Arbitrage pricing theory-based Gaussian temporal factor analysis for adaptive portfolio management

Kai-Chun Chiu*, Lei Xu

Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, PR China

Available online 17 July 2003

Abstract

Ever since the inception of Markowitz's modern portfolio theory, static portfolio optimization techniques were gradually phased out by dynamic portfolio management due to the growth of popularity in automated trading. In view of the intensive computational needs, it is common to use machine learning approaches on Sharpe ratio maximization for implementing dynamic portfolio optimization. In the literature, return-based approaches which directly used security prices or returns to control portfolio weights were often used. Inspired by the arbitrage pricing theory (APT), some other efforts concentrate on indirect modelling using hidden factors. On the other hand, with regard to the proper risk measure in the Sharpe ratio, downside risk was considered a better substitute for variance. In this paper, we investigate how the Gaussian temporal factor analysis (TFA) technique can be used for portfolio optimization. Since TFA is based on the classical APT model and has the benefit of removing rotation indeterminacy via temporal modelling, using TFA for portfolio management allows portfolio weights to be indirectly controlled by several hidden factors. Moreover, we extend the approach to some other variants tailored for investors according to their investment objectives and degree of risk tolerance.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Temporal factor analysis; Arbitrage pricing theory; Portfolio optimization; Sharpe ratio; Downside risk; Upside volatility

1. Introduction

Portfolio management has evolved as a core decision-making activity for investors and practitioners in the financial market nowadays. Prior to the inception of Markowitz's modern portfolio theory [11], theoretical research on investments has concentrated on modelling expected returns [2].

During the early stage of its development, portfolio optimization was often constrained by its static im-

plementation. Unlike dynamic portfolio optimization by which the optimal portfolio weights were tracked over time based on updated market information, the weights determined using static optimization techniques could not adapt to market changes within the investment horizon.

Despite dynamic portfolio optimization being powerful, it turned out to be a problem that required intensive computation. Recall that the most natural technique for solving dynamic portfolio optimization problems was stochastic dynamic programming. However, this approach was often compromised by several factors such as the curse of dimensionality when too many state variables were involved [7]. In

* Corresponding author.
E-mail addresses: kcchiu@cse.cuhk.edu.hk (K.-C. Chiu),
lxu@cse.cuhk.edu.hk (L. Xu).

general, practical considerations such as taxes and transactions costs also increased the number of state variables in the objective function.

In fact, this problem could be better solved via some popular machine learning approaches [3,12,13,21] which required the optimal parameters to be adaptively learned over time, and consequently, we have the term adaptive portfolio management. Among the various methodologies suggested, the most popular one is based on maximizing the well-known Sharpe ratio [17]. In implementation, trading could be based on training a trading system on labelled data [12] or directly maximizing the expected profit via the so-called adaptive supervised learning decision networks [8,21]. In this paper, these approaches were generally referred to as return-based portfolio management because they either explicitly treated the weights as constants or depend directly on the security price or returns.

Inspired by the arbitrage pricing theory (APT) in finance, which assumes that the cross-sectional expected returns of securities is linearly related to k hidden economic factors, typical statistical techniques such as principal component analysis (PCA), independent component analysis (ICA) [1,22] and maximum likelihood factor analysis [10] have been used. However, should we adopt either PCA or ICA for estimating the hidden factors, we have to compromise on the terms of zero noise. Likewise, we have to make a compromise on rotation indeterminacy if we use conventional factor analytic techniques.

In fact, many researchers also realized that variance was not appropriate for quantifying risk in the Sharpe ratio because it counted positive returns as risk. For instance, Fishburn used the lower partial moment (LPM) [5] of returns called downside risk to replace the traditional variance measure. Moreover, similar ideas were adopted for implementing portfolios optimization [8,9].

In this paper, we aim to investigate using the technique temporal factor analysis (TFA) [18] for portfolio optimization. Since TFA is based on the classical APT model and has the benefit of removing rotation indeterminacy via temporal modelling, using TFA for portfolio management allows portfolio weights to be indirectly controlled by several hidden factors. Moreover, we can extend the approach to some other variants tailored for investors according to their risk and return objectives.

