Comput Econ
DOI 10.1007/s10614-015-9490-y

Using a Genetic Algorithm to Improve Recurrent Reinforcement Learning for Equity Trading

Jin Zhang · **Dietmar Maringer**

Accepted: 24 January 2015

Abstract Recurrent reinforcement learning (RRL) has been found to be a successful machine learning technique for building financial trading systems. In this paper, we use a genetic algorithm (GA) to improve the trading results of a RRL-type equity trading system. The proposed trading system takes the advantage of GA's capability to select an optimal combination of technical indicators, fundamental indicators and volatility indicators for improving out-of-sample trading performance. In our experiment, we use the daily data of 180 S&P stocks (from the period January 2009 to April 2014) to examine the profitability and the stability of the proposed GA-RRL trading system. We find that, after feeding the indicators selected by the GA into the RRL trading system, the out-of-sample trading performance improves as the number of companies with a significantly positive Sharpe ratio increases.

Keywords Artificial intelligence · Algorithmic trading · Recurrent reinforcement learning · Genetic algorithm · Indicator selection · Sharpe ratio

1 Introduction

Research on algorithmic trading is ongoing. In this context, recurrent reinforcement learning (RRL), an online learning technique which finds approximate solutions to stochastic dynamic programming problems, is used by researchers to tune financial trading systems for the purpose of utility maximization (see Moody et al. 1998). Much of the research on RRL trading has focused on high-frequency FX trading (see Gold 2003; Dempster and Leemans 2006), where the RRL technique has been found to be particularly good at making profits by discovering autocorrelations in the price

J. Zhang (✉) · D. Maringer
Faculty of Economics and Business Administration, University of Basel, Peter Merian-Weg 6, Basel, Switzerland
e-mail: jin.zhang@unibas.ch

Published online: 04 February 2015

changes. In addition to FX currencies, researchers also applied the RRL technique to trade other financial assets. For example, discussions on RRL applications for daily trading of stock indices can be found in the literature (see Bertoluzzo and Corazza 2007; Gorse 2011). However, there has been little research on whether RRL is a profitable strategy for low frequency trading (e.g. daily trading) of public corporations' common stocks.

In the computer science literature, artificial intelligence (AI) techniques have been increasingly applied to technical analysis for equity trading. For example, Allen and Karjalainen (1999) applied a genetic algorithm (GA) to technical trading rule optimization. Dempster et al. (2001) tested a reinforcement learning approach, a genetic programming approach, a Markov-Chain linear programming method, and a simple heuristic to find a profitable trading rule defined in terms of a set of technical indicators for FX trading. Dempster and Leemans (2006) presented an adaptive reinforcement learning system using 14 commonly-used technical indicators as a part of the system inputs; however, they found that using such technical indicators did not increase trading profits, except when a small number of lagged returns were fed into the system. Creamer (2012) applied a Logitboost method to technical trading rule optimization for stocks and futures trading.

Although AI techniques have been widely used in research on financial trading platform design, studies have tended to focus on the optimization of trading signals in technical analysis, in isolation from financial fundamentals and other available tools. In the real world, however, information such as financial news, company earnings, pre-market and after-hours data, market research, and company analysis are commonly used to analyze price movements. Various approaches based on fundamental analysis (Graham and Zweig 2003) have been developed to forecast future trends in stock prices. Tools from financial engineering have also been used to facilitate price movement analysis. Therefore, the trading results of RRL-type trading systems which use only return information to produce trading signals may be improved by feeding additional information other than just returns into the trading system.

In this paper, we therefore present a GA-RRL trading system which includes ten indicator variables (in addition to return information) as a part of the inputs of the trading system. As input variables have a direct impact on trading signals, in order to reduce the noise at the RRL input terminal, we feed only those indicators which are likely to improve trading performance into the RRL trading system. The proposed system consists of a GA selector and a RRL trading system. The GA finds a combination of Boolean numbers, in which a Boolean number indicates the inclusion/exclusion of an indicator. The RRL trading system produces long/short signals based on the selected indicators and the return information. In addition to the GA selection, we use a population-based parameter update scheme which is different from the parameter update scheme as suggested in the literature, to reduce over-fitting on out-of-sample trading profits. Our research considers whether the inclusion of selected indicator variables as a part of the system inputs improves trading performance of the GA-RRL trading system.

The paper is organized as follows. Section 2 describes the proposed GA-RRL trading system. Section 3 describes the data and parameter settings. Section 4 provides the results of our experiment; and Sect. 5 concludes the paper.

2 A Proposed GA-RRL Trading System

2.1 Recurrent Reinforcement Learning

Recurrent reinforcement learning has been used by Moody et al. (1998) to tune financial trading systems for the purpose of utility maximization. The RRL technique of Moody et al. (1998) is a stochastic gradient ascent algorithm which continuously optimizes a utility measure by using instant market information. In most discussions of RRL-type trading systems, the market information used is usually a series of lagged returns, despite the fact that RRL trading systems can easily incorporate technical indicators and financial fundamentals into the system inputs (Zhang and Maringer 2013).

The RRL trading system of Moody et al. (1998) is designed to trade a single asset with a two-position action (long/short), which is produced using linear combinations of returns and a tanh function. In Fig. 1, x, v and F_t refer to the inputs of the trading system, and θ_t denotes a parameter set of the input signals (hereafter referred to as the signal parameter set).

