The Collective2 statistics program for equity curves

by
Jules L. Ellis[1]

[1] Grotestraat 63, 6511 VB, Nijmegen, Netherlands. E-mail: julesellis@xs4all.nl

Content

Introduction

This is not an introduction to statistics. The objective of this document is to give a detailed description of the statistics that are produced by the program. It is assumed that the reader already knows and understands most of these statistics, and that it is sufficient to describe which statistics are computed exactly and where they are displayed in the output. This is mainly done in the sections that are labelled as *Definition*. However, you will probably find that these sections are rather densely written. In an attempt to make the document more readable, most *Definition* sections are preceded by a *Theory* section that contains an introduction to the underlying theory, and followed by a set of *Remarks* that point out important conclusions and possible mistakes in the interpretation. These remarks are followed by an *Example* section where the most important definitions and remarks are illustrated with a simple data set. Furthermore, the next section contains a brief description of all output. Depending on your knowledge, however, you may still find it necessary to find other sources of information about the statistics. Please remember that the purpose of this document is not to explain what a Sharpe ratio is, or why some people prefer a Sortino ratio. The purpose is rather to explain what the difference is between the 24 different estimates of the Sharpe ratio that the program can yield…

The program has certain limitations that are inherent to its purpose. Firstly, the program analyzes only the account values (i.e. the open equity curve) and not the trades underlying it. This evidently limits the possibilities of risk assessment. Secondly, only the best-case equity curve is analyzed; slippage data are not used. So your real returns can be much less (see the last section for more detailed comments). Thirdly, the program produces only statistics that can be computed straightforwardly. Bootstrap and Monte Carlo simulation methods, factor analysis and autoregressive models are not implemented.

Quick glance at the output

Below you see a sample output of the program at the left side, with a brief interpretation at the right side. A more detailed discussion is given in the subsequent sections. If you read only this section, please be aware that statistics cannot replace common sense. Particularly, “significant” statistical results are always based on the past and thus still no guarantee for future performance. For example, a system that has long positions in a bull market during a few weeks may have excellent statistics that are ‘significant’, but this significance means merely that it is reasonable to believe that the system is profitable *in a bull market*.

Output		**Brief interpretation**
INSUFFICIENT DATA FOR ANALYSIS ON MONTHLY VALUES		Since the data are about 10 days, it is impossible to analyze monthly account values.
ANALYSIS BASED ON DAILY VALUES, FULL HISTORY		But we can analyze the daily account values.
RATIO STATISTICS		
Ratio statistics of excess return rates		A return rate is defined as an account value divided by the previous account value. The excess return rate subtracts the risk-free rate (T-bill note) from that.
Statistics related to Sharpe ratio		The Sharpe ratio is a measure for profit / risk.
Mean	85.037	This is the profit part. It is the annualized mean excess return rate. That would be 8503.7% per year.
SD	15.943	This is the risk part. That would be 1594.3% per year.
Sharpe ratio (Glass type estimate)	5.334	This is a simple estimate of the Sharpe ratio: profit / risk. *But it can also be computed on the log return rates, and then the outcome is very different (see many rows below).*
Sharpe ratio (Hedges UMVUE)	4.815	This is a better estimate of the Sharpe ratio. Note that it is somewhat more conservative.
df	8.000	This is the number of return rates minus 1.
t	0.838	This is a test statistic, used in the calculation of p below.
p	0.213	If p is smaller than 0.05, then there is enough evidence to conclude that the true Sharpe ratio is positive.
Lowerbound of 95% confidence interval for Sharpe Ratio	-7.566	The above estimates of the Sharpe ratio are quite unreliable. This confidence interval ranges from -7.566 to 17.921, which means that the true Sharpe ratio can be anywhere between these two bounds.
Upperbound of 95% confidence interval for Sharpe Ratio	17.921	
Lowerbound of 95% CI (Gibbons, Hedeker & Davis approximation)	-7.888	These outcomes are only useful if the program was unable to compute the bounds exactly above.
Upperbound of 95% CI (Gibbons, Hedeker & Davis approximation)	17.518	
Statistics related to Sortino ratio		Sortino criticized the Sharpe ratio because large profits also increase the “risk” as it is measured in the Sharpe ratio.
Sortino ratio	13.795	The Sortino ratio is a measure for mean performance / downside risk. Only losses enter into the downside risk (unlike the Sharpe ratio). Both profits and losses enter in the numerator. *But it can also be computed on the log return rates, and then the outcome is very different (see many rows below).*

	Upside Potential Ratio	24.794	The Sortino ratio is a measure for mean profit / downside risk. Only losses enter into the downside risk (unlike the Sharpe ratio). Only profits enter in the numerator. *But it can also be computed on the log return rates, and then the outcome is very different (see many rows below).*
	Upside part of mean	152.839	This is the part of the mean that is due to profits.
	Downside part of mean	-67.802	This is the part of the mean that is due to losses.
	Upside SD	14.413	This is the upside "risk", due to profits. So it is not really risk.
	Downside SD	6.164	This is the downside risk, due to losses.
	N nonnegative terms	6.000	This is the number of profits.
	N negative terms	3.000	This is the number of losses.
Statistics related to linear regression on benchmark			The 'benchmark' is the S&P. The question is whether the system performs better than simply buy & hold S & P with some leverage.
	N of observations	9.000	
	Mean of predictor	114.682	The predictor is the series of return rates of the benchmark.
	Mean of criterion	85.037	The criterion is the series of return rates of the system.
	SD of predictor	5.448	Intermediate result.
	SD of criterion	15.943	Intermediate result.
	Covariance	-11.929	Intermediate result.
	r	-0.137	The correlation between the return rates. The negative value indicates that a positive return of the benchmark tends to go with a negative return of the system. However, the correlation is close to 0, which means that the relation is weak.
	b (slope, estimate of beta)	-0.402	Systematic risk of the system, due to variations that it has in common with the benchmark. The benchmark is the S & P.
	a (intercept, estimate of alpha)	131.128	The market advantage in comparison with a benchmark portfolio with the same systematic risk.
	Mean Square Error	285.008	Intermediate result.
	DF error	7.000	Intermediate result.
	t(b)	-0.367	Intermediate result.
	p(b)	0.638	If p is smaller than 0.05 then there is enough evidence that the true *b* is positive. Then *r* is positive too.
	t(a)	0.793	Intermediate result.
	p(a)	0.227	If p is smaller than 0.05 then there is enough evidence that the true *a* is positive.
	Lowerbound of 95% confidence interval for beta	-2.993	The true value of *b* (beta) can be anywhere between -2.993 and 2.189.
	Upperbound of 95% confidence interval for beta	2.189	
	Lowerbound of 95% confidence interval for alpha	-259.895	The true value of a (alpha) can be anywhere between -259.895 and 522.151.
	Upperbound of 95% confidence interval for alpha	522.151	
	Treynor index (mean / b)	-211.587	
	Jensen alpha (a)	131.128	This is *a* again.
Ratio statistics of excess log return rates			All these statistics have a similar counter part above. The difference is that they are now computed on the logarithm of the return rates instead of the return rates themselves. This is more appropriate if you and the system are
Statistics related to Sharpe ratio			

			compounding.
	Mean	-0.049	
	SD	13.376	
	Sharpe ratio (Glass type estimate)	-0.004	
	Sharpe ratio (Hedges UMVUE)	-0.003	
	df	8.000	
	t	-0.001	
	p	0.500	
	Lowerbound of 95% confidence interval for Sharpe Ratio	-12.485	
	Upperbound of 95% confidence interval for Sharpe Ratio	12.478	
	Lowerbound of 95% CI (Gibbons, Hedeker & Davis approximation)	-12.485	
	Upperbound of 95% CI (Gibbons, Hedeker & Davis approximation)	12.478	
Statistics related to Sortino ratio			
	Sortino ratio	-0.005	
	Upside Potential Ratio	10.954	
	Upside part of mean	99.602	
	Downside part of mean	-99.651	
	Upside SD	8.739	
	Downside SD	9.093	
	N nonnegative terms	6.000	
	N negative terms	3.000	
Statistics related to linear regression on benchmark			
	N of observations	9.000	
	Mean of predictor	93.334	
	Mean of criterion	-0.049	
	SD of predictor	3.626	
	SD of criterion	13.376	
	Covariance	-10.262	
	r	-0.212	
	b (slope, estimate of beta)	-0.781	
	a (intercept, estimate of alpha)	72.816	
	Mean Square Error	195.324	
	DF error	7.000	
	t(b)	-0.573	
	p(b)	0.708	
	t(a)	0.469	
	p(a)	0.327	
	Lowerbound of 95% confidence interval for beta	-4.003	
	Upperbound of 95% confidence interval for beta	2.442	
	Lowerbound of 95% confidence interval for alpha	-294.284	
	Upperbound of 95% confidence interval for alpha	439.916	
	Treynor index (mean / b)	0.062	
	Jensen alpha (a)	72.816	

Risk estimates for a one-period unit investment (parametric)			
assuming log normal returns and losses (using central moments from Sharpe statistics)			
	VaR(95%)	0.684	Assuming a lognormal distribution for the returns, there is a probability of 5% of a loss of 68.4% or more in one period.
	Expected Shortfall on VaR	0.757	If the above event happens, then the expected loss is 75.7%.
assuming Pareto losses only (using partial moments from Sortino statistics)			
	VaR(95%)	0.338	Assuming a Pareto distribution for the losses, there is a probability of 5% of a loss of 33.8% or more in one period.
	Expected Shortfall on VaR	0.638	If the above event happens, then the expected loss is 63.8%.
ORDER STATISTICS			
Quartiles of return rates			
	Number of observations	9.000	There are 9 return rates.
	Minimum	0.400	The smallest return rate was 0.400, i.e. 60% loss.
	Quartile 1	0.500	The 25% smallest return rates fall below 0.500, i.e. 50% loss.
	Median	1.125	The 50% smallest return rates fall below 1.125, i.e. 12.5% profit.
	Quartile 3	1.200	The 75% smallest return rates fall below 1.200, i.e. 20% profit.
	Maximum	2.667	The largest return rate was 2.667, i.e. 166.7% profit.
	Mean of quarter 1	0.443	The 25% smallest return rates had a mean of 0.443.
	Mean of quarter 2	1.118	The next 25% smallest return rates had a mean of 1.118.
	Mean of quarter 3	1.183	Similar.
	Mean of quarter 4	2.583	Similar.
	Inter Quartile Range	0.700	Measure of dispersion for the return rates.
	Number outliers low	0.000	There are no losses that are larger than could be expected on basis of the first quartile and the interquartile range.
	Percentage of outliers low	0.000	
	Mean of outliers low	NA	
	Number of outliers high	2.000	There are two profits that are larger than could be expected on basis of the third quartile and the interquartile range.
	Percentage of outliers high	0.222	That is 22.2%.
	Mean of outliers high	2.583	Their mean is 2.583, i.e. 158.3% profit.
Risk estimates for a one-period unit investment (based on Extreme Value Theory)			
	Extreme Value Index (moments method)	-20.296	The negative value indicates that the losses are bounded.
	VaR(95%) (moments method)	0.583	There is a probability of 5% that a loss will be greater than 58.3% in one period, *according to the moments estimation method.*
	Expected Shortfall (moments method)	0.583	If this happens, the average loss is 58.3%.
	Extreme Value Index (regression method)	-2.322	The negative value indicates that the losses are bounded.
	VaR(95%) (regression method)	0.655	There is a probability of 5% that a loss will be

			greater than 65.5% in one period, *according to the regression estimation method.*
	Expected Shortfall (regression method)	0.657	If this happens, the average loss is 65.7%.
DRAW DOWN STATISTICS			
Quartiles of draw downs			
	Number of observations	3.000	There were 3 draw downs.
	Minimum	0.500	The smallest draw down was 0.500 (= 50%).
	Quartile 1	0.536	The 25% smallest draw downs fall below 0.536 (= 53.6%).
	Median	0.571	The 50% smallest draw downs fall below 0.571 (= 57.1%).
	Quartile 3	0.586	The 75% smallest draw downs fall below 0.586 (= 58.6%).
	Maximum	0.600	The largest draw down was 60%.
	Mean of quarter 1	0.500	The mean of the 25% smallest draw downs was 0.500 (= 50%).
	Mean of quarter 2	0.571	See the corresponding statistics of one-period return rates for a description. In this case they are based on draw downs that may extend over many periods.
	Mean of quarter 3	NA	
	Mean of quarter 4	0.600	
	Inter Quartile Range	0.050	
	Number outliers low	0.000	
	Percentage of outliers low	0.000	
	Mean of outliers low	NA	
	Number of outliers high	0.000	
	Percentage of outliers high	0.000	
	Mean of outliers high	NA	
Risk estimates based on draw downs (based on Extreme Value Theory)			
	Extreme Value Index (moments method)	NA	See the corresponding statistics of one-period return rates for a description. In this case they are based on draw downs that may extend over many periods.
	VaR(95%) (moments method)	NA	
	Expected Shortfall (moments method)	NA	
	Extreme Value Index (regression method)	NA	
	VaR(95%) (regression method)	NA	
	Expected Shortfall (regression method)	NA	
COMBINED STATISTICS			
	Annualized return (arithmetic extrapolation)	0.000	Annualized return is the system is not compounding.
	Compounded annual return (geometric extrapolation)	0.000	Annualized return if the system is compounding.
	Calmar ratio (compounded annual return / max draw down)	0.000	
	Compounded annual return / average of 25% largest draw downs	0.000	
	Compounded annual return / Expected Shortfall lognormal	0.000	

Continued example of data

Most examples will be based on this set of data, that represent the daily values of the account (the equity) and the benchmark:

> Account values = (5, 2, 5, 6, 7, 3, 8, 9, 10, 5);
> Benchmark values = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10);

These data are deliberately chosen small and easy, rather than realistic, so that you can easily do most computations yourself if you want.

