Random Stock Picking and Portfolio Variance

Hans Rau-Bredow
May 2008

Contact: hans.rau-bredow@mail.uni-wuerzburg.de

Abstract

This paper derives an explicit formula for portfolio variance then stocks are randomly picked. This is done by considering the portfolio weights as random variiables. The main result is that the variance is roughly proportional to the Herfindahl index. In the special case of an evenly weighted portfolio, variance is proportional to the reciprocal value of portfolio size. The simultaneous influence of random stock picking and random stock returns is also analyzed by applying the law of total variance.

1. Introduction

Where is no evidence in the literature that mutual funds systematically outperform the market[1]. Malkiel 2003a p. 24 famously argued that "a blindfolded monkey throwing darts at a newspaper's financial pages could select a portfolio that would do just as well as one carefully selected by experts". Most studies of US mutual funds have found under performance relative to the market between 1% and 2% or roughly in the order of the yearly fees these funds charge to their customers (e.g. Daniel et al 1997, Carhart 1997, Chevalier and Ellison 1999, Wermers 2000, Malkiel 2003b, Elton et al 2007). Results for markets outside the US are a little bit more ambiguous, see Otten and Bams 2002; Ferreira et al 2006.

However, the return of a particular fund could deviate by a large amount from the average mutual fund return. Therefore, not only the mean of the distribution of mutual funds returns – which seems to be the market return minus fees - is of interest, but also the variance or standard deviation of the distribution. If the variance is low, an investor in a mutual fund knows that the chances are high that the fund return is close to the market return, and vice versa. The variance of mutual fund returns depends on the correlation between stocks, on the number of stocks in the portfolio and on how evenly the investment is distributed among the stocks in the portfolio. The higher the number of different stocks in the portfolio, the less possible is a large deviation from the market return. The overview given in Exhibit 1 by Newbould and Poom 1993 shows that the traditional literature recommends that a portfolio of between eight and twenty stocks is the minimum necessary to eliminate diversifiable risk.

The literature on how much stocks are enough for diversification starts with Evans and Archer 1968. In a simulation analysis, they randomly select stocks out of the 470 securities listed in the Standard and Poor's Index for the year 1958. The portfolio size varies from 1 to 40 stocks. In each case, they determine the standard deviation for the 19 half year portfolio returns in the period from January 1958 to July 1967. Evans and Archer conclude that much of the unsystematic variation is eliminated by the time the 8th security is added to the portfolio.

Elton and Gruber 1977 develop an analytical expression for the relationship between the size of a portfolio and the expected value of its variance. The analysis is build on the assumption that the

1 It should be noted that all market participants as a group cannot outperform the market. So assuming that mutual funds systematically outperform necessarily implies that all other private or institutional investors systematically underperform.

investment is evenly distributed among randomly selected stocks. Since the variance of a particular portfolio could be much higher or lower as the expected value, they also give an analytical expression for the variance of variance in order to construct confidence limits for the variance of a portfolio. In a further step, they construct another risk measure which they call total risk and which also accounts for the possibility that the mean of the portfolio return is different from the market return.

Statman 1987 considers an investment in the market portfolio - with the S&P 500 index as a proxy - which is levered with debt so that the risk is the same as an investment in a random portfolio with only a relatively small number of stocks. The return differential between these two investments can be interpreted as the benefit that an investor derives from increasing the number of stocks in the portfolio. He shows that a well-diversified portfolio of randomly chosen stocks must include at least 30 stocks for a borrowing investor and 40 stocks for a lending investor.

In Newbould and Poom 1993, portfolios were also build by randomly selecting stocks out of the S&P 500 index. The number of stocks in the portfolio is increased from 1 to 80. For each portfolio size, 1,000 random portfolios were constructed. They perform a historical simulation of portfolio returns using monthly returns in the period January 1988 to December 1990. Curves are then constructed that would statistically accommodate 99 percent, 95 percent, and 90 percent confidence intervals of the returns for the simulated portfolios. The results indicate that the minimum number of stocks needed to achieve diversification is much higher than 20 stocks. O'Neal 1997 considers portfolios which do not consist of individual stocks, but of more the one mutual fund. He investigates how risk can be reduced through diversification by holding more than one mutual fund.

