Independent Factor Reinforcement Learning for Portfolio Management

Jian Li, Kun Zhang, and Laiwan Chan

Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
{jli,kzhang,lwchan}@cse.cuhk.edu.hk

Abstract. In this paper we propose to do portfolio management using reinforcement learning (RL) and independent factor model. Factors in independent factor model are mutually independent and exhibit better predictability. RL is applied to each factor to capture temporal dependence and provide investment suggestion on factor. Optimal weights on factors are found by portfolio optimization method subject to the investment suggestions and general portfolio constraints. Experimental results and analysis are given to show that the proposed method has better performance when compare to two alternative portfolio management systems.

1 Introduction

During the past decade, there have been growing number of researches that apply reinforcement learning (RL) [1, 2] techniques to solve problems in financial engineering [3]. What is of particular interest would be using RL to design financial trading systems. Neuneier used Q-learning algorithm to design a system for trading single asset [4]. The system was enhanced in [5] to enable multi-asset trading. However, the size of action space increases exponentially with the number of assets, hence it requires substantial amount of training data to determine policy for such a huge action space. Ormoneit and Glynn applied a kernel-based RL approach on single-asset trading problem [6]. In [7] Dempster and Leemans proposed a layered single-asset trading system where recurrent RL algorithm is used to offer trading recommendation. The recommendation is then evaluated by risk management overlay to make final decision.

One problem of current researches is that, most works only addressed the simple single-asset allocation problem, i.e. capital can either be kept in cash or invested in a risky asset. In practice, however, investors rarely adopt such an extreme strategy. Instead they normally make investments in a number of assets and take advantage of *diversification* to reduce investment risk. In this paper we aim to provide a competitive portfolio management strategy exploiting RL.

A simple "divide-and-conquer" approach using RL for portfolio management can be divided into two steps. First, RL is run separately on each available asset to obtain the Q-values of different actions; Second, asset weights are generated based on these Q-values. However, this approach is subject to two questions. First, the obtained weights may not be optimal in terms of profit as RL neglects the inter-relations between the returns of different assets. Second, whether it is good to apply RL directly on assets is still under question. This is because the prediction of future reward is important to RL's performance while it is well known that asset return is difficult to predict.

The independent factor model in finance [8, 9, 10, 11] may help to address the above two questions. In the independent factor model, the observed asset returns are believed to be linear mixtures of some hidden independent factors. Motivated by these works, we proposed to use RL on independent factors instead of on asset returns as in past works. As the factors are as independent as

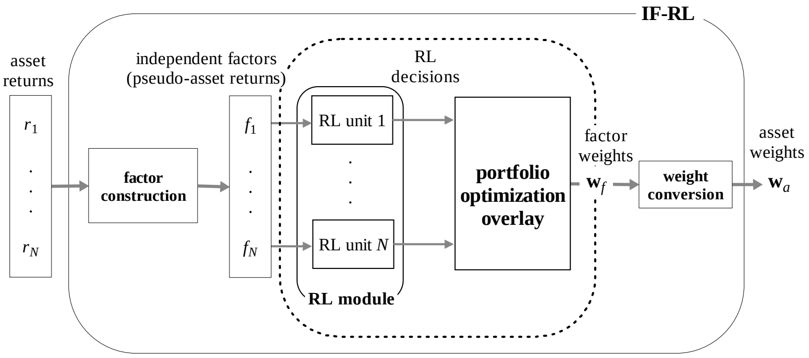


Fig. 1. The structure of IF-RL portfolio management system

possible, the inter-relations between them are almost negligible. Hence we no longer need to consider the inter-relations. Also there have been many research works believing that the independent factors are more structured and regular, and consequently can be predicted better [12, 13] than asset returns. In this way using independent factors can be expected to strengthen the usefulness of RL in portfolio management.

