Dynamic portfolio management with transaction costs

Alberto Suárez
Computer Science Department
Universidad Autónoma de Madrid (Spain)
alberto.suarez@uam.es

John Moody, Matthew Saffell
International Computer Science Institute
Berkeley, CA 94704, USA
moody,saffell@icsi.berkeley.edu

The selection of optimal portfolios is a central problem of great interest in quantitative finance, one that still defies complete solution. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. A drawback of the standard framework formulated by Markowitz [1] is that only one period is used in the evaluation of the portfolio performance. In fact, no dynamics are explicitly considered. Like in many other financial planning problems, the potential improvements of modifying the portfolio composition should be weighed against the costs of the reallocation of capital, taxes, market impact, and other state-dependent factors. The performance of an investment depends on a sequence of portfolio rebalancing decisions over several periods. This problem has been addressed using different techniques, such as dynamic programming [2, 5] stochastic network programming [3], tabu search [4], reinforcement learning [7], sequence prediction [10], online learning [11], and Monte Carlo methods [8, 9].

In this work, we develop a recurrent reinforcement learning (RRL) system that directly induces portfolio management policies from time series of asset prices and indicators, while accounting for transaction costs. The RRL approach learns a direct mapping from indicator series to portfolio weights, bypassing the need to explicitly model the time series of price returns. The resulting policies dynamically optimize the portfolio Sharpe ratio, while incorporating changing conditions and transaction costs. A key problem with many portfolio optimization methods, including Markowitz, is discovering "corner solutions" with weight concentrated on just a few assets. In a dynamic context, naive portfolio algorithms can exhibit switching behavior, particularly when transaction costs are ignored. We extend the RRL approach to produce better diversified portfolios and smoother asset allocations over time. The solutions proposed are to include realistic transaction costs and to shrink portfolio weights toward the prior portfolio.

The architecture of the learning system is depicted in Figure 1. The portfolio weights predicted by the policy are a convex combination of $\tilde{\mathbf{F}}_n$, the composition of the portfolio at t_n^-, prior to rebalancing, and $\mathbf{F}_n^{(S)}(\mathbf{w})$, the output of a softmax network whose inputs are a constant bias term, the information set, $\mathbf{I}_n$ (either lagged information from the time series of asset returns or external economic and financial indices) and the current portfolio weights $\tilde{\mathbf{F}}_n$:

$$\mathbf{F}_n(\lambda, \mathbf{w}) = \lambda \tilde{\mathbf{F}}_n + (1-\lambda)\mathbf{F}_n^{(S)}(\mathbf{w}) \tag{1}$$

The relative importance of these two terms in the final output is controlled by a hyperparameter $\lambda \in [0,1]$. For $\lambda = 0$, the final prediction is directly the output of the softmax network. In the absence of transaction costs, a new portfolio can be created at no expense. In this case, the currently held portfolio need not be used as a reference, and $\lambda = 0$ should be used. If transaction costs are nonzero, it is necessary to ensure that the expected return from dynamically managing the investments outweighs the cost of modifying the composition of the portfolio. The costs are deterministic and can be calculated once the new makeup of the portfolio is established. By contrast, the returns expected from the investment are uncertain. If they are overestimated (e.g. when there is overfitting) the costs will dominate and the dynamic management strategy seeking to maximize the returns by rebalancing will have a poor performance. A value $\lambda > 0$ causes the composition of the portfolio to vary smoothly, which should lead to improved performance in the presence of transaction costs.

The parameters of the portfolio management system are learned via RRL by either directly maximizing the wealth accumulated over the training period or by optimizing an exponentially smoothed Sharpe ratio, while taking into account transaction costs. The training algorithm is a variant of gradient ascent with learning parameter ρ extended to take into account the recurrent terms in (1) (see [6] for further details). The hyperparameters of the learning system (ρ, η, λ) can be determined by either holdout validation or cross-validation.

