International Journal of Theoretical and Applied Finance Vol. 4, No.3 (2001) 535–543

HOW TO PRICE INFORMATION BY KULLBACK-LEIBLER ENTROPY AND A MOMENT-RETURN RELATION FOR PORTFOLIOS

ANDREAS DE VRIES
*WestLB, D-40199 Düsseldorf, Märkische FH University of Applied Sciences, Haldener Straße 182, D-58095 Hagen, Germany**

Received

A connection between the notion of information and the concept of risk and return in portfolio theory is deduced. This succeeds in two steps: A general moment-return relation for arbitrary assets is derived, thereafter the total expected return is connected to the Kullback-Leibler information. With this result the optimization problem to maximize the expected return of a portfolio consisting of n subportfolios by moment variation under a given value-at-risk constraint is solved. This yields an ansatz to price information.

1. Introduction

"How much does a byte cost?" — Information does not only play an important role in theoretical economics as, e.g., in the context of the efficiency of markets or in game theory. Also from the management point of view it is of high relevance being one of the production factors. So the evaluation of information is of prominent interest in business as well as in theoretical economics.

The aim of this article is to indicate a deep relation between information and the concept of risk and return in portfolio theory, and to develop on this basis an ansatz to price information. To be concrete, a moment-return relation of a general asset portfolio is deduced. Thereafter, the total expected return of a portfolio is shown to be related to the Kullback-Leibler information, or relative entropy, measuring the deviation from maximum return of a portfolio consisting of n securities. The expected return turns out to be

$$R = [\ln s - H(p) - K(p,q)]\, W_\Pi / t, \tag{1.1}$$

where W_Π is the total sum of the investments W_i, s is a function of the moments of each security in the portfolio Π, and t is the investment horizon. Moreover, $H(p)$ and $K(p,q)$ are the information and the Kullback-Leibler information, respectively; here p is the probability vector p representing the entropy of the investment distribution, whereas q is the probability distribution yielding the moment information. Modifications of the moments, due to market or portfolio changes, cause variations of q. In fact this yields

$$K(p;q) = (R - R^*)\, t / W_\Pi, \tag{1.2}$$

*e-mail: de-vries@mfh-iserlohn.de

where R^* is the maximum return given the investment distribution p. Thus information costs, namely the price of the expected return diminition relative to the sum of the investments.

As an application the optimization problem to maximize the expected return of a portfolio consisting of n subportfolios by variation of the moments under a value-at-risk constraint is solved.

The present article is divided into five sections: After this introduction some aspects from information theory are recapitulated in the second section, whereas in the third one the expected return of a general asset is deduced as a mathematical function of the first two moments. In the fourth section the central return-information relation is deduced and the effect of investment and moment variation is studied, resulting in stating and solving a moment optimization problem. The article is concluded with a short discussion of the results.

2. Information

The mathematically precise definition of information first was given by Shannon[18] in 1948 and founded a whole discipline, information theory. Shannon's pioneering idea was it to consider information as a function of a given probability distribution. It yields a measure for the uncertainty that a given event *will* occur, or for the surprise that it *has* occurred.

In the original sense of information theory the probability distribution refers to the special case of a source sending a finite set of messages, such that message number i is sent with probability p_i. Evidently, this takes into account only the *syntactic* aspect[a] of information.

Soon this point of view was generalized to an arbitrary probability distribution, especially by Jaynes.[10,11] The basis for this vision is the idea that a probability distribution in essence represents *a priori* knowledge, i.e. yet acquired information.

In this way the semantic aspect of information can be made more precise. However, this point of view did not remain uncontroversial, although it even traces back to the founders of probability theory, Laplace and Bernoulli.[12] In this article information is pragmatically defined as a function H of a given probability distribution. The interpretation is depending on the context and will be made precise in the sequel.

Let Φ be the *phase space*, i.e. the set of "microstates" or "elementary events." In this article we restrict ourselves to the discrete case, i.e. the case of a countable phase space: $\Phi = \{\varphi_1, \varphi_2, \dots\}$. (However, the following considerations can be generalized to the continuous case in a straightforward manner.) Moreover we suppose a probability space $(\Phi, \mathcal{A}, P)$, where $\mathcal{A}$ is a σ-algebra over Φ, and P is a probability measure P: $\mathcal{A} \to [0,1]$ given by $P(A) = \sum_{\varphi_i \in A} p_i$ with $p_i = p(\{\varphi_i\})$.[3]

[a]For the notions of syntactic and semantic information see van der Lubbe[14]

Then the *information* is defined by the relation[8,14,18]

$$H(p) = -\sum_{i=1}^{n} p_i \ln p_i, \tag{2.3}$$

Here 'ln' means the natural logarithm to the basis e. (The basis of the logarithm determines the unit of the information. For the basis 2 it is the 'bit', for the basis e the 'nat' or simply a pure number.)

