Trend-Following Trading Using Recursive Stochastic Optimization Algorithms

D. Nguyen, G. Yin, and Q. Zhang

Abstract—This work develops with trend following trading strategies under a bull-bear market switching model. The asset model is assumed to be geometric Brownian motion type of process, in which drift of the stock price is allowed to switch between two parameters corresponding to an up-trend (bull market) and a downtrend (bear market) corresponding to a partially observable Markov chain. Our objective is to buy and sell the underlying stock to maximize an expected return. It is shown in [6], [7] that an optimal trading strategy can be obtained in terms of two threshold levels, but finding the threshold levels is a difficult task. In this paper, we develop a stochastic approximation algorithm to approximate the threshold levels. The main advantage of our method is that one need not solve the associated Hamilton-Jacobi-Bellman (HJB) equations. We establish the convergence of the algorithm and provide numerical examples to illustrate the results.

I. INTRODUCTION

In this work, we focus on a trend following trading strategy. The idea of trend following is to go long at the beginning of a bull market and exit when the trend reverses. A trend following trader purchases shares when the prices go up to a certain level from the bottom and sells when the prices start to fall.

There are extensive studies in literature devoted to equity trading. For example, Merton [24] pioneered the portfolio selection with utility maximization, which was subsequently extended to incorporate transaction costs by Magill and Constantinides [22]; see also Davis and Norman [8], Shreve and Soner [27], Liu and Loewenstein [20], Dai and Yi [5], and references therein. Zhang and Zhang [34] initiated a study on optimal trading strategy in a mean reverting market. Other work relevant to mean reverting strategies are available to trade. In addition, to keep things simple, we study only one round trip trading (buying the stock and then selling it). The objective is to maximize the expected percentage gain. We construct a stochastic approximation algorithm and establish its convergence. In addition, we also carry out extensive Monte Carlo simulations and real market tests.

In the next section, we present the problem formulation. In Section 3, we carry out convergence analysis. Then in Section 4, we report numerical experiment results. This paper is an extended summary of [25]. The proofs of results are omitted and can be found in [25].

II. PROBLEM FORMULATION

The asset price \( S_t \) at time \( t \) satisfies the following equation

\[
\frac{dS_t}{S_t} = \mu(\alpha_t)dt + \sigma dW_t, \quad S_0 = x, \tag{1}
\]

where \( \mu(i) = \mu_i, i = 1, 2, \) is the expected rate of return, \( \sigma > 0 \) is the volatility, \( W_t \) is a standard Brownian motion, and \( \alpha_t \) is a continuous-time Markov chain taking values in \( \mathcal{M} = \{1, 2\} \). Here \( \alpha_t \) represents the market mode at time \( t \). That is, \( \alpha_t = 1 \) indicates a bull market and \( \alpha_t = 2 \) a bear market.

We assume that \( \mu_1 > 0 > \mu_2 \). Let \( Q = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix} \) be the generator of \( \alpha_t \) for some \( \lambda_1 > 0 \) and \( \lambda_2 > 0 \); see Yin and Zhang [33] on how to generate \( \alpha_t \) through its generator \( Q \). In this paper, the values for \( \mu_i, \lambda_i, (i = 1, 2) \), and \( \sigma \) are given. To implement the results in practice, one needs to calibrate the model. We refer to Dai et al. [6] for an estimation method. Throughout the paper, we assume that \( \{\alpha_t\} \) and \( \{W_t\} \) are independent.
Let $0 \leq \tau \leq \nu$ be stopping times that one buys the stock at $\tau$ and sells at $\nu$. In this paper, we only consider one round-trip trading (one buying and selling cycle). In addition, we trade all available funds. Our objective is to choose $\tau$ and $\nu$ to maximize the reward function

$$J(\tau, \nu) = E \left[ \log \left( \frac{S_\nu}{S_\tau} \right) \right].$$

(2)

We assume that only $S_t$ is observable at time $t$; the market trend $\alpha_t$ is not completely observable. As a result of this assumption, the associated control problem is one that only has incomplete information. To solve this problem, it is necessary to convert the problem into a completely observable one. One way to achieve this is to use the well-known Wonham filter [30]; see also [10].