The simulation analysis of Domian et al 2007 uses an investment universe of 1,000 firms which consists of the 100 stocks with the largest market capitalization in each of 10 different industries. The number of stocks in the portfolio increases stepwise from 10 to 200. For each number of stocks in the portfolio, one million random portfolios are build in order to get the distribution of possible returns and ending wealth for the 20-year investment horizon 1985-2004. They show that in this period 93 stocks were needed to reduce the chance to only 5% that the performance of a portfolio is worse than that of treasury bonds and 163 stocks to reduce this chance to 1%.

Most of the previous literature on the relationship between portfolio size and risk, with Elton and Gruber 1977 being an exception, relies on a simulation analysis. In this paper, an explicit formula for the variance or standard deviation of randomly build portfolios is developed. Analytical expressions not only allow to quickly calculate the riskiness of a portfolio for different market conditions, but are also helpful to identify those factors which influence the risk of a portfolio and to determine the sensitivity with respect to these factors. In the first part of this paper, returns of individual stocks are taken as given and the weights assigned to the different stocks in the portfolio are instead considered as random variables. The main result is that in a large investment universe the variance is roughly proportional to the Herfindahl index, a commonly used measure of concentration. It follows herefrom that in the special case of an evenly weighted portfolio, the variance is proportional to the reciprocal value of the number of stocks in the portfolio. Therefore, if the number of stocks in the portfolio is doubled, the variance reduces to one half and the standard deviation reduces by approximately 30%. This is a new result and the approach differs from that taken in Elton and Gruber 1977.

In the second part of the paper, returns of individual stocks are no longer taken as given. Assuming fixed individual stock returns corresponds to an analysis of the distribution of returns for randomly build portfolios over a predefined period in the past. However, an investor in a mutual fund not only faces the risk whether the portfolio manager will pick the right stocks, but he also does not know what the future returns of individual stocks will turn out to be. Therefore, the simultaneous influence of random portfolio weights and random stock returns has to be studied. By applying the law of total variance, it is possible to calculate the portfolio variance then returns of individual stocks and portfolio weights are both stochastic.

2. Portfolio variance with weights as random variables

2.1 Overview

In this chapter returns of individual stock are taken as given. This assumption corresponds to the comparison of the performance of mutual funds for a given period in the past. In retrospect, it is known how individual stocks have performed. As a result of different investment strategies and different stock picking abilities, some funds have deliver a better performance than others. If a large number of funds are compared, the comparison of funds results in a whole distribution of different investment returns. This distribution can be characterized by its mean and by its variance or standard deviation. As already mentioned, the average return of mutual funds is approximately the market return minus fees.

However, less attention has been given to the variance or standard deviation of mutual funds performance. An investor in a particular fund is exposed to the risk (or chance) that his fund will perform worse (or better) than the average fund. Generally, the variance of fund performance depends on two facts: First, on the correlation between individual stock returns. If individual stock returns are highly correlated, then different investment strategies and stock picking abilities would have only a small impact on the performance of a fund. Secondly, the variance of fund performance depends on in how many different stocks the fund invests and on how evenly the investment is distributed among these stocks. For example, a typical mutual funds holds between 40 and 100 different stocks and top ten holdings consistently represent one third of the portfolio investment[2].

Malkiel 2003b p.4 compares the performance of broadly diversified mutual funds in the period from 1970 to 2001. In 1970, there were 355 funds, of which 158 have survived until 2001. The distribution of under- or over performance after costs of these 158 funds relative to the S&P 500 index is shown in Figure 1. Mean under performance is about -1.2%, approximately the average yearly fee charged by mutual funds. Malkiel has not considered the standard deviation for this distribution, but a roughly estimation shows that this figure is approximately just below 2%.

2 Shawky and Smith 2005.

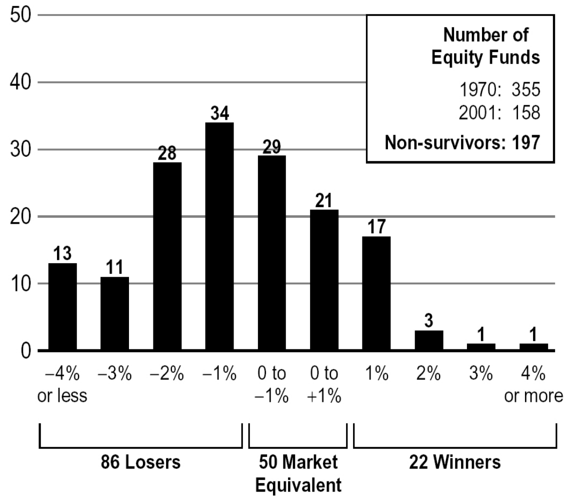

Figure 1: Over- and under performance of survivor funds 1970 – 2001.