The probability distribution $p = (p_1,\ p_2,\ \dots)$ is a *probability vector*, and we have especially $0 \leqq p_i \leqq 1$. Hence, the information $H(p)$ of a discrete probability distribution always is a non-negative quantity, $H(p) \geqq 0$.

Information thus solely depends on the given probability distribution and does not refer to the contents of the underlying events or states. The *probabilities* with which the events occur are important, not the events themselves.

Let p and q be two probability vectors referring to Φ with the property that q_i vanishes only if p_i does also. Then the *Kullback-Leibler information* or the *relative entropy* of p with respect to q is defined by[1,6,13]

$$K(p;q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}, \tag{2.4}$$

The Kullback-Leibler information is a measure for the deviation of two distributions. Here the *Bayesian prior distribution*[11] q can be considered as the mathematical description of the *a priori* knowledge, which leads to another probability distribution p caused by new knowledge emerging from a measurement or from learning; the Kullback-Leibler information then is the gain of information supplied by the measurement or the learning. In this article Kullback-Leibler information is generally considered as the information difference between the "knowledge" p and the "prior knowledge" q.

One important property of information is the *integrated Gibbs inequality.* It says that for arbitrary distributions $p_i,\ q_i \geqq 0$

$$\sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p_i \ln p_i \ \leqq\ \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i \ln q_i. \tag{2.5}$$

An elementary proof can be found, e.g., in Mackey.[15] Because for any probability distribution p we have the identity $\sum_{i=1}^{n} p_i = 1$, there is always a gain of information for two differing probabilities: By inequality (2.5), $K(p;q) > 0$ for $p \neq q$. Moreover $K(p;p) = 0$, i.e. $K(p;q)$ is minimal for $q = p$ ("No new knowledge by learning p").

3. The moment-return relation

In this section we will derive a general formula relating the moments of a stochastic process to its first two moments.

We assume a standard filtrated probability space $(\Omega, \mathcal{F}, P; \mathcal{F}_{t\in\mathbb{R}_+})$. Ω denotes the phase space of the states of the world, the filtration $\mathcal{F}_t$ the measurable states, i.e. the available information evolving with time, $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$. Then for a continuous semimartingale X and an adapted pathwisely left continuous and pathwisely local process F the stochastic integral $\int_0^t F \, \mathrm{d}X$ is well-defined.[7] For a rigorous introduction to the mathematics of stochastic processes the reader may be referred to Bauer[3] or Øksendal.[17]

Suppose that "now" is given by $t = 0$. Let S be a given security price process, i.e., $S(t)$ is the price of a security at time t. Let moreover $\mu(t)$ denote the expected rate of change of $S(t)$ at time t and $\sigma(t)$ its conditional standard deviation. Usually, μ is referred to as the "drift" of $S(t)$, and σ as its "diffusion."[5,17] We assume X to be a continuous semimartingale, and μ and σ to be adapted processes such that $\mu \in L^1([0,t] \times \Omega)$, and $\sigma \in L^2([0,t] \times \Omega)$. Then

$$\frac{S(t)}{S(0)} = 1 + \int_0^t \mu(\tau)\, \mathrm{d}\tau + \int_0^t \sigma(\tau)\, \mathrm{d}X(\tau). \tag{3.6}$$

Usually X_t denotes the standard Brownian motion B_t, and S is an Ito process.[5,16,17] Formula (3.6) applies to any sort of securities:

Asset investments with $S(0) > 0$

- A bond price process is represented by $\beta(t) = \beta(0)\, \mathrm{e}^{rt}$, with the continuously compounding interest rate r and the invested capital $\beta(0)$; in the general setting (3.6), $S = \beta$, $\mu = r$, and $\sigma = 0$.
- A (lognormally distributed) stock price is given by $S(t) = S(0)\, \mathrm{e}^{\alpha t + \sigma X_t}$, with α and σ constant over the time horizon t; for our general setting (3.6), $\mu = \alpha + \sigma^2/2$.
- The price process S_C of a European long call option on a security $S(t)$ with strike price K is given by $S_C(t) = (S(t) - K)^+$; in case of no arbitrage it is determined, e.g., by the Black-Scholes pricing formulas;[5,9,16] for (3.6) this means μ and σ correspond in a definite manner to the values of the underlying asset.
- The price S_P of a European long put option on a security S with strike price K is analogously given by $S_P(t) = (K - S(t))^+$.