Let $p_t = P(\alpha_t = 1 | S_s : 0 \leq s \leq t)$ denote the conditional probability of $\alpha_t = 1$ (bull market) given the stock price up to time $t$. Then the Wonham filter in terms of $p_t$ satisfies the following stochastic differential equation (SDE)

$$dp_t = \left[ -\left( \lambda_1 + \lambda_2 \right) p_t + \lambda_2 \right] dt + \frac{\left( \mu_1 - \mu_2 \right) p_t (1 - p_t)}{\sigma} d\tilde{W}_t,$$

where $d\tilde{W}_t$ is an innovation process given by

$$d\tilde{W}_t = \frac{d \log(S_t) - \left[ (\mu_1 - \mu_2) p_t + \mu_2 - \alpha^2/2 \right] dt}{\sigma}.$$

(3)

In view of (4), we can rewrite (1) in terms of $p_t$ and $\tilde{W}_t$:

$$dS_t = S_t \left[ (\mu_1 - \mu_2) p_t + \mu_2 \right] dt + S_t \sigma d\tilde{W}_t.$$

It follows that

$$S_t = S_0 \exp \left( \int^t^\nu \left[ (\mu_1 - \mu_2) p_r + \mu_2 - \alpha^2/2 \right] dr \right. \left. + \sigma(\tilde{W}_t - \tilde{W}_r) \right).$$

Therefore, we can write the reward in (2) in terms of $p_t$ and $\tilde{W}_t$ as

$$J(\tau, \nu) = E \left[ \int^\nu^\nu \left[ (\mu_1 - \mu_2) p_r + \mu_2 - \alpha^2/2 \right] dr \right. \left. + \sigma(\tilde{W}_t - \tilde{W}_r) \right].$$

(5)

In [6], it is shown that the optimal strategy is given by threshold levels $\theta^b$ (buy level) and $\theta^s$ (sell level) with $\theta^s < \theta^b$. In this paper, we develop an alternative approach and use a stochastic approximation approach to search for these threshold levels $(\theta^s, \theta^b)$. Let

$$\phi(\theta) = \phi(\theta^b, \theta^s) = E \int^{\tau(\theta)}_{\tau(\theta)} \left[ (\mu_1 - \mu_2) p_r + \mu_2 - \alpha^2/2 \right] dr,$$

(6)

where

$$\tau(\theta) = \inf \{ t : p_t \geq \theta^b \},$$

$$\nu(\theta) = \inf \{ t > \tau(\theta) : p_t \leq \theta^s \}.$$

The idea can be explained as follows. Although $\phi(\theta)$ is not observable due to the involvement of expectation, we can measure $\phi(\theta)$ with noise (for example, the simplest of such measurement may be written in an additive form as $\phi(\theta) + \chi(\xi, \theta)$ where $\chi(\xi, \theta)$ is the noise). In lieu of such a simple form, we consider a much more general form in that the noise corrupted observation or measurement of $\phi(\theta)$ is $\phi(\theta, \xi)$. That is, nonadditive noise is allowed and complex nonlinear function form can be incorporated in our formulation. We also assume that $E_{\xi} [\phi(\theta, \xi)] = \phi(\theta)$ and for the moment let us assume that $\phi(\theta, \xi)$ depends smoothly on $\theta$.

Next, we describe the stochastic approximation procedure and illustrate the use of $\phi(\theta, \xi)$.