Therefore, in this paper, we propose to do portfolio management by virtue of RL and the independent factor model. This proposed system is named independent factor RL (IF-RL) for portfolio management. The independent factors can be estimated by using independent component analysis (ICA) [14], a statistical technique for revealing hidden factors underlying the observed signals with a linear transformation. The system implementation consists of four steps. Firstly, ICA is used to construct independent factors from asset returns. Secondly, as the factors are almost independent from each other, RL is run on all factors in parallel to obtain investment suggestions on factors. Thirdly, portfolio optimization method is used to find factor weights that optimize specific objective function subject to investment suggestions from RL and general portfolio constraints. Lastly, optimal asset weights are obtained by converting optimal factor weights.

The rest of the paper is organized as follows. In Sect.2 the design of the proposed IF-RL system is described in detail. In Sect.3, another RL-based portfolio management system without utilizing independent factors is formulated for comparison with the proposed system in later experiments. Experimental results and analysis are provided in Sect.4 to compare the performances of IF-RL system and two alternative portfolio management systems. Finally Sect.5 discusses some future works and concludes.

2 Proposed system

In this section, we describe in detail the design of the proposed IF-RL system. Fig.1 shows the structure of the system which can be divide into two parts. The inner part within the dotted-line block is a RL-based portfolio management model consisting of a RL module and a portfolio optimization overlay. This model is designed to operate on multiple assets. The outer part is composed of factor construction module and weight conversion module which are in charge of transformation between assets and independent factors. In the following text, we elaborate the two parts respectively.

2.1 Factor construction module

Factor construction module extracts factors from asset returns. In this paper we adopt the FastICA algorithm [15] to extract independent components from returns.

Assume that we can invest in a market consisting of N risky assets and risk-free cash. At time t, let $\epsilon_i\left(t\right)$ be the return of asset i at t, which is defined as

$$\epsilon_i\left(t\right) = \frac{p_i\left(t+1\right) - p_i\left(t\right)}{p_i\left(t\right)}$$

where $p_i\left(t\right)$ is the price of asset i at t. The asset returns at t can be summarized with a vector $\epsilon(t)$= $\left(\epsilon_1\left(t\right), \ldots, \epsilon_N\left(t\right)\right)^T$, and for $t = 1, \ldots, T$, a return matrix $\epsilon = (\epsilon(1), \ldots, \epsilon(T))$ can be formed where each row represents the historical returns of a single asset. In the independent factor model, the returns $\epsilon_1\left(t\right), \ldots, \epsilon_N\left(t\right)$ are assumed to be linear combinations of some independent factors. To recover the independent factors, ICA uses the linear transformation

$$\mathbf{f}(t) = \mathbf{B}\epsilon(t) \tag{1}$$

where $\mathbf{f}(t) = (f_1(t), f_2(t), \cdots, f_N(t))^T$ with $f_i(t)$ being the ith recovered factor at t[1], and the matrix $\mathbf{B}$ is the de-mixing matrix for $\epsilon(t)$.

With a proper de-mixing matrix $\mathbf{B}$, we can implement the factor construction module. In IF-RL system, we consider the factors as returns of some pseudo-assets [2].

2.2 Weight conversion module

The weight conversion module converts the optimal factor weights obtained from the inner part to corresponding asset weights. Let the asset weights and factor weights be $\mathbf{w}_a$ and $\mathbf{w}_f$ respectively. We have

$$\mathbf{w}_a^T\epsilon(t) = \mathbf{w}_f^T\mathbf{f}(t)$$

By substituting Eq.(1) into the above equation, we can have the relation between $\mathbf{w}_a$ and $\mathbf{w}_f$

$$\mathbf{w}_a = \mathbf{B}^T\mathbf{w}_f \tag{2}$$

Like asset weights, factor weights are also subject to some portfolio constraints when utilized in portfolio management task. The general portfolio constraints on asset weights can be stated as[3]

$$\sum_{i=1}^{N} w_{ai} \leq 1 \text{ and } \forall i = 1, \ldots, N \ w_{ai} \geq 0 \tag{3}$$

where w_{ai} is the asset weight on asset i. The sum of asset weights is set to be no bigger than 1 as there may be some capital allocated in cash. The equation can be rewritten in matrix form as

$$[1]_N^T\mathbf{w}_a \leq 1 \text{ and } \mathbf{w}_a \geq [0]_N$$

where $[1]_N$ and $[0]_N$ are respectively N-dimension vector of all 1's and all 0's. By utilizing Eq.(**??**) and Eq.(2), general portfolio constraints on factor weights can be specified as

$$[1]_N^T\mathbf{w}_f \leq 1 \text{ and } \mathbf{B}^T\mathbf{w}_f \geq [0]_N \tag{4}$$

These constraints will be used in portfolio optimization with respect to factor weights (see Sect.2.4).