Period and span of the analyses

Definition

The program analyzes subsequently:

1. The monthly account values, over the full history of the system.
2. The daily account values, over the full history of the system.
3. The daily account values, over the last 6 months.

If the history of the system is shorter than 6 months, then the third analyses will be omitted. If the history of the system is shorter than 3 months, then the first analysis will be omitted too.

Remarks

1. In an analysis of monthly values, only the last account value of each month is used. So daily account value variations within a month cannot affect the outcomes of this analysis.

2. The data are counted from the first day that the system started. It is assumed that there are 12 months in a year and 365 days in a year[2], and 365 / 12 days in each month. If the last month is not complete, then the data of that month are not used in the analysis of monthly values.

Example

Suppose a system is 100 days old. Then the analysis of monthly values will be based on the account values of day 0, 30, 61 and 91.

---

[2] In the actual program this has been changed into the expected number of trading days in a year, but this guide was written before it changed and uses 365 everywhere.

Return rates and log return rates

Definition

Many analyses are conducted not on the account values themselves, but rather on the return rates and the log return rates. Suppose the account value on day i is V_i, then the *return rates* are

$$R_i = V_i / V_{i-1}$$

The *returns* are R_i - 1 and the *log return rates* are

$$L_i = \log R_i$$

where log is the natural logarithm. If the risk-free rate of return is r (e.g. r = 1.05 means 5% risk-free return) then the *excess return rates* are

$$ER_i = R_i - r$$

and the *excess log return rates* are

$$EL_i = \log R_i - \log r.$$

Here, the *risk-free rate* is adapted to the periods that are analyzed. For example, if the yearly risk-free rate is 1.05, and the analysis is for daily values, then $r = 1.05^{1/365} =$ 1.00013368061711 is used for each day.

Remarks

1. Note that an analysis of the log return rates corresponds to an analysis of the logarithm of the account values, since $L_i = \log V_i - \log V_{i\text{-}1}$ and $EL_i = \log V_i - \log V_{i\text{-}1} - \log r$.

2. The word **return rate** is used here and throughout the document in the meaning of a multiplication factor, while the word return is used for the corresponding percentage. So

return rate = 1 + *return*

For example, a return rate of 1.23 corresponds to a return of 0.23 = 23%. However, the statistics in the Sharpe, Sortino and regression part of the output are based on the *excess* return rates R_i - r. In that case the 1 included in R_i and the 1 included in r cancel out against each other, so the difference in return rates is the same as the difference in the returns. E.g. 1.23 – 1.05 has the same outcome as 0.23 - 0.05.

Example

For a yearly risk-free rate of 1.05, Table 1 shows a possible series of daily account values in row 2, and the corresponding rates in the rows below it.

Table 1

Day	0	1	2	3	4	5	6	7	8	9
Account value	5	2	5	6	7	3	8	9	10	5
Return rate		0.400	2.500	1.200	1.167	0.429	2.667	1.125	1.111	0.500
Log return rate		-0.916	0.916	0.182	0.154	-0.847	0.981	0.118	0.105	-0.693
Excess return rate		-0.600	1.500	0.200	0.167	-0.572	1.667	0.125	0.111	-0.500
Excess log return rate		-0.916	0.916	0.182	0.154	-0.847	0.981	0.118	0.105	-0.693

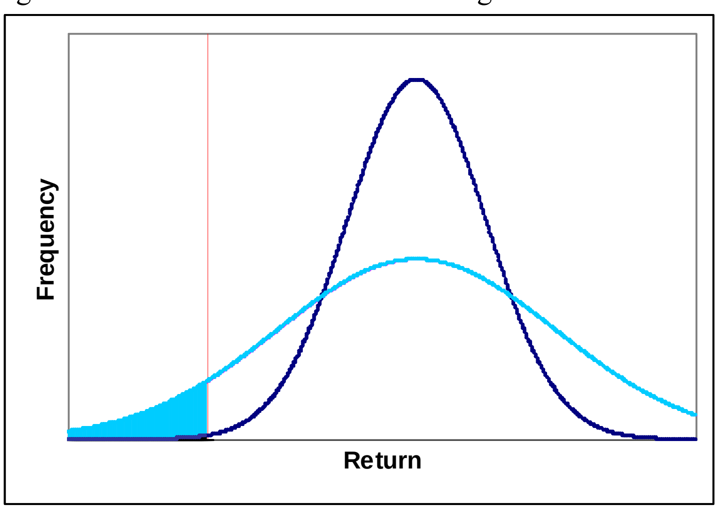
Statistics related to the Sharpe ratio

Theory – Introduction

The Sharpe ratio is one of the central measures of performance in the *mean-variance* paradigm of the now classical "Modern Portfolio Theory" developed by Markowitz (1952, 1991), Sharpe (1964, 1991) and others. One of the basic assumptions underlying this paradigm is that investors want to avoid unnecessary risk. For example, given the choice between winning $100 for sure and a gamble where you win either $50 or $150 with 50% chance, the first alternative would be preferred. It is assumed that investors are averse of *unnecessary* risk; that is, all other things being equal they will choose the option with less risk, but it is not assumed that they want to sacrifice profits to reduce risk. The theory does not assume a specific level of risk aversion.

The basis assumption of the mean-variance paradigm is that the mean and variance of one-period returns are sufficient statistics for evaluating the prospects of an investment portfolio (Sharpe, 1994). The mean is viewed as a measure of the profitability, while the variance is interpreted as a measure of the risk.

In the Capital Asset Pricing Model (CAPM), introduced by Sharpe (1964) within the mean-variance paradigm, it is usually also assumed that the return rates are normally distributed. Under this assumption it is clear that risk *must* be measured by the variance. This is so because a normal distribution is completely characterized by its mean and variance alone. So given the mean, it is only the variance that determines what the probability of a loss is. This is illustrated in the figure below. These two distributions have the same means but the variance of the light curve is greater than the variance of the dark curve. The returns to the left of the red line are losses. The percentage of losses, indicated by the shaded area under the curve, is greater for the distribution with the larger variance.



For interpretation purposes it is usually more convenient to use the standard deviation instead of the variance. The standard deviation is the square root of the variance. If the variance increases, then the standard deviation increases too. So the standard deviation is also a measure of risk, but on a more convenient scale.

Another important assumption of the CAPM is that there exists a *risk-free asset*. Usually this is interpreted as buying a T-bill note. Since it is riskless, it outperforms every risky portfolio with the same return. The model therefore focuses on the *differential* or *excess* returns: the difference between the return of a risky portfolio and the return of the risk-free asset. This can also be viewed as the payoff obtained from a unit investment, financed by borrowing (a short position in the risk-free asset).

It is logical then to express the performance of the portfolio as the ratio of the mean to the standard deviation of the differential returns. This is the *Sharpe ratio*. This is a reward / risk measure within the mean-variance paradigm.

Theory – Optimizing properties of the Sharpe ratio for a one-period investment

If the returns have a normal distribution, then the probability of a return smaller than the riskless return is a strictly decreasing function of the Sharpe ratio alone. So if you want to minimize the probability that you will regret not having bought a T-bill note, then you should *maximize the Sharpe ratio* if the returns are normally distributed.

The importance of the Sharpe ratio stems from this insight of the mean-variance paradigm (e.g., Sharpe, 1994): If you have a portfolio that is a combination of a risky fund and the risk-free asset, where the weights are such that the portfolio has the risk that you want to tolerate, then the expected return after one period is

Expected return = r + k SR

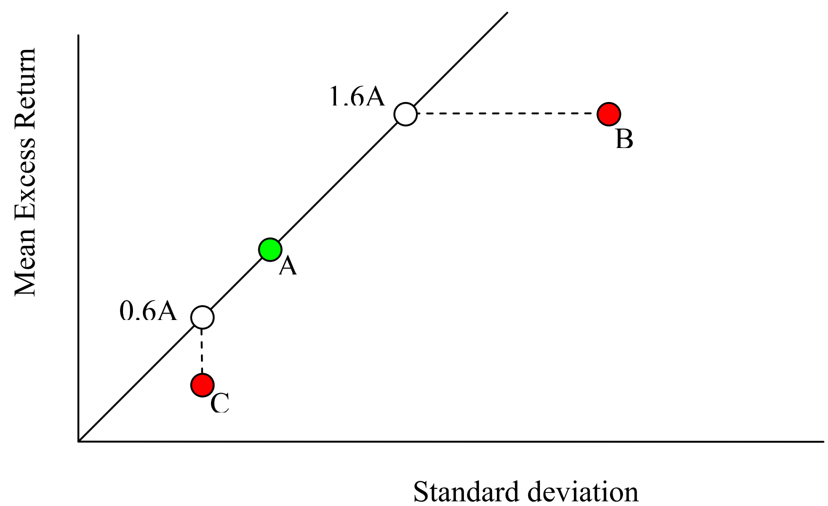
where *r* is the return of the risk-free asset, *SR* is the Sharpe ratio, and *k* is the level of risk that you tolerate, measured in standard deviations. This is an intuitively very appealing formula: The expected return increases with product of the risk that you want to tolerate and the Sharpe ratio of the fund. If you tolerate more risk then you can have a higher leverage in the fund, and then the expected return will be higher, provided that the Sharpe ratio is positive. If the Sharpe ratio increases, then the expected return will be higher too, provided that your risk tolerance is larger than zero (otherwise you will not take a position in the fund). The important fact about this formula is that the Sharpe ratio is the *only* property of the fund that determines the expected return.

Another way to interpret this formula is:

> *Every combination of the portfolio that has the highest Sharpe ratio with borrowing or lending the risk-free asset yields higher returns than every other portfolio with the same risk, and lower risk than every other portfolio with the same return.*

The consequence of this is, again, that maximizing the Sharpe ratio is a rational strategy *regardless of your risk aversion level*. Suppose that you can choose between a portfolio *A* with a high Sharpe ratio and a portfolio *B* that has lower Sharpe ratio but a higher expected return. If you don't care very much about risk you might be tempted to prefer *B* because of its higher returns, but according to the mean-variance paradigm it is even better to buy more of portfolio *A* and finance it by a short position in the risk-free asset (i.e. leverage *A*) such that the expected return is the same as *B*. This will entail the same expected return with less risk. Similarly, if you have a high risk aversion, and there is a portfolio *C* with a smaller Sharpe ratio and a smaller risk, then you might be tempted to prefer *C*, but it would still be more rational to have a combination of *A* and the risk-free asset such that the risk is the same as *C*. This will entail the same risk as *C*, but with a higher return.

The reasoning is illustrated by the figure below. Here, portfolio *A* is the green dot and the portfolios *B* and *C* are the red dots. The white dots represent the portfolios that are derived from *A* by having a partial or a leveraged position in *A*. The standard deviation is on the horizontal axis and the mean return on the vertical axis. So the Sharpe ratio of a portfolio is equal to the *slope* of the line that connects the portfolio with the origin. Here, that line is the solid line through *A*. Its slope is larger than the slope of similar lines through *B* and *C* (not drawn in the figure). So *A* has the largest Sharpe ratio. Portfolio *B* has a larger mean return than *A* because it lies higher. However, 1.4*A* has the same return (horizontal dashed line) with less risk. Portfolio *C* has less risk than *A* because it lies more to the left, but 1.6*A* has also less risk (vertical dashed line) while it has higher returns than *C*. So however you put it, regardless of whether your emphasis is on high returns or low risk, it is always more rational to chose the portfolio with the highest Sharpe ratio and adjust the leverage to your personal wishes of return and risk, than to chose a portfolio with a lower Sharpe ratio. At least, that is what the mean-variance paradigm implies.



One of the limitations of the mean-variance paradigm is that the variance is not a good measure of risk if the distribution of returns is asymmetrical (e.g. many small losses and a few large profits). In that case the Sharpe ratio loses its optimizing properties too, and then indices like the Sortino ratio and the Upside Potential Ratio can be better performance indicators.

Another problem is that distributions of returns can be heavy-tailed (i.e. extreme losses or extreme profits are more likely than predicted by a normal distribution) (Mandelbrot, 1963). This does not necessarily invalidate the theory of the mean-variance paradigm, but the usual statistical estimates will be biased.

It is not uncommon to reduce both of the two above limitations by applying the CAPM to the logarithm of the return rates (e.g. Kim, 2002; Schröder, 2006), thus assuming a lognormal distribution for the returns. The lognormal distribution is asymmetric and has heavier tails than the normal distribution. There is a third, more important reason to use the logarithm, which will be considered now.

Theory – The use of log return rates in compounding systems

Remember that the Sharpe ratio maximizes the *one-period* expected returns. It does not maximize the returns over a long horizon if the returns are compounded. In that case it is better to use the Sharpe ratio based on log return rates. This will be discussed now.

The effect of compounding can be understood easily by considering the extreme situation of a trading system that has each month either 200% profit or 100% loss, each with 50% chance. The Sharpe ratio of the return rates is positive, because the expected return is (200% + -100%) / 2 = 50%. This reflects the fact that for a one-period investment the fund would have a positive expected return. The problem is that if you are compounding, then after a single 100% loss you can never recover anymore.

The underlying problem is that it can be irrational to maximize the expected value in a multi-period investment. Consider this example: you are given the opportunity to play a game that consists of a series of gambles. On each gamble you either triple your capital or lose it all, each with 50% probability. After each gamble are allowed to quit or to continue the game, but you can play the game only once. How long would you continue? This is a variation of the St. Petersburg paradox, which was originally published by Bernoulli in 1738. If you maximize the expected value then you will always continue to play, because the expected value is positive at each point of time. At the same time however, this strategy leads to almost sure ruin: The probability that you keep winning forever is 0.