0. Initialization: Choose an initial threshold estimate $\phi_0 = (\theta_0^b, \theta_0^s)$.

2. Iteration: With $n > 0$ and $\theta_n = (\theta_n^b, \theta_n^s)$ computed, carry out one step stochastic approximation to find updated threshold $\phi_{n+1} = (\theta_{n+1}^b, \theta_{n+1}^s)$. Let $c_1 = (1, 0), c_2 = (0, 1), c_n = n/\sqrt{n^2/3}$, and $\xi_n, \eta_n$ are noise sequences.

(a) Find $\tau(\theta_n + c_n e_1) < \nu(\theta_n)$, compute $\phi(\theta_n + c_n e_1, \xi_n, \eta_n)$.

(b) Find $\tau(\theta_n - c_n e_1) < \nu(\theta_n)$, compute $\phi(\theta_n - c_n e_1, \xi_n, \eta_n)$.

(c) Find $\tau(\theta_n + c_n e_2) < \nu(\theta_n + c_n e_2)$, compute $\phi(\theta_n + c_n e_2, \xi_n, \eta_n)$.

(d) Find $\tau(\theta_n) < \nu(\theta_n + c_n e_2)$, compute $\phi(\theta_n + c_n e_2, \xi_n, \eta_n)$.

(e) Use (a)-(d), find the gradient estimate $\nabla \phi(\theta_n, \xi_n) = (\nabla_1 \phi(\theta_n, \xi_n))$ of $\phi(\theta)$ by

$$\nabla_i \phi(\theta_n, \xi_n) = \frac{1}{2c_n} \left[ \phi(\theta_n + c_n e_i, \xi_n) - \phi(\theta_n - c_n e_i, \xi_n) \right]$$

for $i = 1, 2$.

(f) Update one step the parameter estimate by using stochastic approximation method (SA) given by

$$\theta_{n+1} = \theta_n + \frac{1}{n} \nabla \phi(\theta_n, \xi_n)$$

(7)

3. Repeat step 2 with $n \rightarrow n + 1$ until $|\theta_n - \theta_{n+1}| < Tol$ with a prescribed tolerance level Tol or with $n = N$ for some large $N$.

To proceed, we use the techniques developed in Kushner and Yin [19] to analyze the algorithm.

III. ASYMPTOTIC PROPERTIES

To show the convergence, we use the idea that on each “small” interval, the noise $\xi$ varies much faster than the “state” $\theta$. Thus with $\theta$ “fixed”, the noise will be eventually averaged out resulting in an averaged system that can be characterized by a system of ordinary differential equations. Define

$$t_n = \sum_{j=1}^{n-1} \frac{1}{j},$$

$$m(t) = \max \{ n : t_n \leq t \}$$

$$\theta^b = \theta_n \text{ for } t \in [t_n, t_{n+1}) \text{ and } \theta^s(t) = \theta^b(t_n + t)$$
Note that $\theta^0(\cdot)$ is a piecewise constant interpolation of $\theta_n$ on the interval $[t_n, t_{n+1})$ and $\theta^n(\cdot)$ is its shift. Next, for $i = 1, 2$, we define

\begin{align*}
    b^i_n &= \frac{\phi(\theta_n + c_ne_i) - \phi(\theta_n - c_ne_i)}{2c_n} - \frac{\partial \phi(\theta_n)}{\partial \theta^0}, \\
    \rho^i_n &= \left[ \phi(\theta_n + c_ne_i, \xi^+_{n,i}) - \phi(\theta_n - c_ne_i, \xi^-_{n,i}) \right], \\
    E_n &= \left[ \phi(\theta_n + c_ne_i, \xi^+_{n,i}) - \phi(\theta_n - c_ne_i, \xi^-_{n,i}) \right], \\
    \psi^i_n &= \left[ \phi(\theta_n + c_ne_i, \xi^+_{n,i}) - \phi(\theta_n + c_ne_i) \right], \\
    \psi^i_n &= \left[ \phi(\theta_n - c_ne_i, \xi^-_{n,i}) - \phi(\theta_n - c_ne_i) \right],
\end{align*}