The rest of the paper is organized in the following way. Sections 2 and 3 briefly review the APT and the Gaussian TFA models, respectively. Section 4 illustrates how the APT-based adaptive portfolio management can be effected with algorithms proposed in this paper. Three variants of the APT-based Sharpe ratio maximization technique are studied in Section 5. Section 6 concludes the paper.

2. Review on arbitrage pricing theory

The APT begins with the assumption that the $n \times 1$ vector of asset returns, R_t, is generated by a linear stochastic process with k factors [14–16]:

$$R_t = \bar{R} + A f_t + e_t \qquad (1)$$

where f_t is the $k \times 1$ vector of realizations of k common factors, A is the $n \times k$ matrix of factor weights or loadings and e_t is an $n \times 1$ vector of asset-specific risks. It is assumed that f_t and e_t have zero expected values so that $\bar{R}$ is the $n \times 1$ vector of mean returns. The model addresses how expected returns behave in a market with no arbitrage opportunities and predicts that an asset's expected return is linearly related to the factor loadings or

$$\bar{R} = R_f + A p \qquad (2)$$

where R_f is an $n \times 1$ vector of constants representing the risk-free return, and p is $k \times 1$ vector of risk premiums. Similar to the derivation of CAPM, Eq. (2) is based on the rationale that unsystematic risk is diversifiable and therefore should have a zero price in the market with no arbitrage opportunities.

3. Overview of temporal factor analysis

Suppose the relationship between a state $y_t \in \mathbb{R}^k$ and an observation $x_t \in \mathbb{R}^d$ is described by the first-order state-space equations as follows [18,19]:

$$y_t = B y_{t-1} + \varepsilon_t, \qquad (3)$$

$$x_t = A y_t + e_t, \qquad t = 1, 2, \ldots, N \qquad (4)$$

where ε_t and e_t are mutually independent zero-mean white noises with $E(\varepsilon_i \varepsilon_j^T) = \Sigma_\varepsilon \delta_{ij}$, $E(e_i e_j^T) = \Sigma_e \delta_{ij}$,

$E(\varepsilon_i e_j^T) = 0$, Σ_ε and Σ_e are diagonal matrices and δ_{ij} is the Kronecker delta function:

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

We call ε_t the driving noise upon the fact that it drives the source process over time. Similarly, e_t is called measurement noise because it happens to be there during measurement. The above model is generally referred to as the TFA model.

In the context of APT analysis, Eq. (1) can be obtained from Eq. (4) by substituting $(\tilde{R}_t - \bar{R})$ for x_t and f_t for y_t. The only difference between the APT model and the TFA model is the added Eq. (3) for modelling temporal relation of each factor. The added equation represents the factor series $y = \{y_t\}_{t=1}^T$ in a multichannel autoregressive process, driven by an i.i.d. noise series $\{\varepsilon_t\}_{t=1}^T$ that are independent of both y_{t-1} and e_t. Specifically, it is assumed that ε_t is Gaussian distributed. Moreover, TFA is defined such that the k sources $y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(k)}$ in this state-space model are statistically independent. The objective of TFA is to estimate the sequence of y_t's with unknown model parameters $\Theta = \{A, B, \Sigma_\varepsilon, \Sigma_e\}$ through available observations.

3.1. A learning algorithm

In implementation, an adaptive algorithm has been suggested. At each time unit, factor loadings are estimated by cross-sectional regression, and factor scores are estimated by maximum likelihood learning. Xu proposed an algorithm in Ref. [19] as shown below.

Step 1: Fix A, B, Σ_ε and Σ_e, estimate the hidden factors y_t by
$\hat{y}_t = (\Sigma_\varepsilon^{-1} + A^T \Sigma_e^{-1} A)^{-1} (A^T \Sigma_e^{-1} x_t + \Sigma_\varepsilon^{-1} B \hat{y}_{t-1})$,
$\varepsilon_t = y_t - B\hat{y}_{t-1}$,
$e_t = x_t - A\hat{y}_t$.
Step 2: Fix y_t, update A, B, Σ_ε and Σ_e by the gradient ascent approach as follows:
$B^{\text{new}} = B^{\text{old}} + \eta_0 \text{diag}[\varepsilon_t y_{t-1}^T]$,
$A^{\text{new}} = A^{\text{old}} + \eta_0 e_t y_t^T$,
$\Sigma_\varepsilon^{\text{new}} = (1 - \eta) \Sigma_\varepsilon^{\text{old}} + \eta_0 \text{diag}[\varepsilon_t \varepsilon_t^T]$,
$\Sigma_e^{\text{new}} = (1 - \eta) \Sigma_e^{\text{old}} + \eta_0 \text{diag}[e_t e_t^T]$.

where η_0 denotes the learning rate.