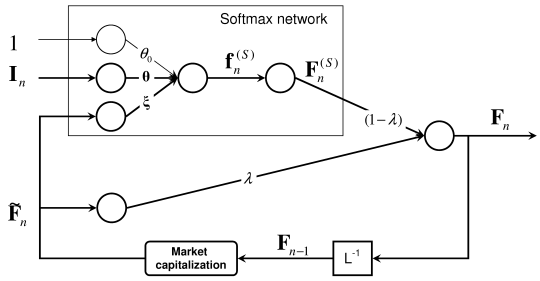

Figure 1: Architecture for the reinforcement learning system.

The performance of the reinforcement learning system is assessed on real market data and compared to the market portfolio (optimal if the market were ideally efficient) and to the tangency portfolio computed using the Markowitz framework portfolio, which is optimal in terms of the Sharpe ratio, assuming zero transaction costs. The experiments are carried out using the MSCI International Equity Indices (gross level) that measure the performance of different economic regions (indices for the Pacific, North America and Europe) and of the global market (the World index) [12].

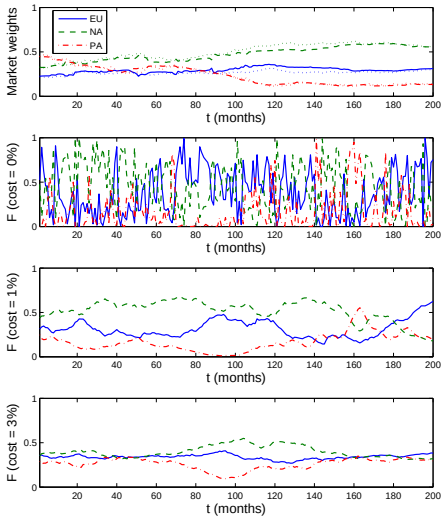

Figure 2: Evolution of portfolio weights for the market portfolio (top) and for optimal investment policy discovered by the reinforcement learning systems for different transaction costs (0%, 1% and 3% from the top down).

From the results obtained (see Figure 2), several important observations can be made. As anticipated, the policy learned in the absence of transaction costs involves a large amount of portfolio rebalancing. The switching observed for the portfolio weights is clearly undesirable in real markets, where transaction costs make this type of behavior suboptimal. By contrast, the policies learned by the RRL system when transaction costs are considered to be smoother and require much less rebalancing. Furthermore, the portfolios selected outperform the market portfolio (except for large transaction costs $\gtrsim 5\%$), and are well-diversified, which is in agreement with financial good practices. In conclusion, the use of the current portfolio composition as a reference in the reinforcement learning architecture considered in Fig. 1 is crucial to the identification of robust investment policies in the presence of transaction costs.

References

[1] Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[2] Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246, aug 1969.

[3] J. M. Mulvey and H. Vladimirou. Stochastic network programming for financial planning problems. *Management Science*, 38:1642–1664, 1992.

[4] F. Glover, J. M. Mulvey, and K. Hoyland. Solving dynamic stochastic control problems in finance using tabu search with variable scaling. In I. H. Osman and J. P. Kelly, editors, *MetaHeuristics: Theory and Applications*, pages 429–448. Kluwer Academic Publishers, 1996.

[5] Ralph Neuneier. Optimal asset allocation using adaptive dynamic programming. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 952–958. The MIT Press, 1996.

[6] John Moody, Lizhong Wu, Yuansong Liao, and Matthew Saffell. Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting*, 17(1):441–470, 1998.

[7] John Moody and Matthew Saffell. Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4):875–889, 2001.

[8] J.B. Detemple, R. Garcia, and M. Rindisbacher. A monte carlo method for optimal portfolios. *The Journal of Finance*, 58(1):401–446, 2003.

[9] Michael W. Brandt, Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud. A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies*, 18(3):831–873, 2005.

[10] Allan Borodin, Ran El-Yaniv, and Vincent Gogan. Can we learn to beat the best stock. *Journal of Artificial Intelligence Research*, 21:579–594, 2004.

[11] Amit Agarwal, Elad Hazan, Satyen Kale, and Robert E. Schapire. Algorithms for portfolio management based on the newton method. In *Proceedings of the 23rd international conference on Machine learning, ICML 2006*, pages 9 – 16, New York, NY, USA, 2006. ACM.

[12] MSCI Inc. http://www.mscibarra.com/products/indices/equity/.