where $E_n$ denotes the conditional expectation with respect to $\mathcal{F}_n$ - the $\sigma$-algebra generated by $\{\xi^+_j: j < n\}$. Write $b_n = (b^1_n, b^2_n)^T$. In the above, $\psi^i_n$ and $b^i_n$ for $i = 1, 2$ represent the noise and bias, and the $\{\rho^i_n = (\rho^i_n, \rho^i_n)^T\}$ is a martingale difference sequence. It is also reasonable to assume that after taking the conditional expectation, the resulting function is smooth. Thus we have separated the noise into two parts, uncorrelated noise $\{\rho^i_n\}$ and correlated noise $\{\psi_n = (\psi^1_n, \psi^2_n)^T\}$. With these notations, the algorithm in (7) can be written as following

$$
\theta_{n+1} = \theta_n + \frac{1}{n} \nabla \phi(\theta_n) + \frac{1}{n} \psi^1_n + \frac{1}{n} \rho^1_n + \frac{b_n}{n}
$$

Note that in fact, both $\psi^i_n$ and $b_n$ are $\theta$ dependent. If in what follow, when it is needed, we write $\psi_n = \psi(\theta_n, \xi_n)$ where $\xi_n$ includes $\xi^\pm$.

To proceed with the analysis of the algorithm (7), we assume the following conditions hold. (A1) For each $\xi$, $\tilde{\phi}(\xi, \xi)$ is a continuous function. (A2) For each $0 < N, \nu < \infty$ and each $0 < T < \infty$, the set $\{\sup_{|\theta| \leq N} \phi(\theta, \xi_n) : n \leq m(T)\}$ is uniformly integrable. (A3) The sequences $\{\xi^+_n\}$ are bounded. For each $\theta$ in a bounded set and for each $0 < T < \infty$

$$
\sup_n \sum_{j=n}^{m(T+t_n)-1} \frac{1}{j} \left| E[\psi^1_j(\theta, \xi_j)] \right| < \infty,
$$

$$
\lim_{n\to\infty} \sup_{0 \leq i \leq m(T+t_n)} E[|\psi^i_i|] = 0,
$$

where

$$
\psi^i_i = (n+i) \sum_{j=n+i}^{m(T+t_n)} \frac{1}{2j} E[\psi^1_j(\theta_{n+i+1}, \xi_j) - \psi^1_j(\theta_{n+i}, \xi_j)]
$$

for $i \leq m(T+t_n)$

(A4) The second derivative of $\phi(\cdot)$ is continuous. Before we proceed to the convergence of the algorithm, let us recall the definition of weak convergence. We say a sequence of $\mathbb{R}^2$-valued random variables $Z_n$ converges weakly to a random variable $Z$ if and only if for any bounded and continuous function $h(\cdot)$,

$$
E_h(Z_n) \to E_h(Z) \text{ as } n \to \infty,
$$

and $Z_n$ is said to be tight if and only if for each $\eta > 0$, there exist a compact set $K_\eta \subset \mathbb{R}^2$ such that

$$
P(Z_n \in K_\eta) \geq 1 - \eta \text{ for all } n.
$$

Let $S_\nu = \{x \in \mathbb{R}: |x| < \nu\}$ be the $\nu$-sphere. We say that the process $\zeta^{n,\nu}$ is a $\nu$-truncation of the process $\zeta^n$ if $\zeta^{n,\nu}(t) = \zeta(t)$ up until the first exist from $S_\nu$ and

$$
\lim_{m \to \infty} \limsup_{n \to \infty} P\left( \sup_{t \leq T} |\zeta^{n,\nu}| \geq m \right) = 0 \text{ for each } T < \infty.
$$

Next let $q^\nu$ be a smooth function such that $q^\nu(\theta) = 1$ when $|\theta| \leq \nu$ and $q^\nu(\theta) = 0$ when $|\theta| \geq \nu + 1$. We define $\{\theta^n\}$ recursively by $\theta^n_1 = \theta_1$ and

$$
\theta^n_{i+1} = \theta^n_i + \frac{1}{n} \nabla \phi(\theta^n_i) + \frac{1}{n} \psi^1_i + \frac{1}{n} \rho^1_i + \frac{b_i}{n}
$$

for $p_0$. Moreover, since $(\mu_1 - \mu_2)p(t)(1-p(t))/\sigma^2 \geq 0$, $p(t)$ moves in the same direction as the stock prices.

