An Online Algorithm for Multi-Strategy Trading Utilizing Market Regimes

Hynek Mlnařík [1]

Subramanian Ramamoorthy [2]

Rahul Savani [1]

[1]Warwick Institute for Financial Computing
Department of Computer Science
University of Warwick

[2]School of Informatics
University of Edinburgh

The Portfolio Allocation Problem

Dynamically allocate working capital in a portfolio of instruments – over time, as market conditions continually change.

- Classic problem with established theory, e.g., mean-variance optimization and modern extensions.
- Traditional techniques are "model-based" - one makes assumptions (e.g., model of expected returns) that may turn out to be troublesome.
- This issue spurred research into "model-free" approaches.

“Model-free” Portfolio Allocation

- Point of departure: Classic work on optimal bet sizing (Kelly 1956, Breiman 1961) - how much to bet given odds?
- Constantly rebalanced portfolios (Thorp 1971, Markovitz 1976, Bell+Cover 1988, Algoet+Cover 1988) - keep relative allocation of capital constant (still assuming known market return distributions).
- Universal portfolio (Cover 1991) - Sequential portfolio allocation to match the best constantly rebalanced portfolio in hindsight (for an arbitrary market process).
- Many extensions and follow-on work: multiplicative updates (Helmbold et al. 1998), efficient online computation (Kalai et al. 2002), Anticor (Borodin et al. 2004), kernel-weighted allocation (Györfi et al. 2006).

Utilizing Market Context

Market processes are not *entirely* arbitrary – how to utilize odds without overly restrictive assumptions?

- Statistical view of Universal Portfolios (Belentepe 2005): Weights (constrained to a partition of unity) are conditional expectation of a multivariate normal distribution, $\mathbf{w} \sim \mathbf{N}\left(\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t, \frac{1}{t}\bar{\boldsymbol{\Sigma}}_t\right)$.
  - Unconstrained version is the standard log-optimal investment.
- Major contribution of universal algorithms is an online procedure to solve this problem, within a target portfolio class.
- We seek online procedures that also allow us to utilize context in the spirit of (non-parametric) statistics.

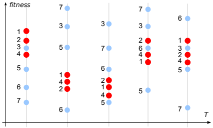
Portfolio Allocation – Our Approach

Dynamically allocate capital in a portfolio of trading strategies.

- Use a set of primitives, i.e., simple strategies such as might be used by traders in practice.
  - Individually, no primitive strategy is well suited (i.e., reliably profitable) under changing market contexts.
- Represent changing market context by regimes - loosely, subsets of strategies that are successful under this context.
  - Use historical data to non-parametrically model these regimes.
- Devise online algorithm for dynamically rebalancing portfolio, shaped by contextual information.

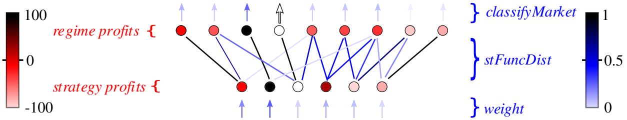
Describing Market State: Our Notion of Regime

- Characterize market state by relative profitability of primitive strategies.
  - A latent switching dynamics induces clusters of similarly performing primitive strategies (of course, this could vary over time).
  - Instead of modelling the latent dynamics in market time series (hard in on-line setting), we seek to model correlation structure in the ensemble of primitive strategies.
  - Identify candidate regimes using a permutation test - perform nonparametric test, over a training horizon, using the sample variance as test statistic, for similarity of a strategy subset versus its complement.



Regimes - Layered Graph of Strategies

Represent market state in terms of the probability that a particular weighted combination of primitive strategies will be the most profitable.



Use multiplicative weight updates to identify possible states from historical data. Over a historical interval,

- Iteratively update weights within candidate regimes according to normalized performance of primitive strategies
- Similarly, generate mixture over candidate regimes

Note: See Appendix 1 for a symbolic description of the same.

Regimes - Interpretation

This architecture is analogous to a particle filter - estimating the probability that a particular (mixture of) primitive strategies maximizes expected performance.

- Iterative update over an interval converges* to a distribution, under the current market context
- While a universal portfolio represents a single weighted sum of underlying assets, we maintain a multi-modal distribution over primitive strategies (i.e., trading rules)

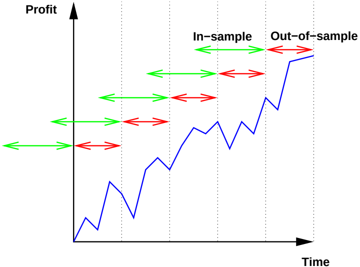
∗ It can be shown that, in a stationary context, this only depends on relative ordering between primitive strategies - see Appendix 2.

