Identification of basic criteria of portfolio analysis based on the rolling verification principle

D A Gercekovich[1], E Yu Gorbachevskaya[2], I S Shilnikova[1]

[1] Irkutsk State University, Karl Marx St. 1, Irkutsk, Russia, 664003
[2] Irkutsk National Research Technical University, 83, Lermontov St., Irkutsk, Russia, 664074

eugorbachevskaya@mail.ru

Abstract. The problem of synthesizing the optimal sizes of training samples specific to each of the considered financial instruments is considered and tested in the article, using real examples. The sample size is selected according to the quality criterion which is based on the accuracy of the generated forecasts. The stated algorithm, which serves as the basis for the synthesis of widely diversified portfolios can significantly increase the efficiency of investment decisions. It is facilitated by, taking into account the characteristics of the markets under study.

1. Introduction

The theory of investment during its development has overcome several stages. Until 1952, such economists as Fisher [1], Keynes [2-4] and many others believed that the size of investment is fully determined by the rate of return. The theory of investments changed radically after the publication of the work by Markowitz [5]. The fundamental difference of the Markowitz model was the definition of a new criterion in the theory of investment decision making - the level of risk through the level of scatter of returns from their expected values. Further development of portfolio theory is associated with the works of Sharp [6]. He proposed the so-called one-factor model of the capital market where the formation of an investment portfolio is carried out on the regression analysis basis. Later Tobin proposed to include risk-free assets in the analysis [7]. For example, government bonds. The works of Sharp, Lintner, Mossin [8, 9] opened the next stage in investment theory, associated with the so-called capital asset valuation model, or Capital Asset Price Model - CAPM. The main result of the CAPM was to establish a relationship between the return and risk of an asset for an equilibrium market. Further, Miller proposed an options model. This model was based on the possibility of a risk-free transaction with the simultaneous use of a share and an option written on it [10].

The task of forming a widely differentiated investment portfolio in the classical point of view involves calculating estimates of their expected returns, risk level, covariances and other statistical characteristics based on historical data of the same length. The task is to form an optimal portfolio from the considered set of securities: ordinary shares, financial derivatives, Real Estate Investment Funds - REIT, Exchange-Traded Fund - ETFs, Fixed Income Securities (preferred shares, bonds, etc.); cash in a certain currency with a given interest rate of increase, exchange commodities, such as, for example, gold, oil, etc. [5-8, 11-16].

The assets listed above represent various industries and finance. Most of the well-known technical indicators: moving average, chart patterns, Moving Average Convergence/Divergence -

MACD, Average Directional Movement - ADX, etc. [17, 18] are equally effective for a wide range of instruments. The key indicator that determines the specifics of various markets and their individuality is the indicator that reflects the "lifetime" of an asset. The most typical are the following: pork meat, beef meat, copper, wheat, Japanese yen, options, futures or forward contracts, and many others. Differences in their dynamics determine the difference in such parameters as the order of the autoregressive equation, the number of averaging points for calculating the moving average [12, 19]. These differences in the parameters of the listed models determine the variability of the optimal training lengths for various financial instruments. The proposed dynamic formation of the historical database, based on the principle of rolling verification of the investment portfolio. It has undoubted practical significance, since it allows us to link together the size of the training sample and the specifics of the corresponding market.

2. Research Methods

Critics of portfolio theory speak negatively about two of its basic provisions [20]:

1. The taken investment decisions assume the values of profitability and risk calculated from past data can be the basis for the formation of investment policy at the present time. Even though in numerous studies the possibility of successful predicting the dynamics of financial instruments is questioned, the authors of this article believe:

a) It is possible to enumerate a fairly extensive class of examples of models that make it possible, with a satisfactory efficiency for practical investors, to synthesize forecasts of the dynamics of financial instruments representing various markets, with a lead time from one month to a year. And indicate a number of successful examples demonstrating the effectiveness of portfolio analysis on independent material, both at educational examples and practical tasks.

b) Even if we take an extreme point of view about the complete impossibility of predicting the dynamics of financial instruments, then in this case, the basic principles of portfolio analysis will help to form such an investment portfolio, the characteristics of "unsinkability" of which (such as: exclusion from consideration of financial instruments that in the recent past showed negative investment results, diversification, etc.) will allow the investor to "survive" an unfavorable period of time with minimal losses.