Source: Malkiel 2003b, Fig. 4 p. 5

The aim of the following analysis is to derive an analytical formula for the variance of funds performance based on the assumption that stocks are selected completely at random. The main idea can be illustrated by the following simple example: Assume that the investment universe consists of only five well known stocks: General Electric, IBM, Coca-Cola, Microsoft and Wal-Mart. The performance of these stocks for the period from May 15, 2007 to May 16, 2008 is given by table 1. The average return or market return of these stocks is μ_m = -5.84%, the population variance[3] is σ_m^2 = 1.45% and the population standard deviation is σ_m = 12.04%.

Name	Symbol	r_i
GENERAL ELECTRIC	GE	-23.60%
INTL. BUSINESS MACHINES	IBM	5.80%
COCA-COLA	KO	-4.80%
MICROSOFT	MSFT	-14.70%
WAL-MART STORES	WMT	8.10%

$$\mu_m = \frac{1}{5}\sum_{i=1}^{5} r_i = -5.84\,\%$$

$$\sigma_m^2 = \frac{1}{5}\sum_{i=1}^{5}\left(r_i - \mu_m\right)^2 = 1.45\,\%, \qquad \sigma_m = 12{,}04\,\%$$

Table 1: Returns for 5 stocks May 15, 2007 to May16, 2008

3 In this context, the population variance is used with denominator *n* rather than *n* − 1.

Now assume that 3 stocks are selected at random and one third of the total investment is allocated to each of them. $\binom{5}{3}=10$ different evenly weighted portfolios are possible with k = 3 different stocks out of an investment universe with n = 5. The portfolios are listed in table 2. Since stocks have been selected at random, it is clear that the average performance of all funds is the already known market return μ_m = -5.84%. However, the variation of funds returns now is much lower compared to the returns of individual stocks: Population variance is σ_m^2 = 0.24% and population standard deviation is σ_m = 4.9%. The aim of the following analysis is to develop a formula which allows to calculate variance and standard deviation as a function of n and k. The analytical result derived below implies that in the present example variance and standard deviation have been reduced compared to the single stock case in table 1 by the factors 6 and $\sqrt{6}$.

i	Portfolio			$r_{P(i)}$
1	GE	IBM	KO	-7.53%
2	GE	IBM	MSFT	-10.83%
3	GE	IBM	WMT	-3.23%
4	GE	KO	MSFT	-14.37%
5	GE	KO	WMT	-6.77%
6	GE	MSFT	WMT	-10.07%
7	IBM	KO	MSFT	-4.57%
8	IBM	KO	WMT	3.03%
9	IBM	MSFT	WMT	-0.27%
10	KO	MSFT	WMT	-3.80%

$$\mu(r_P)=\frac{1}{10}\sum_{i=1}^{10} r_{P(i)}=-5.84\,\%$$

$$\sigma^2(r_P)=\frac{1}{10}\sum_{i=1}^{10}\left(r_{P(i)}-\mu_m\right)^2=0.24\,\% \qquad \sigma(r_P)=4{,}9\,\%$$

Table 2: Returns for 10 randomly constructed portfolios

2.2 An analytical result

Assume that the investment universe consists of *n* stocks and that the returns $r=(r_1, \ldots, r_n)$ of these stocks are taken as given and therefore non-stochastic. Random portfolios are constructed with the weights $\tilde{w}=(\tilde{w}_1, \ldots, \tilde{w}_n)$ of individual stocks considered as random variables. These random variables always fulfill $\sum_{i=1}^{n} \tilde{w}_i=1$. In practice, most of the weights will assume the value zero because only a small fraction of the stocks in the investment universe will be selected for the portfolio. Since portfolio managers have no bias towards particular stocks, all weights $\tilde{w}_i$ are assumed to have the same distribution. Pairwise correlation $\rho(\tilde{w}_i; \tilde{w}_j)=\rho$ are also considered to be the same for all $i \neq j$.