[1] For simplicity, we assume that the number of factors is equal to that of the assets.

[2] In the rest of this paper, for simplicity, we use factor in the place of pseudo-asset unless noted otherwise.

[3] Please be noted that in this paper we assume non-negative asset weights to disallow short-selling.

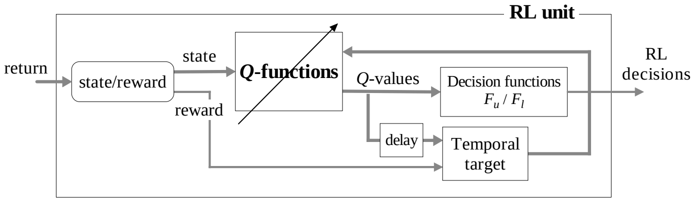


Fig. 2. The structure of RL unit

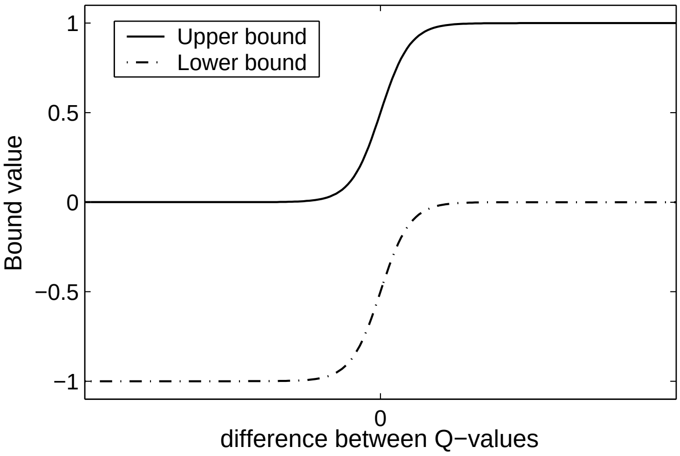


Fig. 3. Example of upper and lower bound on factor weight generated with RL

2.3 RL module

The RL module consists of many RL units, which structure is shown in Fig.2, with each unit operating on one factor. RL units are run in parallel, and output RL decisions on corresponding factors. In practice, it is common for investors to set constraints on the proportion of specific asset in the portfolio. Similarly, in this paper we interpret RL decisions as suggestions on degree of investment of factors, i.e. constraints on factor weights.

In each RL unit, we solve a single-asset allocation problem trading with one factor. The available actions are -1 and 1, representing respectively short and long position. We choose this action setting because we can see from Eq.(4) that factor weights may be negative, which indicates short-selling of the factors. At time t, for factor i, the state $s_{it} = (\$_{it}, k_{it})$ consists of two parts: $\$_{it}$ describes the market impact which is independent of investor's decision; while $k_{it} \in \{-1, 1\}$ represents the current investment position (short or long). Within each RL unit, we maintain Q-values of the binary actions. Q-values represent the expected future return of applying specific action at given state, and can be updated with the following formula during training:

$$Q_a(s) \leftarrow (1-\eta)\, Q_a(s) + \eta\, (r + \gamma \max(Q_{-1}(s'), Q_1(s')))$$