Denoting a utility function as U_t which depends on the realized return R_t at time t, the aim of the trading system is to maximize the wealth function U_t by continuously adjusting the signal parameter set θ_t:

$$\max U_t(R_t; \theta_t). \tag{1}$$

By denoting the closing price as P_t, the price change of an asset is defined as $r_t = \ln \frac{P_t}{P_{t-1}}$. In this study, we assume that the cost of daily risk-free lending and borrowing is zero. At time t, the realized trading profit can be written as (see Moody et al. 1998)

$$R_t = \nu \cdot \left(\operatorname{sgn}(F_{t-1}) \cdot r_t - \delta \cdot \left| \operatorname{sgn}(F_t) - \operatorname{sgn}(F_{t-1})\right|\right), \tag{2}$$

where ν is the number of shares, δ is the transactions cost rate, $\operatorname{sgn}(F_{t-1})$ refers to the current holding position, $\operatorname{sgn}(F_t)$ denotes the holding position on the following day, and F_t is written as:

$$F_t = \tanh\left(\theta_t \times I_t\right), \tag{3}$$

where I_t denotes a set of input signals, which usually includes a prior trading signal F_{t-1}, a constant v with a value of 1, and the market information, i.e. a set of lagged

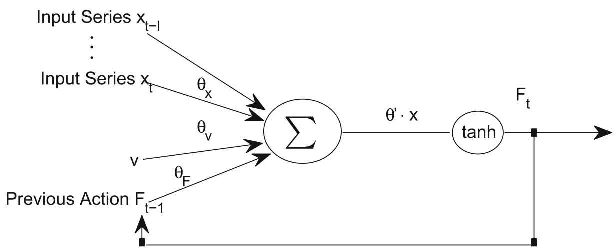

Fig. 1 Recurrent reinforcement learning

returns $r_t, r_{t-1}, r_{t-2}, \ldots, r_{t-l+1}$ ($t = 1, \ldots, T$, l is an integer number representing the length of the return lags). As the signal parameters θ_t are updated using the stochastic gradient ascent, the gradients of U_t with respect to the signal parameter set θ_t is written as:

$$\frac{dU_t(\theta_t)}{d\theta_t} = \frac{dU_t}{dR_t}\left\{\frac{dR_t}{dF_t}\frac{dF_t}{d\theta_t} + \frac{dR_t}{dF_{t-1}}\frac{dF_{t-1}}{d\theta_{t-1}}\right\}, \tag{4}$$

with

$$\frac{dF_t}{d\theta_t} \approx \frac{\partial F_t}{\partial \theta_t} + \frac{\partial F_t}{\partial F_{t-1}}\frac{dF_{t-1}}{d\theta_{t-1}}, \tag{5}$$

$$\frac{dR_t}{dF_{t-1}} = \nu \cdot (r_t + \delta \cdot \operatorname{sgn}(F_t - F_{t-1})), \tag{6}$$

and

$$\frac{dR_t}{dF_t} = -\nu \cdot \delta \cdot \operatorname{sgn}(F_t - F_{t-1}). \tag{7}$$

The differential Sharpe ratio (DSR) which is closely related to the exponential moving average Sharpe ratio (EMSR), has been widely used as a utility measure in RRL-type trading systems. The EMSR was defined by Moody et al. (1998) as:

$$EMSR_t = \frac{A_t}{K_\eta \cdot (B_t - A_t^2)^{1/2}}, \tag{8}$$

where $A_t = A_{t-1} + \eta \cdot \left(R_t - A_{t-1}\right)$, $B_t = B_{t-1} + \eta \cdot (R_t^2 - B_{t-1})$, and $K_\eta = \left(\frac{1-\eta/2}{1-\eta}\right)^{1/2}$. In other words, the RRL trading system aims to maximize the increment of the EMSR at t in a continuous manner.

The utility function U_t, or the DSR (i.e. the first-order term from the Taylor series expansion of EMSR at η approaching zero, see Moody et al. 1998) is defined as:

$$DSR_t = \frac{B_{t-1} \cdot (R_t - A_{t-1}) - \frac{1}{2} \cdot A_{t-1} \cdot (R_t^2 - B_{t-1})}{(B_{t-1} - A_{t-1}^2)^{\frac{3}{2}}}. \tag{9}$$

Therefore, the derivative of U_t with respect to R_t in Eq. (4) is written as:

$$\frac{dU_t}{dR_t} = \frac{B_{t-1} - A_{t-1} \cdot R_t}{\left(B_{t-1} - A_{t-1}^2\right)^{3/2}}. \tag{10}$$

At time t, the signal parameter set θ_t is updated by using $\theta_t = \theta_{t-1} + \rho \cdot \frac{dU_t(\theta_t)}{d\theta_t}$, where ρ is the learning step.

2.2 An Average Elitist Update Scheme

A major concern with RRL-type trading systems is whether the signal parameters are well tuned before real trading. In the literature, designers of RRL-type trading systems have suggested that the trading system (a single trader system) should be tuned first during a training period. The trades made and the performance of the trading system in the training period, are then used to update the parameters of the trading system. The training of the trading system is repeated for numbers of epochs before out-of-sample trading. In the out-of-sample period, the trades are the actual trades for the period, and the update of the parameters is continuously guided by RRL. Unlike high-frequency trading (where tick-by-tick data sets are available), with daily trading, the use of large quantities of daily historical data to tune the trading system, however may hamper the effective prediction of short-term price movements. Only a limited number of recent observations may carry information which is valuable to traders according to the efficient market hypothesis (see Fama 1970).

Maringer and Ramtohul (2012) suggested that, in order to fully search the solution space of signal parameters, a RRL-type trading system should be population-based (i.e. a multiple-trader system). Therefore, the RRL trading system in Maringer and Ramtohul (2012) consists of a group of traders whose signal parameters are initialized using random numbers. These simulation traders are trained with approximately 10 years of daily data. A simulation trader which produces the highest Sharpe ratio (SR) in the training period (the 'elitist' trader) is then selected from the group for out-of-sample trading. However, according to Maringer and Ramtohul (2012), the elitist trader does not necessarily outperform other simulation traders in terms of out-of-sample trading profits, a fact which may be explained by the over-fitting problem (see Dempster et al. 2001).