First, these assumptions imply that $\mu(\tilde{w}_i)=1/n$ for all *i*. In appendix A it is shown that these assumptions also imply the following general result for the correlation between the weights of any two stocks

(1) $$\rho=\frac{\sigma(\tilde{w}_i; \tilde{w}_j)}{\sigma(\tilde{w}_i)\sigma(\tilde{w}_j)}=\frac{-1}{n-1} \quad \text{for all } i, j.$$

This result reflects the fact that weights are generally negatively correlated and that the correlation approximates zero if the number *n* of stocks in the investment universe increases. In the special case of $n=2$, the correlation between w_1 and w_2 is -1 , since an increase of w_1 implies that w_2 decreases exactly by the same amount.

It is now possible to calculate the portfolio variance conditional on given individual stock returns

(2) $$\sigma^2(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{r}=r)=\sum_{i=1}^{n}\sum_{j=1}^{n} r_i r_j \sigma(\tilde{w}_i, \tilde{w}_j)$$

$$=\sigma^2(\tilde{w}_i)\left(\sum_{i=1}^{n} r_i^2 + \sum_{i=1}^{n} \sum_{j=1 \; i\neq j}^{n} r_i r_j \rho\right)$$

Inserting $\rho = \dfrac{-1}{n-1}$ into (2) yields

$$\sigma^2\left(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{r} = r\right) = \sigma^2(\tilde{w}_i)\left(\sum_{i=1}^{n} r_i^2 + \sum_{i=1}^{n} \sum_{j=1\ i \neq j}^{n} \frac{-r_i r_j}{n-1}\right) \tag{3}$$

$$= \sigma^2(\tilde{w}_i)\frac{n^2}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} r_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} r_i^2 + \sum_{i=1}^{n} \sum_{j=1\ i \neq j}^{n} r_i r_j\right)\right]$$

$$= \frac{\sigma^2(\tilde{w}_i)n^2}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} r_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} r_i^2\right)^2\right]$$

$$= \frac{\sigma^2(\tilde{w}_i)n^2}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} (r_i - r_m)^2\right]$$

where $r_m = \frac{1}{n}\sum_{i=1}^{n} r_i$ denotes the market return.

The variance of returns for randomly constructed portfolios is proportional to the population variance of individual stock returns and the variance $\sigma^2(\tilde{w}_i)$ of the portfolio weights. The calculation of $\sigma^2(\tilde{w}_i)$ requires more specific assumptions. The case of an evenly distributed portfolios might serve as an example.

Example: There are $\binom{n}{k}$ possibilities of choosing *k* stocks out of an investment universe with *n* stocks. If the investment is evenly distributed among these *k* stocks, the value of the portfolio weights $\tilde{w}_i$ then is either $\frac{1}{k}$ or zero. On condition that a certain stock has already been chosen for the portfolio, there are $\binom{n-1}{k-1} = \frac{k}{n}\binom{n}{k}$ possibilities how another $k - 1$ stocks out of the remaining $n - 1$ stocks in the investment universe can be selected for the portfolio. Therefore, the probability that a particular stock will be included into the portfolio therefore is $\frac{k}{n}$ and the distribution of the weights $\tilde{w}_i$ is given as follows

$$\text{(4)} \qquad \tilde{w}_i = \begin{cases} \frac{1}{k} & \textit{with probability } \frac{k}{n} \\ 0 & \textit{with probability } 1-\frac{k}{n} \end{cases}$$

This implies $\mu(\tilde{w}_i)=1/n$ and $\sigma^2(w_i)=\frac{1}{k^2}\cdot\frac{k}{n}-\frac{1}{n^2}=\frac{n-k}{k\,n^2}$. Inserting the variance into (3) yields

$$\text{(5)} \qquad \sigma^2(\sum_{i=1}^{n}\tilde{w}_i\tilde{r}_i|\tilde{r}=r)=\frac{n-k}{k(n-1)}\left[\frac{1}{n}\sum_{i=1}^{n}(r_i-r_m)^2\right]$$

The introductory example in chapter 2.1 corresponds to the case n = 5 and k = 3.