The expected value increases under compounding, but the risk increases at a different rate, and as a result *the Sharpe ratio is not invariant under compounding*. For example, consider two trading systems A and B. Suppose A has either a profit of 10% or a loss of 6.6%, each with 50% probability. Suppose B has profits and losses that are 10 times larger, that is either a profit of 100% or a loss of 66%, each with 50% probability. Assume for simplicity that the risk-free return is 0% or that the risk-free return is already subtracted from the above returns. Then the Sharpe ratio of both systems is 0.2048. If the returns for 10 periods are considered, then there are 32768 possible sequences for each system. If the end returns of these sequences are used, then system A has Sharpe ratio 0.6905 while system B has Sharpe ratio 0.04260. Indeed, system A has a probability of 69.6% of a positive end return, while system B has probability 15.1%. Thus both systems have the same Sharpe ratio for a one-period investment but not for a 10-period investment with compounding.

A solution to this has already been suggested by Bernoulli, namely to assume that each account value V has a utility $U = f(V)$, and to maximize the *expected utility* instead of the expected return. He then argues "Now it is highly probably that *any increase in wealth, no matter how insignificant, will always result in an increase in utility which is inversely proportional to the quantity of goods already possessed"*. There is only one function f that has this property, namely a logarithmic function. So this would encourage the use of log return rates.

Utility is often viewed as the psychological value of the return. For example, Bernoulli stated "Thus there is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount". The logarithmic function, because it is concave (decelerating), can be viewed as a utility function that captures the differential psychological impact of losses and profits. It puts more weight on the losses and even more weight on large losses; less weight on profits and even less weight on large profits. This corresponds to the psychological phenomenon that "good things satiate and bad things escalate" (Coombs and Avrunin 1977).

The log return rate can also be interpreted as *the time it will save you to attain a certain investment goal when you are compounding*. If the log return rate of one investment is two times larger than the log return rate of another investment, then any desired account value will be achieved two times faster with the former than with the latter investment. So instead of maximizing the return per unit of time, one minimizes the time needed for a desired return.

However, it is not merely psychology that leads to the logarithmic function, as can be seen in the above variation on the St. Petersburg paradox. A fund manager who maximizes the momentary expected return at the expense of sure ruin in the long run will not be considered rational by any standard. According to the 'Kelly criterion', attributed to Kelly (1956), *the rational strategy for an investor is the one that maximizes the probability of being ahead of other investors in the long run*. Maximization of the expected utility satisfies this criterion, while maximization of the expected return fails to do so. That is, "a rational long-run investor should maximize the expected growth rate of his wealth share and, therefore, should behave as if he were endowed with a logarithmic utility function" (Chen & Huang, 2006). These authors found in a simulation of an artificial stock market that only the investors with a logarithmic utility function survived in the long run, while all investors with other utility functions, including CAPM, were driven to extinction.

Use of the logarithmic function is only advisable in a compounding system, not in a system that trades with a fixed capital. As Kelly (1956, pp.925-926) notes,

> "The gambler introduced here follows an essentially different criterion from the classical gambler. At every bet he maximizes the expected value of the logarithm of his capital. The reason has nothing to do with the value function which he attached to his money, but merely with the fact that it is the logarithm which is additive in repeated bets and to which the law of large numbers applies. Suppose the situation were different; for example, suppose the gambler's wife allowed him to bet one dollar each week but not to reinvest his winnings. He should then maximize his expectation (expected values of capital) on each bet. He would bet all his available capital (one dollar) on the event yielding the highest expectation. With probability one he would get ahead of anyone dividing his money differently."

The Sharpe ratio of the log return rates has analogue optimizing properties as the ordinary Sharpe ratio. Firstly, in the returns have a lognormal distribution, then maximizing the Sharpe ratio of log return rates minimizes the probability of a shortfall under the risk-free asset, just as the original Sharpe ratio does in case of a normal distribution. Secondly, if you consider the log return rate as the utility (i.e. the psychological value or its utility in generating future returns) of the return, then a similar equation can be written:

Expected utility = $= r' + k'\, SR'$

where r' is the log of the risk-free rate, k' is the risk that you want to tolerate, measured in standard deviations of the utilities, and SR' is the Sharpe ratio computed on the log return rates. Here, k' is also an increasing function of the leverage (this is not immediately obvious, but a proof is given in the appendix). So the same reasoning can be applied here, leading to the conclusion that SR' is the only property of the fund that should be maximized in the fund selection if you want to maximize the expected utility with a specified risk tolerance.

In summary, while the Sharpe ratio is usually reported on basis of the return rates, a long-run compounding investor is better served with a Sharpe ratio that is based on the log return rates. This does not mean that the difference is necessarily large. If the returns are close to 0 then they will be very similar to the corresponding log return rates (Sharpe, 1992), and in such cases it can be expected that the fund rank order on basis of the Sharpe ratio of log return rates is often very similar to the rank order on basis of the Sharpe ratio of the returns.

Theory – Sharpe ratio and the t-test

The Sharpe ratio is closely related to Student's *t*-test. The *t*-statistic for the null-hypothesis that the expected excess return is positive is simply the Sharpe ratio multiplied by $\sqrt{T}$, where T is the number of periods. The advantage of this is that much statistical theory is known about *t*, which can be applied directly to the Sharpe ratio (Sharpe, 1998). Particularly, the Sharpe ratio is know as the *effect size* in the theory of *t*-tests, and the estimation of this has recently received much attention (e.g. Kelley, 2005).

Theory – The Sharpe ratio and future performance

Because of its firm theoretical basis, the Sharpe ratio is still one of the most widely used performance indicators, despite its limitations. Nevertheless, you must be aware that it is always computed on past results and that this may not be very predictive of future performance. Sharpe (1966) found for mutual funds a positive but small correlation ($r = 0.36$) between their ranks based on the Sharpe ratio over the period 1944-1953 versus the period 1954-1963 (see e.g. Gerrans, 2006 for a more contemporary analysis).

Definition

The results in the output are all annualized. Consider first the non-annualized statistics. The mean and the standard deviation are computed on either the excess return rates or the excess log return rates. The Glass type estimate for the Sharpe ratio is then obtained as

$$SR = Mean\ /\ Standard\ deviation$$

Although this is the estimate that is typically reported, it is in fact biased (Hedges, 1981), particularly when the sample size is small. The unbiased estimate can be computed by Hedges correction. Hedges proved that this is the uniform minimum variance unbiased estimator (UMVUE):

$$SR' = SR * c$$

where $c = \Gamma(df/2)(df/2)^{-1/2}(\Gamma((df-1)/2))^{-1}$ and $df = n - 1$ and n is the number of ER_i values. Here, Γ is Euler's gamma function. The correction factor c is approximated by $1 - 3/(4*df-1)$ if $df > 50$.

Next, the degrees of freedom are computed as $df = n - 1$, and the *t*-value as $t = SR\ \sqrt{n}$. The p-value is the area under the right tail of the central t-distribution with df degrees of freedom. Under the usual assumptions this yields a test of the null hypothesis that the mean excess return rate is larger than 0. That is, if the population mean of the excess return rates is μ then the null hypothesis is that $\mu \leq 0$ and the alternative hypothesis is that $\mu > 0$, and the null hypothesis should be rejected for small values of p (e.g. $p < 0.05$).

The 95% confidence interval for the population value of the Sharpe ratio is computed by means of the noncentral *t*-distribution. The tail probabilities are computed by Lenth's (1989) algorithm AS 243 and the initial estimates are iteratively refined until the bounds correspond to the required probability. However, the convergence of this refinement is not sure, and therefore the program also reports an approximation of the 95% confidence interval, suggested by Gibbons, Hedeker, and Davis (1993), based on a result of Johnson and Welch (1939).

Finally, where applicable the outcomes are annualized in this way: If the number of periods in a year is T (e.g., in an analysis of monthly values, $T = 12$) then the mean is multiplied by T, the standard deviation is multiplied by $\sqrt{T}$, and the Sharpe ratio estimates (including the bounds of the confidence interval) are multiplied by $\sqrt{T}$. The *df*, *t* and *p* values are not annualized.

Remarks

1. Note that the statistics related to the Sharpe ratio are computed twice for each time period and span: Once on the excess return rates and once on the excess log return rates. The results are not reported adjacently, but are separated by the statistics about the Sortino ratio and the regression statistics.

2. It is important to recognize that this part of the output is not computed on the account values directly, but rather on the excess return rates or the excess log return rates. That is, the return rate at time i represents the value at time i of a unit investment at time $i - 1$, and not the value at time i of a unit investment at time 0. The utilization of excess return rates rather than compounded account values is in agreement with the definition of Sharpe (1994, particularly equations 12 and 13; 1997), Goetzmann (1996, Ch.I.III) and the formulas used by Morningstar (Sharpe, 1997, 1998). Morningstar does, however, in other respects use a slightly different formula for the Sharpe ratio, where it uses the geometric mean rather than the arithmetic mean in the numerator. The empirical outcomes are very similar to the conventional Sharpe ratio, with a correlation of .995 (see Sharpe, 1997, for a discussion). Contemporary financial theories often consider log return rates (e.g. McNeil & Frey, 2000). (Indeed, Morningstar's use of the geometric mean is equivalent to using the arithmetic mean of the log return rates, but this is not true for their version of the standard deviation). However, on many internet sites the Sharpe ratio is explained erroneously in that the 'average return' in the numerator is computed as the end value of the account after many periods, divided by the number of periods. This yields entirely different outcomes! See the example below.

3. The use of the excess return rates can lead to counterintuitive results, however, because the rates are not symmetrical in profits and losses. The asymmetry is that a loss of 50% must be followed by a profit of 100% to break even. For example, suppose that the equities are 10, 5, 10. Assume for simplicity that the risk-free rate is 1. Then the return rates are 0.5 and 2. So the mean excess return rate is $(0.5 + 2) / 2 - 1 = 0.25$, i.e. positive while the sequence would better be described as neutral for a compounding investor. This asymmetry can be solved by using the excess log return rates. In this example they are -0.693 and 0.693, and their mean is exactly 0.

4. Many values in the output are annualized. So in an analysis of daily values, the reported mean is 365 times the mean of the excess (log) return rates, the standard deviation is $\sqrt{365}$ times the standard deviation of the excess (log) return rates, and the Sharpe ratio estimates are $\sqrt{365}$ times the initial estimates. Similarly, in the analysis of monthly values the factors 12 and $\sqrt{12}$ are used.

5. This choice is for annualization is made because it is more or less a convention. It means something different for the return rates than for the log return rates. For example, consider a fixed return rate of 1.01 (i.e. 1%) per day with a risk-free rate of 1.00 (i.e. 0%). Applied to the return rates, the annualization formula says that a return of 1% per day is equivalent to 365% per year. That is, the profits are withdrawn and the account grows linearly. Applied to the log return rates, the annualization formula says that a return of 1% per day is equivalent to 3778% per year; the profits are reinvested and the account grows exponentially.

6. The p-value and the confidence interval are computed on basis of the ordinary assumptions, which implies that the excess (log) return rates should have a normal distribution and that they should be uncorrelated. These assumptions are likely to be violated. Nevertheless, the p-value and the confidence interval give you a crude estimate about the size of the unreliability of the estimates. The accuracy is generally better for large sample sizes (see Kelley, 2005).

7. Even if the sample size is large and the p-value small, suggesting a significant positive Sharpe ratio, please be aware of the fact that an essential assumption of this statistical test is that the observations are drawn randomly. In the present application, however, they are drawn from the past and definitely not random from the future…

8. The mean of the excess log return rates has a simple relation with the last account value. For the non-annualized mean we have

$$Mean(EL_i) = \frac{\sum_{i=1}^{n} EL_i}{n} = \frac{\sum_{i=1}^{n} (\log(V_i) - \log(V_{i-1}) - \log(r))}{n} = \log((\frac{V_n}{V_0})^{1/n} / r) .$$

So the mean of the excess log return rates depends only on the ratio of the last account value to the initial account value, and not on the intermediate account values. If the last account value is smaller than the initial account value then the mean will be negative (assuming that r ≥ 1) and then the Sharpe ratio estimates on basis of the excess log return rates will be negative too.

9. The mean of the excess log return rates is also closely related to the compounded annual return. Suppose that M is the annualized mean of the excess log return rates as it is reported in the output, and r_{annual} is the compounded annual risk-free return rate. (So $M = T$ * Mean(EL_i) and $r_{annual} = r^T$). Then the formula of the previous remark implies that the compounded annual return is $C = r_{annual} e^M - 1$. To put it differently,

$$e^M = \frac{C+1}{r_{annual}} .$$

So e^M is equal to the ratio of the compounded annual return rate to the compounded annual risk-free rate, and thus it may be called the ‘excess compounded annual return rate’. (Remember that the word *rate* is here used in the meaning of 1 + *percentage*/100).

10. The relation discussed in the previous remark also allows you to compute a confidence interval for the compounded annual return under the assumption that the return rates are independent and have a lognormal distribution. The compounded annual return itself is given

in the output (see below), but not the confidence interval. The lower bound of the 95% confidence interval can be computed as

$$r_{annual} \exp(M - t_{0.95} S \sqrt{\frac{T}{n}}) - 1$$

and the upper bound can be computed analogously if you replace "-" by "+". Here $r_{annual} = r^T$ is the compounded annual risk-free return rate, exp means exponentiation (i.e. exp(x) is another way to write e^x), M is the annualized mean that is reported in the output, $t_{0.95} \approx 1.96$ is the 97.5th percentile of a t-distribution with degrees of freedom equal to the df in the output (you can obtain the exact value with the function T.INV in Excel, if you substitute 0.05 for the probability; the outcome should be approximately 1.96 if df is large), S is the annualized standard deviation that is reported in the output, T is the number of periods in a year (12 if the analysis was based on monthly values, 365 if the analysis was based on daily values), and n is the number of periods (days or months) on which the analysis was based.