To reduce the impact of over-fitting on out-of-sample trading profits, we use a new parameter update scheme (the 'average elitist' scheme) to update the signal parameters. In this paper, our RRL trading system is also population-based, i.e. a multiple-trader trading system. The trading system consists of a group of simulation traders (sim-traders) rather than a single trader to derive the signal parameters for out-of-sample trading. Because real world traders usually have different levels of information asymmetry and heterogeneous expectations of price movements, we initialize the signal parameters of these sim-traders using random numbers. These sim-traders are then trained with the data from a training period. The training period includes an evaluation period which contains latest market information before out-of-sample trading. At the end of the training period, we set up an elitist set $\mathcal{E}$, in which members are selected according to their SR rankings in the evaluation period.

For the average elitist update scheme, the signal parameters are the average of the elitist members' signal parameters in the set $\mathcal{E}$. More specifically, assuming that the elitist set $\mathcal{E}$ consists of a number of N elitist members, at the beginning of the out-of-sample period $\mathrm{t}_0 = T_{train} + 1$, the signal parameter set under the average elitist update scheme is defined as:

$$\widehat{\theta}_{\mathrm{t}_0} = \frac{1}{N} \sum_{i} \theta_{\mathrm{t}_0,i}, \qquad i \in \mathcal{E}. \tag{11}$$

The update of signal parameters in the out-of-sample trading period follows:

$$\widehat{\theta}_{t+1} = \widehat{\theta}_t + \rho \frac{1}{N} \sum_i \frac{dU_t(\theta_{t,i})}{d\theta_{t,i}}, \qquad i \in \mathcal{E}. \tag{12}$$

In this paper, we refer to the trader whose signal parameters' update follows Eqs. (11) and (12) as the average elitist trader. Based on earlier research (see Zhang and Maringer 2013; Maringer and Zhang 2014), we find that the average elitist trader outperforms the single elitist trader as suggested by Maringer and Ramtohul (2012) in terms of both profitability and stability.

2.3 A Genetic Algorithm for Indicator Selection

In the computer science field of artificial intelligence, genetic algorithms (GA) which were first proposed by Holland (1975) in the 1970s, have been successfully applied to trading rule optimization. GAs are search algorithms inspired by biological evolution which involves three stages: natural selection; crossover; and mutation. In this study, each gene of a chromosome is a Boolean number indicating the inclusion/exclusion of the corresponding indicator variable. In the natural selection process, we use a roulette wheel selection. As the SR can be negative, to guarantee a positive fitness value in the fitness proportionate selection process, we write the fitness function as:

$$\max_{\mathcal{C}} \widetilde{SR} = 1 + \frac{1}{N} \sum_i \frac{\bar{R}_i - r_f}{\sigma_i}, \qquad i \in \mathcal{E}, \tag{13}$$

where $\mathcal{C}$ denotes the indicator combination and N is the number of selected sim-traders in the elitist set $\mathcal{E}$.

As a group of traders, not just a single trader will benefit from extracting useful information from the indicators, we define the fitness value in terms of a set of 'elitist' traders $\mathcal{E}$ rather than a single elitist trader. In Eq. (13), $\bar{R}_i$ represents the average daily return of the i-th elitist member, σ_i is the standard deviation of the i-th elitist trader's returns over the evaluation period, and r_f is the risk-free rate. In this paper, the risk-free rate r_f is assumed to be zero for daily trading. Denoting $\widetilde{SR}_q$ as the fitness of the q-th individual chromosome in the population, its probability of being selected is decided by

$$\pi_q = \frac{\widetilde{SR}_q}{\sum_{n=1}^{S} \widetilde{SR}_n}, \tag{14}$$

where S is the population size. It should be noted that, the population size S is different from the number N of the selected sim-traders in the elitist set.

After the selection, the GA performs the crossover. We use a uniform crossover scheme, which exchanges corresponding genes in the parent chromosomes with probability π_c in this study. After the crossover, genes in each chromosome undergo a mutation process which introduces gene variation into the population. For the binary

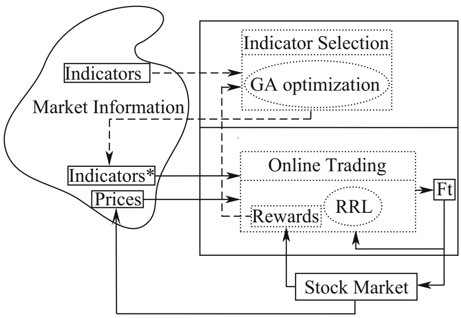

Fig. 2 The integrated trading system of GA and RRL

representation of a gene, the mutation can be implemented by flipping the binary value. After the mutation, we replace the current population with the children which have better fitness values to form the next generation. The above processes are repeated until a fixed number of generations is reached. The GA algorithm is described in Algorithm 1 below. Figure 2 presents the integrated trading system which consists of the GA selector and the RRL trading system. After determining the optimal combination $\mathcal{C}$, the chosen indicators are exported to the RRL trading system for out-of-sample trading.