For example, on May 27, 2008, where were n = 494 stocks which have been included into the S&P 500 index for at least one year. The population standard deviation of returns over the past twelve months for these 494 stocks was approximately 30%. If k = 9 stocks were chosen at random and with the investment evenly distributed among them, according to formula (5) the standard deviation of the returns of so constructed portfolios would be approximately 10%. If k = 100 stocks were selected, the standard deviation would be around 2.7%.

2.3 The case of an unevenly weighted portfolio

In order to calculate $\sigma^2(\tilde{w}_i)$ in the case of an unevenly weighted portfolio, assume that there are $k \leq n$ numbers $a=a_1,\dots,a_k$ with $\sum_{i=1}^{k}a_i=1$. The selection process is as follows: First, one stock is chosen at random out of an universe of n stocks and a fraction a_1 of total investment is allocated to that stock. The probability that a particular stock is chosen in the first round is $\frac{1}{n}$. Then, a second stock is chosen at random out of the remaining $n-1$ stocks and a fraction a_2 is allocated to that stock. The probability that a particular stock is chosen not in the first round but in the second round is again $(1-\frac{1}{n})(\frac{1}{n-1})=\frac{1}{n}$ and so on. Portfolio weights are therefore distributed as follows

(6) $$\tilde{w}_i = \begin{cases} a_1 & \textit{with probability } \frac{1}{n} \\ \dots & \\ a_k & \textit{with probability } \frac{1}{n} \\ 0 & \textit{with probability } 1-\frac{k}{n} \end{cases}$$

Variance of random portfolio weights then is

(7) $$\sigma^2(\tilde{w}_i)=\frac{1}{n}\sum_{j=1}^{k} a_j^2-\left(\frac{1}{n}\sum_{j=1}^{k} a_j\right)^2=\frac{1}{n}\sum_{j=1}^{k} a_j^2-\frac{1}{n^2}$$

Inserting into (3) finally yields

(8) $$\sigma^2\left(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i \middle| \tilde{r}=r\right)=\frac{nH-1}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(r_i-r_m\right)^2\right]$$

where H is the Herfindahl index[4] $H=\sum_{j=1}^{k} a_j^2=\sum_{i=1}^{n} w_i^2$. In case of an evenly weighted portfolio with $a_i=1/k$ for all i, the Herfindahl index is given by $H = 1/k$.

If the number n of stocks in the investment horizon tends toward infinity, portfolio variance is approximately given by the product of the Herfindahl index and the population variance of individual stock returns

(9) $$\sigma^2\left(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i \middle| \tilde{r}=r\right)=H\left[\frac{1}{n}\sum_{i=1}^{n}\left(r_i-r_m\right)^2\right] \quad \text{if} \quad n\to\infty$$

Note that the Herfindahl index is generally bounded by the minimal and maximal proportion of a stock in the portfolio

4 Note that the realization of the weights $\tilde{w}_i$ is either a_j, $j=1,\dots,k$ or zero.

(10) $$a_{min}=\sum_{i=1}^{k} a_i a_{min} \leq H=\sum_{i=1}^{k} a_i^2 \leq \sum_{i=1}^{k} a_i a_{max}=a_{max}$$

If for example no stock accounts for more than $a_{max}=5\%$ of all holdings, formula (8) with $H=5\%$ would give an upper bound of the portfolio variance.

2.4 Extension: Dividing the investment universe into subcategories

Until now, it has been assumed that every stock has the same probability of being included into the portfolio. This assumption can be relaxed by assuming that stocks can be divided into different categories. For instance it may be possible to assign each stock to a certain industry. The portfolio manager first decides which fraction to invest in each industry. In a second step, within each industry group, stocks are selected at random. Formula (8) makes it possible to calculate the variance for each subportfolio. Since it is plausible that the picking of stocks in each subcategory is independent from the stock picking in the other subcategories, portfolio variance is simply given as the sum of the variances of the subportfolios multiplied by the square of their weights.