2. The estimation of the variation in returns is carried out according to the generally accepted formula, while for the investor only those cases are unfavorable when the expected returns exceed the real ones. The approbation of practical investors showed that the use of semi-dispersion does not improve the quality of the resulting risk level estimates, but, in some cases, has a negative impact [14].

Based on the above and based on the foundations of portfolio theory, a procedure for synthesizing an investment portfolio on a variable training sample is proposed, which is based on the principle of sliding verification.

Let the set of securities under study be given by the corresponding time series of the same length, which are the realizations of the quotes of the corresponding securities. In other words, the data that the investor has is a two-dimensional array, where the number of rows is equal to the number of securities under consideration m, and the number of columns n is the amount of historical data, i.e. number of time intervals (bars). Let us introduce the following designations: $D_R(i, j)$ is the realized profitability of the i – th security, j is the number of the time interval (bar), $D_A(i, k)$ is the expected (calculated by the investor) profitability of the i – th security, k is the number time interval (bar) for which this value is calculated (Table 1).

The expected result of diversification is an optimal portfolio, i.e. a list of securities from the initial set and the corresponding weights which they are included in the resulting investment portfolio with. Since portfolio investors, making investment decisions, rely on two basic parameters: expected return and level of risk, portfolio theory is called a two-parameter model. That is why, in order to accomplish this task, it is necessary to estimate the future profitability and the level of risk unique for each security individually and for the investment portfolio as a whole using the time series of the returns of the corresponding securities that the investor has at his disposal.

Table 1. Historical data on the profitabilities of *m* assets.

Asset Number	Time slot number						
	1	2	...	q_1	q_{1+1}	...	n
1	$D_R(1, 1)$	$D_R(1, 2)$	...	$D_R(1, q_1)$	$D_R(1, q_{1+1})$	...	$D_R(1, n)$
2	$D_R(2 ,1)$	$D_R(2, 2)$	...	$D_R(2, q_1)$	$D_R(2, q_{1+1})$	...	$D_R(2, n)$
...	...	...	...	...	...	...	...
m	$D_R(m, 1)$	$D_R(m, 2)$	...	$D_R(m, q_1)$	$D_R(m, q_{1+1})$	...	$D_R(m, n)$

According to Markowitz [5], the risk of investing in a certain type of securities is determined by the probability of deviation of the realized profitability from the expected value. Suppose that during a certain nearest time interval the trends existing in the market will continue, then the predicted value of profitability can be determined based on the processing of historical data on the dynamics of quotations of these assets in the past, thus, the expected profitability is the result that the investor can expect during the nearest time interval (for example, one month).

Then the difference between the investor's expectations and the actual profitability of the analyzed time interval for each of the considered assets will fully reflect the quality of the forecast.

In other words, if $D_R (i, j) D_A (i, j)$, then the investor's expectations are met. Conversely, if $D_R (i, j) D_A (i, j)$, then the investor's expectations turned out to be overestimated, and the real profitability is less than the expected one. Let us take this quantitative expression as a basis for the formation of an optimality criterion.

In its turn, the expected return is the average which can be calculated for the last one, two, three, etc. previous values of the return on the asset or group of assets under study. Continuing to increase the number of historical data to calculate the expected return from the current point in time to the past, the investor, according to the a priori chosen quality criterion, will receive the optimal length of the training sample (i.e. the most acceptable number of historical data that will be used to calculate the expected return).

To improve the quality of developed investment strategies, a procedure is proposed that is designed to restore the slowly changing in time basic criteria of portfolio analysis: profitability and risk, considering the effect of information "aging". For the first time, the problem of selecting the optimal length of the training sequence when choosing the best empirical model was formulated by Gershengorn [21].

To illustrate the effect of a slow change in time of the basic criteria of portfolio analysis of the expected return D_R and risk R_S, the known data for 1926 1993 were used [6]. The comparison was carried out according to the sliding verification principle [16]. For this purpose, the expected profitability, and the level of risk for 1926 1974 were calculated. On the graph (abscissa), this is the time point "1975". Further, the "oldest" data of 1926 were excluded from consideration, and the "fresh" data of 1975 (on the graph the time point "1976"), etc. were included in the analysis until 1993 inclusively. Based on the values of the expected return and risk, the ratio of return to risk is calculated, which most informatively reflects the changes that occur in the US stock market (Figure 1). From Figure 1, the abscissa of which is time, and the ordinate is the ratio of return to risk, it follows that since the mid-70s, the investment attractiveness of the US stock market has been steadily decreasing, while in the mid-80s there is a confident trend reversal and the rise begins, which continues until 1993. A sufficiently high variation of the criterion on the considered time interval indicates that the criterion of the ratio of the expected return to the level of risk is not only informative at a sufficiently high level, but also allows predicting the threat of an unfavorable investment time period with sufficient advance for practical purposes.