Define the interpolation of $\theta^n_i$ as $\theta^{n,\nu} = \theta^n_i$ for $t \in [t_n, t_{n+1})$ and $\theta^{n,\nu}(t) = \theta^{n,\nu}(t + t_n)$. Thus $\theta^{n,\nu}$ is a $\nu$-truncation of $\theta^n$ (see [19, p.278]).

In what follows, we first show that the truncated process $\{\theta^{n,\nu}(\cdot)\}$ is tight in $D^2[0, \infty)$ - the space of $\mathbb{R}^2$-valued functions that are right continuous, have left-hand limits, and endowed with the Skorohod topology. We then obtain the weak convergence of $\theta^{n,\nu}$ and characterize the limit as a solution of an ODE. Finally, by letting $\nu \to \infty$, we conclude that the untruncated process $\{\theta^n(\cdot)\}$ also converges.

**Lemma 1:** Under conditions (A1)-(A4),

$$
\sum_{k=m(t+t_n)-1}^{m(t+t_n)+i-1} \frac{1}{k} \left[ \psi^1_k + \psi^2_k \right] q^\nu(\theta^n_k)
$$

converges in probability as $n \to \infty$, and the convergence is uniform in $t$.

**Theorem 1:** Under conditions (A1)-(A4), $\theta^n(\cdot)$ converges weakly to $\theta(\cdot)$, which is a solution of

$$
\theta = \nabla \phi(\theta)
$$

provided that the ordinary differential equation above has a unique solution for each initial condition.

**IV. NUMERICAL EXAMPLES**

In this section, we demonstrate the numerical performance of our algorithm. First, we carry out Monte Carlo simulations to illustrate the method. Then we test how the algorithm works in real markets.

Using the SDE in (3), we rewrite $p_t$ in term of the stock price $S_t$ by

$$
dp_t = f(p_t)dt + \frac{(\mu_1 - \mu_2)p_t(1-p_t)}{\sigma^2} d\log(S_t),
$$

where $f$ is a third-order polynomial of $p(t)$ given by

$$
f(p) = -(\lambda_1 + \lambda_2)p + \lambda_2 \frac{(-\mu_1 - \mu_2)p(1-p)((\mu_1 + \mu_2)p + \mu_2 - \sigma^2/2)}{\sigma^2}.
$$

It is easily checked that $f(0) = \lambda_2 > 0$ and $f(1) = -\lambda_1 < 0$. Furthermore, it is readily seen that $f(p)$ approaches $-\infty$ as $p \to \infty$ and $\infty$ as $p \to -\infty$. Therefore, $f$ has exactly one root $\xi \in (0, 1)$. When the stock prices stay as a constant, $p(t)$ is attracted to $\xi$. This attractor is unbiased choice for $p_0$. Moreover, since $(\mu_1 - \mu_2)p(t)(1-p(t))/\sigma^2 \geq 0$, $p(t)$ moves in the same direction as the stock prices.
This is also intuitive since the movement of stock price forms trends. We will estimate \( p(t) \) simply by replacing the differential equation in (10) with a difference using the trading day as the step size on the finite horizon \([0, Ndt]\)

\[
p_{t+1} = p(t) + f(p(t))dt + \frac{(\mu_1 - \mu_2)p(t)(1 - p(t))}{\sigma^2} \log \left( \frac{S_{t+1}}{S(t)} \right)
\]

where \( t = 0, dt, 2dt, \ldots, Ndt \).