In the most simple case, the portfolio manager can distinguish between two subcategories, for instance between growth and value stocks. The portfolio manager decides to invest a fraction *f* of total investment into growth stocks and the remaining into value stocks. If variables are marked with a superscript (g) for growth and (v) for value, respectively, and if weights $\tilde{w}_i^{(g)}$ and $\tilde{w}_i^{(v)}$ are stochastic independent and fulfill $\sum_{i=1}^{n^{(j)}} \tilde{w}_i^{(j)}=1$ for *j = v, g,* then

(11) $$\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i \;=\; f \sum_{i=1}^{n^{(g)}} \tilde{w}_i^{(g)} \tilde{r}_i^{(g)} \;+\; (1-f) \sum_{i=1}^{n^{(v)}} \tilde{w}_i^{(v)} \tilde{r}_i^{(v)}$$

and

(12) $$\sigma^2(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{r}=r)=f^2 \sigma^2(\sum_{i=1}^{n^{(g)}} \tilde{w}_i^{(g)} \tilde{r}_i^{(g)} | \tilde{r}=r) \;+\; (1-f)^2 \sigma^2(\sum_{i=1}^{n^{(v)}} \tilde{w}_i^{(v)} \tilde{r}_i^{(v)} | \tilde{r}=r)$$

The variance of the two subportfolios can be calculated by applying formula (8).

3. Calculation of the unconditional portfolio variance

The previous results have been derived conditional on given individual stock returns. If instead both portfolio weights and stock returns are random variables, the situation can be described by the following two-step model:

Step 1: Portfolio weights are chosen at random.
Step 2: Realization of individual stock returns

The unconditional variance can then be calculated by applying the law of total variance[5]

(13) $$\sigma^2(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i) = \mu[\sigma^2(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{w})] + \sigma^2[\mu(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{w})]$$

Total variance is the sum of the mean of the conditional variance plus the variance of the conditional mean. Here, conditional mean and conditional variance depend on the realization of the weights $\tilde{w} = (\tilde{w}_1, \ldots, \tilde{w}_n)$. Since no specific value is assigned to the weights, conditional mean and conditional variance are functions of random variables and therefore themselves random variables with a certain mean and a certain variance.

However, it is more convenient to change the order of step 1 and 2. This is possible since it is plausible that the selection of the portfolio weights and the realization of stock returns are stochastic independent[6]:

Step 1: Realization of individual stock returns.
Step 2: Portfolio weights are chosen at random.

Law of total variance then states the following

(14) $$\sigma^2(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i) = \mu[\sigma^2(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{r})] + \sigma^2[\mu(\sum_{i=1}^{n} \tilde{w}_i \tilde{r}_i | \tilde{r})]$$

5 See e. g. Weiss 2005.
6 Otherwise, the stochastic influence of the portfolio weights on stock returns must be incorporated into the model. If for instance Warren Buffet publicly announces that he will buy a certain stock, this will probably push up the price of the respective stock. In this case, portfolio selection and stock returns are not independent. The law of total variance itself does not require stochastic independence.

Mean and variance are now conditional on the stock returns $\tilde{r}=(\tilde{r}_1,\ldots,\tilde{r}_n)$. Because the mean of randomly build portfolios is the market return (if fees are neglected)

(15) $$\mu(\sum_{i=1}^{n}\tilde{w}_i\tilde{r}_i|\tilde{r})=\tilde{r}_m$$

the second part of the sum on the right hand sight of (14) is the variance σ_m of the market portfolio. The first part of the sum on the right hand sight in (14) is the mean of formula (8), with individual stock returns now being random variables. Inserting (8) into (13) gives

(16) $$\sigma^2(\sum_{i=1}^{n}\tilde{w}_i\tilde{r}_i) = \frac{nH-1}{n-1} \mu\left[\frac{1}{n}\sum_{i=1}^{n}(\tilde{r}_i-\tilde{r}_m)^2\right] + \sigma_m^2$$

In appendix B it is shown that (15) can be transformed into

(17) $$\sigma^2(\sum_{i=1}^{n}\tilde{w}_i\tilde{r}_i)= \frac{nH-1}{n-1} \left[\frac{1}{n}\sum_{i=1}^{n}\sigma^2(\tilde{r}_i)-\sigma_m^2+\frac{1}{n}\sum_{i=1}^{n}(\mu(\tilde{r}_i)-\mu_m)^2\right]+\sigma_m^2$$

The squared brackets now contain the difference between the average variance of all stocks and the variance of the market portfolio plus the variance of mean returns. If the CAPM

(17) $$\mu(\tilde{r}_i)=r_f+\beta_i(\mu_m-r_f)$$

is applied, it is shown in appendix C that this can also be written as

(18) $$\sigma^2(\sum_{i=1}^{n}\tilde{w}_i\tilde{r}_i)= \frac{nH-1}{n-1} \left[\frac{1}{n}\sum_{i=1}^{n}\sigma^2(\tilde{r}_i)-\sigma_m^2+(\mu_m-r_f)^2(\frac{1}{n}\sum_{i=1}^{n}\beta_i^2-1)\right]+\sigma_m^2$$

In this formula, the sum of all squared beta-factors is used instead of the means of individual stock returns.

