Why the return notion matters

Gregor Dorfleitner

Dr. Gregor Dorfleitner
Institut für Statistik und Mathematische Wirtschaftstheorie
der Universität Augsburg, D-86135 Augsburg.
Email: gregor@dorfleitner.de

Why the return notion matters

Abstract

Returns in finance can be defined as log returns or as simple returns. Whereas on a numerical level the difference between these two terms is small as long as the return values are close to zero, there can be non-negligible differences if we look at expected values and (co)variances in a stochastic context. This paper examines the consequences of mixing up the two return terms when variances and convariances are considered. Two examples show that these consequences can be severe in the sense of suboptimal portfolio selection or invalid betas. The paper argues that more awareness of the suited return term is necessary. (JEL G11, G12)

A. Introduction

The term "return" is a very basic component of finance. Thus, it turns up in about every second publication in the field of empirical capital market research. No discussion seems to be necessary on this basic and simple issue.

It is well known that returns can be defined in two different ways: as simple returns and as log returns. The cases in which "return" means the simple return are surely more numerous than those where it refers to the log return. Thus, many papers are based on the simple return without even mentioning this fact or defining the returns explicitly. Usually, this is not critical unless μ and σ values stemming from log returns are used. For instance, this is the case when an implicit volatility is used as an estimation of σ.

This paper is a short note on the importance of the return notion for everyday finance. It illustrates the differences between the two return terms. These differences mainly relate to the stochastic and statistical behavior of returns.

After a short definition of both return terms in Section B, this paper examines several stochastic aspects of the difference between the two return terms in Section C. Section D provides two important examples with normal distribution that make clear how relevant the discussed problem can be in the field of portfolio selection. Section E concludes the paper.

B. Two definitions and two additivity properties

Let S_t be the price of a certain security (adjusted for dividends etc.) or the value of a stock index. The *log return* (or continuously compounded return) with respect to period t

is defined as

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \ . \tag{1}$$

From (1), we have the (time-related) additivity property

$$r_1 + \cdots + r_t = r_{0t} \ . \tag{2}$$

Due to this property log returns are favorable if developments along the time axis are the topic of interest. The log return R_t^P of a portfolio is related to the log returns $R_t^{(i)}$ of the single stocks by the nonlinear formula

$$r_t^P = \ln\left(\sum_{i=1}^{n} w_i e^{r_t^{(i)}}\right) \ , \tag{3}$$

where $w_1, \ldots, w_n$ are the fractions of the invested capital with $\sum w_i = 1$. Because of the nonlinearity (3) capital market researchers and finance practitioners prefer to work with *simple returns* (or percentage price changes or discrete returns)

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} = \frac{S_t}{S_{t-1}} - 1 = e^{r_t} - 1 \ . \tag{4}$$

With this notion of returns, we have a linear relation between the returns of the single stocks and the portfolio return, i.e.

$$R_t^P = \sum_{i=1}^{n} w_i R_t^{(i)} \ . \tag{5}$$

This relation is also referred to as the additivity property within portfolios. Many portfolio models make full use of this simple linear structure.

Obviously, the support of R_t is the entire real line $\mathbb{R}$, whereas R_t is subject to the restriction

$$-1 \leq R_t < \infty \ . \tag{6}$$

Since for small values around zero $\ln(1 + R_t)$ is approximately equal to R_t (for instance $\ln(1 + 0.02) = 0.0198$) it is often claimed that simple returns can generally be approximated by log returns (and thus both additivity properties can be used). This statement may be true from a numerical point of view. But as soon as distribution assumptions and statistical estimation procedures for distribution parameters come into consideration, one has to be very cautious, as will be shown in the rest of this paper.

The textbook of Fama (1976) spends relatively much attention on the question whether simple or log returns should be used. However, it does not focus on the question, what errors can turn up in the context of the return notion. As many authors, Fama works with simple returns distributed according to the stable or the normal law, an assumption which is actually only apt for log returns (because of the validity of (6)). In this paper we do not follow this approach. We take the view that rather log returns than simple returns are distributed according to a law defined on the entire real line. This has consequences for the simple returns and the μ and σ values used in portfolio theory. These consequences are examined in the following.

C. The difference

In the previous section we stated that on a numerical level simple returns and log returns can be used as an approximation of each other—as long as the absolute value is no too high. Figure 1 illustrates this fact. From this figure, one can also see how the difference between the two return notions increases when the absolute value becomes higher.