where $\langle s, a, r, s' \rangle$ is the observed tuple of current state, applied action, perceived reward and next state, η is learning rate, $0 \le \gamma \le 1$ is the discount factor in RL, and $Q_a(s)$ is estimated Q-value of applying action a at state s. In trading system, generally the reward can be represented as the capital gain subtracted by the transaction cost, i.e. at time t, for factor i, we have $r_{it} = g_{it} + c_{it}$, where g_{it} is the change of total capital during $[t, t+1]$ due to the price variations, and c_{it} the commission charge for traing at t , if applicable. For convenience, transaction cost is assumed non-positive to denote the *paid* charge. In the context of factor trading, the capital gain and transaction cost can be computed with

$$
\begin{aligned}
g_{it} &= \log\left(1 + a_{it} \cdot f_i(t)\right) \\
c_{it} &= \log\left(1 - \delta \cdot \sum_{j=1}^{N} |b_{ij}| \cdot |k_{it} - a_{it}|\right)
\end{aligned}
$$

where δ is transaction cost rate of asset trading.

At time t, for factor i, the optimal action a_{it}^* can be determined via

$$a_{it}^* = \operatorname{sgn}\left(d\left(s_{it}\right)\right) \tag{5}$$

where $d(s_{it}) = Q_1(s_{it}) - Q_{-1}(s_{it})$ is the difference between Q-values at s_{it}, and sgn() is the sign function. In RL module, we use two sigmoid-shape functions $F_u()$ and $F_l()$ (see Eq.(6)) to generate decisions on upper bound and lower bound of factor weight to control the weight from approaching boundary values of -1 and 1.

$$
\begin{aligned}
F_u\left(d\left(s_{it}\right)\right) &= \frac{1}{2}\left(1 + \tanh\left(N_u \cdot d\left(s_{it}\right)\right)\right) \\
F_l\left(d\left(s_{it}\right)\right) &= \frac{1}{2}\left(-1 + \tanh\left(N_l \cdot d\left(s_{it}\right)\right)\right)
\end{aligned} \tag{6}
$$

where N_u and N_l are respectively upper/lower bound parameter. The outputted RL decisions can be stated as

$$\forall i = 1, \ldots, N \quad F_l\left(d\left(s_{it}\right)\right) \leq w_{fi} \leq F_u\left(d\left(s_{it}\right)\right) \tag{7}$$

An example of upper and lower bound is shown in Fig.3, we can see from the figure that when RL prefers action 1 (or -1), the difference between Q-values is a positive (or negative) value, the upper and lower bound approaches to 1 and 0 (or 0 and -1) respectively with greater preference, i.e. bigger absolute value of difference between Q-values.

2.4 Portfolio optimization overlay

The portfolio optimization overlay can be implemented with various portfolio optimization methods. In this paper's experiments, we use mean-variance optimization (MVO) model, which was initialized by Markowitz in his landmark paper [16] and may be the most renowned portfolio optimization method. In MVO, the objective function to be maximized can be expressed as

$$U\left(\mathbf{w}_a\right) = \mathbf{w}_a^T \overline{\epsilon} - u \mathbf{w}_a^T \mathbf{\Sigma} \mathbf{w}_a$$

where $\overline{\epsilon}$ and $\mathbf{\Sigma}$ are respectively expected return and covariance matrix, $u \geq 0$ is the risk aversion. By utilizing Eq.(2), we can state the optimization problem in IF-RL's portfolio optimization overlay as maximizing the following objective function subject to constraints in Eq.(4) and Eq.(7)

$$U\left(\mathbf{w}_f\right) = \mathbf{w}_f^T \mathbf{B} \overline{\epsilon} - u \mathbf{w}_f^T \mathbf{B} \mathbf{\Sigma} \mathbf{B}^T \mathbf{w}_f$$

