the next x is: $\hat{x}_n(1) = \phi_1 x_n$
1-step ahead forecast error: $e_n(1) = x_{n+1} - \hat{x}_n(1) = a_{n+1}$
where a_{n+1} is the random shock at time $n+1$
Variance of 1-step ahead forecast error: $\mathrm{var}[e_n(1)] = \mathrm{var}(a_{n+1}) = \sigma_a^2$
Similarly, 2-step ahead forecast: $\hat{x}_n(2) = \phi_1 \hat{x}_n(1) = \phi_1^2 x_n$
2-step ahead forecast error: $e_n(2) = x_{n+2} - \hat{x}_n(2) = a_{n+2} + \phi_1 a_{n+1}$
Variance of 2-step ahead forecast error: $\mathrm{var}[e_n(2)] = (1 + \phi_1^2)\sigma_a^2$

In general, for the m-step ahead forecast at time n, we have $\hat{x}_n(m) = \phi_1^m x_n$

m-step ahead forecast error: $e_n(m) = a_{n+m} + \phi_1 a_{n+m-1} + \cdots + \phi_1^{m-1} a_{n+1}$

Variance of forecast error $\mathrm{var}[e_n(m)] = (1 + \phi_1^2 + \cdots + \phi_1^{2(m-1)})\sigma_a^2$

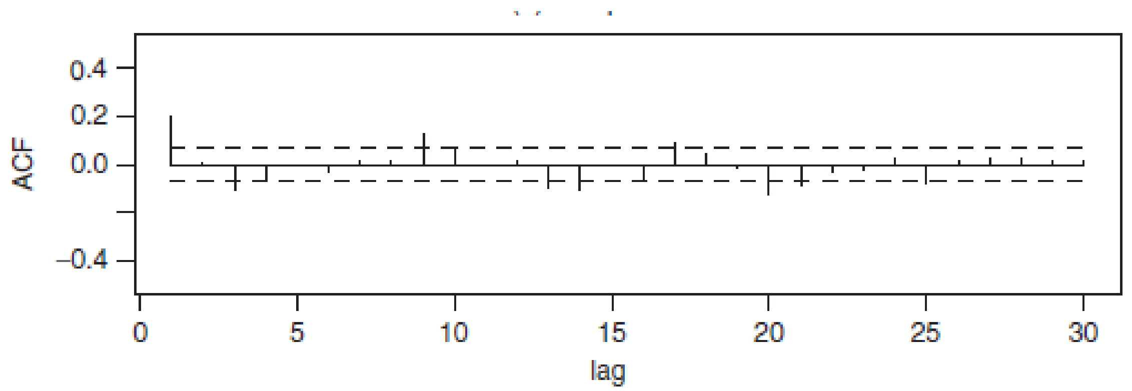
As $m \to \infty, \hat{x}_n(m) \to 0, i.e.\ \hat{r}_n(m) \to \mu$. Therefore we can say that the AR(1) is a mean-reverting process. Variance of forecast error approaches

$$\mathrm{var}[e_n(m)] = \frac{\sigma_a^2}{1 - \phi_1^2} = \mathrm{var}(r_t)$$

A.4.4. Estimation and forecasting with an MA process

Determining the order of an MA process

The main tool that can be used is the ACF. From an earlier result, we notice that the ACF at lag x of a times series r_t following an MA(q) model is 0 for $x>q$.



The above illustration is taken from Tsay (2005), pp.52-53. The ACF is plotted against different time lags. In this example, the two dashed lines denote the 2 standard-error limits. A plausible model is MA(9), as ACFs at orders 10 or above mostly lie within the dash lines (although ACFs at higher lags can be significant, such as order 20).

Estimating the parameters in an MA model

Estimation is more difficult than an AR process because efficient explicit estimators cannot be found, and a non-linear optimization method is required. For example, in an MA(1) model of the form $r_t = \mu + a_t - \theta_1 a_{t-1}$
The ACF is given by: $\rho_1 = \hat{\theta}_1 / (1 + \hat{\theta}_1^2)$
Starting with estimates of μ and θ_1 and take $a_0 = 0$, we can calculate $a_1 = r_1 - \mu$, and then $a_2 = r_2 - \mu + \theta_1 a_1$ and so on until $a_N = r_N - \mu + \theta_1 a_{N-1}$.

We can calculate the residual sum of squares $S = \sum_{t=1}^{N} a_t^2$, and then adjust μ and θ_1 until S is minimized.

Forecasts using MA(1) models

The point forecasts go to the mean of the series quickly. Starting at time *h*, the 1-step ahead forecast of an MA(1) model is: $r_{h+1} = \mu + a_{h+1} - \theta_1 a_h$

Taking expectation: $\hat{r}_h(1) = E(r_{h+1}) = \mu - \theta_1 a_h$
The error at time 1 is: $e_h(1) = r_{h+1} - \hat{r}_h(1)$
The variance of the 1-step ahead forecast error = σ_a^2
For the 2-step ahead forecast,

$$r_{h+2} = \mu + a_{h+2} - \theta_1 a_{h+1}$$
$$\hat{r}_h(2) = E(r_{h+2}) = \mu \quad \text{[at time } h,\ E(a_{h+1}) = E(a_{h+2}) = 0]$$
$$e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}$$

Variance of the forecast error = $(1 + \theta_1^2)\sigma_a^2$
More generally, the mean of the *p*-step ahead forecast

$$\hat{r}_h(p) = \mu,\ p \geq 2$$

,i.e. mean-reverting is already achieved after 1 time period.