11. Note that the Sharpe ratio is not affected by the order in which losses occur. Indeed, the Sharpe ratio is basically a measure for a one-period investment (Sharpe, 1994).

Example

The various non-annualized and annualized means, standard deviations and ratios corresponding to Table 1 are displayed in Table 2. These are computed separately in Excel. The boldface numbers return in the output.

Table 2

	Mean	*Mean * 365*	*SD*	*SD * sqrt(365)*	*Mean / SD*	*sqrt(365) * Mean / SD*
Account value	6.111	2230.556	2.667	50.947	2.292	43.782
Return rate	1.233	450.086	0.834	15.943	1.478	28.231
Log return rate	0.000	0.000	0.700	13.376	0.000	0.000
Excess return rate	0.233	**85.037**	0.834	**15.943**	0.279	**5.334**
Excess log return rate	0.000	**-0.049**	0.700	**13.376**	0.000	**-0.004**

These are the corresponding parts of the output:

Table 3

Based on excess return rates	*Based on excess log return rates*
Mean 85.037 SD 15.943 Sharpe ratio (Glass type estimate) 5.334 Sharpe ratio (Hedges UMVUE) 4.815 df 8.000 t 0.838 p 0.213 Lowerbound of 95% confidence interval for Sharpe Ratio -7.566	Mean -0.049 SD 13.376 Sharpe ratio (Glass type estimate) -0.004 Sharpe ratio (Hedges UMVUE) -0.003 df 8.000 t -0.001 p 0.500 Lowerbound of 95% confidence interval for Sharpe Ratio -12.485

Upperbound of 95% confidence interval for Sharpe Ratio 17.921 Lowerbound of 95% CI (Gibbons, Hedeker & Davis approximation) -7.888 Upperbound of 95% CI (Gibbons, Hedeker & Davis approximation) 17.518	Upperbound of 95% confidence interval for Sharpe Ratio 12.478 Lowerbound of 95% CI (Gibbons, Hedeker & Davis approximation) -12.485 Upperbound of 95% CI (Gibbons, Hedeker & Davis approximation) 12.478

The first three numbers in each part of the output correspond to the boldface numbers in the table. The *df* is simply the number of rates minus 1. All other values are derived from these outcomes.

Consider first the results for the excess return rates. In this case, the ordinary estimate for the Sharpe ratio is 5.334. A better estimate is Hedges UMVUE, which is 4.815. The difference will vanish if the sample size increases. Note that the Sharpe ratio is positive according to both estimates, even though the last account value was equal to the initial account value. See remark 3.

The Sharpe ratio seems large, but note that it is extremely unreliable. This is caused by the small sample size. The first way to see this is by looking at the p-value. The p-value is 0.213, which is well above the ordinary level of significance of 5%. So the Sharpe ratio is not significantly larger than 0. This means: There is not enough evidence to be reasonably sure that the population value of the Sharpe ratio is positive. This does not necessarily mean that the Sharpe ratio is 'small', only that the number of data is small. The other way to see this is to look at the confidence interval. It ranges from -7.566 to 17.921. So the 'true' Sharpe ratio might be as low as -7.566. That is, the hypothesis that the Sharpe ratio is actually -7.556 would still be acceptable with these data. However, the hypothesis that it is actually 17.518 cannot be rejected either. Given the fact that the assumptions may be violated, the p-value and the confidence interval should be interpreted as no more than crude estimates. See remark 6.

Next, consider the results for the excess log return rates. In this case, the ordinary estimate for the Sharpe ratio is -0.004. A better estimate is Hedges UMVUE, which is -0.003. The difference will vanish if the sample size increases. Note that the Sharpe ratio is slightly negative according to both estimates. This reflects nicely the fact that the last account value is somewhat less than what could be obtained by the risk-free rate. See remarks 3 and 8.

Note that this Sharpe ratio is also extremely unreliable. This is caused by the small sample size. The p-value is 0.5, which is well above the ordinary level of significance of 5%. So the Sharpe ratio is not significantly larger than 0 – which should not surprise you given the fact that the sample value is negative. However, the confidence interval ranges from -12.485 to 12.478. So the 'true' Sharpe ratio might be as low as -12.485 or as high as 12.478. Given the fact that the assumptions may be violated, the p-value and the confidence interval should be interpreted as no more than crude estimates. See remark 6. Note that in this case, the assumption is that the return rates have a *lognormal* distribution.

Many people compute the Sharpe ratio not on the return rates but rather on the account values directly. I believe that this is an error if the profits are compounded, and therefore the program will not produce these outcomes. Nevertheless, to make the difference clear, consider how the computations would be in this example. Then the computations would be like this:

Table 4

Account value	5	2	5	6	7	3	8	9	10	5
Value of risk-free investment	5.000	5.001	5.001	5.002	5.003	5.003	5.004	5.005	5.005	5.006
Excess account value		-3.001	-0.001	0.998	1.997	-2.003	2.996	3.995	4.995	-0.006

Table 5

	Mean	*Mean * 365*	*SD*	*SD * sqrt(365)*	*Mean / SD*	*sqrt(365) * Mean / SD*
Account value	6.111	2230.556	2.667	50.947	2.292	43.782
Value of risk-free investment	5.003	1826.220	0.002	0.035	2731.681	52188.689
Excess account value	1.108	404.335	2.666	50.926	0.416	7.940

So according to these computations the Sharpe ratio would be 7.940, which is different from the results based on the return rates and the log return rates.

Finally, it is worthwhile to recognize that the Sharpe ratio does not depend on the order of the return rates. For example, consider the account values of Table 6. Here, the initial account value and the last account value are the same as in Table 1. The return rates are also the same as in Table 1, but their order is different. As a result, the maximum draw down is larger (91% versus 50%), which is caused by a series of consecutive losses from 58.3 to 5. The Sharpe ratio is is insensitive to this order, and so it is exactly the same as in the example of Table 1, for both the return rates (5.334) and the log return rates (-0.004). Computed on the account values directly it is even larger in this example (20.624) despite the larger draw down.

Table 6

Account value	5.000	12.500	15.000	17.500	46.667	52.500	58.333	23.333	10.000	5.000
Value of risk-free investment	5.000	5.001	5.001	5.002	5.003	5.003	5.004	5.005	5.005	5.006
Excess account value		7.499	9.999	12.498	41.664	47.497	53.329	18.329	4.995	-0.006
Return rate		2.500	1.200	1.167	2.667	1.125	1.111	0.400	0.429	0.500
Log return rate		0.916	0.182	0.154	0.981	0.118	0.105	-0.916	-0.847	-0.693
Excess return rate		1.500	0.200	0.167	1.667	0.125	0.111	-0.600	-0.572	-0.500
Excess log return rate		0.916	0.182	0.154	0.981	0.118	0.105	-0.916	-0.847	-0.693

Table 7

	Mean	Mean * 365	SD	SD * sqrt(365)	Mean / SD	sqrt(365) * Mean / SD
Account value	26.759	9767.130	20.153	385.027	1.328	25.367
Value of risk-free investment	5.003	1826.220	0.002	0.035	2731.681	52188.689
Excess account value	21.756	7940.909	20.153	385.029	1.080	20.624
Return rate	1.233	450.086	0.834	15.943	1.478	28.231
Log return rate	0.000	0.000	0.700	13.376	0.000	0.000
Excess return rate	0.233	85.037	0.834	15.943	0.279	5.334
Excess log return rate	0.000	-0.049	0.700	13.376	0.000	-0.004

Statistics related to the Sortino ratio

Theory

As was noted in the theory of the Sharpe ratio, the Sharpe ratio uses the standard deviation as risk measure, but this is a problem is the distribution of returns is skewed. For example, if there many small losses but a few large profits, then these profits will lead to overestimation of the risk and to a lower Sharpe ratio. For this reason Sortino (Sortino & Van der Meer, 1991; Sortino & Price, 1994) introduced a version of the Sharpe ratio that uses only the *downside risk* in the denominator. Sortino assumes that the investor has specified a private Minimum Acceptable Return (MAR). This can be the risk-free return, but investor may as well choose another level. Only returns that fall below this threshold are considered risk. Therefore, the adapted standard deviation is computed on shortfalls of the returns relative to the MAR:

$$\text{Shortfall}_i = R_i - MAR \text{ if } R_i < MAR$$
$$\text{Shortfall}_i = 0 \text{ if } R_i \geq MAR$$

Next, the downside deviation is obtained as the average of the squared shortfalls. If M is the average return, and D^- is the downside deviation, then the Sortino ratio is M / D^-. Most of the theory that has been discussed previously in the context of the Sharpe ratio applies equally well to the Sortino ratio, provided that you replace 'risk' by 'downside risk' and 'standard deviation' by 'downside deviation' everywhere. So this theory will not be repeated here.

Definition

The results in the output are all annualized. Consider first the non-annualized statistics. The mean and the standard deviation are computed on either the excess return rates or the excess log return rates. Denote these by X_i, and let n be the number of rates. The upside part of the mean is $m^+ = \Sigma\{X_i \mid X_i \geq 0\}/n$. The numerator is the sum of the nonnegative values X_i. Similarly, the downside part of the mean is $m^- = \Sigma\{X_i \mid X_i < 0\}/n$. The numerator is the sum of the negative values of X_i. The upside part of the second moment is $v^+ = \Sigma\{X_i^2 \mid X_i \geq 0\}/n$ and the downside part of the second moment is $v^- = \Sigma\{X_i \mid X_i < 0\}/n$. So the first and the second moment are $m = m^+ + m^-$ and $v = v^+ + v^-$ respectively. Note that the variance is not v but rather $v - m^2$. The upside SD and the downside SD are then computed as $s^+ = \sqrt{v^+}$ and $s^- = \sqrt{v^-}$ respectively. The Sortino ratio is computed as m / s^- and the Upside Potential Ratio is computed as m^- / s^-. Finally, these values are annualized: If there are T periods in a year, then the partial means are multiplied by T, the partial standard deviations are multiplied by $\sqrt{T}$ and the ratios are multiplied by $\sqrt{T}$.

Remarks

1. Note that the statistics related to the Sortino ratio are computed twice for each time period and span: Once on the excess return rates and once on the excess log return rates. The results are not reported adjacently, but are separated by the statistics about the regression statistics and the Sharpe ratio. See the discussion of the Sharpe ratio statistics to understand the difference between analyses of the return rates versus the log return rates.

2. Sortino defined his coefficients with respect to the Minimum Acceptable Return, and the risk-free rate of return is substituted for that here.

3. Sortino currently advocates the use of the Upside Potential Ratio rather than the index known as the "Sortino ratio" (see www.sortino.com).

4. See the remarks about annualization of the Sharpe ratio statistics. The same reasoning is used here.

Example

In the example the only negative excess return rates are -0,600; -0,572; and -0,500. So the downside part of the mean is [(-0.600 -0.572 -0.500) / 9] * 365 = -67.802. Similarly, the downside SD is $\sqrt{365} * \sqrt{[(0.600^2 + 0.572^2 + 0.500^2) / 9]} = 6.164$. The other results are displayed in Table 8. Here you can see the same pattern as for the Sharpe ratio: The Sortino ratio is positive when based on the return rates, but slightly negative when based on the log return rates. The Upside Potential Ratio, however, is positive in both cases. It cannot be negative because the numerator involves only the profits and not the losses.

Table 8

Based on excess return rates	*Based on excess log return rates*
Sortino ratio 13.795	Sortino ratio -0.005
Upside Potential Ratio 24.794	Upside Potential Ratio 10.954
Upside part of mean 152.839	Upside part of mean 99.602
Downside part of mean -67.802	Downside part of mean -99.651
Upside SD 14.413	Upside SD 8.739
Downside SD 6.164	Downside SD 9.093
N nonnegative terms 6.000	N nonnegative terms 6.000
N negative terms 3.000	N negative terms 3.000

Statistics related to the linear regression on a benchmark

Theory

The *Capital Asset Pricing Model* (*CAPM*) assumes, among other things, that the excess return rates of the portfolio can be predicted from those of the benchmark by

$$ER_i = \alpha + \beta(EB_i) + \varepsilon_i \qquad \text{(CAPM)}$$

where ER_i is the systems excess return rate and EB_i is the benchmark excess return rate (e.g. Bartholdy & Peare, 2003; Sharpe, 1991; www.mathworks.com). The ε_i are the errors of prediction, which are assumed to be normally distributed with expectation 0 and independent of EB_i. The model implies that in a plot of ER_i and EB_i, the points are distributed around a straight line with slope β and intercept α. The slope β is interpreted as the system's risk relative to the benchmark, and the intercept α is interpreted as the return in excess of a benchmark portfolio with the same risk.