Algorithm 1 The Genetic Algorithm for Indicator Selection

1: Initialize population of vectors $\mathbf{s}_q$, $q = 1...S$
2: Evaluate the initial population
3: **while** a fixed number of generations is not met **do**
4: Select parent chromosomes based on the roulette wheel selection with probability $\pi_q = \frac{\widetilde{SR}_q}{\sum_{n=1}^{S} \widetilde{SR}_n}$
5: Generate $\mathbf{s}_c$ by applying the uniform crossover and the mutation scheme at probabilities π_c and π_m
6: Evaluate the offspring $\mathbf{s}_c$
7: **if** Fitness($\mathbf{s}_c$) is better than Fitness($\mathbf{s}_q$)
8: **then** $\mathbf{s}_q \leftarrow \mathbf{s}_c$
9: **end while**

2.4 The Indicator Variables

In the context of equity investment, technical analysis is a method for forecasting price movements by studying previous market data (primarily price and volume) of an equity (Kirkpatrick and Dahlquist 2006). Technical analysis is now widely accepted as one of many viable and useful analytical tools for price movement prediction (see Achelis 2001). One of the most popular technical analysis indicators is the relative strength index (RSI) which is an oscillator that measures current price strength in relation to previous prices. Investors use the RSI to generate buy and sell signals and to confirm price movements. Chen et al. (2001) pointed out that trading volume and return volatility contribute information to the returns processes of stocks. In addition to technical analysis, financial news, market research and fundamental analysis are commonly used today for price movement analysis. Fundamental analysis (see Graham and Dodd 2004) is a useful and valuable method for determining the intrinsic value of an equity by examining and measuring related economic, financial and other qualitative and quantitative factors. For example, in a study by Lev and Thiagarajan (1993), it was found, for the 1980s, that the fundamental variables identified in their research added approximately 70 %, on average, to the explanatory power of earnings with respect to excess returns.

In this study, the indicator variables for the RRL trading system include two volume indices (a negative and a positive), an oscillator index (relative strength indices using 3, 9, 14, 30 days), and a conditional volatility of returns from a generalized autoregressive conditional heteroskedasticity (GARCH) model (see Bollerslev (1986)). The indicator variables also include two price multiple indicators and one leverage ratio indicator. Thus, price-to-cash-flow and price-to-earnings ratios are used to check the overvaluation/undervaluation of an equity; and a debt-to-equity-market ratio is used to measure the financial fitness of a company's capital structure. In our trading system, indicator variables do not include return information; instead the lagged daily returns constitute the basic input stream of the RRL trading system.

As the focus of this paper is to study value-added trading profits from the GA selection, we consider only a limited number of indicators in this study. The following paragraphs provide a general overview of the above indicators.

- Positive volume index (PVI) and negative volume index (NVI)
  Negative and positive volume indicators were introduced by Fosback (1991) as signals of bull markets:
  - If $VOL_t > VOL_{t-1}$, $PVI_t = PVI_{t-1} + \frac{P_t - P_{t-1}}{P_{t-1}} \cdot PVI_{t-1}$, otherwise $PVI_t = PVI_{t-1}$,
  - If $VOL_t < VOL_{t-1}$, $NVI_t = NVI_{t-1} + \frac{P_t - P_{t-1}}{P_{t-1}} \cdot NVI_{t-1}$, otherwise $NVI_t = NVI_{t-1}$.

  Rules associated to the above indicators are:

$$IPVI_t = \begin{cases} 1 & \text{if } PVI_{t-1} \leq EMA_t(PVI, l) \text{ and } PVI_t > EMA_t(PVI, l) \\ -1 & \text{if } PVI_{t-1} \geq EMA_t(PVI, l) \text{ and } PVI_t < EMA_t(PVI, l) \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

and

$$INVI_t \begin{cases} +1 & \text{if } NVI_{t-1} \leq EMA_t(NVI, l) \text{ and } NVI_t > EMA_t(NVI, l) \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

In this study, we use a window size of $l = 250$ to compute the exponential moving average (EMA) values of PVI and NVI.

- Indicator of relative strength index (IRSI)
  The relative strength index which was introduced by Wilder (1978) is one of the most popular oscillator indices.

$$RSI(n)_t = 100 - \frac{100}{1 + U(n)/D(n)} \tag{17}$$

U and D are the averages of upward and downward price changes of the last n periods (see Achelis 2001).

$$IRSI(n)_t = \begin{cases} +1 & \text{if } RSI(n)_{t-1} \geq 30 \text{ and } RSI(n)_t < 70 \\ -1 & \text{if } RSI(n)_{t-1} \leq 30 \text{ and } RSI(n)_t > 70 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

In this study, we use four different values of n, i.e. RSI 3, RSI 9, RSI 14 and RSI 30.

- Price to cash flow (P–CF), price to earnings (P–E), debt to equity market (D–M)
  - P–CF is the ratio of a stock's price divided by the cash flow per share.
  - P–E is the ratio of the price of a stock and the company's earnings per share.
  - D–M is the ratio used to measure a company's debt against its market capitalization.
- Conditional volatility (CVOL)
  We use a GARCH model to retrieve the conditional volatility as auto-regressive (AR) GARCH models have been widely used by people to study equity returns (e.g. Baillie and DeGennaro 1990). A detailed introduction to the GARCH models can be found in Bollerslev (1986). For example, an AR(2)-GARCH(1,1) process is written as follows.

$$\begin{cases} r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t \\ \varepsilon_t \sim \mathcal{N}(0, h_t) \\ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{cases} \tag{19}$$

3 The Data and Parameter Settings

3.1 The Data Sets

The companies selected for this study are S&P 500 American companies traded on the New York Stock Exchange and Nasdaq. We download daily prices [PX_LAST], trading volume [PX_VOLUME], price-earning [PE_RATIO], price-cash flow [PX_TO_CASH_FLOW], debt-market value [DEBT_TO_MKT], RSI 3, RSI 9, RSI 14 and RSI 30 of the 500 companies from the Bloomberg Excel Terminal for the period 1st January 2009–1st May 2014 (headers of these fields in Bloomberg are reported in square brackets). There are 1,340 observations in each series. We compute values of the NVI, the PVI and the conditional volatility of each stock in Matlab. Where a company's information is insufficient for analysis (e.g. where there is an incomplete price-earning series), we exclude the company from the S&P 500 list. There are 180 S&P companies left for this study. The 1,340 observations are partitioned into a training set consisting of the first 1,215 samples ($T_{train} = 1,215$) and a trading set containing the subsequent 125 observations ($T_{trade} = 125$). The training set includes an evaluation set ($T_{eva} = 125$, i.e. the last 125 observations of T_{train}) for the elitist member selection.