Forecasts using MA(q) models

A similar derivation can be obtained for MA(2) model

$$r_{h+p} = \mu + a_{h+p} - \theta_1 a_{h+p-1} - \theta_2 a_{h+p-2}$$
$$\hat{r}_h(1) = \mu - \theta_1 a_h - \theta_2 a_{h-1}$$
$$\hat{r}_h(2) = \mu - \theta_2 a_h$$
$$\hat{r}_h(p) = \mu \quad \text{for} \quad p > 2$$

Thus the multi-step ahead forecasts and the variance of forecast errors of an MA(2) model go to the mean and variance of the series after two steps. In general, for an MA(q) model, the forecasts go to the mean after the first q steps

A.4.5. Estimation for ARMA process

Unfortunately, ACF and PACF are not informative in determining the order of an ARMA model. One has to use the AIC / BIC, or adopt more sophisticated measures such as the extended auto-correlation function proposed by Tsay and Tiao.

The forecasting behavior of an ARMA(p,q) model has characteristics similar to those of an AR(p) model.

A.4.6. Random walk with drift

Many financial time series are non-stationary, especially for price series of assets or foreign exchange rates. As a special case of the AR(1) process, when we set $\phi_1 = 1$, we have

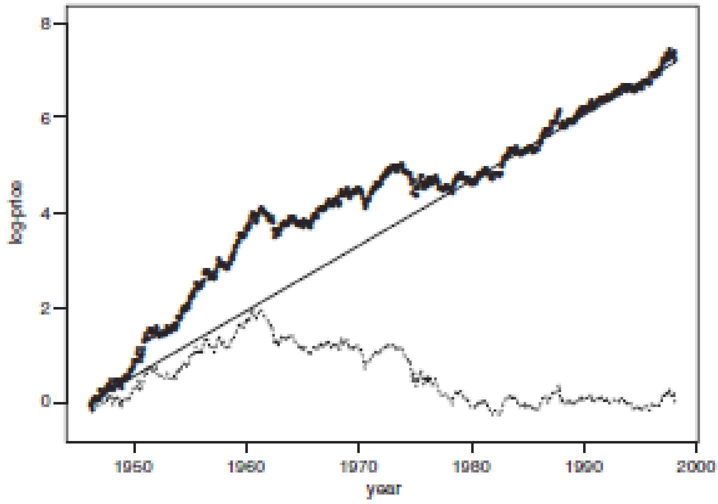
$$x_t = x_{t-1} + a_t = a_t + a_{t-1} + a_{t-2} + \cdots$$

This is different from a typical AR(1) process. For example, if one of the a_i is large and subsequent values of a_t are small, the path taken by x_t can be quite unpredictable. As such, this process is known as the random walk process.

A related process is the random walk with drift μ: $x_t = \mu + x_{t-1} + a_t$

$$\begin{aligned} x_1 &= \mu + x_0 + a_1 \\ x_2 &= \mu + x_1 + a_2 = 2\mu + x_0 + a_2 + a_1 \\ &\vdots \\ x_t &= \mu t + x_0 + a_t + a_{t-1} + \cdots + a_1 \end{aligned}$$

The expected value of $x_t = x_{t-1} + \mu = x_0 + \mu t$, thus it would not oscillate around a fixed level, but would show trending behavior.



The above example of a random walk with drift model is taken from Tsay (2005), p.66.

Properties of the random walk process

Because we don't assume stationarity for the variable x_t, its variance is:

$$\text{var}(x_t) = \text{var}(a_t) + \text{var}(a_{t-1}) + \cdots = t\sigma_a^2$$

i.e. it depends on the time instant, which increases linearly with time *t*. In practice, it means that the variable can possibly take any value when the forecast horizon *t* increases.

The ACFs at all lags do not decay over time, as the weight $\psi_i = 1$ for all *i*, and therefore all past shocks have impacts on the current value.

Finally, we would like to comment on the meanings of a constant term in a time series model. In the MA(q) model (section A.3.2.), the constant μ is the mean of the series. In the AR(p) model (e.g. AR(2) in section A.3.1.), the constant term is related to the mean via

$$\mu = \phi_0/(1 - \phi_1 - \ldots - \phi_p)$$

However, in the random walk with drift model, the constant term is the slope of the price/time diagram.

A.5. Concluding remarks

In this chapter we have introduced the basic models used for financial time series estimation and forecasting. However, it should be noted that the model fitting process is not an exact procedure - no clear cut rule exists as to how to choose a model. One must understand the assumptions and limitations of the models. In fact, both statistical and financial assumptions are important, and models should not just be used as a kind of "black box."

References:

Chris Chatfield, *Time Series Forecasting*, Chapman (2000).
Ruey Tsay, *Analysis of Financial Time Series*, Chapters 1–2, 2nd ed. Wiley (2005) or 3rd ed., Wiley (2010).