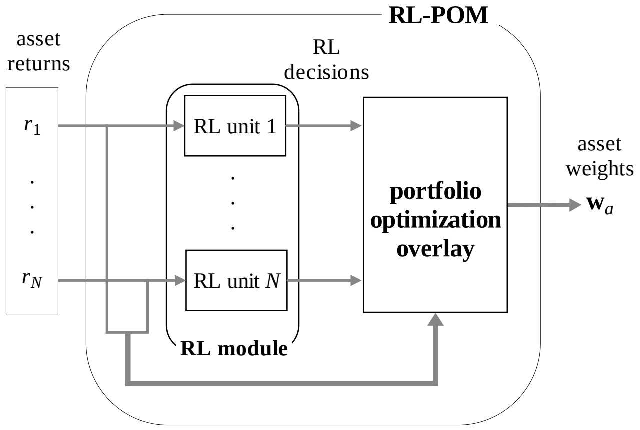


Fig. 4. The structure of RL-POM system

3 Another RL-based portfolio management system

In this section, we formulate another RL-based portfolio management system as shown in Fig.4. We can find from the figure that this system has a structure similar to IF-RL system, except that RL is used directly on asset returns instead of independent factors. The system is named RL-POM as it combines RL module and portfolio optimization overlay. In the sysem, RL is run separately on inter-related asset returns, which may lead to suboptimal portfolio. Also the poor predictability of asset returns will limit the usefulness of RL in portfolio management. The RL-POM system will be used for comparison with the IF-RL system in experiments to investigate the advatange of applying RL on independent factors.

In the RL-POM system, the RL module is also composed of many RL units. RL units provide RL decisions on degree of investment of assets. Since the asset weight lies in the range $[0, 1]$, the available actions are set as 0 (invest in cash) and 1 (invest in risky asset). At time t, for asset i , let the Q-values of binary actions be $Q_0\left(s_{it}\right)$ and $Q_1\left(s_{it}\right)$ respectively, the optimal action a_{it}^* is

$$a_{it}^* = H\left(d\left(s_{it}\right)\right) \tag{8}$$

where $d\left(s_{it}\right) = Q_1\left(s_{it}\right) - Q_0\left(s_{it}\right)$ is the difference between Q-values at s_{it}, and $H\left(\right)$ is heaviside step function. In RL-POM we only need to provide decision on upper bound of asset weights with

$$F_u\left(d\left(s_{it}\right)\right) = \frac{1}{2}\left(1 + \tanh\left(N_u \cdot d\left(s_{it}\right)\right)\right)$$

The outputted RL decisions are

$$\forall i = 1, \ldots, N \quad w_{ai} \leq F_u\left(d\left(s_{it}\right)\right) \tag{9}$$

The portfolio optimization overlay directly find the asset weights $\mathbf{w}_a$ that optimize specific objective function while subject to constraints in Eq.(3) and Eq.(9).

4 Experimental results and analysis

Experiment on real stock data in Hong Kong market is provided in this section to illustrate the performance of proposed IF-RL portfolio management system. The experiment is carried out by

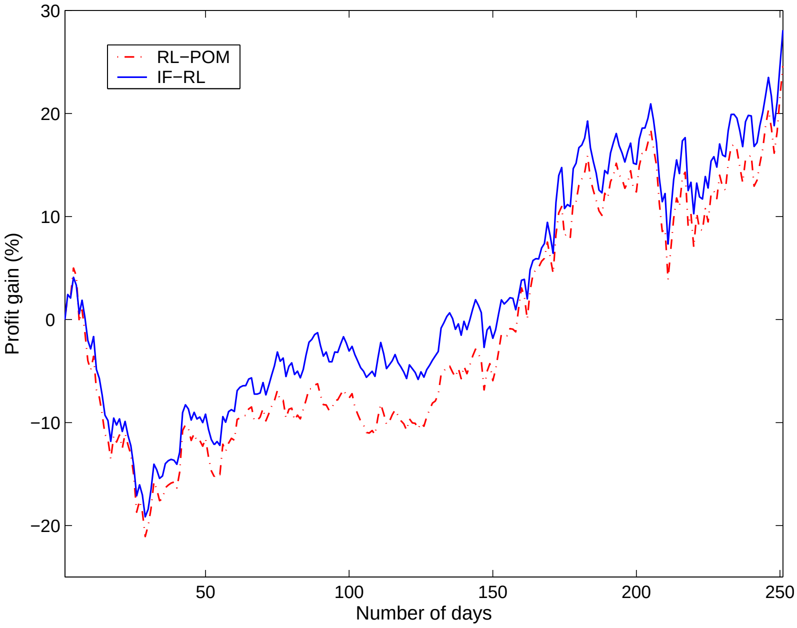


Fig. 5. Profit gains of the two portfolio management systems in the testing stage

investing in a portfolio of 8 stocks[4]. The experimental data consists of 1000 data points from April 22, 2003 to May 4, 2007. The first 750 data points are used as training data set, whereas the remaining 250 days compose the testing stage. The transaction cost rate δ is set as 0.5%.