Liabilities with $S(0) < 0$

- A raised credit of amount C with discounted interest rate r, $S(t) = -C\, \mathrm{e}^{rt}$.
- A sold bond where $S(t) = -\beta(t) = -\beta(0)\, \mathrm{e}^{rt}$, with the continuously compounding interest rate r and the "borrowed" capital $-\beta(0)$.
- The price process S_C of a short call option on a security $S(t)$ with strike price K is given by $S_C(t) = -(S(t) - K)^+$.

- The price S_P of a European short put option on a security S with strike price K is analogously given by $S_P(t) = -(K - S(t))^+$.

In general, the continuously compounding interest rate r, also called *short rate*, is given by[5,9,21] $S(t) = S(0)\mathrm{e}^{rt}$, or $r = \frac{1}{t} \ln \frac{S(t)}{S(0)}$. The expected return R of the investment $S(0)$ is given by $R = rS(0)$. It is a function depending on the time horizon, the investment, and the moments, $R = R(t; S(0), \mu, \sigma)$. With the assumption $X_0 = 0$ (almost surely), and supposing μ and σ to be constant over the investment horizon t, we thus have by (3.6) the *moment-return relation*

$$\boxed{R = \frac{W}{t}\ \ln[1 + \mu t + \sigma x(t)],} \tag{3.7}$$

where $x(t) = \int_\Omega X_t \mathrm{d}t$. For a risk-free bond we have $\sigma = 0$, and R is equal to its continuously compounded return.[16,19]

The return function (3.7) obeys the law of marginal return for growing risk, for it is a concave function with respect to μ and σ.[4,20] By the logarithm series we have with $u = \mu t + \sigma x(t)$ that $\ln(1 + u) \approx u$ for $u \ll 1$; moreover $\ln(1 + u) \approx \ln u$ for $u \gg 1$. This means that for small mean values and volatilities $u \ll 1$ the return R behaves approximately proportional to u, whereas for extremely high moments $u \gg 1$ a further augmentation of u causes a lower augmentation of the return.

3.1. *The case of n subportfolios*

Suppose a portfolio Π consisting of n subportfolios $\Pi_1, \ldots, \Pi_n$ ($n \in \mathbb{N}$). For each subportfolio Π_i we denote the return R_i on the invested capital W_i with time horizon t. In the sequel we will often call W_i simply the investment of Π_i.

We assume that each subportfolio Π_i denotes an aggregate investment W_i with

$$W_i \geqq 0. \tag{3.8}$$

This is immediately clear for a portfolio consisting solely of active assets. But what about passive liability transactions with $S(0) < 0$? We deduce from the considerations above that there are two general classes of such deals: (i) raising credits or borrowing capital, and (ii) taking short positions on derivative products. In the first case the borrowed capital in itself does not make sense for a bank or a financial institution, it is always reinvested in one or more active asset investments. So we take the subportfolio Π_i to include at least as many investments such that the net aggregate investment capital is non-negative. In the second case of short positions on derivatives, the bank actually invests no capital, i.e. $W_i = 0$, because such deals can be considered as mere bets on future security prices. Winning the bet just means pocketting the "bet provision" $S(0)$, losing it means redemption of the pool.

Applying the considerations above to the expected returns R_i for each subportfolio Π_i with fixed investment horizon t, we achieve with (3.7) that

$$R_i = \frac{W_i}{t}\ \ln s_i \qquad \text{with } s_i = 1 + \mu_i t + x_i \sigma_i. \tag{3.9}$$

Here we denote the mean value by μ_i, the volatility by σ_i, and the expected value of the stochastic process at time t by $x_i = E[X_i(t)]$. The total expected return R is given as $R = \sum_{i=1}^{n} R_i$. We note that the volatility is an aggregate quantity of