3.2. TFA driven by ARCH(p) process

In the finance literature, effects of autoregressive conditional heteroscedasticity (ARCH) were considered in modelling unobserved components [6] as well as hidden factors [4]. In fact, the TFA model can be directly extended so as to explicitly consider the presence of ARCH effect. For example, we may just assume that each factor series has ARCH(p) effect. Mathematically, we have

$$\varepsilon_t^{(j)} = v_t^{(j)} \psi_t^{(j)}, \quad v_t^{(j)} \sim N(0, 1)$$

$$\psi_t^{(j)^2} = a_0^{(j)^2} + \sum_{\tau=1}^{p} a_\tau^{(j)^2} \varepsilon_{t-\tau}^{(j)^2}$$

To accommodate for the learning of ARCH effect, updating of Σ_ε at time t can be alternatively done via updating $a_0^{(j)}$ and $\{a_\tau^{(j)}\}_{\tau=1}^{p}$ as shown below:

$$a_0^{(j)\text{new}} = a_0^{(j)\text{old}} + \frac{\eta a_0^{(j)}}{a_0^{(j)^2} + \sum_{\tau=1}^{p} a_\tau^{(j)^2} \varepsilon_{t-\tau}^{(j)^2}}$$
$$\times \left(\frac{\varepsilon_t^{(j)^2}}{a_0^{(j)^2} + \sum_{\tau=1}^{p} a_\tau^{(j)^2} \varepsilon_{t-\tau}^{(j)^2}} - 1 \right) \qquad (6)$$

$$a_\tau^{(j)\text{new}} = a_\tau^{(j)\text{old}} + \frac{\eta a_\tau^{(j)} \varepsilon_{t-\tau}^{(j)^2}}{a_0^{(j)^2} + \sum_{\tau=1}^{p} a_\tau^{(j)^2} \varepsilon_{t-\tau}^{(j)^2}}$$
$$\times \left(\frac{\varepsilon_t^{(j)^2}}{a_0^{(j)^2} + \sum_{\tau=1}^{p} a_\tau^{(j)^2} \varepsilon_{t-\tau}^{(j)^2}} - 1 \right) \qquad (7)$$

$$\Sigma_\varepsilon = \begin{pmatrix} \psi^{(1)^2} & 0 & \cdots & 0 \\ 0 & \psi^{(2)^2} & \cdots & 0 \\ \vdots & \ddots & \cdots & 0 \\ 0 & \cdots & 0 & \psi^{(k)^2} \end{pmatrix}$$

$$\psi^{(j)^2} = a_0^{(j)^2} + \sum_{\tau=1}^{p} a_\tau^{(j)^2} \varepsilon_{t-\tau}^{(j)^2}, \qquad j = 1, 2, \cdots, k$$

4. Gaussian TFA for adaptive portfolio management

When the APT-based Gaussian TFA model is adopted for portfolio management, portfolio weights adjustment can be made under the control of independent hidden factors that affect the portfolio. In the sequel, we illustrate how this can be achieved under the following four scenarios:

	Transaction cost	Short sale permission
Scenario I	no	no
Scenario II	yes	no
Scenario III	no	yes
Scenario IV	yes	yes

4.1. Scenario I: No transaction cost and short sale not permitted

The assumptions underlying this scenario are no transaction cost and short sale not permitted. Consequently, we consider the return of a typical portfolio which is given by Ref. [19]

$$R_t = (1 - \alpha_t)r^f + \alpha_t \sum_{j=1}^{m} \beta_t^{(j)} x_t^{(j)},$$

$$\text{subject to} \begin{cases} \alpha_t > 0, \\ 0 \le \beta_t \le 1, \\ \sum_{j=1}^{m} \beta_t^{(j)} = 1 \end{cases} \quad (8)$$

where r^f denotes the risk-free rate of return, x_t denotes returns of risky securities, α_t the proportion of total capital to be invested in risky securities and $\beta_t^{(j)}$ the proportion of α_t to be invested in the jth risky asset.

