Is log ratio a good value for measuring return in stock investments

Alfred Ultsch

Databionics Research Group, University of Marburg, Germany,

Contact: ultsch@informatik.uni-marburg.de

Measuring the rate of return is an important issue for theory and practice of investments in the stock market. A common measure for rate of return is the logarithm of the ratio of successive prices (LogRatio). In this paper it is shown that LogRatio as well as arithmetic return rate (Ratio) have several disadvantages. As an alternative relative differences (RelDiff) are proposed to measure return. The stability against numerical and rounding errors of RelDiff is much better than for LogRatios and Ratio). RelDiff values are identical to LogRatios and Return for small absolutes. The usage of RelDiff maps returns to a finite range. For most subsequent analyses this is a big advantage. The usefulness of the approach is demonstrated on daily return rates of a large set of actual stocks. It is shown that returns can be modeled with a very simple mixture of distributions in great precision using Relative differences.

1 Introduction

The daily rate of return for stock is an important figure, not only for practical people, who want to see how their portfolio performs, but also in many theories on market risk. A model of the distribution of daily stock returns is a prerequisite for many theories. For example the Black and Scholes' formula for options (Black /Scholes, 1973) relies on the assumption that daily returns are log normal distributed. Markowitz portfolio theory is built on the assumption that returns follow a Gaussian normal distribution (Markowitz 1952). It is known, however, that actual returns measured in the market do not follow these model distributions (Aas 2004, Nawroth/Peinke 2007). Return may be calculated using different formulas. The most common used are arithmetic return ratio and LogRatio. Both measures have the disadvantage of an unsymmetrical and unbound range. In this paper relative differences are introduced to measure returns. It is shown that this measure leads to a simple but very precise model of actual observed returns. The approach is tested on a large database of daily stock prices.

2 The Data

In this paper data from stock markets in the US was used. During the period Jan 1st 2000 until March 1st 2008 daily stock prices were extracted from a public source (Yahoo® finance). The stock prices are the daily closing prices adjusted for splits in the past. This gave 2047days of (adjusted) closing prices. During this period the Standard and Poor's 500 Index (S&P500) had a range between 776 and 1.566. The S&P500 showed periods of positive and negative slope of about the same length. In order to avoid effects from stocks with very small prices (penny-stocks) and very high priced stocks, only such stocks were used that had during 99% of the time period a price in the interval from 2$ to 25$. A set of 7030 stocks fulfilled this condition. Overall the following empirical results are based on 7030 stocks *2047 prices. This is more than 14 million numbers.

3 Measuring Daily Return

A straight forward measurement of return is the Ratio R (also known as arithmetic return): $R = \frac{P(today) - P(yesterday)}{P(yesterday)}$ (i)

where P(d) is the closing price of a stock at day d. This is, however not the only way to measure a sock's performance. The next common measure is LogRatio LR (e.g. Aas 2004): $LR = \ln\left(\frac{P(today)}{P(yesterday)}\right)$

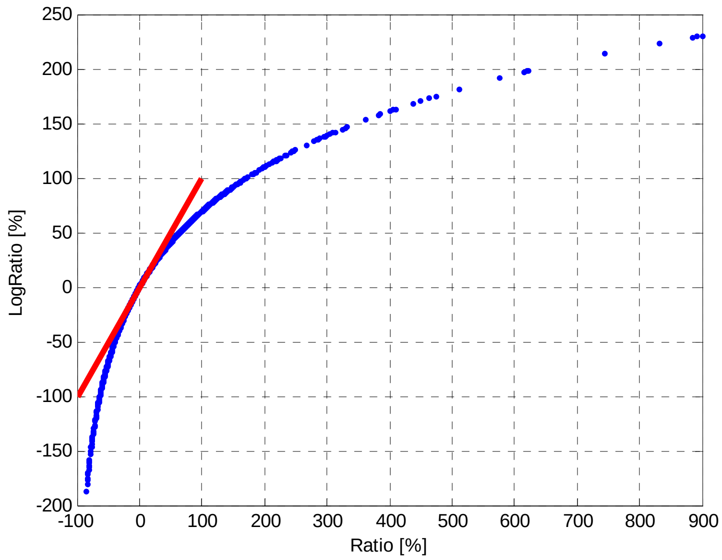

Figure 1: Ratio and LogRatio for empirical data

LogRatios have the advantage that returns of longer periods can be simply calculated by multiplying the LogRatios of the intermediate periods. Furthermore, if stock pricing is assumed to be a time continuous process LogRatios are the infinitesimal limit of the arithmetic returns. Equation (i) can be rewritten to:

$$R+1=\frac{P(today)}{P(yesterday)} \; .$$

Using a Taylor series approximation for ln(R+1) gives ln(R+1) = R for small R. Figure 1 compares Ratio and LogRatio for the data described below. It can be seen, that for |R| < 10% Ratio and LogRatio give almost the same value. Fig 1 also shows one of the problems of Ratio and LogRatio: the range is not symmetrical for positive and negative values. Positive Ratios (gains) of bigger than 1000% were observed in practice. Total loss (ruin) is, however limited to -100% for Ratio. For LogRatios this problem is worse. If the price of today or yesterday is close to zero LogRatio is numerically instable towards infinity. For a model, this has the consequence that gains and losses must be described differently.