4. Conclusion

The number of stocks in a portfolio depends on the amount of risk the investor is willing to accept. In this paper, an explicit formula has been derived which allows to calculate portfolio variance as a function of portfolio size or, more generally, as a function of the Herfindahl index, a commonly used measure of concentration. It has also been shown how to account for the simultaneous influence of random stock picking and random stock returns.

The use of the variance or standard deviation as a risk measure might be questioned. An often proposed alternative risk measure is shortfall risk, the possibility that ending wealth will be less than a certain target amount. However, under the assumption of a Gauss normal distribution, the distribution is fully characterized by its mean and variance or standard deviation. It is always possible to calculate shortfall risk with these two parameters. For instance, the probability that the result is not more than $x = 2$ standard deviations below the mean is $N^{-1}(x) \approx 98\%$.

The validity of a normal distribution at least for sufficiently large portfolios may be justified because of the law of large numbers, since the portfolio composition is then the sum of many independent stock picks. The case may be different for smaller portfolios, in particular if individual stock returns itself are not normal distributed. It is well known that time series of individual stock returns are fat tailed. A fat tailed distribution could also apply then the time period is fixed but returns for a large number of different stocks are compared. The investment universe may contain a few extreme outliers with a spectacular performance. For instance, in the 10 years from January 1990 to January 2000, shares of Dell Computer increased 700 times in value, which corresponds to an average annual return of more than 92%. In particular for smaller portfolios it could then make a big difference whether the few superstocks in the investment universe – Dell, Microsoft or Google – are included into the portfolio or not. It might be argued that this risk is not fully covered by the variance or standard deviation. Further research is obviously needed.

Appendix A

Propostion:

If $\tilde{w} = (\tilde{w}_1, \ldots, \tilde{w}_n)$ are identical distributed random variables with pairwise identical correlation

and if always $\sum_{i=1}^{n} \tilde{w}_i = 1$, then

$$\rho(\tilde{w}_i ; \tilde{w}_j) = \frac{-1}{n-1} \quad \text{for all } i, j.$$

Proof:

$$1 = \mu[(\sum_{i=1}^{n} \tilde{w}_i)^2]$$

$$= \sum_{i=1}^{n} \mu(\tilde{w}_i^2) + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mu(\tilde{w}_i \tilde{w}_j)$$

$$= n\mu(w_i^2) + n(n-1)\mu(\tilde{w}_i \tilde{w}_j)$$

This implies

$$\mu(\tilde{w}_i \tilde{w}_j) = \frac{1 - n\mu(\tilde{w}_i^2)}{n(n-1)}$$

Then together with $\mu(\tilde{w}_i) = \frac{1}{n}$

$$\rho(\tilde{w}_i \tilde{w}_j) = \frac{\mu(\tilde{w}_i \tilde{w}_j) - \mu^2(\tilde{w}_i)}{\mu(\tilde{w}_i^2) - \mu^2(\tilde{w}_i)} = \frac{\left(\frac{1 - n\mu(\tilde{w}_i^2)}{n(n-1)}\right) - \frac{1}{n^2}}{\mu(\tilde{w}_i^2) - \frac{1}{n^2}} = \frac{-1}{n-1}$$

Appendix B

Proposition:

$$\mu\left[\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{r}_i-\tilde{r}_m\right)^2\right]=\left[\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2-\sigma_m^2+\frac{1}{n}\sum_{i=1}^{n}[\mu(\tilde{r}_i)-\mu_m]^2\right]$$

Proof:

$$\mu\left[\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{r}_i-\tilde{r}_m\right)^2\right]=\mu\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{r}_i^2-\tilde{r}_m^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mu(\tilde{r}_i^2)-\mu(\tilde{r}_m^2)$$

$$=\frac{1}{n}\sum_{i=1}^{n}[\sigma^2(\tilde{r}_i)+\mu^2(\tilde{r}_i)]-\sigma_m^2-\mu_m^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}\sigma^2(\tilde{r}_i)-\sigma_m^2+\frac{1}{n}\sum_{i=1}^{n}\mu^2(\tilde{r}_i)-\mu_m^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}\sigma^2(\tilde{r}_i)-\sigma_m^2+\frac{1}{n}\sum_{i=1}^{n}[\mu(\tilde{r}_i)-\mu_m]^2$$