4.1 Experimental results

In the experiment, we compare the performances of two portfolio management systems, i.e. RL-POM system and IF-RL system, described respectively in previous two sections. For both systems, MVO is chosen to implement portfolio optimization overlay with risk aversion $u = 1$, and the upper/lower bound parameter N_u and N_l are both set as 100.

For both systems, the 250-day testing stage is divided into 5 segments, and asset weights are rebalanced at the beginning of each segment. In the testing stage, the tested segment will be added to the training data set before moving forward to next segment; and the RL modules in RL-POM and IFRL system continue to train their trading policies with newly-observed data. This mechanism enables the systems adaptive to the changes in the dynamic market.

Fig.4 shows the profit gains of the three systems in the testing stage. The proposed IF-RL system can be found to outperform the RL-POM systems in terms of profitability. We also notice that IF-RL system can control the loss better when asset prices decrease, e.g. around day 50 to 60.

Table 1 provides more performance statistics including mean return, risk, Sharpe Ratio, etc. All the performance statistics are measured on the testing stage, and risk is defined as variance of the trading return. In the table, dod stands for *degree of diversification* [17]. This measure tells how well

[4] The 8 stocks are all constituent stocks of Hang Seng Index in Hong Kong market. They are 0002-0003.HK, 0005-0006.HK, 0011.HK, 0013.HK, and 0016-0017.HK.

Table 1. Performances of the three portfolio management systems

system name	profit (%)	mean return	risk	Sharpe Ratio	dod
RL-POM	24.59	9.92e-4	2.25e-4	6.61e-2	0.4978
IFRL	28.10	1.10e-3	2.12e-4	7.54e-2	0.5976

the system diversifies its investments. It is computed with

$$\text{dod} = \frac{1}{N} \sum_{m=1}^{N} \mathbf{w}_m^T \left[[1]_N - \mathbf{w}_m \right]$$

where N is the number of segments, and $\mathbf{w}_m$ is the asset weights determined at the beginning of segment m.

From the table we can have two observations. First, the IF-RL system achieves a higher mean return, lower risk and better Sharpe Ratio when compared to the RL-POM system. These results can provide some positive evidences for the conjecture that independent factors may have better predictability than asset returns. Second, in terms of dod, the IF-RL system can achieve a more diversified portfolio than the RL-POM system does.

4.2 Analysis on portfolio formation

Besides the performance measures discussed above, we are also interested in the optimal asset weights found by the three systems. In Fig.6 we show the asset weights of the two systems determined at the beginning of the 5 segments. Among all the 8 assets, there are 7 assets selected by at least one system during the testing stage. We use a clustering of 7 bars to represent asset weights, where the bar height is equal to the weight on the corresponding asset. For those bars with non-zero height, asset indexes are marked on top of them.

It can be observed from Fig.6 that asset 8 is consistently selected by both systems as a major component. This can be contributed to this asset's significant profitability when compared to other 7 assets. Despite this similarity in constructing portfolio, the portfolios found by the IF-RL system are still different from those found by RL-POM system in terms of the minor portfolio components. This may be the reason why IF-RL system can achieve better performance than the RL-POM system. To further demonstrate this, we pick some example segments to show how different asset selections by IF-RL and RL-POM system lead to different performances.