The unconfined range of both measures has another nasty consequence for small portfolios. Consider for example a small portfolio of three stocks. At one time d1 these stocks may have LogRatios (or Ratios) of {5%, 8%, 600%} at another day d2 LogRatios of {5%, 8% , 6%} . While for d2 the calculation of an average return (6.33%) makes sense, the same calculation is biased by the extreme gain of the third stock in d1. The same holds for an estimation of the variances. In particular, if a comparison of days in terms of returns is wanted, one might want to use the sum of differences (or Euclidean distance) on the days. This is also extremely biased by the unbound gain.

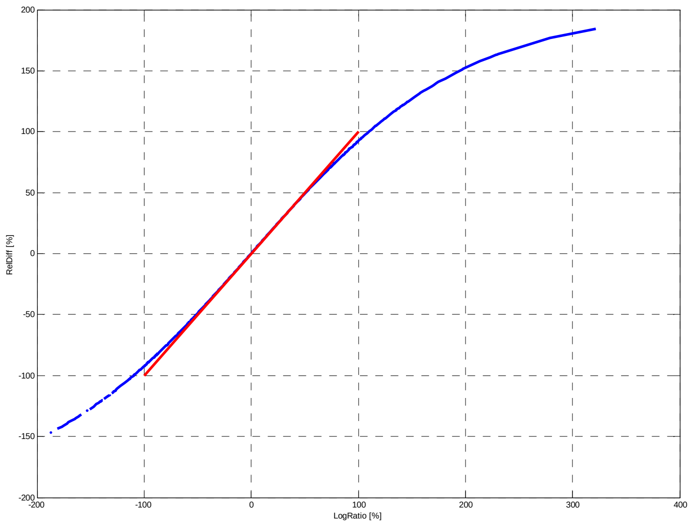

Figure 2: Comparison of LogRatio and RelDiff for the emprical data

A different definition of returns in the form of Relative Differences (RelDiff) alleviates these problems. Define

$$\text{RelDiff} = \frac{P(today) - P(yesterday)}{\frac{1}{2}\left(P(today) + P(yesterday)\right)} = 2 \cdot \frac{(P(today) - P(yesterday))}{\left(P(today) + P(yesterday)\right)} \quad (ii)$$

RelDiff means a comparison of gains and losses to the average price of both days, yesterday and today. For the ranges |Ratio| < 25% and |LogRatio| < 60% RelDiff is numerical identical to Ratio resp. LogRatio, see Figure 2.

Ruin is not a numerical problem for RelDiff: P(today) = 0 in equation (ii) gives a value of -200%. For extreme gains P(today) + P(yesterday) ≈ P(today) + P(yesterday) equation (ii) results in Reldiff = 200%. This means RelDiff has a symmetrical and limited range of -200% < RelDiff < 200%.´

4 The distribution of daily returns

A model of the distribution of daily returns is important for theoretical and practical purposes. Markowitz' theory assumes normal distribution of Ratios. Portfolio theory according to Black & Scholes assumes log normal distribution of Ratios (Black and Scholes, 1973) which is equivalent to a normality assumption for LogRatios. It is well known, however, that both assumptions do not hold in practice. Figure 3 shows a quantile/quantile (QQ-) plot of Ratios, LogRatios and RelDiffs.