Note that it is possible that \( p(t) > 1 \) or \( p(t) < 0 \) for some \( t \). To keep \( p(t) \in [0, 1] \), we truncate the process and follow the truncated equation instead:

\[
p_{t+1} = \min \left( \max \left( p(t) + f(p(t))dt + \frac{(\mu_1 - \mu_2)p(t)(1 - p(t))}{\sigma^2} \log \left( \frac{S_{t+1}}{S(t)} \right), 0 \right), 1 \right).
\]

**Example 1. (Monte Carlo simulations)** In this example, we choose the parameters with the following specifications:

<table>
<thead>
<tr>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( \mu_1 )</th>
<th>( \mu_2 )</th>
<th>( \sigma )</th>
<th>( K )</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.36</td>
<td>2.53</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.001</td>
<td>0.0679</td>
</tr>
</tbody>
</table>

Table 1. Monte Carlo parameter values

These parameters can be obtained following classification of bull and bear markets in terms of rising 24% from the bottom and falling 19% from the top. They were used in [7] for DJIA index. We generate sample paths of \( S(t) \) and \( p(t) \) using equations (1) and (3) with \( N = 50000 \) steps, step size \( dt = 1/N \), and \( T = 1 \). One sample path of \( S(t) \) is given in Figure 1. The corresponding \( p(t) \) is given in Figure 2 and \((\theta^b_0, \theta^s_0)\) in Figure 3. We start with initial guess \((\theta^b_0, \theta^s_0) = (0.918, 0.722)\) and perform stochastic approximation (7) with 1000 iterations. We then use different seeds to perform the algorithm using 500 replications. After taking the sample mean we obtain \((\theta^b, \theta^s) = (0.916, 0.725)\), and \( \phi = 0.042817 \) which is equivalent to 4.2817% gain obtained in one round-trip transaction.

We next perturb the parameters to see the dependence of the threshold levels. Our tests show that the threshold \((\theta^b, \theta^s)\) is not sensitive to changes in these parameters. These are summarized in Table 2.

<table>
<thead>
<tr>
<th>( \lambda_2 )</th>
<th>( \lambda_3 )</th>
<th>( \mu_1 )</th>
<th>( \mu_2 )</th>
<th>( \sigma )</th>
<th>( \theta^b )</th>
<th>( \theta^s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.36</td>
<td>2.53</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.916</td>
<td>0.725</td>
</tr>
<tr>
<td>0.36</td>
<td>2.53</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.917</td>
<td>0.725</td>
</tr>
<tr>
<td>0.36</td>
<td>2.53</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.916</td>
<td>0.725</td>
</tr>
<tr>
<td>0.36</td>
<td>2.53</td>
<td>0.19</td>
<td>-0.83</td>
<td>0.184</td>
<td>0.917</td>
<td>0.731</td>
</tr>
<tr>
<td>0.36</td>
<td>2</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.915</td>
<td>0.724</td>
</tr>
<tr>
<td>0.36</td>
<td>3</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.916</td>
<td>0.725</td>
</tr>
<tr>
<td>0.3</td>
<td>2.53</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.915</td>
<td>0.733</td>
</tr>
<tr>
<td>0.42</td>
<td>2.53</td>
<td>0.18</td>
<td>-0.77</td>
<td>0.184</td>
<td>0.916</td>
<td>0.718</td>
</tr>
</tbody>
</table>

Table 2. Thresholds with different parameters

Next we simulate this trend following strategy and compare it with the buy-and-hold strategy by using a large number of sample paths. Again, the parameter values given in Table 1 are used. The average returns of the trend following (TF) strategy on one unit invested on simulated paths are listed on Table 3 together with the average number of trades on each path. We also list the average return of the buy and hold (BH) strategy for comparison. These results are provided in Table 3.