Assuming a log return distribution defined on the whole real line, the probability of a high absolute difference between the two return terms is not zero. Therefore, if we consider moments of the log return resp. the simple return distribution, we cannot take for granted that the moments of both return terms still are approximately the same.

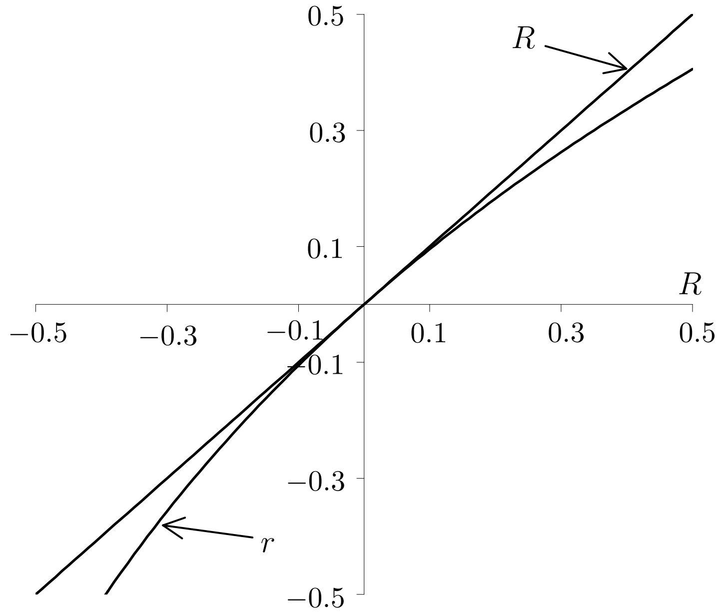
For many researchers fat-tailed log return distributions are beyond doubt.[1] Bamberg/Dorfleitner (2001) show that if this claim is true, then neither any moments of the simple return distribution nor an expected price of the risky asset can exist. Traditional capital market theory then does not make sense anymore. But maybe truncated fat-tailed distributions, which still have a fat tail behavior up to a certain point but vanish beyond that point,[2] are a better model to cope with the empirically "proven" fat-tailedness.

A main case in the textbook literature are normally distributed log returns. In this case the moments of the simple return distribution do exist, they can even be calculated: If μ

[1] See for instance Mittnik/Paolella (2000), p. 313.

[2] See e.g. Matacz (2000) for an introduction to modeling financial data with truncated Levy processes.

Figure 1: The log return (r) against the simple return (R) over the interval $[-0.5, 0.5]$



represents the expected log return and σ^2 the variance of the log return, then the expected simple returns is:

$$\mu_R = \mathrm{E}(R) = \mathrm{E}(e^r - 1) = \mathrm{E}(e^r) - 1 = e^{\mu + \frac{\sigma^2}{2}} - 1\,. \tag{7}$$

The covariance of the simple returns R^A and R^B can also be expressed with the parameters of the corresponding log return distribution. We have:[3]

$$\mathrm{Cov}\left(R^A\,,\, R^B\right) = \mathrm{Cov}\left(e^{r^A}\,,\, e^{r^B}\right) = e^{\mu_A + \mu_B + \frac{1}{2}\left(\sigma_A^2 + \sigma_B^2\right)}\left(e^{\sigma_{AB}} - 1\right)\ . \tag{8}$$

In this formula μ_A and μ_B represent the expected log returns, σ_A^2 and σ_B^2 the log return variances of the assets A and B. The symbol σ_{AB} stands for the covariance of the log returns. From (8) the variance of a log-normally distributed discrete return can be derived as:

$$\mathrm{Var}\,(R) = e^{\left(2\mu + \sigma^2\right)}\left(e^{\sigma^2} - 1\right)\ . \tag{9}$$

[3]The formula is generally valid for the covariance of log-normally distributed random variables. The proof works conventionally by integrating the density function of a bivariate log-normal distribution. Cf. Poirer (1995), p. 129 and 130.

Besides the existence of moments and distribution properties, we have another issue where log returns and simple returns differ: the question of annualizing. Since returns by definition relate to a certain time span, annualization is necessary when expected values and (co)variances of periods with different lengths have to be made comparable. This procedure is very easy for expected log returns and (co)variances of log returns: It is done by multiplying with an appropriate factor. In the case of simple returns annualizing is much more uncomfortable since annualized variances also depend on the expected return value. Unfortunately sometimes this fact is ignored and annualization is done by multiplying a factor even for simple returns. Proceeding in such a way can not be recommended, of course.