To implement the algorithm for optimizing the size of the training sample, we will divide the entire series of returns in chronological order into two non-intersecting sequences: training – q_1 and test – q_2, then: $q_1 + q_2 = n$,
$q_1, q_2 > 0$.
According to the data of previous yields, from the subsystem q_1 for each asset i, the optimal interval of length S is selected, at which

$$D_R(i, t_{q1-s}), D_R(i, t_{q1-s+1}), ..., D_R(i, t_{q1}), \tag{1}$$

the most accurate forecasts for
$$D_R(i, t_{q1+1}), D_R(i, t_{q1+2}), ..., D_R(i, t_n), \tag{2}$$

The forecast based on the test sequence data is carried out on based the sliding verification principle.
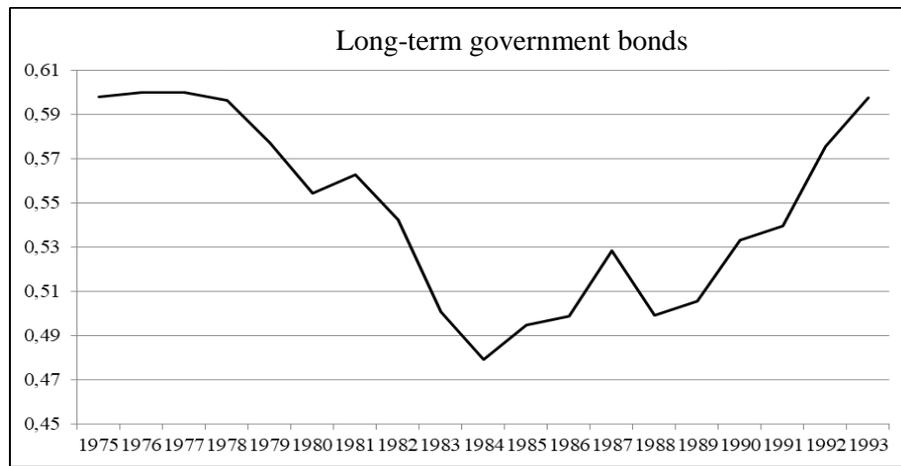
Figure 1. Drift in the ratio of expected return to risk level

Data (1) is used to "train" the model. Then, according to the received criteria, an "exam" is made on the data (2). The algorithm for finding the optimal training length for an investment portfolio with the adaptation of basic criteria works as follows.

Let there be observations in the training sequence. According to the latest observations, the expected return is calculated from $q_1 – D_A (i, q1 + 1)$. Based on the found value, a forecast is made for the next time point – this is $D_R (i, q1 + 1)$. Forecast error:
$E_R(i, q_1+1) = (D_R(i, q_1+1) – D_A(i, q_1+1))$,

remembered. Next, the first time point from the training sequence is discarded and the first point from the test sequence is added (where the adaptive algorithm was tested). The expected profitability is restated according to the updated data. The model with a given number of observations on training n_{tr} is sequentially tested at all time points q_2 from the test sequence.

Sequentially increasing the length of the training sequence from 1 to q_1, it is possible to find the optimal length of the training sequence (n_{opt}) for each i – th asset from the condition of the maximum of the following criterion $S(n_{tr})$:

$S_1 = \sum_{j=q_1+1}^{n}(D_R(i,j) – D_A(i,j))$, for cases, when $E_R(i, q_1+1) > 0$.
$S_2 = \sum_{j=q_1+1}^{n}(D_R(i,j) – D_A(i,j))$, for cases, when $E_R(i, q_1+1) < 0$.
Then the optimality criterion can be written as follows:
$\max (S(n_{tr})) = S_1+S_2$.
When limiting the maximum permissible level of risk:

$$\sqrt{\frac{1}{q_2}\sum_{j=q_1+1}^{n}(D_R(i,j) – D_A(i,j))^2} \leq V. \tag{3}$$

Where (3) is the standard deviation of the results of testing the model with a fixed number of historical data on its training, *V* is the maximum risk level specified by the investor. Then, the essence of the last inequality is to exclude from consideration those variants of the model from among those considered that do not meet the set task of synthesizing the optimal investment portfolio.