Instead of focussing on the mean variance efficient frontier, we seek to optimize the portfolio Sharpe ratio (S_p) [8] with $S_p = M(R_T)/\sqrt{V(R_T)}$ given by Ref. [19]. In other words, the objective function to maximize is:

$$\max_{\psi,\phi} S_p = \frac{M(R_T)}{\sqrt{V(R_T)}}$$

$$\text{subject to} \begin{cases} \alpha_t = \exp(\zeta_t), \\ \zeta_t = g(y_t, \psi), \\ \beta_t^{(j)} = \exp\left(\xi_t^{(j)}\right) \Big/ \sum_{r=1}^{m} \exp\left(\xi_t^{(r)}\right), \quad (9) \\ \xi_t = f(y_t, \phi). \end{cases}$$

where $M(R_T) = \frac{1}{T}\sum_{t=1}^{T} R_t$ is the conditional expected return and $V(R_T) = \frac{1}{T}\sum_{t=1}^{T}[R_t - M(R_T)]^2$ is a measure of risk or volatility, $\{y_t\}_{t=1}^{N}$ is the time series of independent hidden factors that drives the observed return series $\{x_t\}_{t=1}^{N}$, $g(y_t, \psi)$ and $f(y_t, \phi)$ are some nonlinear functions that map y_t to, respectively, ζ_t and ξ_t which, in turn, adjusts the portfolio weights α_t and $\beta_t^{(j)}$, respectively.

Maximizing the portfolio Sharpe ratio in effect balances the trade-off between maximizing the expected return and at the same time minimizing the risk. In implementation, we can simply use the gradient ascent approach. The time series $\{y_t\}_{t=1}^{N}$ can be estimated via the Gaussian TFA algorithm in Ref. [19]. Although the functions $g(y_t, \psi)$ and $f(y_t, \phi)$ are not known a priori, it may be approximated via the adaptive extended normalized radial basis function (ENRBF) algorithm in Ref. [20].

Like radial basis function (RBF) network, ENRBF is one of the popular models adopted for function approximation. The general form of RBF is

$$f_k(x) = \sum_{j=1}^{k} w_j \varphi([x - \mu_j]^T \Sigma_j^{-1}[x - \mu_j]) \quad (10)$$

ENRBF is an improved modification of RBF by replacing w_j with a linear vector function $W_j^T x + c_j$ and dividing the term $\varphi([x - \mu_j]^T \Sigma_j^{-1}[x - \mu_j])$ over the aggregate of all terms to arrive at

$$f_k(x) = \frac{\sum_{j=1}^{k} (W_j^T x + c_j)\varphi([x - \mu_j]^T \Sigma_j^{-1}[x - \mu_j])}{\sum_{j=1}^{k} \varphi([x - \mu_j]^T \Sigma_j^{-1}[x - \mu_j])} \quad (11)$$

where W_j is a parameter matrix.

Basically, each $W_j^T x + c_j$ represents a local linear segment. The ENRBF network approximates a globally nonlinear function by joining all piecewise linear segments weighted by probability. The set of parameters to be estimated is $\Theta = \{\mu_j, \Sigma_j, W_j, c_j\}_{j=1}^{k}$.

Specifically, $g(y_t, \psi)$ and $f(y_t, \phi)$ can be modelled by the ENRBF shown below.

$$g(y_t, \psi) = \sum_{p=1}^{k} (W_p^T y_t + c_p)\varphi(\mu_p, \Sigma_p, k) \quad (12)$$

$$f(y_t, \phi) = \sum_{p=1}^{\hat{k}} (\hat{W}_p^T y_t + \hat{c}_p)\varphi(\hat{\mu}_p, \hat{\Sigma}_p, \hat{k}) \quad (13)$$

where

$$\varphi(\mu_p, \Sigma_p, k) = \frac{\exp(-0.5(y_t - \mu_p)^T \Sigma_p^{-1}(y_t - \mu_p))}{\sum_{r=1}^{k} \exp(-0.5(y_t - \mu_r)^T \Sigma_p^{-1}(y_t - \mu_r))}.$$