D. Examples

To illustrate the severe consequences that a confusion of discrete returns and log returns can have, two examples are presented in the following. The first concerns the classical portfolio selection problem, the second deals with betas.

Both examples work with the assumption of normally distributed log returns. This is not supposed to reflect the author's opinion on the distribution properties of stock returns. It is rather assumed for two practical reasons: Firstly, the μ and σ values of the simple returns do exist and can easily be computed with formulae (7), (8) and (9). Secondly, normal distribution is not unrealistic to such an extend that an example based on this distribution means nothing at all for real world applications. Short-term returns may be not, but at least long-term log returns are more or less normally distributed.[4]

Example 1: A false and the correct market portfolio

We consider a market with four risky assets which have normally distributed log returns, i.e.

$$(r_1, r_2, r_3, r_4)' \sim \mathrm{N}\,(E, V)$$

with the mean vector

$$E = (0.16, 0.14, 0.12, 0.08)'$$

[4]Cf. for instance Akgiray/Booth (1988).

and the matrix of variances and covariances

$$V = \begin{pmatrix} 0.20 & -0.05 & 0.00 & 0.00 \\ -0.05 & 0.15 & 0.01 & 0.02 \\ 0.00 & 0.01 & 0.10 & 0.00 \\ 0.00 & 0.02 & 0.00 & 0.04 \end{pmatrix} .$$

We now calculate the efficient frontier and the market portfolio of this setting according to classical mean-variance analysis. Let w_1, w_2, w_3 and w_4 denote the weights of the four assets in a portfolio. We restrict to

$$w_1 + w_2 + w_3 + w_4 = 1 \ ,$$

as the sole constraint, i.e. to "Black's model", where short sales are possible.[5]

For the well known solution of this problem we refer the reader to e.g. Kaduff (1996). Here, we only apply this solution. But before one can do that, one needs to have the correct—i.e. the simple-return related—mean vector E_R and the correct (co)variance matrix V_R. Applying formulae (7), (8) and (9) we get:

$$E_R = (0.2969, 0.2399, 0.1853, 0.1052)'$$

and

$$V_R = \begin{pmatrix} 0.3724 & -0.0784 & 0.0000 & 0.0000 \\ -0.0784 & 0.2488 & 0.0148 & 0.0277 \\ 0.0000 & 0.0148 & 0.1478 & 0.0000 \\ 0.0000 & 0.0277 & 0.0000 & 0.0498 \end{pmatrix}$$

(values are rounded to 4 digits).

Obviously, the differences to E and V are enormous, whereas the relations between the elements of V_R resp. E_R do not deviate to the same extend from the corresponding relations between elements of V resp. E.

The set of (σ_R, μ_R) pairs forming the efficient frontier (where μ_R resp. σ_R is the expected value resp. the standard deviation of the simple return of a possible efficient portfolio) can

[5] Cf. e.g. Markowitz (1987), p. 11.

now be described by the functional relation:

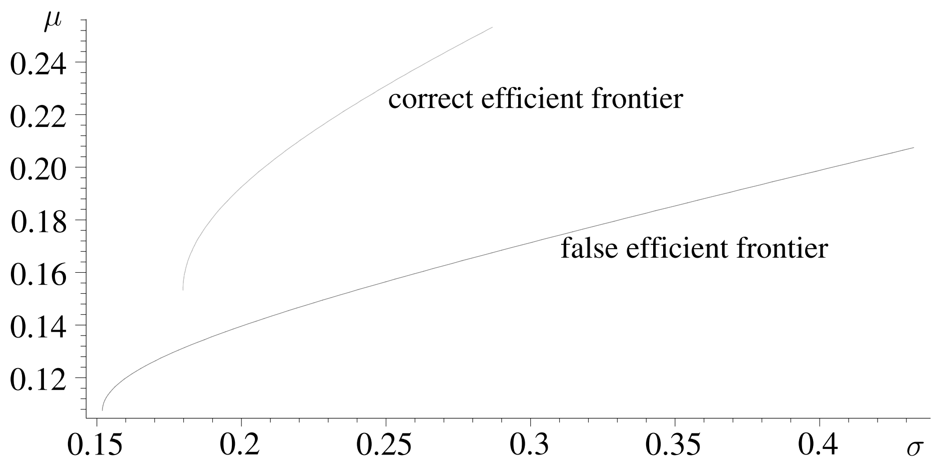
$$\sigma_R(\mu_R) = \sqrt{\frac{c\mu_R^2 - 2b\mu_R + a}{ac - b^2}} \qquad \text{for } \mu_R > b/c \tag{10}$$

where a, b and c depend on the vector E_R and the matrix V_R. The correct values of these parameters are

$$a = 0.9270, \quad b = 4.743, \quad c = 30.95 \ .$$

Figure 2 shows the correct efficient frontier.