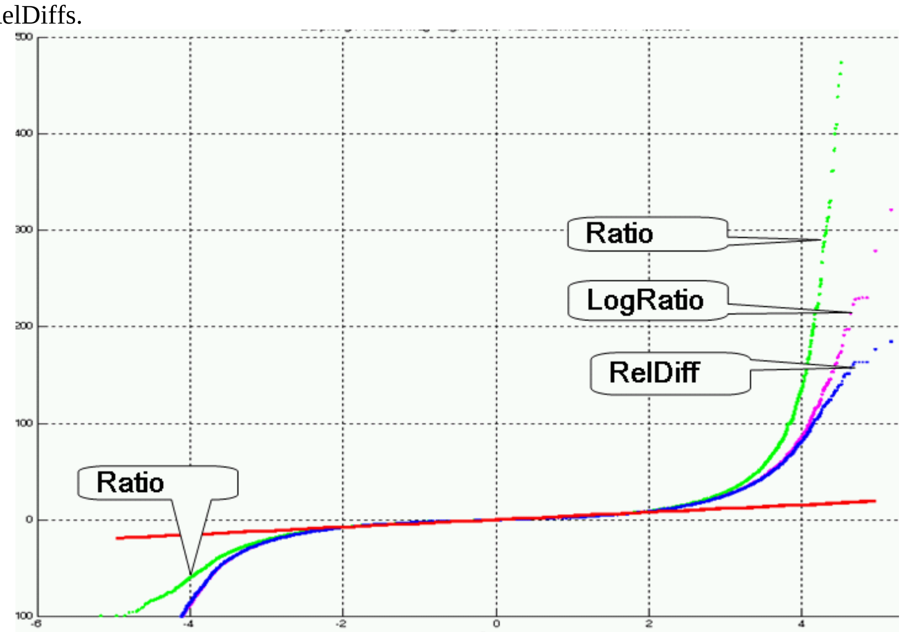

Figure 3: Q/Q Plot of Ratio, LogRatio and RelDiff vs a Gaussian

It can be seen that the normality assumption is appropriate for small absolute Returns. This holds for about 90 to 95% of the data. For larger gains or losses the distributions are leptocurtic. RelDiff with its limited range has the best potential for a precise model.

5 Modeling the Distribution of Returns

The suitability of the different measures for return was tested via a model of the distributions. To address the leptokurtic nature of the distribution a central Gaussian plus a LogNormal distribution at each side (gains and losses) was used. These mixtures of Log-Gauss-Log were optimized for Ratio, LogRatio and RelDiff using the EM algorithm (Bilmes 1979). As quality measure the linearity of a Q/Q plot vs. the model distribution is used. For the central Gaussian all three measures showed complete linearity. This can be seen in figure 3. For losses LogRatio and RelDiff the Q/Q plot showed a good linear relation. Ratio measures for losses did not show a linear relation to the LogNormal distribution (see Fig 3). For the gain side of returns the situation is as shown in figure 4. A linear function has been interpolated for LogRatio and RelDiff in figure 4.

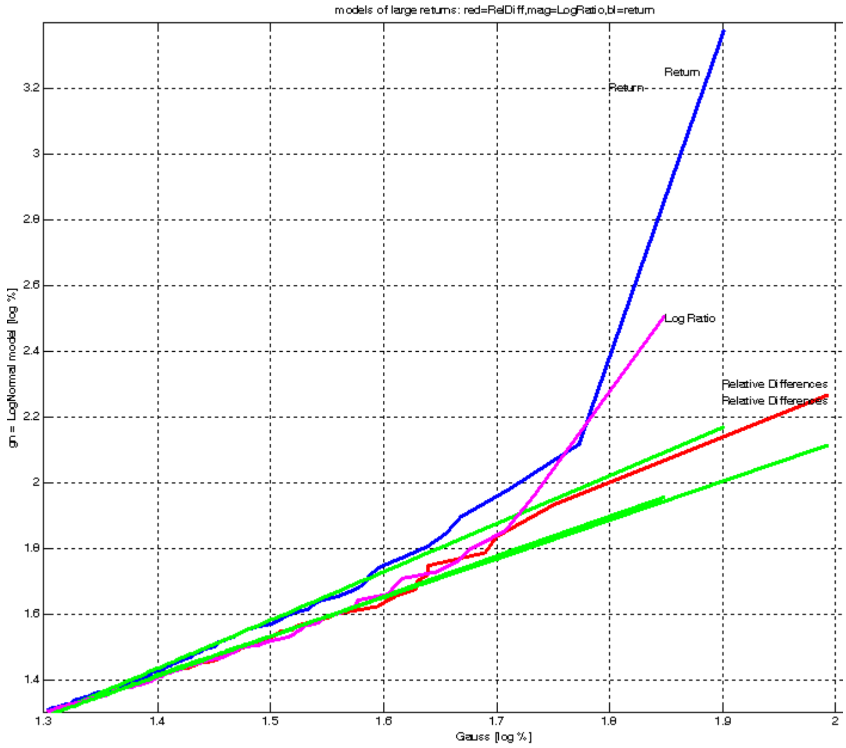

Figure 4:Q/Q plot of winner returns w.r.t. LogNormal compared to a linear fit

For the Ratio measure there a linear interpolation is not appropriate. This indicates that a Log-Gauss-Log model is not suitable for this measure. For LogRatio and RelDiff the linear model is appropriate for values up to about 15%. For larger returns only RelDiff can be reasonably modeled with this type of distribution.