The original, stricter form of the CAPM implies moreover that $\alpha = 0$. Rewriting the equation then yields

$$R_i = r + \beta(B_i - r) + \varepsilon_i. \qquad \text{(strict CAPM)}$$

If E denotes the expected value (i.e. the mean) then this in turn implies

$$E(R_i) = r + \beta(E(B_i) - r).$$

Many internet sites describe the model by the last equation, which is also given in Sharpe (1966). This equation is much weaker than the strict CAPM because it pertains only to the means while both preceding equations to the individual returns too. However, the last equation is not sufficient to estimate the parameters. For any application one must assume the stronger equations, which are the equations that are actually used by Sharpe. Therefore it is more appropriate to consider these as the characterizing equations of the model

Under the CAPM the variance of the returns will be $\beta^2\sigma_B^2 + \sigma_\varepsilon^2$, where σ_B^2 is the variance of the returns of the benchmark and σ_ε^2 is the variance of the error in the predictions. The variance of the benchmark is caused by variation in the general economic activity and cannot be diversified away. It is common to all portfolios and so it is no basis to distinguish them. *The coefficient* β *(or rather* $|\beta|$*) reflects how sensitive the portfolio is to variations in the benchmark,* and is called the *systematic risk*. The errors ε_i are by definition not correlated with variations in the benchmark and therefore the term σ_ε is called the *unsystematic risk*. For example, using more leverage will cause a higher systematic risk, while random entries and exits as opposed to a fixed diversified position will cause higher unsystematic risk.

The *Treynor Index* (Sharpe, 1966) is defined as the expected excess return divided by the systematic risk β. So it is conceptually similar to the Sharpe ratio as a reward to risk measure. Sharpe found a correlation of 0.974 between the ranks on basis of the Sharpe ratio and the ranks based on the Treynor ratio for 34 mutual funds in the period 1954 -1963, where the Dow Jones Industrial Average was used as benchmark.

The strict CAPM implies that $\alpha = 0$ because it assumes that all investors have an efficient portfolio, i.e. that no one has a market advantage. That is, according to the CAPM all investors would essentially have the same optimal portfolio, only with different risk positions (leverage). Obviously this is not exactly true in reality, and so α can actually be positive. *A positive value of* α *implies that the portfolio has a market advantage above a benchmark portfolio with the same systematic risk* β. Such a benchmark portfolio would have returns βB_i while the portfolios returns are $\beta B_i + \alpha$ plus a random error component ε_i with mean 0.

In the analysis of the Sharpe ratio it was concluded that the compounding long-run investor is better served with an analysis based on the log return rates, and the same arguments apply here. For this reason the analysis is also done on log return rates.

The choice of the benchmark has a large effect on the parameters. A simple possibility is to define the benchmark as a pre-existing market index. For example, Sharpe (1966) uses the DJIA as benchmark for a selection of mutual funds. More sophisticated methods use factor analysis to estimate the weight of each asset in the benchmark empirically instead of using the a priori eights of a pre-existing index. This method requires the input of the data of many funds simultaneously, which falls outside the scope of this project.

Definition

The results in the output are all annualized. Consider first the non-annualized statistics. The mean and the standard deviation are computed on either the excess return rates or the excess log return rates. Denote these by Y_i, and let n be the number of rates. Similarly the excess return rates or the excess log return rates are computed for an array with benchmark values. Denote these by X_i. An ordinary least square linear regression analysis is conducted where Y (the criterion) is predicted from X (the predictor). This yields the estimates a for α, b for β and r for ρ. Specifically, the sample covariance between X and Y is $S_{XY} = \sum (X - \overline{X})(Y - \overline{Y}) \big/ (n-1)$. Denote the standard deviations S_X and S_Y. Then the correlation is $r = S_{XY} / (S_X S_Y)$, the slope is $b = S_{XY} / S_X^2$ and the intercept is $a = \overline{Y} - b\overline{X}$. The degrees of freedom for the error is $dfe = n - 2$, and the *Mean Squared Error* (*MSE*) is $S_Y^2 (1 - r^2) / dfe$. The standard error of a is $s(a) = (MSE\ (1/n + \overline{X}^2 / ((n-1) S_Y^2)))^{1/2}$ and the standard error of b is $s(b) = (MSE / (S_Y^2 (n-1)))^{1/2}$. (e.g. Neter et al., 1996). The t-values are $t(a) = a / s(a)$ and $t(b) = b / s(b)$. The p-values $p(a)$ and $p(b)$ are p-values for the one-sided test of the hypothesis that the corresponding parameter is larger than 0, on basis of ordinary normal theory. Jensen's alpha is just a, and the Treynor ratio is $\overline{Y} / b$.

Finally, these results are annualized in this way: If there are T periods in a year, then the means, the intercept, the covariance and MSE are multiplied by T, the standard deviations are multiplied by $\sqrt{T}$. The other values (r, b, $t(a)$, $t(b)$, $p(a)$, and $p(b)$) are not annualized.

Remarks

1. Note that the statistics related to the linear regression are computed twice for each time period and span: Once on the excess return rates and once on the excess log return rates. The results are not reported adjacently, but are separated by the statistics about the Sharpe ratio and the Sortino ratio. See the discussion of the Sharpe ratio statistics to understand the difference between analyses of the return rates versus the log return rates.

2. The value of this analysis depends for a large part on the data that are used as benchmark. This is not under influence of the program; the benchmark data are provided by the calling program of C2. For now, the decision was made to use the S&P as benchmark for all systems, even though it is clear that more relevant benchmarks are possible for specific systems.

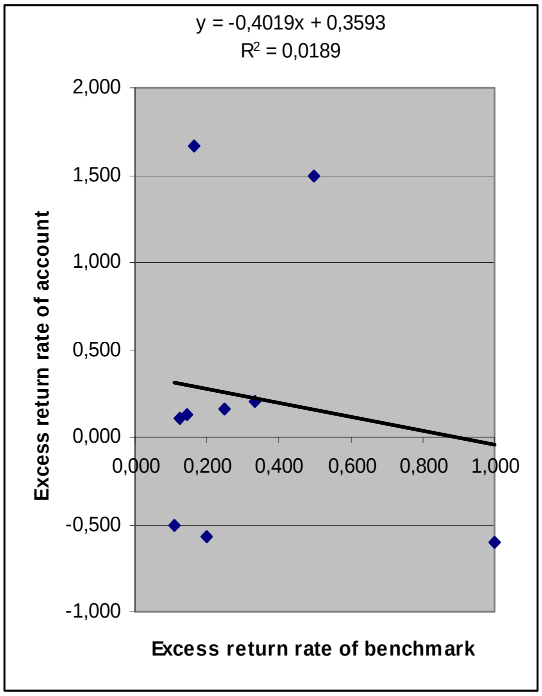
3. See the remarks about annualization of the Sharpe ratio statistics. The same reasoning is used here.

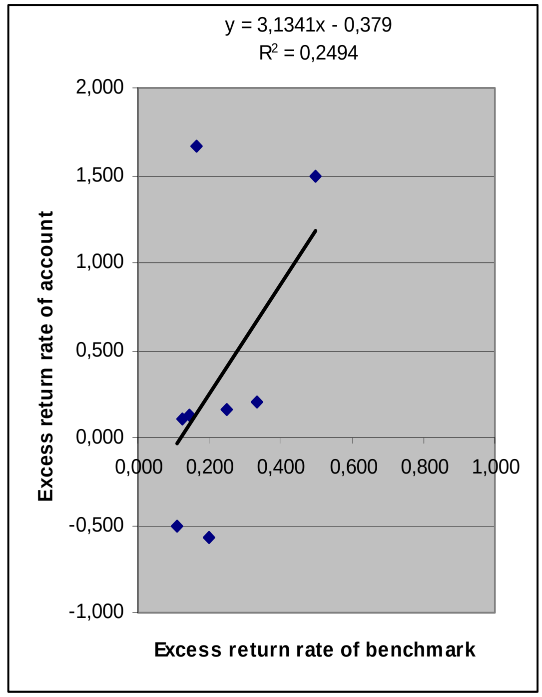
4. The *p*-values and the confidence intervals are computed on basis of the ordinary assumptions for linear regression. These imply that the excess (log) return rates should not have a serial correlation, that *Y* depends on *X* by a linear relation, and that $Y|X$ has a normal distribution for each value of *X* (but not necessarily that *Y* and *X* are normally distributed). These assumptions are likely to be violated. Nevertheless, the *p*-values and the confidence intervals give you a crude estimate about the size of the unreliability of the estimates.

5. In the computation of the Treynor ratio, *b* is used in the denominator even if *b* is negative. So a fund that has a continuous short position in the benchmark during a bull market will have a negative mean excess return and a negative *b*, yielding a positive Treynor index. Perhaps it would be more logical to use $|b|$ in that case, but I found no indication that this is commonly done. So I left it as it is, but I recommend the user to always change the sign into the sign of the mean of the criterion.

Example

Suppose the account values are the same as in Table 1 and the values of the benchmark are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. So the excess return rates of the benchmark are 1, 0.5, … The left figure below shows how the excess return rates of the account are related to those of the benchmark. This figure, including the regression equation above it, is obtained by Excel. The slope is b = -0.4019 and the intercept is a = 0.3593. Note that the regression line is heavily influenced by the outlier right below. Without this observation (the first one) we would get the figure on the right, with b = 3.134, a = -0.379 and r^2 = 0.249.





The output of the program is shown in Table 9. Here, the slope is the same as in the left figure, but the intercept is annualized (0.3593 * 365 = 131.128). The correlation is the same, but squared in the figure ($-0.137^2 = 0.0189$).

Table 9

Based on excess return rates	*Based on excess log return rates*
N of observations 9.000	N of observations 9.000
Mean of predictor 114.682	Mean of predictor 93.334
Mean of criterion 85.037	Mean of criterion -0.049
SD of predictor 5.448	SD of predictor 3.626
SD of criterion 15.943	SD of criterion 13.376
Covariance -11.929	Covariance -10.262
r -0.137	r -0.212
b (slope, estimate of beta) -0.402	b (slope, estimate of beta) -0.781
a (intercept, estimate of alpha) 131.128	a (intercept, estimate of alpha) 72.816
Mean Square Error 285.008	Mean Square Error 195.324
DF error 7.000	DF error 7.000
t(b) -0.367	t(b) -0.573
p(b) 0.638	p(b) 0.708
t(a) 0.793	t(a) 0.469
p(a) 0.227	p(a) 0.327
Lowerbound of 95% confidence interval for beta -2.993	Lowerbound of 95% confidence interval for beta -4.003
Upperbound of 95% confidence interval for beta 2.189	Upperbound of 95% confidence interval for beta 2.442
Lowerbound of 95% confidence interval for alpha -259.895	Lowerbound of 95% confidence interval for alpha -294.284
Upperbound of 95% confidence interval for	Upperbound of 95% confidence interval for

alpha 522.151 Treynor index (mean / b) -211.587 Jensen alpha (a) 131.128	alpha 439.916 Treynor index (mean / b) 0.062 Jensen alpha (a) 72.816

The statistics *r*, *b* and *a* are the most essential ones in this part of the output. The correlation *r* says how strongly the performance of the system is related to the performance of the benchmark. This is expressed on a scale between -1 and 1, where 1 indicates that the systems account values always changed in the same direction as the benchmark and -1 indicates that they always changed in opposite directions. Although the correlation measures the strength of the relation, it does not say anything about the *cause* of the relation – as you can learn in every introductory statistics books. So a high correlation with the benchmark does not imply that the return rates of the system were caused by a long position in the benchmark, but it says that they could be mimicked by a long position in the benchmark.

The slope *b* is an estimate for β, which is the systematic risk that reflects how sensitive the portfolio is to variations in the benchmark. (If *b* is negative, then one should consider $|b|$). It is influenced by the average leverage of the portfolio. It is in fact the same as *r*, but on a different scale: It is *r* multiplied by the ratio of the standard deviations (which can be viewed as measures of the volatilities). So the causal limitation of *r* applies to *b* as well.

The intercept *a* is an estimate for α which is the average return in excess of the average return of a position in the benchmark with the same systematic risk. In causal terms you might be tempted to think that this is the part of the system's performance that is independent of the benchmark, caused by the fund manager's ability to outperform the market. Obviously, since *a* is computed via *b*, this interpretation in subject to the same limitations.

According to the CAPM it is desirable to have a low value of $|b|$ and a large value of *a*. Even outside the CAPM this can be considered as a reasonable rule of thumb, but the conclusions are less decisive. For example, you cannot infer from a high value of *r* and *b* that the systems performance was caused by the performance of the benchmark, but you can sometimes do the opposite – infer from the trade style what the values of *r* and *b* should be. For example:

- For a stock system that holds a continuous long position in the benchmark with its entire capital but no leverage, you will get $r = 1$, $b = 1$ and $a = 0$.
- For a system with a continuous long position in the benchmark with leverage 2, and no other positions, you will find $r = 1$, $b = 2$ and $a = 0$.
- For a system that at random times opens and closes a long position in the benchmark with leverage 2, you can expect $0 < r < 1$, $0 < b < 2$ and $a \approx 0$.
- For a system that at random times opens and closes randomly a long or a short position in the benchmark, you can expect $r \approx 0$, $b \approx 0$ and $a \approx 0$.

So if $b \approx 0$ you can infer that the first three scenarios didn't happen, and if $a > 0$ you can infer that none of the above scenarios happened.

So a small value of *b* combined with a high value of *a* is perhaps a good sign, but another pattern is not necessarily a bad sign. One reason is that the correlation can depend strongly on the time periods and span that are analyzed. For example, it is in theory possible that $r = b = 1$ and $a = 0$ when you analyze monthly data, but $r = b = 0$ and $a > 0$ when you analyze daily data. This can happen if the benchmark has a long term bull market with intra-month draw downs, while the system exploits both these intra-month draw downs (by taking short positions) and their recovery (by taking long positions), and no positions in the rest of the month. In this example the benchmarks monthly profits are caused by the days outside the draw-down-and-recovery periods, whereas the systems profits are made within these periods.

The same paradox can happen with daily data versus intra-day data, but you will not know that because intra-day data are not analyzed.

