Sequential investment strategies

Hershel Safer

Presented to
Bar Ilan Conference on
Financial Mathematics

21/6/2016

Outline

The setting

The model

Calculating a portfolio allocation

Empirical results

Lessons about modeling

The setting

The challenge: Given an initial capital stake, invest it in a given set of assets to maximize long-term wealth.

Decisions to be made: At the beginning of each day, allocate the current wealth among the assets in the portfolio. Positions are not closed out each day.

Modeling framework: Use a data-driven approach with minimal assumptions about market behavior.

Based on Gyorfi, Lugosi, and Udina (2006), "Nonparametric kernel-based sequential investment strategies," Mathematical Finance 16:2, 337.

Assumptions 1: Market behavior

This approach relies on two assumptions about market behavior:

- Stationarity: The joint probability distribution of the relevant variables does not change when shifted in time.
- Ergodicity: Statistical properties of the market's behavior can be deduced from a sufficiently long sample.

These assumptions allow a data-driven process to find hidden patterns in the distribution of variables that describe market behavior by looking at increasingly long samples as investments are made and evaluated.

With these assumptions, the maximum asymptotic growth rate can be achieved if the distribution of price relatives is known.

This work describes a "universal strategy" that achieves the same asymptotic growth *without* knowing the distribution.

Assumptions 2: The model

Assets are arbitrarily divisible.

Transaction costs are ignored.

Transactions are long only; no shorting or purchasing on margin.

The strategy is self financing; no capital injections are allowed after the initial investment.

Consumption of capital is not allowed; capital cannot be withdrawn during the strategy's implementation.

Assets are available in unlimited quantities at the current price on any trading day.

Market behavior is not affected by these investments.

Outline

The setting

The model

Calculating a portfolio allocation

Empirical results

Lessons about modeling

The model: Components

The investor has a portfolio of n assets.

The *price relative* of asset j on day i =
closing price / opening price = 1 + return =
Amount that \$1 invested in the a.m. yields at the closing bell

Characterize the market on day i by the vector of price relatives $p_i = (p^{(1)}, \ldots, p^{(n)})$. These are the data.

- Weight returns by inverse volatility of returns?

The proportion of current capital that the investor allocates to each asset on day i is captured by $x_i = (x^{(1)}, \ldots, x^{(n)})$.

- $x_j \geq 0$ (no shorting)
- $\sum x_j = 1$ (self financing, no consumption)

The model: Evolution over time

Start with an initial stake of S_0. Select the initial allocation x_1; place buy and sell orders for execution during the day. Finish the day with capital

$$S_1 = S_0 \sum_{\text{assets } j} x^{(j)} * p^{(j)} = S_0 <x , p>$$

Repeat the next day. After D days, the capital is

$$S_D = S_0 \prod_{\text{days } i} <x_i , p_i> = S_0 \, e^{D\, W(D)}$$

where W(D) is the mean growth rate of capital.

THE BIG QUESTION: How should the investor select portfolio allocations (x)? In particular, how can historical market behavior be used to find good allocations?

Outline

The setting

The model

Calculating a portfolio allocation

Empirical results

Lessons about modeling

Overview of the algorithm

To determine the allocation on a new day:

- Find days in the past that are “similar” to today. In retrospect, the price relatives for those days are known.
- Compute the portfolio allocation that would, in retrospect, have been optimal for those days as a group.
- Use that optimal allocation for today.

Open issues:

- How to calculate the optimal allocation for the similar days
- How to select similar days

Calculating the optimal retrospective allocation

Suppose that m similar days have been identified. The goal is to choose a single allocation x over the n assets that maximizes the average growth rate on those days, using the (retrospectively known) actual price relatives p[day][asset] for those days.

$\max_x \, {}^1/_m \sum_{\text{days } i} \ln(\sum_{\text{assets } j} x_j \, p_{ij})$

s.t. $\sum_{\text{assets } j} x_j = 1$ (self financing, no consumption)

$0 \leq x_j$ (long only)

Solve using a package that can do nonlinear optimization with a single linear constraint and non-negativity constraints.

Identifying similar days: The measure

We model the market with price relatives. Consider two days similar if the price relatives over the preceding k days are close.

Consider the matrix P containing the last k days of price relatives:

	Asset 1	…	Asset n
Day d - 1			
…		p[day][asset]	
Day d - k			

For each previous day i, consider the corresponding matrix P_i. Say that day i is similar to today if P_i is close to P as measured by the Euclidean norm: $|| P - P_i ||_2$

Open issues:

- How close is close enough?
- How many days of history (k) to use?

Identifying similar days: How close is close enough?

Option 1: Use the T most similar days (e.g., T=20).

Option 2: Use a fixed threshold that increases with the number of historical days (k) and the number of assets (n). This defines a uniform, moving-window kernel for evaluating candidates.

Option 3: Try various non-uniform kernels, e.g., use a different norm, or give greater weight to P_i matrices that are closer to P or that are more recent.