<table>
<thead>
<tr>
<th>No. of sample paths</th>
<th>TF</th>
<th>BH</th>
<th>No. of trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>11.51%</td>
<td>4.64%</td>
<td>36.53</td>
</tr>
<tr>
<td>50000</td>
<td>11.55%</td>
<td>4.71%</td>
<td>36.58</td>
</tr>
<tr>
<td>100000</td>
<td>11.51%</td>
<td>4.67%</td>
<td>36.64</td>
</tr>
</tbody>
</table>

Table 3. Monte Carlo simulation: 20 years

We next investigate the performance of our strategy on each sample path. We use the parameters in Table 1 and the
buy-sell threshold \((\theta^b, \theta^s) = (0.916, 0.725)\). The result of simulation is collected in the Table 4. We can see that the result is very sensitive to each individual sample path, but our strategy clearly outperforms the buy and hold strategy.

<table>
<thead>
<tr>
<th>Trend Following</th>
<th>Buy and Hold</th>
<th>No. of trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.61%</td>
<td>4.42%</td>
<td>22</td>
</tr>
<tr>
<td>7.10%</td>
<td>2.40%</td>
<td>40</td>
</tr>
<tr>
<td>6.70%</td>
<td>2.92%</td>
<td>46</td>
</tr>
<tr>
<td>5.50%</td>
<td>2.11%</td>
<td>50</td>
</tr>
<tr>
<td>13.89%</td>
<td>6.31%</td>
<td>32</td>
</tr>
<tr>
<td>14.72%</td>
<td>9.32%</td>
<td>30</td>
</tr>
<tr>
<td>13.50%</td>
<td>10.70%</td>
<td>40</td>
</tr>
<tr>
<td>12.78%</td>
<td>5.35%</td>
<td>32</td>
</tr>
<tr>
<td>9.96%</td>
<td>6.56%</td>
<td>40</td>
</tr>
<tr>
<td>12.33%</td>
<td>3.70%</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 4. Ten single-path simulations

Table 4 shows that the average of number of trades in the trend following trading strategy is inversely correlated to the spreads of the threshold. However, the relative advantage of the trend following strategy over the buy and hold strategy is not sensitive to the perturbation of the thresholds. Next we examine the dependence of the performance on the threshold levels \((\theta^b, \theta^s)\). The results are summarized in Table 5. We can see from the Table 5 that shifting the thresholds has little impact on the performance.

<table>
<thead>
<tr>
<th>(\theta^b)</th>
<th>(\theta^s)</th>
<th>TF</th>
<th>BH</th>
<th>No. of trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.906</td>
<td>0.718</td>
<td>11.53%</td>
<td>4.71%</td>
<td>36.16</td>
</tr>
<tr>
<td>0.915</td>
<td>0.724</td>
<td>11.54%</td>
<td>4.69%</td>
<td>36.58</td>
</tr>
<tr>
<td>0.916</td>
<td>0.727</td>
<td>11.59%</td>
<td>4.70%</td>
<td>36.78</td>
</tr>
<tr>
<td>0.917</td>
<td>0.731</td>
<td>11.44%</td>
<td>4.58%</td>
<td>37.39</td>
</tr>
<tr>
<td>0.925</td>
<td>0.733</td>
<td>11.51%</td>
<td>4.72%</td>
<td>36.25</td>
</tr>
</tbody>
</table>

Table 5. Shifting the thresholds

Next we test the robustness of the algorithm. We examine the ratio of changes in \(\phi\) and a parameter in \(\{\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma\}\). We fix \((\lambda_1, \lambda_2, \mu_1, \mu_2, \sigma) = (0.36, 2.53, 0.18, -0.77, 0.184)\). Then we vary one of these parameters. The result is summarized in Tables 6-10 below. It is clear that the changes in \(\phi\) with respect to these variables are within normal ranges.