The set of parameters in Eqs. (12) and (13) to be estimated is Θ where $\Theta = \psi \cup \phi$, $\psi = \{\mu_p, \Sigma_p, W_p, c_p\}_{p=1}^{k}$ and $\phi = \{\hat{\mu}_p, \hat{\Sigma}_p, \hat{W}_p, \hat{c}_p\}_{p=1}^{\hat{k}}$. In general, for each $\theta \in \Theta$, updating takes place adaptively in the following form:

$$\theta^{\text{new}} = \theta^{\text{old}} + \eta_0 \nabla_\theta S_{\text{p}} \quad (14)$$

where η_0 is the learning step size, $\nabla_\theta S_{\text{p}}$ denotes the gradient with respect to θ in the ascent direction of S_{p}. Typically, the adaptive algorithm shown in Table 1 can be adopted for implementation.

4.2. Simulation

4.2.1. Data considerations

All simulations in this paper are based on the past average fixed deposit interest rate, stock and index data of Hong Kong. Daily closing prices of the 1-week bank average interest rate, 3 major stock indices as well as 86 actively trading stocks covering the period from January 1, 1998 to December 31, 1999 are used. The number of trading days throughout this period is 522. The three major stock indices are, respectively, Hang Seng Index (HSI), Hang Seng China-Affiliated Corporations Index (HSCCI) and Hang Seng China Enterprises Index (HSCEI). Of the 86 equities, 30 of them are HSI constituents, 32 are HSCCI constituents and the remaining 24 are HSCEI constituents. The index data are directly used for adaptive portfolio management while the stock prices are used by Gaussian TFA for recovering independent hidden factors y_t.

4.2.2. Methodology

We consider the task of managing a portfolio which consists of four securities, the average fixed deposit interest rate and the three major stock indices in Hong Kong. The fixed deposit interest rate is used as the proxy for the risk-free rate of return r^f. The first 400 samples are used for training and the last 121 samples for testing. In the test phase, we first make prediction on $\hat{y}_t$ and $\hat{x}_t$ with $\hat{y}_t \approx By_{t-1}$ and $\hat{x}_t \approx A\hat{y}_t$. Moreover, learning is carried out in an adaptive fashion such that the actual value of x_t at time t is used to extract y_t and modify the parameters once it is known (i.e., once the current time t is passed into $t+1$). The APT-based algorithm in Table 1 is adopted that uses hidden independent factors extracted by TFA for controlling portfolio weights. We refer to this approach APT-based portfolio management. Both TFA algorithms with or without ARCH effect consideration could be used for this purpose. For simplicity, in the following experiments, we only adopt the one

Table 1

An adaptive algorithm for implementation of the APT-based portfolio management

---

Updating rules for the parameter set ψ

$\mu_p^{new} = \mu_p^{old} + \eta(\nabla_{\zeta_T}S_p)\varphi(\mu_p,\Sigma_p,k)\tau(\mu_p,\Sigma_p,W_p,c_p,k)(y_T - \mu_p)$

$\Sigma_p^{new} = \Sigma_p^{old} + \eta(\nabla_{\zeta_T}S_p)\varphi(\mu_p,\Sigma_p,k)\tau(\mu_p,\Sigma_p,W_p,c_p,k)\kappa(\mu_p,\Sigma_p)$

$W_p^{new} = W_p^{old} + \eta(\nabla_{\zeta_T}S_p)y_T\varphi(\mu_p,\Sigma_p,k)$

$c_p^{new} = c_p^{old} + \eta(\nabla_{\zeta_T}S_p)\varphi(\mu_p,\Sigma_p,k)$

Updating rules for the parameter set ϕ

$\hat{\mu}_p^{new} = \hat{\mu}_p^{old} + \hat{\eta}(\nabla_{\xi_T^{(j)}}S_p)(y_T - \hat{\mu}_p)\varphi(\hat{\mu}_p,\hat{\Sigma}_p,\hat{k})\chi(\hat{\mu}_p,\hat{\Sigma}_p,\hat{W}_{p,q},\hat{c}_{p,q},\hat{k})$