Identifying similar days: How many days of history?

The question is somewhat incongruous. The modeling so far has been data-driven, with minimal assumptions. Specifying the number of days for defining similarity makes a strong statement.

Instead, consider multiple values of k, and find the optimal allocation X[k] for each.

Ultimately, one allocation is needed for today. Which one?

Let the data decide: Use a weighted sum of the allocations. On a given day, the weight for X[k] is the overall return that would have been achieved by using k over some chunk of history.

Note: Calculating distances for all k is a bottleneck. Nesting calculations can save an order of magnitude in time.

Outline

The setting

The model

Calculating a portfolio allocation

Empirical results

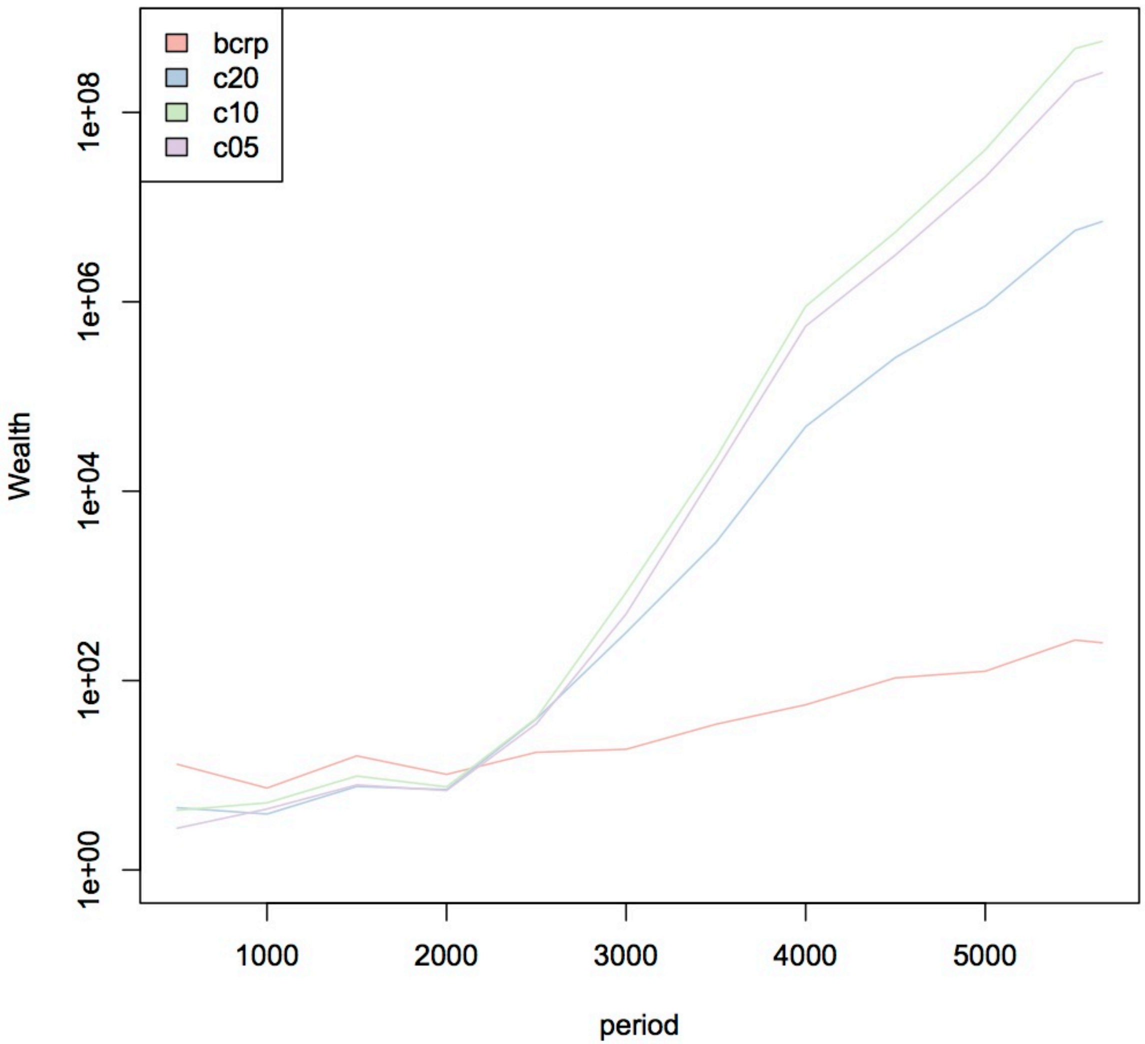
Lessons about modeling

NYSE data set

The dataset was used in a variety of previous papers by other authors. It includes daily prices of 36 assets for a 22-year period (5,651 trading days) ending in 1985.

The results are compared to the performance of the best constantly rebalanced portfolio (BCRP). A single allocation vector is determined for the entire period, and the portfolio is rebalanced accordingly each day. The allocation fractions could only be determined *ex post facto*.