Algorithm: **RE**gime **D**etection and **ST**rategy **OP**timization

Training phase - Use above procedure to acquire, from historical data, regimes and possible market states (expressed as weighted sum over regimes)

Trading phase - Allocate capital based on regime-level performance

- In-sample period (Estimate current state): Multiplicative weight update to compute weighted sum of strategy fitnesses
- Out-of-sample period: Online adjustment of asset allocation, multiplicative weight update

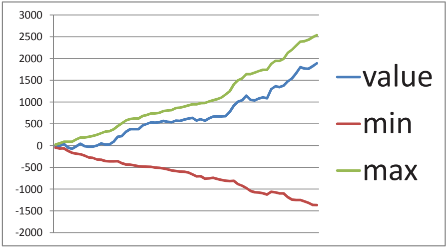


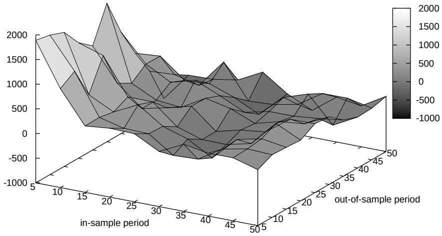
Experiments

We have implemented this algorithm and we report the following preliminary results (using NASDAQ E-mini Futures contracts data from Jan 2006 - Jan 2009):

- Performance of algorithm compared against constituent primitive strategies and robustness w.r.t. some parameter settings
- Comparison against two baseline architectures:
  1. Max: Allocate funds to the best historical strategy
  2. **k**-NN:
     - Identify **k** historical states with similar profitability vector.
     - Use a forest of kd-trees (number of trees equals number of regimes/contexts)
     - Allocate funds as weighted average based on past out-sample performance

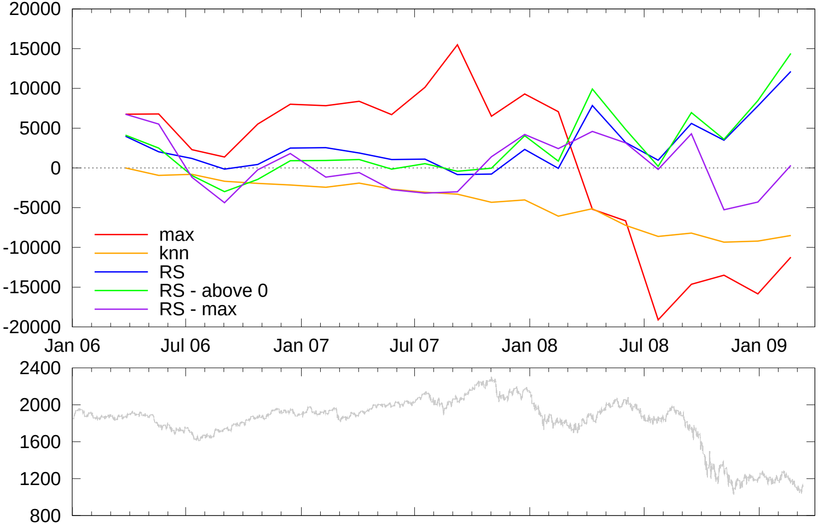
Performance of RED-STOP Algorithm Experiment 1





Out-of-sample profits over the period 2006-11-01 to 2008-08-28

Performance of RED-STOP Algorithm
Experiment 2



Discussion

- Relationship to alternate regime-switching models:
  - We could have directly modelled the switches in time-series using EM/MCMC techniques, but we find the models to be fragile in an on-line setting.
  - We claim that there are benefits in a more direct *'action-oriented'* state representation.
- What is the role of historical data? What happens in novel out-of-sample situations?
  - We use data to identify possible correlation patterns within strategy space
    – structure induced by latent dynamics
    – few parametric assumptions about details of latent dynamics
  - Structure in this space (e.g., low-dimensional regime subspaces) may be exploited to devise more efficient strategies.

Conclusions

- Framework for on-line multi-strategy trading.
- Utilization of market context:
  - Inferred from data
  - Represented in terms of directly measurable/diagnosable quantities

Future Work:

- Systematic empirical evaluation (across multiple markets)
- Explore alternatives for clustering primitive strategies and incorporate into probabilistic model of state estimation
- Risk-sensitive optimization and predictive-modelling

Appendix 1: Multiplicative Updates

Definition *Let Ω be a sample space, E, F its subsets, $Name$ a set of names. We define the following (partial) functions: a* quantifier[2]*: $q_E : E \rightarrow \mathbb{R}$; a* quantitative property*: $qp_E : State \times Time \rightharpoonup (E \rightarrow \mathbb{R})$. Then a* 0-property *is any quantitative property, and for $n > 0$, an* n-property *np_F is a probability distribution on m-properties $qp_{E_1}, \ldots, qp_{E_k}$ where $0 < m < n$, $F = \bigcup_i E_i$.*