$\hat{\Sigma}_p^{new} = \hat{\Sigma}_p^{old} + \hat{\eta}(\nabla_{\xi_T^{(j)}}S_p)\kappa(\hat{\mu}_p,\hat{\Sigma}_p)\varphi(\hat{\mu}_p,\hat{\Sigma}_p,\hat{k})\chi(\hat{\mu}_p,\hat{\Sigma}_p,\hat{W}_{p,q},\hat{c}_{p,q},\hat{k})$

$\hat{W}_{p,q}^{new} = \hat{W}_{p,q} + \hat{\eta}(\nabla_{\xi_T^{(j)}}S_p)y_T\varphi(\hat{\mu}_p,\hat{\Sigma}_p,\hat{k})$

$\hat{c}_{p,r}^{new} = \hat{c}_{p,r}^{old} + \hat{\eta}(\nabla_{\xi_T^{(j)}}S_p)\varphi(\hat{\mu}_p,\hat{\Sigma}_p,\hat{k})$

where η and $\hat{\eta}$ are learning rates,

$M(R_T) = 1/T\Sigma_{t=1}^T R_t$, $V(R_T) = 1/T\Sigma_{t=1}^T [R_t - M(R_T)]^2$

$$\nabla_{\zeta_T}S_p = \frac{\left[V(R_T) - M(R_T)\left(R_T - M(R_T) - \frac{1}{T}\sum_{t=1}^T (R_T - M(R_t))\right)\right]}{T\sqrt{[V(R_T)]^3}}\left(\frac{\sum_{r=1}^m \exp(\xi_T^{(r)})x_T^{(r)}}{\sum_{r=1}^m \exp(\xi_T^{(r)})} - r^f\right)\exp(\zeta_T),$$

$$\nabla_{\xi_T^{(j)}}S_p = \frac{\left[V(R_T) - M(R_T)\left(R_T - M(R_T) - \frac{1}{T}\sum_{t=1}^T (R_t - M(R_t))\right)\right]\exp(\zeta_T)x_T^{(j)}\left(\sum_{r=1}^m \exp(\xi_T^{(r)}) - \exp(\xi_T^{(j)})\right)\exp(\xi_T^{(j)})}{T\sqrt{[V(R_T)]^3}\left(\sum_{r=1}^m \exp(\xi_T^{(r)})\right)^2},$$

$$\varphi(\mu_p,\Sigma_p,k) = \frac{\exp(-0.5(y_T - \mu_p)^T\Sigma_p^{-1}(y_T - \mu_p))}{\sum_{r=1}^k \exp(-0.5(y_T - \mu_r)^T\Sigma_r^{-1}(y_T - \mu_r))},$$

$$\kappa(\mu_p,\Sigma_p) = \Sigma_p^{-1}(y_T - \mu_p)(y_T - \mu_p)^T\Sigma_p^{-1} - 0.5\text{diag}\,[\Sigma_p^{-1}(y_T - \mu_p)(y_T - \mu_p)^T\Sigma_p^{-1}],$$

$$\tau(\mu_p,\Sigma_p,W_p,c_p,k) = \frac{(W_p^T y_T + c_p) - \sum_{r=1}^k (W_r^T y_T + c_r)\varphi(\mu_r,\Sigma_r,k)}{\sum_{r=1}^k \exp(-0.5(y_T - \mu_r)^T\Sigma_r^{-1}(y_T - \mu_r))},$$

$$\chi(\mu_p,\Sigma_p,W_{p,q},c_{p,q},k) = \frac{(W_{p,q}^T y_T + c_{p,q}) - \sum_{r=1}^k (W_{p,r}^T y_T + c_r)\varphi(\mu_r,\Sigma_r,k)}{\sum_{r=1}^k \exp(-0.5(y_T - \mu_r)^T\Sigma_r^{-1}(y_T - \mu_r))},$$

$W_{p,q}$ denotes the pth column of the qth matrix, diag[M] denotes a diagonal matrix that takes the diagonal part of a matrix M, $\zeta_T = g(y_T,\psi)$ as defined in Eq. (6) and $\xi_T^{(j)}$ is the jth output of $f(y_T,\phi)$ as defined in Eq. (7).

without ARCH consideration. For each y_t under test, we can adaptively get $\zeta_t = g(y_t,\psi)$ and $\xi_t = f(y_t,\phi_t)$, and then the portfolio weights $\alpha_t = \exp(\zeta_t)$ and $\beta_t^{(j)} = \exp(\xi_t^{(j)})/\sum_{r=1}^m \exp(\xi_t^{(r)})$. Finally, returns can be computed via Eq. (8). For the sake of comparison, we also implement a traditional approach that directly uses stock returns x_t instead of hidden factors

y_t [8]. We refer to this approach return-based portfolio management.