Wealth achieved for various thresholds



NYSE data with transaction costs

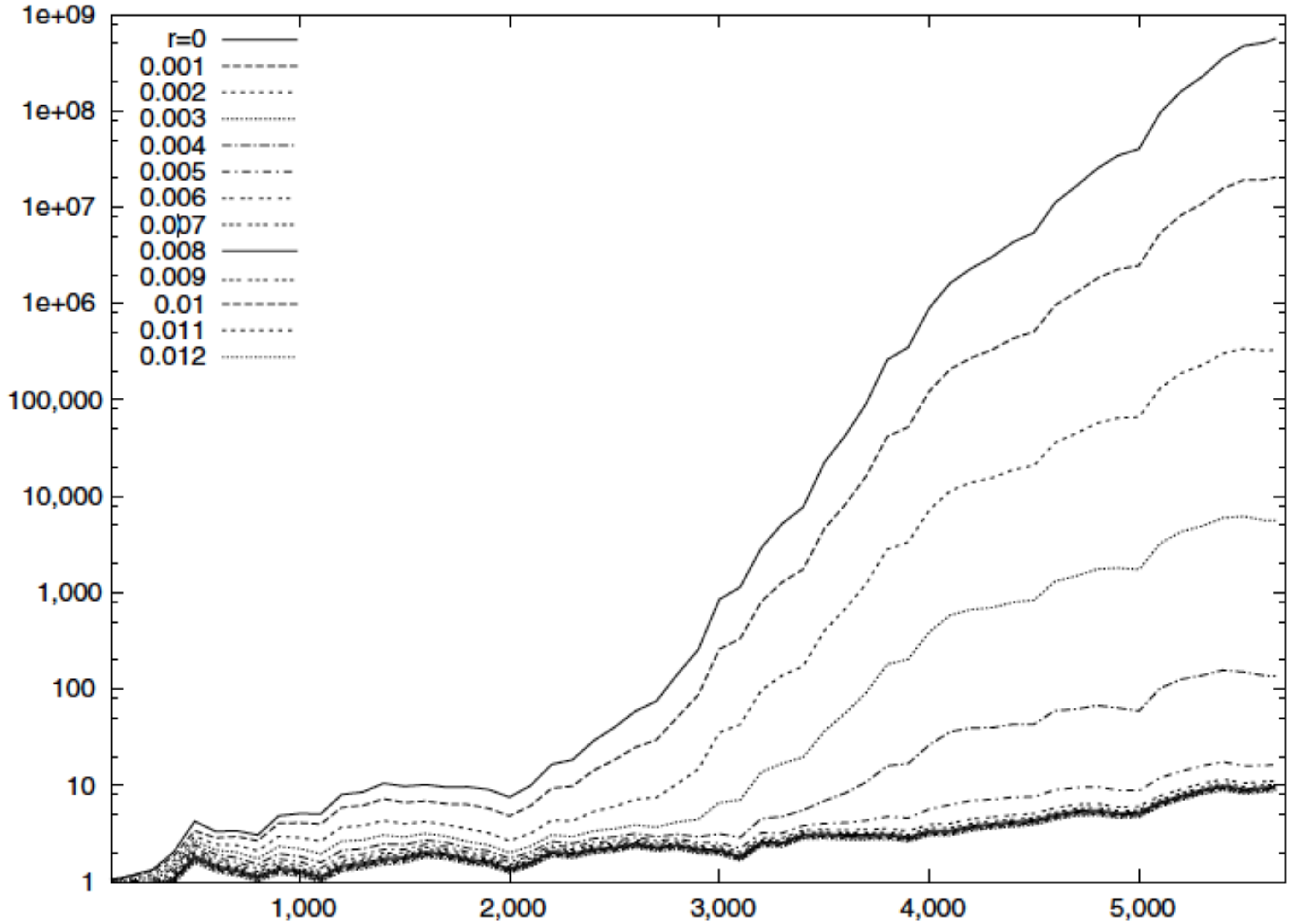


FIGURE 4.2. Wealth achieved by investing one unit uniformly in the 36 NYSE stocks and using the kernel strategy (with constant 1.0) for several values of the transaction costs $r = 0, 0.001, 0.002, \ldots, 0.012$.

Outline

The setting

The model

Calculating a portfolio allocation

Empirical results

Lessons about modeling

Lessons about modeling

Be clear about assumptions. Don't worry about oversimplifying; test to see if the model works anyhow.

A simple model that is implemented is generally better than a complex model that is not implemented.

Don't fixate on a single modeling paradigm. This work combines nearest neighbors (similar days), ensembles (combining results for various values of k), and nonlinear optimization.

Be creative (e.g., using a non-uniform kernel for identifying similar days).

Understand caveats that may affect the validity of your results (e.g., transaction costs).

Implementation matters. Learn how to code.