3.2 The Parameter Settings

With respect to the GA selection, each chromosome contains ten genes representing the number of indicator variables. Based on our earlier research, the following parameters are found to be suitable settings for the indicator selection problem: the population size and the number of generations are set at 50 and 20 respectively; the crossover probability is 0.5; and the mutation probability is 0.2. We find that the convergence of the optimized results is insensitive to these parameter settings. It should be noted that, the fitness value of a chromosome is a function of the average Sharpe ratio of the selected sim-traders. The population size S is different from the number N of the selected sim-traders in the elitist set $\mathcal{E}$. As we want to select the variables which can improve the trading performance of the entire elitist set $\mathcal{E}$, the average Sharpe ratio of the elitist members in $\mathcal{E}$ is used as the fitness value of a solution (see Eq. (13)).

Regarding the RRL settings, each of the RRL trading systems consists of 100 sim-traders which are initialized using random numbers from a Gaussian distribution with a mean of 0 and a standard deviation of 0.05. At the end of the training period, we select a number of N sim-traders (i.e. an elitist set $\mathcal{E}$ comprising the top 5% sim-traders of the 100 sim-traders) based on the Sharpe ratio rankings in the evaluation period. The average Sharpe ratio of the elitist members in the elitist set $\mathcal{E}$ is used to compute the fitness value in Eq. (13). We find that the following parameters are suitable for our trading problem: the number of shares traded $\nu = 1$; the learning rate $\rho = 0.15$; the adaption rate $\eta = 0.05$; and the transaction cost $\delta = 3$ basis points (BPS). As we expect that new information of an equity will be fully reflected in its daily prices in 2 weeks (maximum), we use a value of $l = 10$. The system inputs include the trading signals which are generated based on the technical rules (see Eqs. (15), (16), and (18)), and the values of P–CF, P–E, D–M, CVol, lagged returns which are standardized to their means and standard deviations from the training period.

3.3 The Scenarios

For each asset, the GA selector produces an optimized indicator pattern based on the trading results in the evaluation period T_{eva}. There are 180 indicator patterns in total after the selection. We then check the out-of-sample trading results in four scenarios: a tailored scenario, a general scenario, an all-in scenario, and an all-zero scenario. The tailored scenario is where the optimized indicator pattern of an asset is used when we trade the asset. With the general scenario, the indicators which received more than

Table 1 Indicators under the general scenario

NVI	PVI	P-CF	P-E	D-M	CVol	RSI3	RSI9	RSI14	RSI30
✓		✓			✓	✓	✓		✓

50 % of the 180 votes are included as a part of the inputs when we trade the 180 assets. The all-in scenario includes all of the ten indicator variables as a part of the system inputs, providing a benchmark scenario for assessing the value-added trading profits due to the GA selection. With the all-zero scenario, the system inputs do not include any indicator variables, only return information is used to produce the trading signals as has been suggested in the literature (see Moody et al. 1998; Gold 2003; Dempster and Leemans 2006; Maringer and Ramtohul 2012). Table 1 only provides details of the indicators under the general scenario, the indicator combinations of the 180 stocks in the tailored scenario are available upon request.

4 Results of the Experiment

As profitability and stability are particularly important for financial trading systems, the following discussion focuses on these two issues. In the experiment, we use the daily Sharpe ratio (see Sharpe (1994)) to measure profitability, as that ratio is closely related to the optimization criterion in RRL, i.e. the exponential moving average Sharpe ratio (EMSR, Eq. (8)), which is actually a variation of the conventional Sharpe ratio. Moody et al. (1998) identified a limitation of using the EMSR is that the largest possible improvement in the DSR is limited, therefore the Modigliani risk-adjusted performance ratio (see Modigliani and Modigliani 1997) which measures risk-adjusted excess returns, may be modified and used as an alternative optimization criterion for the EMSR in future research.

Stability in this study refers to the consistency of the Sharpe ratios generated from independent restarts of the trading system. As trading performance is directly related to the initial values of the signal parameters when we use RRL-type trading systems, it is necessary to assess the stability of the GA-RRL trading system. To do this, we restart the GA-RRL trading system 100 times, and save the Sharpe ratio of the elitist trader (i.e. the one which produces the highest Sharpe ratio from the 100 sim-traders in the evaluation period) from each trial. We compute the mean and the standard deviation (SD) of the 100 elitist Sharpe ratios. Then a ratio of the mean to the SD (i.e. the Z value in statistics) is used to check whether a company's Sharpe ratio produced by the GA-RRL trading system is statistically different from zero. In other words, the Z value can be considered as a stability-adjusted performance measure for RRL-type trading systems. In the trading period, we use the average elitist update scheme to produce long/short signals, and we collect these Sharpe ratio statistics based on the 100 restarts.

In addition to the significance check of the Sharpe ratio means, we benchmark the GA-RRL trading system against two of the most common strategies for equity trading: a random trading strategy (i.e. a 'zero-intelligence' strategy); and a 'buy-and-hold' strategy. To use the random trading strategy, we set up 180 random trading systems for the 180 stocks, with each of the random trading systems consisting of 100 random traders. For each stock, a 'best' label is attached to the random trader which produces the highest Sharpe ratio in the evaluation period. For each stock, we restart the random trading system 100 times and save the Sharpe ratio statistics (the means, the SDs and the Z values) of the 'best' random trader in both the evaluation and the trading periods.

We also use a bootstrap approach (i.e. a resampling approach, see Efron 1981; Gilli et al. 2006) to derive the Sharpe ratio statistics in the case where the 180 stocks are held during the evaluation and trading periods. The bootstrap approach keeps the dependence structure among the stocks unchanged, and each time it bootstraps a 125×180 artificial return matrix based on the historical returns of the 180 stocks from the evaluation and the trading periods. We then compute the Sharpe ratio statistics (the means, the SDs and the Z values) based on a bootstrapped iteration number of 2,000 for each stock in the two periods.