4.2.3. Results

Fig. 1 shows the returns of individual securities that make up the portfolio during the test phase, with relevant risk-return statistics given in Table 2. Graph-
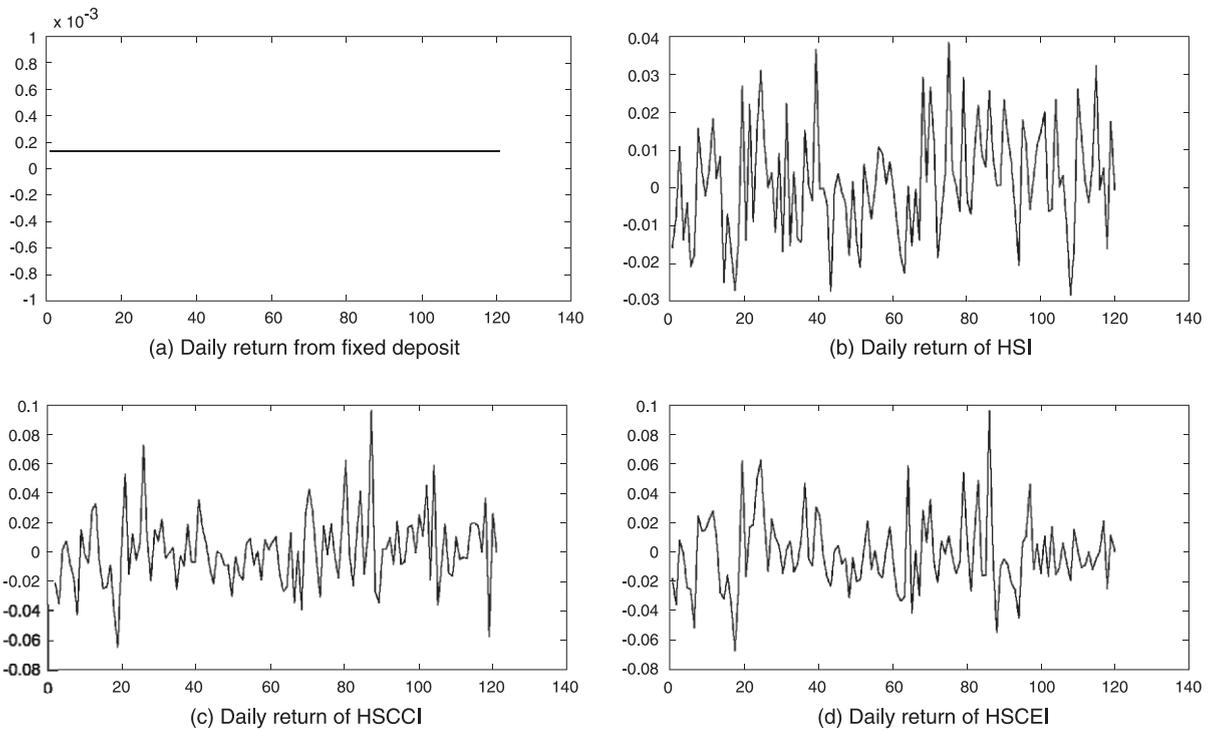
Fig. 1. Returns of individual securities in the portfolio.

ical comparison of profit gain between the two approaches using test data is shown in Fig. 2. Daily risk-return statistics of the portfolios are given in Table 3.