After identifying the optimal training length sequentially for each asset, out of the total number *m*, we obtain a one-dimensional array containing *m* elements (*m* unique training lengths for each financial asset under study) *L*(1*: m*).
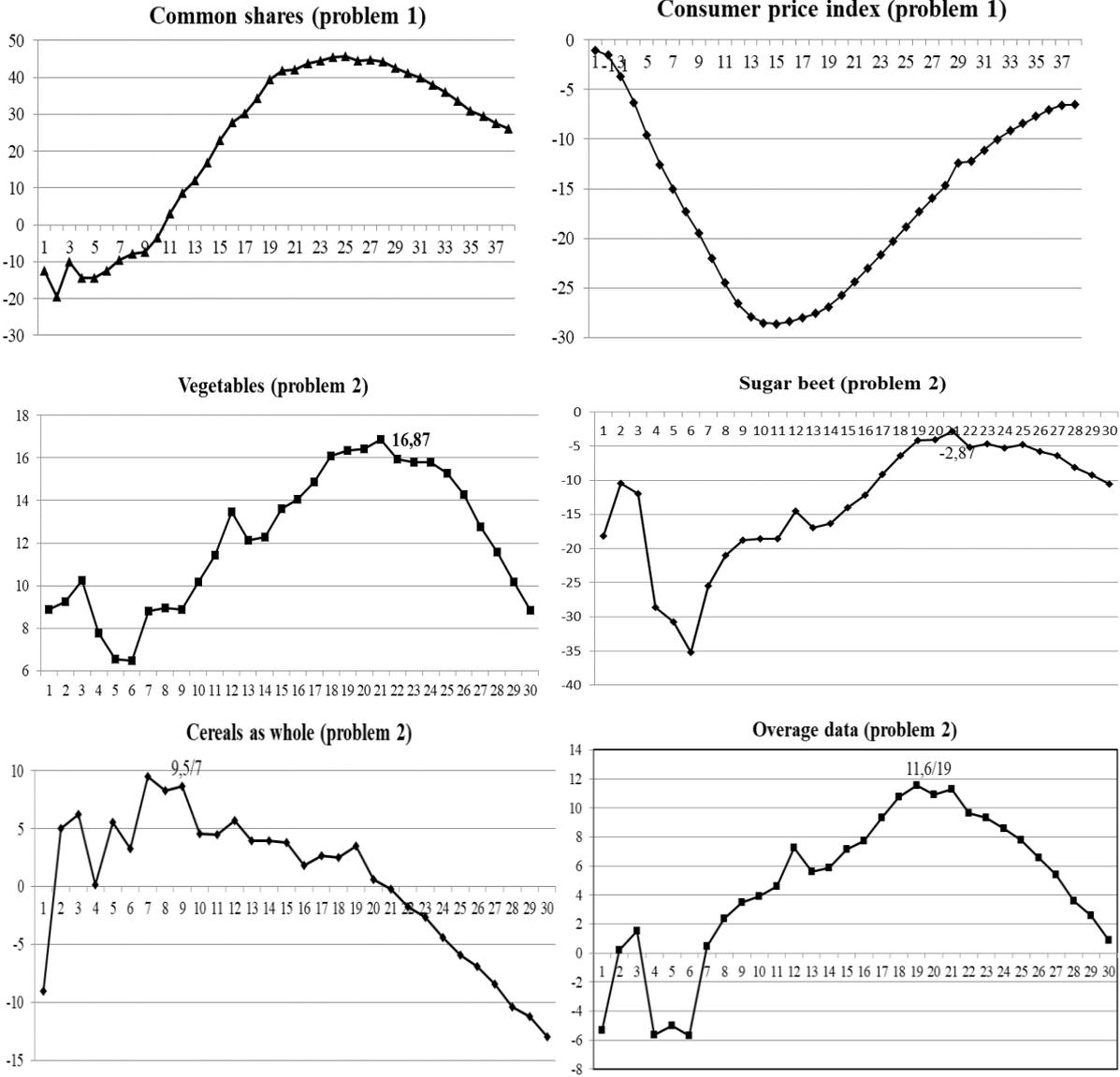
Figure 2. Visualization of synthesis results for optimal training lengths

3. Approbation

The proposed method for synthesizing the optimal training length was tested on two problems: 1. Data on the long-term dynamics of the US stock market for 1926 till 1993 (treasury bills, ordinary shares, long-term government and corporate bonds and changes in the consumer price index) [6]. 2. Data for 1945 till 1996 on the dynamics of crop yields in Russia (cereals in general, sugar beets, vegetables, potatoes, sunflowers) [22].