Figure 2: The correct and the false efficient frontier



With a riskfree interest rate of 4% we also can compute the market portfolio. It is determined by the weights

$$w_1 = 25.36\%, \quad w_2 = 26.92\%, \quad w_3 = 25.37\% \quad \text{and} \quad w_4 = 22.35\% \ .$$

The simple return of the market portfolio has an expected value of 0.2104 and a variance of 0.0486. Note that the variance is even smaller than that of (the least risky) asset number 4, while the expected value lies between the one of asset number 2 and asset number 3: The desired benefit of diversification.

Up to now we have clearly distinguished between μ and σ values stemming either from the simple or the log return. But unfortunately this is not always done. For instance, the Deutsche Börse daily publishes estimated volatilities and correlations (and thus variances and covariances) of the 30 DAX stocks—and the estimation is based on the log returns.[6] This procedure is not wrong per se. But the estimations cannot be used as the correct values in a portfolio framework. Especially the betas can be false, as will be pointed out in the next example. Another possible source of volatility estimates are implicit volatilities derived from option prices.[7] The suitable return notion (of the Black/Scholes world) of course is the log return. One has to be cautious when using these estimates in a portfolio framework, where the apt return notion is the simple return.

So if E_R is mistakenly substituted by E and V_R by V, we end up with completely different values for a, b and c, thus producing the false efficient frontier also depicted in Figure 2. One can see that the false efficient frontier can not be used as an approximation of the correct one.

The resulting false portfolio weights for the market portfolio are

$$w_1 = 27.41\%, \quad w_2 = 27.58\%, \quad w_3 = 24.60\% \quad \text{and} \quad w_4 = 20.41\% \ .$$

Even if the values are not very far away from the ones above, they definitely represent a different portfolio. The expected value and the variance of this portfolio's simple return are:

$$\mu_R = 0.2146 \quad \text{and} \quad \sigma_R^2 = 0.05118 \ .$$

With formula (10), we see that this is not an efficient portfolio.

Note, that the parameter values of this example were not derived in a sophisticated process. They are just any values that are plausible. Summarizing this example, we can say that a confusion of the expected values and (co)variances of simple returns with those of the log returns can lead to non-negligibly suboptimal portfolio selection.

Example 2: False and right betas

Now we use a setting similar to the one of Example 1. But this time we consider the beta of a stock with respect to a market.

We assume that the log returns of a stock (r^A) and of the market (r^B) are normally distributed with

$$\left(r^A, r^B\right)' \sim \mathrm{N}\left(\left(\mu_A, \mu_B\right)', \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}\right) \ .$$

[6] Cf. Deutsche Börse (2000).

[7] A recent paper on this topic is Hafner/Wallmeier (2001).

The definition of the beta factor is based on the simple returns:

$$\beta = \frac{\text{Cov}\left(R^A\,,\;R^B\right)}{\text{Var}\left(R^B\right)}\;.$$

Employing formulae (8) and (9), this can be transformed to

$$\beta = \frac{e^{\mu_A+\frac{1}{2}\sigma_A^2}}{e^{\mu_B+\frac{1}{2}\sigma_B^2}} \cdot \frac{e^{\sigma_{AB}}-1}{e^{\sigma_B^2}-1}\;. \tag{11}$$

Looking at (11) more closely one can see that β strongly depends on the expected log return values μ_A and μ_B. The false beta of the share of course is $\frac{\sigma_{AB}}{\sigma_B^2}$. Even if it is not correct, this false value still can be used as an approximation. Comparing $\frac{\sigma_{AB}}{\sigma_B^2}$ with formula (11) one can see, that an approximation can be recommended if

$$\mu_A + \frac{1}{2}\sigma_A^2 \approx \mu_B + \frac{1}{2}\sigma_B^2$$

holds and σ_{AB} and σ_B^2 are relatively close to zero so that

$$\frac{e^{\sigma_{AB}}-1}{e^{\sigma_B^2}-1}$$

can be approximated by $\frac{\sigma_{AB}}{\sigma_B^2}$. The latter usually is given. The first condition cannot be assumed for any case.