<table>
<thead>
<tr>
<th>(\Delta \lambda_1)</th>
<th>0.10</th>
<th>0.01</th>
<th>0.001</th>
<th>0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi_{\lambda_1} + \Delta \lambda_1)</td>
<td>0.06951</td>
<td>0.0646</td>
<td>0.137</td>
<td>1.79</td>
</tr>
</tbody>
</table>

Table 6. Partial derivative of \(\phi\) w.r.t. \(\lambda_1\)

<table>
<thead>
<tr>
<th>(\Delta \lambda_2)</th>
<th>0.10</th>
<th>0.01</th>
<th>0.001</th>
<th>0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi_{\lambda_2} + \Delta \lambda_2)</td>
<td>0.02225</td>
<td>0.4147</td>
<td>1.42</td>
<td>4.97</td>
</tr>
</tbody>
</table>

Table 7. Partial derivative of \(\phi\) w.r.t. \(\lambda_2\)

<table>
<thead>
<tr>
<th>(\Delta \mu_1)</th>
<th>0.10</th>
<th>0.01</th>
<th>0.001</th>
<th>0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi_{\mu_1} + \Delta \mu_1)</td>
<td>0.38712</td>
<td>0.596</td>
<td>0.783</td>
<td>4.41</td>
</tr>
</tbody>
</table>

Table 8. Partial derivative of \(\phi\) w.r.t. \(\mu_1\)

Example 2. (Real Market Tests) Next, we examine how our algorithm works in real markets. We test on the historical data of SSE (01/03/2000-12/30/2011) and SP500 (01/03/1972-12/30/2011). To run the SA algorithm, we first determine the parameters following the 19%-24% classification. Statistic of bull and bear markets for SSE index in 12 years from 2000-2011 are shown in Table 11. Similar for SP 500, see [6] for more details.

<table>
<thead>
<tr>
<th>Index</th>
<th>(\lambda_1)</th>
<th>(\lambda_2)</th>
<th>(\mu_1)</th>
<th>(\mu_2)</th>
<th>(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE(00-11)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>0.253</td>
</tr>
<tr>
<td>SP500(72-11)</td>
<td>0.353</td>
<td>2.208</td>
<td>0.196</td>
<td>-0.616</td>
<td>0.173</td>
</tr>
</tbody>
</table>

Table 11. Statistics of bull and bear markets

The testing results for trading SSE and SP 500 are summarized in Table 12. The annual return with trend following, and buy and hold strategy is collected along with annual Sharpe ratio. Using our trend following trading strategy, e.g. SSE, one dollar invested in the beginning of 2000 returns $5.30 at the end of 2011 while buy and hold strategy returns $1.61 in the same period. The annual return for trend following and buy and hold are 14.90% and 4.05%, respectively. The story for SP500 is similar: one dollar invested in the beginning of 1972 with trend following returns $15.68 which corresponds to the annual return 7.12% at the end of 2011 while buy hold returns $12.36 which corresponds to the annual return 6.49% in the same period. The equity curves for SSE and SP 500 tests are given in Figures 4 and 5, respectively. For example, in Figure 4, we have two curves: the upper represents the equity curve of trend following strategy while the lower curve is that of buy and hold method. As we can see, the trend following strategy not only outperforms the buy and hold strategy in total return, but also has a smoother equity curve which means higher Sharpe ratio.

<table>
<thead>
<tr>
<th>TF</th>
<th>BH</th>
<th>TF Sharpe</th>
<th>BH Sharpe</th>
<th>10 yr bond</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE(00-11)</td>
<td>14.90%</td>
<td>4.05%</td>
<td>1.159</td>
<td>0.461</td>
</tr>
<tr>
<td>SP500(72-11)</td>
<td>7.12%</td>
<td>6.49%</td>
<td>0.1172</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Table 12. Testing results for trend following with real stock data

Acknowledgement. We thank Professor Qiji Zhu for fruitful discussions that led to the improvement of the computational results.

REFERENCES