Figure 2 is a visualization of the results of the synthesis of optimal training lengths for estimating the expected return in the Markowitz model. The graph, in addition to the original crops, displays the results of averaging the optimization results. On the graph the abscissa shows the value of the training sample, and the ordinate shows the value of the maximum optimism criterion. The last of the six presented graphs shows the optimization results based on the averaged values of the optimality criterion for problem 2. The meaning of such averaging is to demonstrate the effectiveness of the proposed approach. Namely, supporters of the classical approach in the problem of forming an investment portfolio can use the result to assess the expected return and the level of acceptable risk, an investor must and suffice to use historical data for the last 19 years.

Computational experiments carried out on an independent material have shown the practical suitability of the proposed approach. For each of the studied assets, its own specific values of the training sample were obtained, which closely correlate with the "life time" of the corresponding financial instruments. The extremum of the optimality criterion in the studied examples is either within the investigated interval of possible values of the lengths of the training samples, or on its border when the values of the optimality criterion reach a plateau.

4. Conclusions

In the general case, upon completion of a full study on the synthesis of optimal learning values for a sufficiently wide set of financial assets, it is recommended for the investor, in his further practical activities, when calculating the basic criteria of portfolio analysis (expected profitability and risk level), to use training samples of different length for each asset. The right edge of these samples corresponds to the current moment in time, while the left edges will differ.

References
[1]   Fisher I 1930 *The Theory of Interest* (The Macmillan Company, N.Y.).
[2]   Keynes J M 1937 General Theory of Employment. *Economic Journal* **51(2)** 214
[3]   Keynes J M 2007 *General Theory of Employment, Interest and Money* (Moscow: Eksmo)
[4]   Keynes J M 1936 *General Theory of Employment, Interest and Money* (Harcourt, Brace, N.Y.) 154 – 158
[5]   Markowitz H M 1959 *Portfolio Selection: Efficient Diversification of Investment* (New York: Wiley)
[6]   Sharpe W F and Aleksandr J 2016 *Investment* (Moscow: Infra-M)
[7]   Tobin J 1965 *The Theory of Portfolio Selection. Theory of Interest Rates* (London: MacMillan) 3 – 51
[8]   Lintner J 1965 The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics* **47(1)** 13 – 37
[9]   Mossin J 1966 Equilibrium in a Capital Asset Markets. *Econometrica* **34(4)** 768 – 783
[10]   Miller M H 1978 Dividends and Taxes. *Journal of Financial Economics. M. Scholes* **6** 333 – 364
[11]   Merton R C 1971 Optimum consumption and portfolio rules in a continuous time model. *Journal of Economic Theory* **3** 373 – 413
[12]   Gitman L J 1997 *Fundamentals of investing* (Moscow: Delo)
[13]   Ferri R 2013 *All About Asset Allocation. The Easy Way to Get Started* (Moscow: Publisher "Mann, Ivanov and Ferber")
[14]   Bernstein U 2001 *The Intelligent Asset Allocator. How to Build Your Portfolio to Maximize Returns and Minimize Risk* (The McGraw-Hill Companies, Inc.)
[15]   Gertsekovich D A 2008 *Quantitative methods for analyzing financial market* (Irkutsk: ISU Publisher)
[16]   Gertsekovich D A and Babushkin RV 2019 Dynamic portfolio analysis of stock indices. *The world of economy and management* **19(4)** 14 – 30
[17]   Elder A 2003 *How to play and win on stock exchange* (Moscow : Diagramma)
[18]   LeBo Ch 1999 *Computer analysis of Futures markets* (Moscow : Alpina)

[19]   Gertsekovich D, Gorbachevskaya L, Grigorova L and Peshkov V 2019 Return on investment in REIT real estate funds. *IOP Conference Series: Materials Science and Engineering* **667 (1) № 012025** Available at: https://iopscience.iop.org/article/10.1088/1757-899X/667/1/012025 (accessed: 01.12.2020)

[20]   Roll R 1977 A critique of the asset pricing theory`s tests. On past and potential testability of the theory. *Journal of Financial Economics* **4(2) (March)** 129 – 176

[21]   Gershengorn G I 1977 *A software package for constructing empirical differential equations. Long-term forecasts of natural phenomena* (Novosibirsk: Nauka) 133 – 137

[22]   *Investman in Russia.* 2017 Statistical compilation.