6 Discussion

Using relative differences (RelDiff) as a measure of return has several advantages: RelDiff is in a wide range identical to Ratio and LogRatio. RelDiff is easy to understand: RelDiff is the price difference compared to the average price. RelDiff has less numerical problems than the other measures with regard to ruin, penny stocks and exorbitant gains. RelDiff has a confined and symmetrical range for both gains and losses (-200% to 200%). This makes a symmetrical model for gains and losses possible. For clustering and measuring performances of small portfolios the outlier problems are alleviated. An integrated simple model of returns with a mixture of three components was possible. Figure 5 shows a Q/Q-plot of this model for all data.

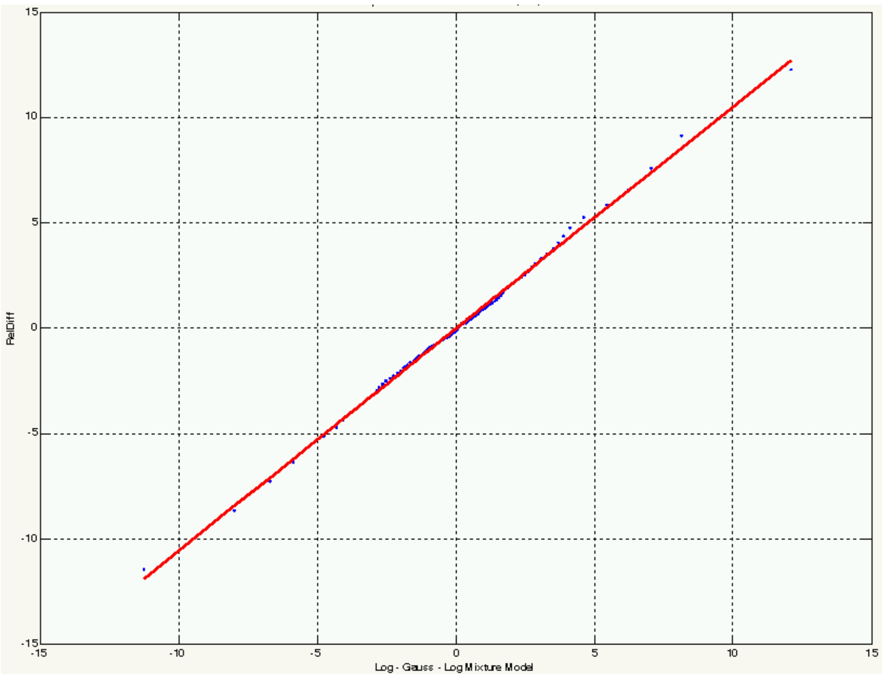

Figure 5: Q/Q plot of returns measured in RelDiff compared to a Log-Gauss-Log model distribution

The huge number of data used for figure 5 allows the statement, that returns can be modeled precisely with a Log-Gauss-Log mixture of distribution on RelDiff. This model can be interpreted as follows: there is one mode of the stock marked that produces random fluctuations in stock prices. These random fluctuations are Gaussian distributed with zero average. Gains and losses are produced by the marked with processes different from the central random walk. The magnitude of gains and losses can be described appropriately by a LogNormal distribution of relative differences. In order to model portfolio and market risks this model can be used to obtain a more precise risk model. The transformation of RelDiff using the distribution model to posterior probabilities for the membership in the classes "Gain", "Loss" "Random Fluctuation" gives the chance for a better characterization and prediction of short time periods and small portfolios. Compare the PUL method for DNA microarray data analysis in this proceedings volume (Ultsch et al 2008).

7 Summary

The daily rate of return for stock is an important figure. Not only for practical people, who want to see how their portfolio performs, but also in many theories of market risks. A model of the distribution of daily stock returns is a prerequisite for many theories. For example the Black & Scholes' formula for options (Black and Scholes, 1973) relies on the assumption that daily returns are log normal distributed.

References

Aas, K.: To log or not to log: The distribution of asset returns, SAMBA/03/04 Norwegian Computing Center, 2004.

Markowitz,H,, Portfolio selection. Journal of Finance, 7(1):77–91, 1952.

Bilmes, J. (1997): A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, University of Berkeley, ICSI-TR-97-021.

F. Black and M. Scholes. The pricing of options and corporate liabilities. Journal of Political Economy, 81:167–179, 1973.

Nawroth, A.P. Peinke,J., Small scale behavior of financial data, Phys. Lett. A 360 (2006), p. 234.

Ultsch, A. et al, A Comparison of Algorithms to find Differentially Expressed Genes in Microarray Data, this proceedings volume.