Knowing about these relations, it is easy to construct an example with a clear effect: With $\mu_A = 0.30$, $\mu_B = 0.10$, $\sigma_A^2 = 0.25$, $\sigma_B^2 = 0.04$ and $\sigma_{AB} = 0.03$ we have a beta value of 1.012, whereas the false beta is

$$\frac{\sigma_{AB}}{\sigma_B^2} = 0.75\;.$$

But, even if we feed the correct and the false formula with less extreme values, a clear bias of the beta factor can arise: Taking the volatility and correlation estimates (on a 250-days basis) of February 26th of 2002 (from the website of the Deutsche Börse) for the Deutsche Telekom AG (stock A) and the DAX (market B), we have

$$\rho(r^A, r^B) = 0.76980\,,\quad \sigma_A = 0.49038 \quad \text{and} \quad \sigma_B = 0.29532\;.$$

With that, the false beta value is:

$$\frac{0.76980 \cdot 0.49038 \cdot 0.29532}{0.29532^2} = 1.2782 \; .$$

This is exactly the value published by the Deutsche Börse.

If the log returns r^A and r^B were normally distributed (which presumably they are not), then with estimates (on a 250-days basis) of μ_A and μ_B the real beta value could be calculated. Estimating μ_A and μ_B by

$$\hat{\mu}_A \approx \ln \frac{15.90}{26.50} = -0.510 \quad \text{and} \quad \hat{\mu}_B \approx \ln \frac{4900}{6200} = -0.235$$

we have a correct beta (under normal distribution) of about 1.06 with formula (11).

Summarizing Example 2, we state that mixing up log return and simple return σ values can lead to very invalid results concerning the beta of a stock.

E. Conclusions

This paper does not claim that all publications are wrong where the return term is not fixed explicitly. However, it wants to direct more attention to the return notion. This is necessary because there exists a certain lack of awareness in this regard. As far as science is concerned, this lack can lead to invalid results. In practical applications like asset allocation it can imply suboptimal portfolio selection. None of both is desirable. One main problem is that estimates of variances or covariances are simply denoted with the symbol σ. As long as the model behind the estimation of σ is not revealed, using these estimates can cause severe mistakes. Therefore, the author would like to suggest that the term "return" always should be defined explicitly at the beginning of an empirical work or a research paper.

A Monte Carlo simulation study could shed some light on the question how the relations between log return and simple return σ values could look like, if we leave the area of normally distributed or fat-tailed log returns. The point of this paper was to show how dangerous a mixing up of both return notions can be. However, a simulation study could also clarify in which cases they can be mixed up without making a major mistake. There surely are such cases.

References

Akgiray, V.; Booth, G.G. (1988): The Stable-Law Model of Stock returns, Journal of Business and Economic Statistics 6, 51–57.

Bamberg, G.; Dorfleitner, G. (2001): Fat Tails and Traditional Capital Market Theory, Arbeitspapiere des Instituts für Statistik und Mathematische Wirtschaftstheorie der Universität Augsburg, Heft 177.

Deutsche Börse AG (Ed.) (2000): Leitfaden zu den Indexkennzahlen der Deutschen Börse (März 2000, Version 1.1), Frankfurt.

Fama, E.F. (1976): Foundations of Finance, New York.

Hafner, R.; Wallmeier, M. (2001): The Dynamics of DAX Implied Volatilities, Quarterly International Journal of Finance 1, 2001, 1-27.

Kaduff, J.V. (1996): Shortfall-Risk-basierte Portfolio-Strategien, Bern/Stuttgart/Wien.

Matacz, A. (2000): Financial Modeling and Option Theory with the Truncated Levy Process, International Journal of Theoretical and Applied Finance 3, 143–160.

Markowitz, H.M. (1987): Mean-Variance Analysis in Portfolio Choice and Capital Markets, Oxford.

Mittnik, S.; Paolella, M.S. (2000): Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates, Journal of Forcasting 19, 313–333.

Poirier, D.J. (1995): Intermediate Statistics and Econometrics: A Comparative Approach, Cambridge/London.