Definition *Let for any $i, n \in \mathbb{N}$, $1 \leq j \leq n$, $L = [l_1, \ldots, l_n]$ be a list of real numbers, $p, p^{(L,i)}$ be probability distributions on L. Let p_j denote the probability of j-th event w.r.t. p. We define the symbols N_L, X_L denote the minimum, and maximum elements of L, and $E^p[L] = \sum_j p_j l_j$ to denote the expected value of L w.r.t. p. For $0 < w \leq 1$, we define the following:*

$$dp_j^{(L,i)} \overset{def}{=} w \frac{l_j - \sum_k p_k^{(L,i)} l_k}{X_L - N_L}, \quad rp_j^{(L,i)} \overset{def}{=} \begin{cases} p_j^{(L,i)} + p_j^{(L,i)} \cdot dp_j^{(L,i)} & \text{if } dp_j^{(L,i)} \leq 0 \\ p_j^{(L,i)} + (1 - p_j^{(L,i)}) \cdot dp_j^{(L,i)} & \text{if } dp_j^{(L,i)} > 0, \end{cases}$$

$$p_j^{(L,i+1)} \overset{def}{=} \frac{rp_j^{(L,i)}}{\sum_k rp_k^{(L,i)}}$$

Since $-1 \leq dp_j^{(L,i)} \leq 1$, $rp_j^{(L,i)} \geq 0$ for any j; hence $p^{(L,i+1)}$ is indeed a probability distribution. It represents the adjusted 1-property.

If the iterative step leads to bad results for a sufficiently long time, the whole algorithm is restarted.

Appendix 2: Convergence of Updates

Our task is to find a probability distribution p which maximizes $E^p[L]$ for a list L of real numbers. As $N_L \leq \sum_k p_k l_k \leq X_L$, any distribution mp such that $mp_k \geq 0$ if $l_k = X_L$, and $mp_k = 0$ otherwise, maximizes the expected value as $E^{mp}[L] = X_L$. We call such a distribution an $E[L]$-*maximizing probability distribution* and denote a set of such probability distributions by $\mathrm{MPD}(L)$.

We show that for any list L of real numbers and any probability distribution $p^{(L,1)}$ on L, the limiting distribution of sequence of probability distributions $\{p^{(L,i)}\}_i$ is an $E[L]$-maximizing probability distribution, *i.e.* $\lim_{i\to\infty} p^{(L,i)} \in \mathrm{MPD}(L)$.

Lemma 7. *Let L be a list of real numbers, and $p^{(L,i)}$ be as above for any $i \in \mathbb{N}$. Then $E^{p^{(L,i)}}[L] \leq E^{p^{(L,i+1)}}[L]$ with equality only in the case $p^{(L,i)} \in \mathrm{MPD}(L)$.*

Corollary 8. *Let L and $p^{(L,i)}$ be as in Lemma 7. Then $\lim_{i\to\infty} p^{(L,i)} \in \mathrm{MPD}(L)$. Furthermore, the sequence $\left\{E^{p^{(L,i)}}[L]\right\}_i$ is monotone and converges to X_L.*

Lemma 9. *Let K, L be lists of the length n of real numbers with the same ordering of their elements, and let $p^{(K,i)}$ be a probability distribution. Then $E^{p^{(K,i)}}[L] \leq E^{p^{(K,i+1)}}[L]$ with equality only in the case $p^{(K,i)} \in \mathrm{MPD}(L)$.*

Corollary 10. *For any two lists K, L of the length n of real numbers with the same ordering of their elements, the sets $\mathrm{MPD}(K)$ and $\mathrm{MPD}(L)$ are equal.*

Lemma 11. *Let K, L be as above, and $p^{(K,i)}, q^{(K,i)}$ be probability distributions on L. Then:*

$$\left|E^{p^{(K,i+1)}}[L] - E^{q^{(K,i+1)}}[L]\right| \leq \left|E^{p^{(K,i)}}[L] - E^{q^{(K,i)}}[L]\right|.$$

Corollary 12. *Let $\mathcal{K} = \{K_h\}_h, \mathcal{L} = \{L_h\}_h$ be two sequences of lists K_h, L_h of the length n with the same ordering of their elements, and let $p^{(K_1,1)}, q^{(K_1,1)}$ be two probability distributions. Then:*

$$\lim_{i\to\infty} E^{p^{(K_i,i)}}[L_i] = \lim_{i\to\infty} E^{q^{(K_i,i)}}[L_i].$$