In the above output example, if we consider the analysis based on the excess log return rates, the first impression would be that the system has a negative correlation (r = -0.212) with the benchmark, i.e. if the benchmark has a positive return then the system tends to have a negative return. However, the correlation is close to 0, so the relation is not very strong. Moreover, the confidence interval of b contains 0; since the signs of b and r are always the same, this implies that the confidence interval of r must contain 0 too. So the correlation is not significantly different from 0. The estimated slope (b = -0.781) suggests that the system's performance can partially be predicted as a 78.1% short position in the benchmark. This prediction is not very accurate, however, since the correlation is close to 0. Similarly, the estimate a of the intercept suggests that the system also has a profitable part that is in excess of the performance of the benchmark, but a look at the confidence interval for alpha makes clear that we actually don't know very much about alpha.

Jensen alpha is equal to the intercept a. The Treynor ratio is similar to the Sharpe ratio, with the denominator replaced by the systematic risk b. It is negative here because the mean excess return is positive while b is negative. I suggest to interpret it as a positive value in these cases.

Risk estimates for a one-period unit investment (parametric)

Theory – Definition of VaR and Expected Shortfall

There are many statistics that are used to measure risk:

- The standard deviation is used as risk measure in the mean-variance paradigm and the Sharpe ratio.
- The downside standard deviation is used as risk measure in the Sortino ratio and the Upside Potential Ratio
- The maximum draw down and the maximum loss are used as risk measure by many investors
- Value-at-risk (VaR) is used as risk measure by brokers and institutions. The value-at-risk at confidence level 95% is defined as the value such that future values of the positions fall below it with a probability of at most 5%.

Artzner, Delbaen, Eber, and Heath (1997; 1999) initiated a more fundamental approach to risk. Instead of defending some ad hoc measure they stated a set of axioms that specify which properties a risk measure should have in order to base rational decisions on it. These properties are:

> *Monotonicity:* If one position has always larger values than another, then the risk of the former is not larger than the risk of the latter.
> *Subadditivity:* The risk of a diversified position is not larger than the sum of the risks of its components.
> *Positive homogeneity:* If the size of a position if multiplied by any positive factor then its risk is multiplied by the same factor.
> *Tranlation invariance:* Adding an amount α to the position and investing it in the risk-free instrument decreases the risk by α.

A measure that satisfies all four axioms is called a *coherent risk measure*.

Artzner, Delbaen, Eber, and Heath (1999) show that the standard deviation and value-at-risk are not coherent risk measures. The *expected shortfall*, on the other hand, is a coherent risk measure under very general conditions (Acerbi & Tasche, 2002, proposition 3.1). See Yamai and Yoshiba (2002) for a very readable introduction to VaR, expected shortfall and coherence. This document will not try to improve their presentation.

Expected shortfall (ES) is defined as the conditional expectation (i.e. the mean value) of the position, given that it falls at or below VaR:

$$ES^{(\alpha)}(X) = -E(X \mid X \leq VaR(\alpha))$$

For example, if the 5% smallest returns have a mean of -20% then the expected shortfall at level 95% is 20%. This definition is valid for continuous distributions only. In non-continuous distributions it can happen that the probability of $X \leq VaR(\alpha)$ is somewhat larger than $1 - \alpha$, and then a small adjustment has to be made to the formula of the expected shortfall (Acerbi &

Tasche, 2002, definition 2.6). In that case the outcome of the above formula is called the *tail conditional expectation* while the adjusted version is called the expected shortfall. Conceptually, however, expected shortfall and tail conditional expectation are the same.

Some authors (notably Acerbi & Tasche, 2002) use the opposite notation VaR(5%) and ES(5%) for what is denoted as VaR(95%) and ES(95%) here.

Theory – Estimation of VaR and Expected Shortfall

The definitions of VaR and ES pertain in principle to future values of the position, but their estimates can only be obtained from past performances. One can distinguish three methods to do so:

1. *Empirical estimation.* In principle, one could estimate VaR(95%) by determining the 5% smallest returns and ES(95%) as their mean. This method is used in the sections Quartiles of return rates and Quartiles of draw downs. However, because the number of observations in the tail can be small, it is not uncommon to try to improve the estimates on basis of theoretical considerations.
2. *Non-parametric estimation.* This method uses the statistical properties of extreme values and the observations in the tail of the empirical distribution. This method is used in the section Risk estimates for a one-period unit investment (based on Extreme Value Theory) and Risk estimates based on draw downs (based on Extreme Value Theory).
3. Parametric estimation. This method assumes that the returns have specific distribution, e.g. a lognormal, and uses all observations to estimate the parameters of that distribution. Next, VaR and ES are computed on basis of the estimated distribution. This method is used in the present section.

Within these methods one can further distinguish between risk estimates which are based on the one-period returns, and risk estimates based on draw downs. A draw down is a peak-to-valley loss in the equity curve that is caused by the accumulation of losses over possibly many periods. The present section deals only with one-period returns and losses.

So this section deals with the parametric estimation of VaR and Expected Shortfall of one-period returns. Within this section there is further a distinction between two different assumptions about the distribution of the returns. The first possible assumption is that the return rates have a lognormal distribution. In the computation of the Sharpe ratio the mean and the variance of the log return rates were estimated, and these are used to compute VaR and ES. So these estimates are influenced by the size of both the positive and the negative returns. Note that the ensuing risk estimates pertain to the returns, not to the log return rates.

The second possible assumption is that the losses have a generalized Pareto distribution. This can be motivated by the fact that extreme values must approximately have a generalized Pareto distribution (e.g., McNeil & Frey, 2000). In the computation of the Sortino ratio the mean and the second moment of the losses were estimated, and these are used to compute VaR and ES. So these estimates are influenced by the frequency and the size of the losses only. These risk estimates also pertain to the returns, not to the log returns.

In the theory of the Sharpe ratio it was argued that the compounding trader is better served with considering the log return rates. The program does not report VaR and ES for them. If you want risk estimates for the excess log return rates under the assumption of a lognormal distribution for the return rates, you can compute them from the mean and standard deviation that are reported in the Sharpe ratio for the excess log return rates. First deannualize them: If there are T periods in a year then the de-annualized mean is M = *reported mean* / T and the de-annualized standard deviation is S = *reported standard deviation* / $\sqrt{T}$. Next,

$$\text{VaR}(95\%) = M - 1.645\, S, \quad \text{and}$$

$$\text{ES}(95\%) = 2.063\, S - M.$$

For 99% confidence, replace the coefficients by 2.326 and 2.665 respectively, and for 99.9% confidence replace them by 3.090 and 3.367 respectively.

Definition

In the calculation of the Sharpe ratio statistics the sample mean M and sample standard deviation S of the excess log return rates EL were obtained. Assuming that the EL are independent normally distributed, the 5% percentile of EL can be estimated as $q = M + z\, S$, where z is the 5% percentile of a standard normal distribution ($z \approx -1.645$). Next, the VaR(95%) is computed as $1 - \exp(q)$. So VaR(95%) = 0.70 means that on a randomly chosen period there is probability of 5% to have a loss of 70% or more. The Expected Shortfall on VaR is computed via the partial moments of a lognormal distribution with parameters M and S.

In the calculation of the Sortino ratio statistics from the excess return rates, the downside mean m^- and the downside standard deviation s^- were obtained. From these the downside variance of the losses is computed as $s^-s^- - m^-m^-$. The downside mean and the downside variance are then used to obtain the moment estimators of a generalized Pareto distribution (Hosking & Wallis, 1987, p. 341). The Expected Shortfall is computed as the conditional expectation in that generalized Pareto distribution (e.g., McNeil & Frey, 2000, formula 16).

Remarks

1. These methods are parametric in that they assume a specific distribution for the returns. Some nonparametric risk estimation methods are used elsewhere in the output.

2. The first method is based on all return rates – including both profits and losses. So the empirical distribution of the profits can affect the estimates for the expected losses. The second method is based on the percentage and the size of the losses only, but not on the size of the profits.

3. The assumption of lognormal return rates is not uncommon in financial risk models. There are however indications that losses can better be described by a generalized extreme value distribution (LeBaron & Samanta, 2005). The tail of a generalized extreme value distribution is approximately a generalized Pareto distribution according to the Gnedenko-Pickands-Balkema-de Haan theorem (see e.g. Pisarenko and Sornette, 2003; Fernandez, 2003). However, the second method used here assumes that a generalized Pareto distribution applies to all losses, not only the extreme losses.

4. The moments method for the generalized Pareto distribution is not optimal, but used to minimize the computational burden.

5. Although risk is conventionally measured by VaR, the recent theories give more attention to Expected Shortfall because it is a coherent risk measure. See Yamai and Yoshiba (2002) for an introduction.

Example

For the example data of Table 1 the output for the parametric risk estimates is shown in Table 10.

Table 10

```
--assuming log normal returns and losses (using central moments from Sharpe statistics)
  VaR(95%)   0.684
  Expected Shortfall on VaR   0.757
--assuming Pareto losses only (using partial moments from Sortino statistics)
  VaR(95%)   0.338
  Expected Shortfall on VaR   0.638
```

Based on lognormal theory, one would can reasonably expect losses up to 68.4% in one period. The mean of the losses that exceed this bound is estimated to be 75.7%. Based on a generalized Pareto distribution, one would expect these numbers to be 33.8% and 63.8% respectively. Note that no confidence intervals are given, so even if you know which model to use, you won't know how reliable these percentages are. The reliability increases with the number of observations, which in case of the second method means that *more losses* will yield more reliable estimates.

Quartiles of return rates

Theory

This section of the output provides a description of the empirical distribution of the return rates, without much theoretical inference. In principle this could be done by drawing a histogram of the return rates, but this would make it difficult to compare different systems. Instead the *quartiles* of the distribution are computed. The first quartile is the value below which 25% of the return rates fall; the second quartile is the value below which 50% of the return rates fall; the third quartile is the value below which 75% of the return rates fall. In addition the minimum and the maximum of the return rates are reported. (These definitions of the quartiles are only valid for a continuous distribution. In a finite sample they are not always uniquely defined, and then one of the commonly applied adaptations is used.)

The difference between the first and the third quartile is called the inter quartile range and can be viewed as a measure of dispersion, like the standard deviation. The quartiles are often used to define outliers, and this is also done here.

The quartiles divide the set of return rates in subsets that each contain of 25% of the return rates. These subsets are called quarters below. The means of the quarters and the mean of the outliers are reported too.

The first quartile can be interpreted as an empirical estimate for the value-at-risk at 75% confidence, VaR(75%). Similarly, the mean of the first quarter can be interpreted as an empirical estimate for the expected shortfall at 75% confidence, ES(75%). See the section Risk estimates for a one-period unit investment (parametric) for more theory about these risk measures.

Definition

The return rates are ranked and ordered. The q^{th} quantile is obtained as element $q * n$ in the rank order, with linear interpolation between the adjacent values if $q * n$ is not an integer. This method is used to compute the quartiles of the return rates. Next, the mean is computed for each quarter of the ordered return rates. The Inter Quartile Range (*IQR*) is *Quartile 3 – Quartile 1*. A return rate R_i is an outlier at the low side if $R_i <$ *Quartile 1* – 1.5 *IQR* and an outlier at the high side if $R_i >$ *Quartile 3* + 1.5 *IQR*. The percentage and mean of each type outlier is computed.

Remarks

1. The quartiles are computed for the return rates; not the excess return rates or the log return rates.

2. Quantiles are generally not uniquely defined when the data set is finite. The present method to calculate quantiles produces the same quartiles as Excel, but not SPSS.

3. Note that the first quartile can be viewed as an empirical VaR with a confidence level of 75%. Similarly, the mean of quarter 1 can be viewed as an empirical Expected Shortfall estimate.

Example

Consider the example of Table 1. The return rates are ordered from small to large in the first row of Table 11. The colours show how they are divided into quarters that each contain approximately 25% of the return rates. The second row contains the bounds of these quarters, which are the minimum, the three quartiles, and the maximum. The third row shows the identification number of each quarter. The means of the return rates are shown per quarter in the last row.

Table 11

Return rate	0.4	0.429	0.5	1.111	1.125	1.167	1,2	2,5	2,667
Quartile	Min = 0.4		Q1 = 0.500		Median = 1.125		Q3 = 1.200		Max = 2.667
Quarter	Quarter 1			Quarter 2		Quarter 3		Quarter 4	
Mean of quarter	Mean of quarter 1 = 0.443			Mean of quarter 2 = 1.118		Mean of quarter 3 = 1.183		Mean of quarter 4 = 2.583	

The program returns the following output.

Table 12

Number of observations	9.000
Minimum	0.400
Quartile 1	0.500
Median	1.125
Quartile 3	1.200
Maximum	2.667
Mean of quarter 1	0.443
Mean of quarter 2	1.118
Mean of quarter 3	1.183
Mean of quarter 4	2.583
Inter Quartile Range	0.700
Number outliers low	0.000
Percentage of outliers low	0.000
Mean of outliers low	NA
Number of outliers high	2.000
Percentage of outliers high	0.222
Mean of outliers high	2.583

The 25% lowest return rates range from the minimum 0.400 (= 60% loss) to the first quartile 0.500 (= 50% loss) with a mean of 0.443 (= 55.7% loss). In other words, VaR(75%) = 50% and the corresponding Expected Shortfall is 55.7%. The next 25% lowest return rates range from 0.500 to 1.125 (= 12.5% profit) with a mean of 1.118. Et cetera. The inter quartile range is 0.700, which means that the middle 50% return rates have differences up to 0.700. The inter quartile range is a measure of variation, like the standard deviation (in a normal distribution, the inter quartile range is always 1.35 times the standard deviation). The return rates have no outliers at the low side (at least not in comparison to the other return rates), but they have 2 outliers at the high side. The outliers at the high side have a mean of 2.583 (= 158.3 % profit).