In Fig. 3, the left column subplots report the probability density curves of the daily Sharpe ratio statistics recorded in the evaluation period (i.e. the means, the SDs and the Z values). The upper subplot shows that most of the Sharpe ratio means produced by the GA-RRL trading system under the tailored, general and all-in scenarios are higher than those produced by using the random trading strategy. Under the all-zero scenario, the elitist trader (i.e. the one which uses only return information to produce long/short signals) cannot beat the best random trader. The middle subplot shows the SD density curves which are used to measure the stability of the Sharpe ratios. It is found that, under the four scenarios the GA-RRL trading system produces lower SDs than those produced by using the random trading and buy-and-hold strategies. The SDs in the all-zero scenario (i.e. where the RRL trading system uses only return information) are slightly higher than those in the tailored, general and all-in scenarios. The lower subplot shows the significance of the Sharpe ratios. It shows that, except for the buy-and-hold strategy, the Sharpe ratios produced by using the other five strategies are statistically higher than zero at the conventional 5 % significance level.

The right column subplots in Fig. 3 report the probability density curves of the daily Sharpe ratio statistics recorded in the trading period (i.e. the out-of-sample period). The upper subplot shows that the Sharpe ratio mean density curves in the tailored and general scenarios are very close to those using the buy-and-hold strategy. In other words, the GA-RRL trading system does not outperform the buy-and-hold strategy in terms of profitability. The random trading strategy produces the worst density curve of the six strategies, as the strategy does not generate any significant and positive Sharpe ratios. The middle subplot shows that the GA-RRL trading system produces lower SDs in the tailored, general and all-in scenarios, as compared to those using the other three strategies (all-zero, random trading, and buy-and-hold). The lower subplot reports the density curves of the Z values. It is found that, in the tailored and general scenarios, the GA-RRL trading system produces better Z values, in terms of a greater number of positive Sharpe ratios than the other four strategies.

Although the differences in the Sharpe ratio means in the four scenarios (tailored, general, all-in and all-zero) are not statistically significant in the out-of-sample period, we find that feeding the selected indicators into the trading system reduces the Sharpe

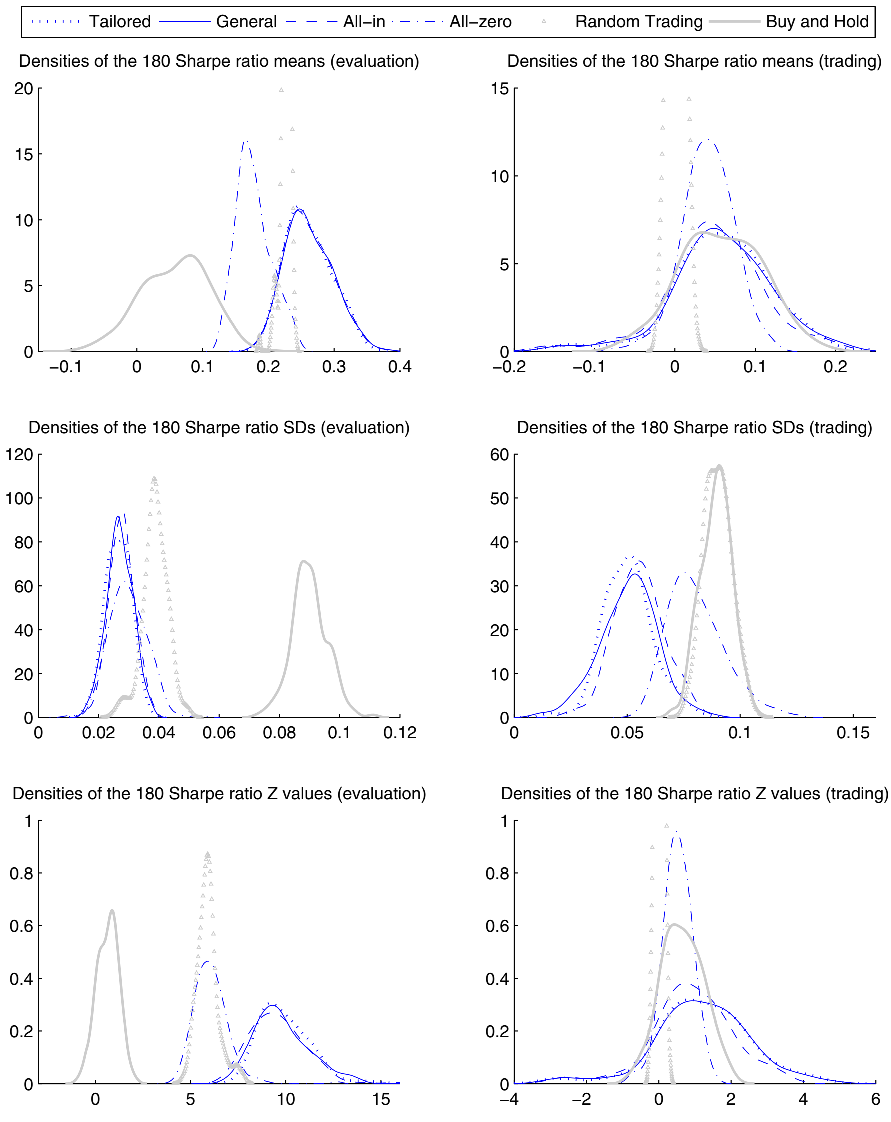

Fig. 3 Density curves of the Sharpe ratio statistics from the evaluation period (*left*) and the trading period (*right*)

ratio SDs (i.e. improves the trading system's stability). For example, the Sharpe ratios in the all-zero scenario (i.e. using only return information), show a higher SD than that in the tailored and general scenarios. It seems that using the return information alone is insufficient to generate positive Sharpe ratios. In the all-zero scenario, the RRL trading system behaves like a random trader, in that the Sharpe ratio means are lower and the Sharpe ratio SDs are higher than they are in the tailored and general scenarios.