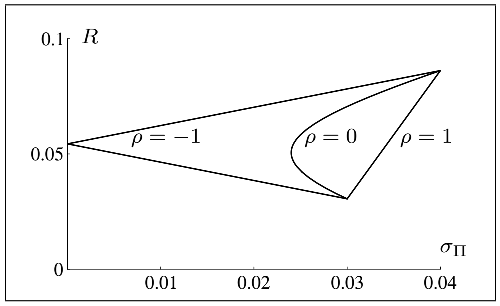

Figure 1: Risk-return efficiency curves in dependence from various correlation coefficients ρ for a two-asset portfolio with $\mu_1 = .001$, $\sigma_1 = .03$, $\mu_2 = .05$, $\sigma_2 = .04$; the return R is given by $R = \sum R_i$, and the portfolio volatility by $\sigma_\Pi^2 = \boldsymbol{W} C \boldsymbol{W}^T$ with the covariance matrix $C = (c_{ij})$, $c_{ii} = \sigma_i^2$, $c_{12} = c_{21} = \rho\sigma_1\sigma_2$. The curves (R, σ_Π) are parametrized by the investments satisfying $W_1 + W_2 = 1$, cf. Steiner and Bruns (1994). Note that $\sigma_\Pi^2 = (\sigma_1 W_1 - \sigma_2 W_2)^2 + 2(\rho+1)\,\sigma_1\sigma_2 W_1 W_2$; because each summand is non-negative, $\sigma_\Pi$vanishes iff $\rho = -1$ and $W_1 = \sigma_2/(\sigma_1 + \sigma_2)$.

n_i assets (or subsubportfolios, respectively), given by $\sigma_i = 1_{n_i} \cdot C^{(i)} \cdot 1_{n_i}^T$, where $C^{(i)}$ denotes the $n_i \times n_i$-covariance matrix of Π_i, and 1_{n_i} the n_i-dimensional vector consisting of 1's, $1_{n_i} = (1, \ldots, 1)$,[2,3,17,19] see figure 1.

4. Moment variation

We start from the following model: A portfolio Π consists of n subportfolios Π_i, in each of which there is made an investment $W_i \geqq 0$ with time horizon t and expected returns R_i $(i = 1, \ldots, n)$. The expected total return R of the total portfolio then simply is $R = \sum_{i=1}^{n} R_i$. Here R_i is given by (3.9). Let us define the quantities p_i and q_i as

$$p_i = \frac{W_i}{\sum_j W_j}, \qquad q_i = \frac{s_i}{\sum_j s_j}. \tag{4.10}$$

By construction p_i, $q_i \geqq 0$, and $\sum_i p_i = \sum_i q_i = 1$. Thus the vectors $p = (p_1, \ldots, p_n)$, and $q = (q_1, \ldots, q_n)$ can formally be considered as probability vectors referring to the abstract phase space $\Omega = \{\Pi_1, \ldots, \Pi_n\}$ of the subportfolios Π_i.

With (4.10) the expected total return $R = \sum_i R_i$ may be rewritten as

$$R = \left[\ln s + \textstyle\sum_i p_i \ln q_i\right] W_\Pi / t \tag{4.11}$$

with $s := \sum_j s_j$ and $W := \sum_j W_j$. We note that $s = s(t; \boldsymbol{\mu}, \boldsymbol{\sigma})$ is a linear function with respect to the moment vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. With (2.3) and (2.4) equation (4.11) can be expressed as

$$\boxed{R = \left[\ln s - H(p) - K(p; q)\right]\, W_\Pi / t.} \tag{4.12}$$

Assuming that W_Π and s are constants given by budget and value-at-risk constraints, the integrated Gibbs inequality (2.5) yields

$$R \leqq \left[\ln s + \sum_i p_i \ln p_i\right] W_\Pi / t. \tag{4.13}$$

The expression in the square brackets in (4.12) shows that the return is maximal if and only if the Kullback-Leibler information vanishes: $K(p; q) = 0$. According to the discussion in the sequel of the Gibbs inequality (2.5), for two probability vectors p and q this true if and only if $p = q$. Thus under the conditions s = const and W_Π = const, the return is maximal if and only if

$$\frac{W_i}{\sum_{j=1}^n W_j} = \frac{s_i}{\sum_{j=1}^n s_j} \tag{4.14}$$

for each $i = 1, \ldots, n$. Geometrically thus the investment vector $\boldsymbol{W} = (W_1, \ldots, W_n)$ points into the same direction as the moment term vector $\boldsymbol{s} = (s_1, \ldots, s_n)$. Hence we have proved the following theorem:

Theorem. *Let be given a portfolio $\Pi = (\Pi_1, \ldots, \Pi_n)$ with aggregate investments $W_i \leqq 0$ and moments μ_i, σ_i for each subportfolio Π_i. Then under the value-at-risk constraint s = const, with s given by (3.9), the expected return R of Π is maximal if and only if the moments are related to the investments W_i by equation (4.14). The maximum return R^* then is determined by (4.13), or*

$$R^* = [\ln s - H(p)]\, W_\Pi / t. \tag{4.15}$$

5. Discussion

In this article a relation between information and expected return of a portfolio Π consisting of n subportfolios $\Pi_1, \ldots, \Pi_n$ is deduced. This succeeds in two steps: First, the moment-return relation (3.9) for the expected returns R_i of each subportfolio Π_i is derived, stating that R_i are functions depending on the corresponding stochastic measures, namely the mean value μ_i, the volatility σ_i, and a constant x_i given by the stochastic processes. Next the total expected return R is related to the Kullback-Leibler information by equation (4.12).