Appendix C

Proposition:

If $\mu(\tilde{r}_i)=r_f+\beta_i(\mu_m-r_f)$, then

$$\frac{1}{n}\sum_{i=1}^{n}(\mu(\tilde{r}_i)-\mu_m)^2=(\mu_m-r_f)^2(\frac{1}{n}\sum_{i=1}^{n}\beta_i^2-1)$$

Proof:

$$(\mu(\tilde{r}_i)-\mu_m)^2=[r_f+\beta_i(\mu_m-r_f)-\mu_m]^2$$

$$=(r_f-\mu_m)^2(1-\beta_i)^2$$

$$=(r_f-\mu_m)^2(1-2\beta_i+\beta_i^2)$$

Since always $\frac{1}{n}\sum_{i=1}^{n}\beta_i=1$, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}(\mu(\tilde{r}_i)-\mu_m)^2=(\mu_m-r_f)^2[\frac{1}{n}\sum_{i=1}^{n}(1-2\beta_i+\beta_i^2)]$$

$$=(\mu_m-r_f)^2[\frac{1}{n}\sum_{i=1}^{n}\beta_i^2-1]$$

Literature

Carhart, M. (1997): On Persistence in Mutual Fund Performance, Journal of Finance, 52(1), pp. 57-82

Chevalier, J. and Ellison, G. (1999): Are Some Mutual Fund Managers Better than Others? Cross-Sectional Patterns in Behavior and Performance, Journal of Finance, 54(3), pp. 875-899.

Daniel, K. M., Grinblatt, M., Titman, S. and Wermers, R (1997): Measuring Mutual Fund Performance With Characteristic Based Benchmarks, Journal of Finance, 52, pp. 1035-1058.

Domian, Dale L.; Louton, David A.; Racine, Marie D. (2007): Diversification in Portfolios of Individual Stocks: 100 Stocks are Not Enough. The Financial Review, Vol. 42, No. 4, pp. 557-570.

Elton, E.J.; Gruber, M.J. (1977): Risk reduction and portfolio size: An analytical solution, The Journal of Business 50, pp. 415–437.

Elton, E.J.; Gruber, M.J.; Brown, S.J.; Goetzmann (2007): Evaluation of Asset Mannagers, Working Paper.

Evans, J. L. Archer, D. H. (1968): Diversification and the Reduction of Dispersion: An Empirical

Analysis, Journal of Finance, 23, pp. 761-67.

Ferreira, Miguel A.; Miguel, António F.; Ramos, Sofia (2006): The Determinants of Mutual Fund Performance: A Cross-Country Study, Swiss Finance Institute Research Paper Series 06-31, Swiss Finance Institute.

Malkiel, Burton G. (2003a): A Random Walk Down Wall Street.

Malkiel, Burton G. (2003b): Passive Investment Strategies and Efficient Markets, European Financial Management, Vol. 9, pp. 1-10.

Newbould, G.D.; Poon, P.S. (1993): The minimum number of stocks needed for diversification, Financial. Practice and Education 3, pp. 85–87.

O'Neal, Edward S. (1997): How Many Mutual Funds Constitute a Diversified Mutual Fund Portfolio? Financial Analysts Journal, March/April 1997: pp. 37-46

Otten, R.; Bams, D. (2002): European mutual fund performance, European Financial Management 8, pp. 75—101.

Shawky, Hany A.; Smith, David (2005): The Optimal Number of Stock Holdings in Mutual Fund Portfolios Based on Market Performance, The Financial Review 40 pp. 481--495

Statman, M. 1987: How Many Stocks Make a Diversified Portfolio? The Journal of Financial and Quantitative Analysis, Vol. 22, No. 3, pp. 353-363

Weiss, Neil (2005): A Course in Probability.

Wermers, R. (2000): Mutual Fund Performance : An Empirical Decomposition into Stock Picking Talent, Style, Transactions Costs, and Expenses, Journal of Finance, 55(4), pp. 1655-1703.