In segment 2, while both systems choose to invest in asset 3 and 8, IF-RL system also invests in asset 1 and prefers it to asset 3. Fig.7(a) shows the normalized prices[5] of asset 1, 3 and 8 during segment 2. We can see that asset 1 outperforms asset 3 in terms of profit, indicating a better choice of IF-RL system. The opposite trends of asset 1 and 3 in the middle part of the segment show that adding asset 1 to portfolio can effectively reduce the portfolio risk. In segment 5, both system invest similar amount of capital in asset 8, but IF-RL system chooses to diversify the remaining capital in asset 1, 4 and 5 while RL-POM invests only in asset 1. Fig.7(b) depicts the normalized prices of asset 1, 4 and 5 in segment 5. We can see that asset 4 and 5 outperform asset 1, which indicates that the diversification decision of IF-RL system is correct.

The analysis of portfolio formation shows that with the assistance of independent factors, IF-RL system can take advantage of the better predictability of factors and find better-performing portfolios.

[5] Here the asset prices are normalized so that prices at the beginning of the segment are 1, the same normalization is also used for segment 5.

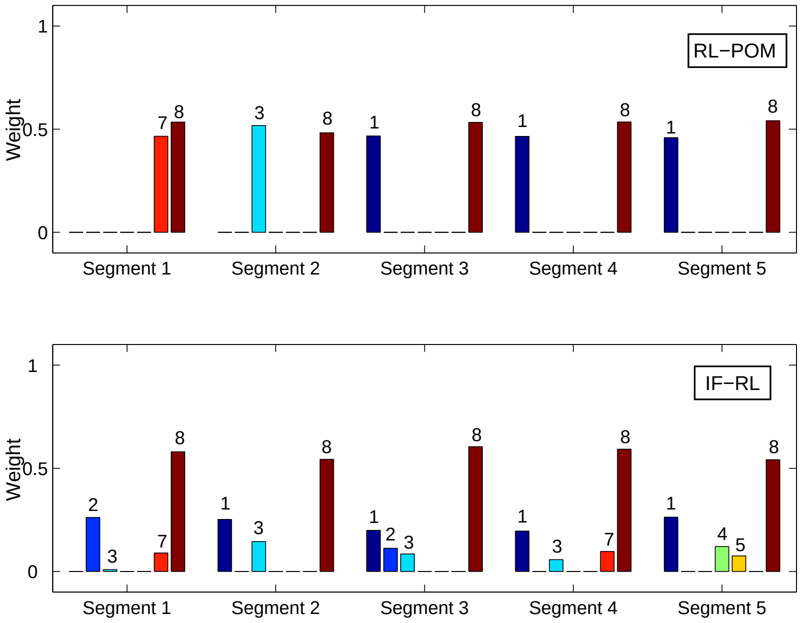
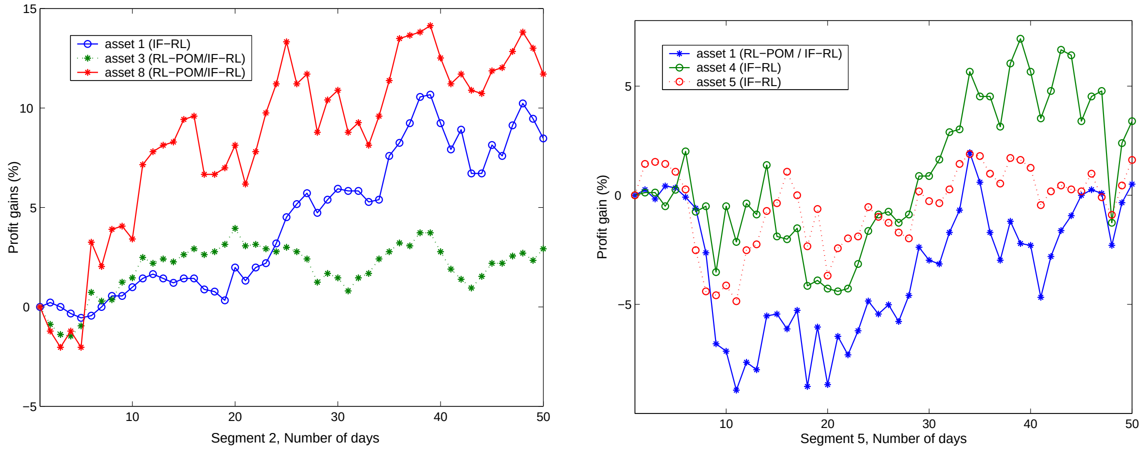