As an application the optimization problem to maximize the expected return of a portfolio consisting of n subportfolios by variation of the moments is solved.

Equation (4.12) allows the remarkable conclusion that any information difference $K(p; q)$ between the knowledge p given by the investment distribution and the prior knowledge q given by the moments diminishes the expected return.

A further aspect is achieved by regarding the case of incomplete information. Usually the moment term distribution only is known incompletely, be it by ignorance or by conscious information hiding of the subportfolio managers. Qualitatively we immediately see that in this case the optimum investment choice must fail, for the optimum condition of the theorem is not achieved. The expected total return will not be maximal.

The result of this article supplies an ansatz to price information: With equation (4.12) the maximum expected return reads $R^* = [\ln s - H(p)]W_\Pi/t$. Changing the moment distribution, e.g. due to modifications of the market or the portfolio, implies $p \neq q$. Changing moments, especially risk, thus causes an information difference K given by (4.12) and (4.15), i.e.

$$K(p;q) = (R - R^*)\, t/W_\Pi. \tag{5.16}$$

Information costs the price of the expected return diminition relative to the sum of the investments.

Acknowledgement. I want to thank J.-C. Curtillet for the many inspiring and fruitful discussions.

References

1. M. Avellaneda, C. Friedman, R. Holmes and D. Samperi: 'Calibrating volatility surfaces via relative-entropy minimization', *Applied Mathematical Finance* **4** (1997)
2. C. Bandelow: *Einführung in die Wahrscheinlichkeitstheorie.* Bibliographisches Institut, Mannheim (1981)
3. H. Bauer: *Wahrscheinlichkeitstheorie.* de Gruyter Berlin New York (1991)
4. T. Bröcker: *Analysis II.* Spektrum Akademischer Verlag Heidelberg Berlin Oxford (1995)
5. D. Duffie: *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton (1996)
6. W. Ebeling, J. Freund and F. Schweizer: *Komplexe Strukturen: Entropie und Information.* B.G. Teubner Stuttgart Leipzig (1998)
7. W. Hackenbroch and A. Thalmaier: *Stochastische Analysis.* B.G. Teubner Stuttgart (1994)
8. W. Heise and P. Quattrocchi: *Informations- und Codierungstheorie. 3. Auflage.* Springer-Verlag Berlin Heidelberg (1995)
9. J.C. Hull: *Options, Futures, and Other Derivatives. Second Edition.* Prentice-Hall, Upper Saddle River (2000)
10. E.T. Jaynes: 'Information theory and statistical mechanics', *Phys. Rev.* **106**, 620-630 (1957)
11. E.T. Jaynes: 'Information theory and statistical mechanics', in K. Ford (Ed.): *1962 Brandeis Summer Institute in Theoretical Physics.* W.A. Benjamin New York (1963)
12. E.T. Jaynes: 'Where do we stand on maximum entropy?', in R.D. Levine and M. Tribus (eds.): *The Maximum Entropy Formalism.* MIT Press Cambridge (1978)
13. J.N. Kapur: *Measures of Information and Their Applications.* Wiley New Delhi (1994)
14. J.C.A. van der Lubbe: *Information Theory.* Cambridge University Press, Cambridge (1997)
15. M.C. Mackey: *Time's Arrow: The Origins of Thermodynamic Behavior.* Springer-Verlag Berlin (1992)
16. S.N. Neftci: *An Introduction to Mathematics of Financial Derivatives.* Academic Press, San Diego (1996)

17. B. Øksendal: *Stochastic Differential Equations. An Introduction with Applications.* 5th edition. Springer-Verlag Berlin Heidelberg New York (1998)
18. C.E. Shannon and W. Weaver: *The Mathematical Theory of Communication.* University of Illinois Press Urbana and Chicago (1949)
19. M. Steiner and C. Bruns: *Wertpapiermanagement.* Schäffer-Poeschel Verlag Stuttgart (1994)
20. H.R. Varian: *Microeconomic Analysis.* 3rd edition. Norton New York (1992)
21. P. Wilmott: *Derivatives. The Theory and Practice of Financial Engineering.* John Wiley, Chichester (1998)