Risk estimates for a one-period unit investment (based on Extreme Value Theory)

Theory

Extreme value theory is concerned with estimating the tails of a distribution. It is often used to model extreme events, such as floods, earth quakes, insurance risks and investment risks. The basic reason for a separate theory is that estimates that are based on the centre of the distribution have little predictive value for the tail of the distribution. For example, the daily variations in sea level bear little information about how high the water can rise in a hurricane. People who are aware of this intuitively tend to base their protection on the maximum value that was previously observed, and add a little bit to it. For example, medieval Dutch dyke builders had the strategy to measure after a breach of the dyke how high the flood has been, and then built the next dyke a little higher. Similarly, private investors tend to assess their risk by finding the maximum draw down in the past, and then add a little to it.

Such intuitive assessments have two limitations. The major limitation is that they do not recognize the fact that some distribution have 'fat' or 'heavy' tails, whereas others have 'light' tails. A light tail decreases at an exponential rate, while a fat tail decreases much slower – making extreme events more likely. The second limitation is that they use only the maximum value (e.g. the flood that breached the last dyke) and not the other observations in the tail (e.g. the floods that breached the previous, lower dykes). These two limitations are related, because the maximum value alone is insufficient to estimate how fast the tail decays.

Extreme value theory attempts to overcome these limitations by estimating the heaviness of the tail on basis of the ratios between the values in the tail. For this purpose the values above a certain threshold are ordered from small to large. If the spacings are large and decay slowly, this indicates a heavy tail. In that case the next maximum is probably not just a little bit larger than the current one, but much larger. This is associated with risk levels well above intuitive estimates.

The first major theorem of extreme value theory is the Fisher-Tippett (1928) theorem. It states that if the distribution of the suitable rescaled maximum of samples converges, then it must converge to a distribution of this form:

$$H(x) = \exp(-(1 + \tau x)^{-1/\tau}, \qquad \tau \neq 0,\ 1 + \tau x > 0$$
$$H(x) = \exp(-\exp(-x)), \qquad \tau = 0.$$

This distribution is called the Generalized Extreme Value Distribution (GEVD). The second major theorem of extreme value theory is the Gnedenko-Pickands-Balkema-de Haan theorem (Pickands, 1975; Balkema & de Haan, 1974), which states that a distribution satisfies the conditions of the Fisher-Tippett theorem if and only if its distribution has tails that are approximately a generalized Pareto distribution with the same shape parameter. That is, for some threshold u, consider the conditional distribution of the excesses over u, given that the variable exceeds u. Then, when the threshold u is raised, this conditional distribution converges to a distribution of the form

$$G(x) = 1 - (1 + \tau x / s)^{-1/\tau}, \qquad \tau \neq 0,\ s > 0,\ s \text{ depending on } u$$
$$G(x) = \exp(-x / s), \qquad \tau = 0,\ s > 0,\ s \text{ depending on } u$$

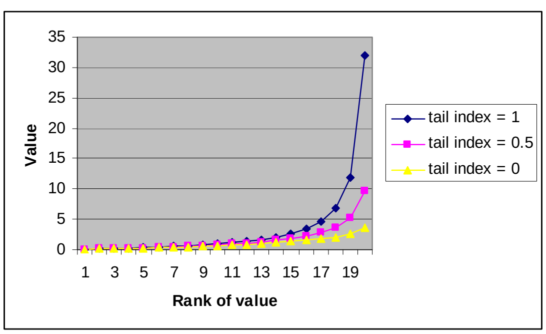
with the same τ as in the GEVD to which the normalized sample maximum converges. A distribution of this form is called a Generalized Pareto Distribution (GPD). Many distributions satisfy the condition, including the normal, lognormal, Students t, and loggamma distributions.

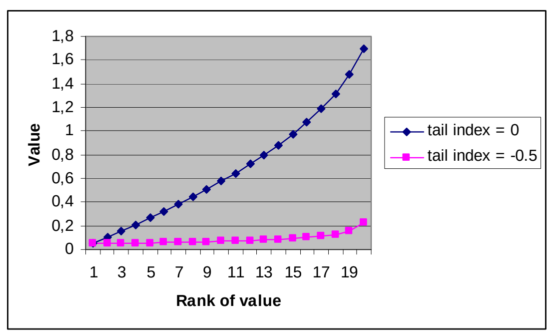
The parameter τ is called is called the *Extreme Value Index* (EVI) or *tail index*. It indicates how heavy the tail is:

- If $\tau > 0$ then the tail decays according to a power law. Such tails are called *heavy*, and the distribution is said to belong to the *Fréchet domain of maximum attraction*. Examples are the Student t, loggamma and F distributions. If $\tau > 0.5$ then their variance is infinite. If $\tau > 1$ then their mean is also infinite.
- If $\tau = 0$ then the tail decays exponentially fast. Such tails are called *light*, and the distribution is said to belong to the *Gumbel domain of maximum attraction*. Examples are the normal and the lognormal distribution.
- If $\tau < 0$ then the tail is bounded. Such distributions are said to belong to the *reverse Weibull domain of maximum attraction*. Examples are the uniform and the beta distribution.

The tail index is usually estimated on basis of the observations that exceed a certain threshold. These estimators are called Peak-over-Threshold (POT) estimators. The threshold is usually chosen as the 75th or the 90th percentile. There are many estimation methods designed. One of the estimation methods used in the program is to order the values that exceed the threshold and compute the differences between the subsequent values. The logarithm of the spacings has approximately a linear regression on the logarithm of their reversed ranks (where the largest value has rank 1, the 2nd largest value rank 2, etc) and the slope of this regression is $-\tau$ (Matthys & Beirlant, 2003). Once the tail index is estimated, it can be used to estimate the VaR and expected shortfall on basis of the GPD.

This estimation method can be used to illustrate the effect of the tail index. The figures below show what kind of relation can be expected between the rank of the 20 largest observations and their value, assuming that the threshold is 0. Note that the figures are on very different scales.





It is also informative to consider what can be expected for the difference between the largest and the second largest value, divided by the difference between the second largest and the third largest. For $\tau = -0.5$, 0, 0.5 and 1 the ratio of these differences is 1.414, 2, 2.828 and 4 respectively. So the differences between the maxima increase much faster if $\tau = 1$ then if $\tau = 0$. For example, if the current maximum is the previous maximum plus 0.2, then the next maximum can be expected to be the current maximum plus 0.3 if $\tau = 0$, but the current maximum plus 0.8 if $\tau = 1$.

More information and examples of extreme value theory in finance can be found in Fernandez (2003), Giesecke, Schmidt and Weber (2005), McNeil (1997, 1999), LeBaron and Samanta (2005).

Definition

First, the loss rates are computed as the complement of the return rates: $\overline{R_i} = 1 - R_i$. The tail index is estimated on basis of the 25% largest loss rates. This is done by two methods. The first method (called the "moments method" here) is the moment estimator of Dekkers, Einmahl and de Haan (1989). This is a Hill estimator with a bias correction. The second method (called the "regression method" here) is based on the second exponential regression model of Matthys and Beirlant (2003). However, this is transformed to an ordinary least squares linear regression via a log transformation.

Remarks

1. These methods are nonparametric in that they do not assume a specific distribution for the returns. Some parametric risk estimates are given elsewhere in the output.

2. The 25% largest losses are used in the estimation. It is generally difficult to assess how much of the tail should be used. If the percentage is too high then the center of the distribution can create a bias in the estimates. If the percentage is too low, then the small number of observations leaves the estimates unreliable (Diebold, Schuermann & Stroughair, 1998; McNeil, 1997). Pickands (1975) recommended 25%, but usually 10% is recommended. Here, 25% is used because the number of observations is often small for C2 systems.

3. The regression method should not be confused with the method to estimate the tail index by regression of the empirical survival function (Diebold, Schuermann & Stroughair, 1998).

Example

Table 13

Extreme Value Index (moments method)	-20.296
VaR(95%) (moments method)	0.583
Expected Shortfall (moments method)	0.583
Extreme Value Index (regression method)	-2.322
VaR(95%) (regression method)	0.655
Expected Shortfall (regression method)	0.657

The Extreme Value Index (also called tail index) indicates how heavy the tail of the distribution of losses is. Distributions with a positive tail index are called heavy-tailed and typically decay as a power function. Examples of this are the t-distribution and the Pareto distribution. A tail index of zero indicates a tail that decays exponentially fast. Examples of this are the normal and the log normal distribution. Distributions with a negative tail index have a finite right end point.

In the example, both estimates of the tail index are negative. This suggests that the losses are bounded, but it doesn't say what the bounds are, so without further information it could be 100%. The moments method indicates that VaR(95%) and the corresponding Expected Shortfall are both a loss of 58.3%. The regression method indicates losses of 65.5% and 65.7% respectively.

Note that there is nothing reported about the reliability of these estimates. The estimates are based on the 25% largest losses, which are only 2 or 3 observations in this example.

Quartiles of draw downs

Theory

All the previous statistics were based on the one-period returns and losses. These are sufficient to describe the profitability and risk if one assumes that subsequent returns are independent. Independence means that the actual returns in the past do not affect the distribution of future returns. For example, the expected return of tomorrow would be the same regardless of whether today's return is a large loss or a large profit. In reality there is often a serial correlation between the return at period i and the return at period $i + 1$ (LeBaron & Samanta, 2005). Common sense also suggests that profits and losses tend to come in clusters.

For this reason it can be important to assess to which extent losses tend to accumulate into larger draw downs. A draw down is defined as the difference between the largest account value up to certain time point (a historic high or peak in the account value curve) and the lowest value after it (a valley). This is not the same as a series of consecutive losses. A draw down can contain multiple series of consecutive losses separated by periods of profits, if these profits are not large enough to produce a new high and a lower value occurs after them. For example, if these are the subsequent changes in the account value ($V_i - V_{i-1}$) on six days,

-1, -1, -1, +1, -1, -1,

then largest series of consecutive losses consists of the first three days whereas the draw down extends to all six days.

It is common practice among traders to consider the maximum draw down as a risk measure. This has certain disadvantages. One reason is that it can only increase as time passes; so it is not a stable statistic. Another reason is that it is only one event and that it does not say how often draw downs of similar size occur.

Another possibility is to consider the mean draw down. The disadvantage of this is that many small draw downs can obscure the occurrence of large draw downs.

So the best is probably to consider the distribution of the draw downs, and for this reason the quartiles of the draw downs are reported, together with the outliers, the means of the quarters and the means of the outliers. See the section Quartiles of return rates for the definition of quartiles. The most interesting of these statistics are the mean of the fourth quarter and the mean of the high outliers.

Definition

The draw downs are defined as follows. Let $V_0, V_1, \ldots, V_n$ be the account values. Let H_i be the historic high up to period i, that is $H_i = \max\{V_j \mid j \leq i\}$. This is necessarily a monotone non-decreasing function. The system is in a draw down or recovering from a draw down if $V_i < H_i$. So define a draw down period as a series of consecutive time points $D = \{d, d + 1, \ldots, d + k\}$ such that $V_i < H_i$ for each $i \in D$ and such that $V_i \geq H_i$ if $i = d - 1$ and if $i = d + k + 1$ or $d + k = n$. Let $m(D) = \min\{V_i \mid i \in D\}$, the minimum account value in D. Then the size of the draw down in D is $(H(d) - m(D)) / H(d)$. In this way the draw downs are defined uniquely. Next, the quartiles, the means and the outliers are computed in the same way as for the one-period return rates.

Remark

1. Unlike all previous statistics, the draw down statistics are sensitive to accumulation of losses. See the example about the order insensitivity of the Sharpe ratio.

Example

Consider first the example that was used to discuss the order insensitivity of the Sharpe ratio, in Table 6. Here, there is only one draw down period of 3 time points.

Table 14

Account value	5	12.5	15	17.5	46.667	52.5	58.333	23.333	10	5
Historic high	5	12.5	15	17.5	46.667	52.5	58.333	58.333	58,333	58,333
Draw down period								1	1	1
Size of draw down										91.4%

Consider the example data of Table 1.

Table 15

Account value	5	2	5	6	7	3	8	9	10	5
Historic high	5	5	5	6	7	7	8	9	10	10
Draw down period		1				2				3
Size of draw down		0.600				0.571				0.500

Here there are three draw down periods, each of one time point. The corresponding output is shown in Table 16. Of course, the division into quartiles is somewhat arbitrary in this cause because there are only three draw downs.

Table 16

Number of observations 3.000
Minimum 0.500
Quartile 1 0.536
Median 0.571
Quartile 3 0.586
Maximum 0.600
Mean of quarter 1 0.500
Mean of quarter 2 0.571
Mean of quarter 3 NA
Mean of quarter 4 0.600
Inter Quartile Range 0.050
Number outliers low 0.000
Percentage of outliers low 0.000
Mean of outliers low NA
Number of outliers high 0.000
Percentage of outliers high 0.000
Mean of outliers high NA

Risk estimates based on draw downs (based on Extreme Value Theory)

Definition

The computation of the draw downs is described in the section Quartiles of draw downs. Next, the nonparametric risk estimates are computed in the same way as described in the section Risk estimates for a one-period unit investment (based on Extreme Value Theory)

Remark

1. The risk statistics are now for an unlimited period, in theory.

Example

Table 17

Extreme Value Index (moments method)	NA
VaR(95%) (moments method)	NA
Expected Shortfall (moments method)	NA
Extreme Value Index (regression method)	NA
VaR(95%) (regression method)	NA
Expected Shortfall (regression method)	NA

There are only 3 draw downs, so the largest 25% of these are insufficient data to estimate the distribution.