Fig. 6. Optimal asset weights found by two portfolio management systems for five segments



(a) Segment 2: asset 1 (IF-RL), 3 and 8 (RL-POM and IF-RL)

(b) Segment 5: asset 1 (RL-POM and IF-RL), 4 and 5 (IF-RL)

Fig. 7. Normalized prices of assets selected by different systems in two example segments

5 Conclusion and future work

In this paper, we propose Independent Factor RL (IF-RL) system for portfolio management. With the assistance of independent factors, we can operate RL on all factors in parallel, which enables an efficient system structure. Also, combining independent factors with RL can take advantages of both techniques: RL has good forecasting power, while independent factors are believed to have better predictability than asset returns. Experimental results on real stock data in Hong Kong market show that IF-RL system achieves better trading performance than the comparative MVO model and RL-POM system. Analysis on portfolio formation shows that IF-RL system attempt to find better-performing portfolio that is different in formation from the portfolios found by other two systems, thus demonstrating the usefulness of independent factors.

Future work may include using other ICA techniques to extract independent factors, as well as applying IF-RL system on more data sets with more optimization criterions such as those related to downside risk.

References

[1] Kaelbling, L.P., Littman, M.L., Moore, A.P.: Reinforcement learning: A survey. Journal of Artificial Intelligence **4** (1996) 237–285

[2] Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press, Cambridge, Massachusetts (1998)

[3] Roy, R.V.: Neuro-dynamic programming and reinforcement learning for finance. In: Proc. Computational Finance. (1999)

[4] Neuneier, R.: Optimal asset allocation using adaptive dynamic programming. In Touretzky, D.S., Mozer, M., Hasselmo, M.E., eds.: NIPS, MIT Press (1995) 952–958

[5] Neuneier, R.: Enhancing q-learning for optimal asset allocation. In Jordan, M.I., Kearns, M.J., Solla, S.A., eds.: NIPS, The MIT Press (1997) 936–942

[6] Ormoneit, D., Glynn, P.: Kernel-based reinforcement learning in average-cost problems: An application to optimal portfolio choice. In: NIPS. (2000) 1068–1074

[7] Dempster, M.A.H., Leemans, V.: An automated fx trading system using adaptive reinforcement learning. Expert systems with applications: special issue on financial engineering **30** (2006) 534–552

[8] Back, A.D.: A first application of independent component analysis to extracting structure from stock returns. International Journal of Neural Systems **8**(4) (August 1997) 473–484

[9] Kiviluoto, K., Oja, E.: Independent component analysis for parallel financial time series. In: Proc. ICONIP'98. Volume 2., Tokyo, Japan (1998) 895–898

[10] Cha, S.M., Chan, L.W.: Applying independent component analysis to factor model in finance. In: Proc. of IDEAL. (2000) 538–544

[11] Chan, L.W., Cha, S.M.: Selection of independent factor model in finance. In: 3rd International Conference on Independent Component Analysis and blind Signal Separation, San Diego, California, USA (Dec 2001)

[12] Pawelzik, K., Müller, K.R., Kohlmorgen, J.: Prediction of mixtures. In: Proc. Int. Conf. on Artificial Neural Networks (ICANN'96), Springer (1996) 127–132

[13] Malaroiu, S., Kimmo, K., Oja, E.: Time series prediction with independent component analysis. In: Proc. Int. Conf. on Advanced Investment Technology, Gold Coast, Australia (2000)

[14] Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, Inc (2001)

[15] Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks **10(3)** (1999) 626–634

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

[17] Hung, K.K., Cheung, Y.M., Xu, L.: An extended asld trading system to enhance portfolio management. IEEE Transactions on Neural Networks **14**(2) (Mar 2003) 413–425