Table 2 Range and significance check of the 180 Sharpe ratio means in the trading period

	[−0.2 0.1]	[−0.1 − 0.05]	[−0.050]	[0 0.05]	[0.05 0.1]	[0.1 0.15]	[0.15 0.2]	[0.2 0.25]	sum
Panel A									
Tailored	5	4	10	62	57	30	11	1	180
General	4	6	10	61	57	32	8	2	180
All-in	4	7	19	68	54	19	8	1	180
All-zero	0	1	15	90	69	5	0	0	180
Random	0	0	85	95	0	0	0	0	180
BuyHold	0	3	15	64	55	40	3	0	180
Panel B									
H0: SR ≤ 0 at 10 %									
Tailored	(5)	(1)	0	2	43	30	11	1	71
General	(4)	(4)	0	4	42	32	8	2	80
All-in	(4)	(5)	0	3	43	19	8	1	65
All-zero	0	0	0	0	1	3	0	0	4
Random	0	0	0	0	0	0	0	0	0
BuyHold	0	0	0	0	0	24	3	0	27
Panel C									
H0: SR ≤ 0 at 5 %									
Tailored	(5)	0	0	0	25	30	11	1	62
General	(4)	(2)	0	1	23	32	8	2	60
All-in	(4)	(3)	0	2	22	18	8	1	44
All-zero	0	0	0	0	0	0	0	0	0
Random	0	0	0	0	0	0	0	0	0
BuyHold	0	0	0	0	0	6	3	0	9

It should be noted that the SDs in the all-in scenario are also higher than those in the tailored and general scenarios. As we use random numbers to initialize the signal parameters, then the more indicators we feed into the system, the greater the noise will be at the RRL input terminal. In order to maintain the trading system's stability, it is necessary to use the GA to select only those indicators which can improve trading performance.

Table 2 provides quantified evidence to support the above findings. Panel A shows the numbers of the 180 Sharpe ratio means in different ranges. Under the tailored, general, all-in and all-zero scenarios, most of the Sharpe ratio means have a value in the range of 0 and 0.15. Panel B shows the numbers of the Sharpe ratio means which are statistically greater than zero at the 10 % significance level (the numbers of the Sharpe ratio means which are statistically less than zero at the 10 % significance level are reported in brackets). The last column shows the numbers of the positive Sharpe ratios, after netting the number of negative Sharpe ratios. Panel C shows the numbers of the Sharpe ratio means which are greater than zero at the 5 % significance level. In the all-zero scenario, the RRL trading system does not produce a Sharpe ratio which is statistically greater than zero at the 5 % significance level. In general, we find that

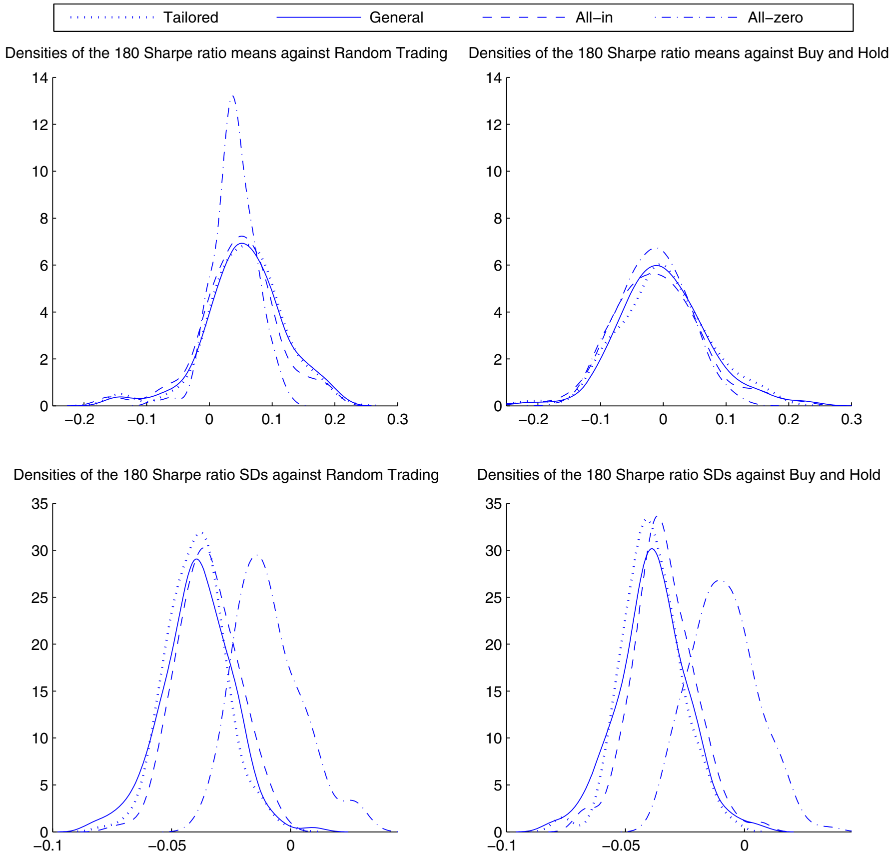

Fig. 4 Density curves of the Sharpe ratio means and SD from GA-RRLs benchmarking against random trading (*left*) and buy-and-hold (*right*)

the GA-RRL trading system produces a higher number of positive Sharpe ratios at the two significance levels. In other words, the GA selection brings value-added profits to the RRL trading system.