Combined statistics

Definition

Suppose that V_0 is the initial account value and V_n the last one, and that there are T periods in a year (i.e. $T = 12$ or $T = 365$). Then the annualized return with arithmetic extrapolation is $(T/n)(V_n / V_0 - 1)$, and the compounded annual return is $(V_n / V_0)^{T/n} - 1$. The Calmar ratio is computed as the compounded annual return divided by the maximum draw down. In addition, two other ratios are computed, namely the compounded annual return divided by the mean of the 25% largest draw downs, and the compounded annual return divided by the Expected Shortfall estimated previously by the lognormal model.

Remarks

1. Compounding is the re-investment of profits. The Annualized return with arithmetic extrapolation estimates what the return percentage will be after one year under the assumption that the performance remains the same and that the system was not compounding and that it will not do so in the future. The Compounded annual return estimates what the return percentage will be after one year under the assumption that the performance remains the same and that the system was compounding and that it will continue to do so in the future.

2. However, if the system is not compounding then the compounded annual return does not tell you what you can expect if you compound in your own account. For example, if the system has 10% profit per month, then both annual return estimates will be equal to 120% after one year (since $T = n$ at that moment). However, if you compound each month then you will get $1.1^{12} = 3.14 = 314\%$.

3. Suppose that fixed slippage percentage applies to each return rate, so that for each period your return rate is not R_i but rather sR_i for some number s < 1. Say A is the annualized return with arithmetic extrapolation, and C is the compounded annual return. Then, for a noncompounding system it would be reasonable to expect a return of sA but for a compounding system it will not be sC but rather $s^T C + s^T - 1$.

Example

For the example of Table 1 the program produces rather trivial output where all estimates are 0 because the last account value is equal to the initial account value. So assume that the last account value is not 5 but 5.1. Then the output is this:

Table 18

Annualized return (arithmetic extrapolation)	0.811
Compounded annual return (geometric extrapolation)	1.232
Calmar ratio (compounded annual return / max draw down)	2.054
Compounded annual return / average of 25% largest draw downs	2.054
Compounded annual return / Expected Shortfall lognormal	1.632

This means that if the system keeps the same performance (turning 5 into 5.1 in 9 days every time) then, if this is a non-compounding system, the annual return would be 81.1%. If it is a compounding system, then the annual return is 123.2%. The Calmar ratio says that this is approximately twice the size of the maximum draw down that was incurred up to date, which was 60%. The expected shortfall on basis of the lognormal model was even larger, namely 75.7%.

The estimate of 123.2% compounded annual return is of course very unreliable because it is based on only 9 days. How unreliable is it? You can compute this with the formula for the confidence interval of the compounded annual return, discussed previously in the section on the Sharpe ratio of the excess log return rates, provided that you are willing to assume that the return rates are independent and have a lognormal distribution. Using the formula of that section, the lower bound of the confidence interval is -1 (= 100% loss) and the upper bound is 2.3×10^{85}. So 9 days performance doesn't say very much, but that is not the only reason for this extreme result. The other reason is that the variation in return rates is extremely large and unrealistic in this example. So let's consider somewhat more realistic values, obtained from the S&P from Dec 8 2006 until Dec 21 2006. The values are:

Table 19

1409.84	1413.04	1411.56	1413.21	1425.49	1427.09	1422.48	1425.55	1423.53	1418.3

The compounded annual return on basis of these values is 0.275 = 27.5%. The program reports a mean of the excess log return rates of M = 0.194 if an annual risk-free rate of 1.05 is assumed. Applying to this the M-to-C formula that relates the mean excess log return rate to the compounded annual return, we get exactly the same outcome for the compounded annual return: 0.275. For the confidence interval you need also the standard deviation and the critical *t*-value. These are 0.071 and 2.306 respectively in this case. Applying the formula for the confidence interval, we get the lower bound -0.551 and the upper bound 2.61. So even with a more realistic variation in the returns we get that after 9 days, the estimate is very unreliable – like common sense suggests. The unreliability at the low side of the interval is about 3 times larger than the compounded annual return itself, and the unreliability at the high side of the interval is 8.5 times larger than the compounded annual return itself. How would this be if the mean and standard deviation were based on one year performance? To give you an impression: Suppose that the mean and standard deviation of the excess log return rates were still the same, then the compounded annual return would also still be the same, and its unreliability would still be 61% of its value at the downside and 70% of its value at the upside. If the mean and standard deviation were both 10 times larger after one year, then the compounded annual return would be 630% and its unreliability would be 87% of its value at the downside and 356% of its value at the upside. In summary: Compounded annual returns can be very unreliable and this does not only depend on the number of days but also on the variation in the return rates.

Now go back to the simple example with a compounded annual return of 123.2%. This seems impressive, but if you apply the slippage correction formula of remark 3 above with s = 0.99 (1% slippage per day) then the outcome is -94.3%. This is negative because 1% slippage per day is more than 2% profit in 9 days. To break even with 1% slippage per day the compounded annual return should be 3819% for a compounding system, which corresponds to a daily return rate of a non-compounding system with an arithmetic annual return of 367%. The following table displays which annualized returns are needed to break even with various daily slippage percentages. Note that *the slippage percentages are relative to the account value*, not relative to the trade sizes.

Table 20

	Return needed to break even	
daily slippage	arithmetic annualized	compounded annual
0.1%	36.5%	44.1%
0.2%	73.1%	107.7%
0.3%	109.8%	199.4%
0.4%	146.6%	331.9%
0.5%	183.4%	523.1%
0.6%	220.3%	799.4%
0.7%	257.3%	1198.7%
0.8%	294.4%	1776.0%
0.9%	331.5%	2610.9%
1.0%	368.7%	3818.8%

Appendix: Proof that the variance of excess log returns increases with leverage.

First, notice that the variance of a variable can be expressed as $Var(X) = (0.5/N^2)\sum_i \sum_j (X_i - X_j)^2$, which follows by elementary algebra. This formula is here stated only for a finite space, but a similar formula holds for continuous variables. So the variance is an increasing function of the pairwise differences. So it is sufficient to show that the difference between any pair of excess log return rates increases if the leverage increases. Suppose the leverage is λ and the returns are R_1 and R_2. Then the excess log return rates after leveraging are $\log(1 + \lambda R_i) - \log r$, so their difference is $\log(1 + \lambda R_1) - \log(1 + \lambda R_2)$. This should be an increasing function of λ if $R_1 > R_2$ and $\lambda R_i > -1$. The derivative to λ is $R_1[1 + \lambda R_1]^{-1} - R_2[1 + \lambda R_2]^{-1}$, which is positive if

$$\frac{R_1}{1+\lambda R_1} > \frac{R_2}{1+\lambda R_2}.$$

So it is sufficient to show that this inequality holds if $R_1 > R_2$ and $\lambda R_\mathrm{i} > -1$. If R_1 and R_2 are either both positive or both negative, then the inequality is equivalent to

$$\frac{1+\lambda R_1}{R_1} < \frac{1+\lambda R_2}{R_2},$$

which is equivalent to $R_1^{-1} + \lambda < R_2^{-1} + \lambda$, which is clearly true if $R_1 > R_2$ and R_1 and R_2 have the same sign. If R_1 and R_2 have different signs or one of them is zero, then the inequality holds because the signs of $R_1[1 + \lambda R_1]^{-1}$ and $R_2[1 + \lambda R_2]^{-1}$ are the same as the signs of R_1 and R_2, respectively.

References

Acerbi, C., and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, Vol. 26, No. 7, pp. 1487–1503.

Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1997). Thinking coherently. *RISK*, Vol. 10, No. 11, pp. 68 – 71.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, Vol. 9, No. 3, pp. 203 – 228.

Balkema, A., and L. de Haan (1974). Residual life time at great age. *Annals of Probability*, Vol. 2, pp. 792-804.

Bartholdy, J. and Peare, P. (2003). Unbiased estimation of expected return using CAPM. *International Review of Financial Analysis*, Vol. 12, No. 1, pp. 69–81.

Chen, S. and Huang, Y. (2006). *On the Role of Risk Preference in Survivability.* AI-ECON Research Center, Department of Economics, National Chengchi University, Taipei, Taiwan

Coombs, C.H. and Avrunin, G.S. (1977). Single-Peaked Functions and the Theory of Preference. *Psychological Review*, Vol. 84, No. 2, pp. 216-230

Dekkers, A. L. M. , Einmahl, J. H. J. and de Haan, L. (1989). A Moment Estimator for the Index of an Extreme-Value Distribution. *Ann. Statist.* 17, pp. 1833-1855.

Diebold, F.X., Schuermann, T., and Stroughair, J.D. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. Pennsylvania: Wharton Financial Institutions Centre.

Fernandez, V. (2003). Extreme Value Theory and Value at Risk. *Revista de Análisis Económico*, Vol. 18, No 1, pp. 57-85.

Fisher, R.A., and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Cambridge Phil. Soc*., Vol. 24, pp. 180-190.

Gerrans, P. Morningstar ratings and future performance. *Accounting and Finance*, 46, Vol. 46 No. 4, pp. 605–628.

Gibbons, R. D. , Hedeker, D. R. , and Davis, J. M. (1993). Estimation of Effect Size from a Series of Experiments Involving Paired Comparisons. *Journal of Educational Statistics*, Vol. 18, No. 3. pp. 271-279.

Giesecke, K., Schmidt, T. and Weber, S. (2005). Measuring the Risk of Extreme Events. *Risk Books*, M. Avellaneda (Ed.), Risk Books. Also available at SSRN: http://ssrn.com/abstract=810886.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, Vol. 44, pp. 423-453.

Goetzmann, W.N. (1996) *An Introduction to Investment Theory.* Yale School of Management, http://viking.som.yale.edu/will/finman540/classnotes/notes.html

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, Vol. 6, No. 2, pp. 107-128.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, Vol. 29, No. 3, pp. 339-349.

Johnson, N. L. and Welch, B. L. (1940). Applications of the Non-Central t-Distribution. *Biometrika*, Vol. 31, No. 3/4, pp. 362-389.

Kelley, K. (2005). The Effects of Nonnormal Distributions on Confidence Intervals Around the Standardized Mean Difference: Bootstrap and Parametric Confidence Intervals. *Educational and Psychological Measurement*, Vol. 65, No. 1, pp. 51-69

Kelly, J. L. (1956). A New Interpretation of Information Rate. *Bell System Technical Journal*, 35, 917–926.

Kim, J. (2002). *The stable long-run CAPM and the cross-section of expected returns.* Frankfurt am Main: Economic Research Centre of the Deutsche Bundesbank. http://www.bundesbank.de/download/volkswirtschaft/dkp/2002/200205dkp.pdf

LeBaron, B. D. and Samanta, R. (2005). Extreme Value Theory and Fat Tails in Equity Markets. *Society for Computational Economics: Computing in Economics and Finance 2005*, No. 140. Available at SSRN: http://ssrn. com/abstract=873656.

Lenth, R. V. (1989). Algorithm AS 243: Cumulative Distribution Function of the Non-Central t Distribution. *Applied Statistics*, Vol. 38, No. 1, pp. 185-189.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, Vol. 36, pp. 394–419.

Markowitz, Harry M. (1952). Portfolio Selection. *The Journal of Finance*, Vol. 7, No. 1, pp. 77-91.

Markowitz, H.M (1991). Foundations of Portfolio Theory. *The Journal of Finance*, Vol. 46, No. 2, pp. 469-477.

Matthys, G. and Beirlant, J. (2003). Estimating the Extreme Value Index and High Quantiles with Exponential Regression Models. *Statistica Sinica*, Vol. 13, pp. 853-880

McNeil, A.J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin,* Vol. 27, pp. 117-137.

McNeil, A. J. (1999). Extreme Value Theory for Risk Managers. Departement Mathematik ETH Zentrum; mcneil@math. ethz. ch

McNeil, A. and Frey, R. . (2000). Estimation of Tail-Related Risk Measures for Heterocedastic Financial Times Series: an Extreme Value Approach. *Journal of Empirical Finance* 7, pp. 271-300.

Neter, J. , Kutner, M. H. , Nachtsheim, C. J. and Wasserman, W. . *Applied linear statistical models. .* Chicago: Irwin. (1996).

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, Vol. 3, pp. 119-131.

Pisarenko, V. F. and Sornette, D. (2003). Characterization of the Frequency of Extreme Earthquake Events By the Generalized Pareto Distribution. *Pure and Applied Geophysics.*

Schröder, M. (2006). Is there a Difference? The Performance Characteristics of SRI Equity Indices. *Journal of Business Finance & Accounting (OnlineEarly Articles).*

Sharpe, W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, Vol.19, No. 3, pp. 425-442.

Sharpe, W.F. Mutual Fund Performance. *Journal of Business*, 1966, Vol. 39, No. 1, pp. 119-138.

Sharpe, W.F. (1991). Capital Asset Prices with and without Negative Holdings. *The Journal of Finance*, Vol. 46, No. 2, pp. 489-509.

Sharpe, W.F. (1992). Asset allocation: management style and performance measurement. An Asset class factor model can help make order out of chaos. *Journal of Portfolio Management*, pp. 7-19.

Sharpe, W. F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*.

Sharpe, W.F. (1997). *Morningstar's performance measures.* http://www.stanford.edu/~wfsharpe/art/stars/stars0.htm.

Sharpe, W.F. (1998). *Morningstar's risk-adjusted ratings*. Financial Analysts Journal, 54, 4; pp. 21-33. Also available at http://www.stanford.edu/~wfsharpe/art/msrar/msrar.htm

Sortino, F. & Van der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management*, pp. 27-31.

Sortino, F. and Price, L. (1994). Performance measurement in a downside risk framework. *The Journal of Investing*, pp. 59-65.

Yamai, Y. and Yoshiba, T. (2002). On the Validity Of Value-at-Risk: Comparative Analyses With Expected Shortfall. *Monetary and Economic Studies.*