Figure 4 shows the density curves of the excess Sharpe ratio means and the excess Sharpe ratio SDs, benchmarking against the random trading strategy and against the buy-and-hold strategy. It is found that the GA-RRL trading system outperforms the random trading strategy, in that a larger number of the positive excess Sharpe ratio means and a greater number of the negative excess Sharpe ratio SDs are found in the out-of-sample period. In comparison with the buy-and-hold strategy, the GA-RRL trading system does not outperform the benchmark in terms of profitability because the Sharpe ratio mean values of these density curves are around zero. However, it is found that the GA-RRL trading system is more stable than the buy-and-hold strategy.

5 Conclusions

In this paper, we have put forward an equity trading system consisting of a GA selector and a RRL online trading system (the 'GA-RRL' trading system) which aims to increase the trading profits of RRL-type trading systems by including indicator variables (not just return information) as a part of the system inputs.

We studied the trading performance of the GA-RRL trading system in an out-of-sample period in four scenarios: tailored; general; all-in; and all-zero. In the all-zero scenario, the trading system used only return information as inputs (as is described in the literature). We found that the profitability and the stability of the trading system, which were measured using the mean value and the standard deviation of the Sharpe ratios, were better in all the scenarios except the all-zero scenario.

We also benchmarked the proposed GA-RRL trading system against two of the most common strategies for equity trading: the 'zero-intelligence' random trading strategy; and the buy-and-hold strategy. The results of the out-of-sample test showed that the GA-RRL trading system outperformed the random trading strategy. Although the GA-RRL trading system did not outperform the buy-and-hold strategy by producing a greater number of positive Sharpe ratio means, it was nonetheless found to be more stable than the buy-and-hold strategy.

In addition to looking at profitability and stability, the significance of the Sharpe ratios from trading the 180 stocks was also examined. It was found, after feeding the selected indicators from the GA into the RRL trading system, that the number of companies with a Sharpe ratio significantly greater than zero increased. It can therefore be concluded that designers of RRL-type trading systems should use tools and indicators from technical analysis, fundamental analysis and financial engineering, in addition to return information in order to achieve utility maximization.

Acknowledgments The authors acknowledge the support provided by the Swiss National Science Foundation (SNSF) grant program. The authors would like to thank three reviewers from the Artificial Intelligence Applications and Innovations Conference (AIAI) 2013, two anonymous referees, and editors of Computational Economics for their valuable suggestions.

References

Achelis, S. B. (2001). *Technical analysis from A to Z*. New York: McGraw Hill.

Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, *51*, 245–271.

Baillie, R. T., & DeGennaro, R. P. (1990). Stock returns and volatility. *The Journal of Financial and Quantitative Analysis*, *25*, 203–214.

Bertoluzzo, F., & Corazza, M. (2007). Making financial trading by recurrent reinforcement learning. In B. Apolloni, R. Howlett, & L. Jain (Eds.), *Lecture notes in computer science* (Vol. 4693, pp. 619–626). Berlin: Springer.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Chen, G., Firth, M., & Rui, O. M. (2001). The dynamic relation between stock returns, trading volume, and volatility. *Financial Review*, *36*, 153–174.

Creamer, G. (2012). Model calibration and automated trading agent for euro futures. *Quantitative Finance*, *12*, 531–545.

Dempster, M. A. H., & Leemans, V. (2006). An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications*, *30*, 543–552.
Dempster, M. A. H., Payne, T. W., Romahi, Y., & Thompson, G. W. P. (2001). Computational learning techniques for intraday FX trading using popular technical indicators. *IEEE Transactions on Neural Networks*, *12*(4), 744–754.
Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, *68*, 589–599.
Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*, 383–417.
Fosback, N. G. (1991). *Stock market logic: A sophisticated approach to profits on wall street*. Fort Lauderdale: Dearborn Financial Publishing.
Gilli, M., Këllezi, E., & Hysi, H. (2006). A data-driven optimization heuristic for downside risk minimization. *Journal of Risk*, *8*, 1–18.
Gold, C. (2003). FX trading via recurrent reinforcement learning. *Proceedings of IEEE international conference on computational intelligence in financial engineering* (pp. 363–370). Los Alamitos: IEEE Computer Society Press.
Gorse, D. (2011). Application of stochastic recurrent reinforcement learning to index trading. In European symposium on artificial neural networks.
Graham, B., & Dodd, D. (2004). *Security analysis*. Burlington: McGraw Hill.
Graham, B., & Zweig, J. (2003). *The intelligent investor: The definitive book on value investing*. New York: HarperCollins.
Holland, J. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan Press.
Kirkpatrick, C. D., & Dahlquist, J. (2006). *Technical analysis: The complete resource for financial market technicians*. Upper Saddle River: Financial Times Press.
Lev, B., & Thiagarajan, S. R. (1993). Fundamental information analysis. *Journal of Accounting Research*, *31*, 190–215.
Maringer, D., & Ramtohul, T. (2012). Regime-switching recurrent reinforcement learning for investment decision making. *Computational Management Science*, *9*, 89–107.
Maringer, D., & Zhang, J. (2014). Transition variable selection for regime switching recurrent reinforcement learning. In 2014 IEEE conference on computational intelligence for financial engineering and economics, (pp. 407–412).
Modigliani, F., & Modigliani, L. (1997). Risk-adjusted performance. *The Journal of Portfolio Management*, *23*, 45–54.
Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting*, *17*, 441–470.
Sharpe, W. (1994). The Sharpe ratio. *Journal of Portfolio Management*, *21*, 49–58.
Wilder, J. W. (1978). *New concepts in technical trading systems*. Greensboro: Trend Research.
Zhang, J., & Maringer, D. (2013). Indicator selection for daily equity trading with recurrent reinforcement learning. *In the genetic and evolutionary computation conference (GECCO)*, *2013*, 1757–1758.