4.3. Scenario II: has transaction cost but short sale not permitted

Scenario II differs from Scenario I in taking into account the effect of transaction cost. Since any

Table 2
Daily risk-return statistics of constituents of portfolios

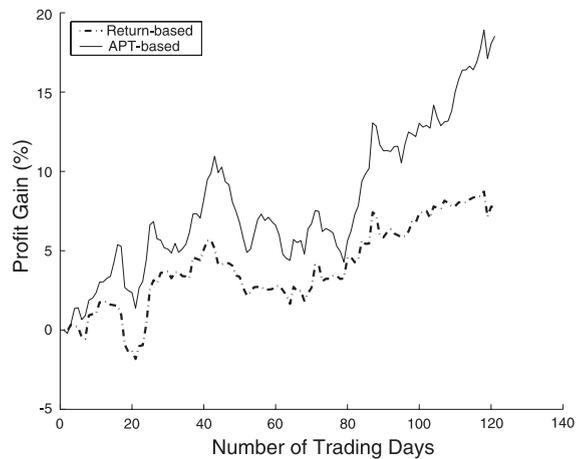| Component name	Mean return (%)	Risk (%)
Average interest rate	0.0148	0.00
HSI	0.18	1.48
HSCCI	0.03	2.51
HSCEI	− 0.20	2.55

Fig. 2. Comparative profit gain of APT-based and return-based portfolios for Scenario I.

Table 3
Daily risk-return statistics of the portfolio for Scenario I

	Return-based portfolio	APT-based portfolio	Change in Sharpe ratio ΔS_p
Mean return	0.06%	0.14%	–
Risk	0.48%	0.81%	–
Sharpe ratio	0.1250	0.1728	↑ 38.24%

change on $\beta_t^{(j)}$ leads to a transaction that incurs a cost on return c_t given by

$$c_t = -\alpha_t \sum_{j=1}^{m} r_c \mid \beta_t^{(j)} - \beta_{t-1}^{(j)} \mid p_t^{(j)}/p_{t-1}^{(j)}$$

$$= -\alpha_t \sum_{j=1}^{m} r_c \mid \beta_t^{(j)} - \beta_{t-1}^{(j)} \mid (1 + x_t^{(j)}) \quad (15)$$

where r_c is a constant denoting the rate of transaction cost. Consequently, we consider the portfolio return adjusted for transaction cost given by Ref. [19]

$$R_t = (1 - \alpha_t)r^f + \alpha_t \sum_{j=1}^{m} [\beta_t^{(j)} x_t^{(j)} - r_c \mid \beta_t^{(j)}$$

$$- \beta_{t-1}^{(j)} \mid (1 + x_t^{(j)})],$$

$$\text{subject to} \begin{cases} \alpha_t > 0, \\ 0 \le \beta_t \le 1, \\ \sum_{j=1}^{m} \beta_t^{(j)} = 1 \end{cases} \quad (16)$$

The APT-based algorithm in Table 1 could still be adopted in this case, except that the two terms $\nabla_{\zeta_T} S_p$ and $\nabla_{\xi_T^{(j)}} S_p$ become, respectively,

$$\nabla_{\zeta_T} S_p = \frac{\left[V(R_T) - M(R_T) \left(R_T - M(R_T) - \frac{1}{T}\sum_{t=1}^{T}(R_t - M(R_t)) \right) \right]}{T\sqrt{[V(R_T)]^3}}$$

$$\times \left(\sum_{j=1}^{m} \left[\frac{\exp(\xi_T^{(j)})x_T^{(j)}}{\sum_{r=1}^{m}\exp(\xi_T^{(r)})} - r_c \left| \frac{\exp(\xi_T^{(j)})}{\sum_{r=1}^{m}\exp(\xi_T^{(r)})} \right. \right.$$

$$\left. \left. - \frac{\exp(\xi_{T-1}^{(j)})}{\sum_{r=1}^{m}\exp(\xi_{T-1}^{(r)})} \right| (1 + x_T^{(j)}) \right] - r^f \right) \exp(\zeta_T),$$

$$\nabla_{\xi_T^{(j)}} S_p = \left[V(R_T) - M(R_T) \left(R_T - M(R_T) \right. \right.$$

$$\left. \left. - \frac{1}{T}\sum_{t=1}^{T}(R_t - M(R_t)) \right) \right] \exp(\zeta_T)$$

$$\times \left[x_T^{(j)} - r_c \text{sign}\left(\exp(\xi_T^{(j)}) - \exp(\xi_T^{(j-1)}) \right) \right]$$

$$\times \left[\sum_{r=1}^{m}\exp(\xi_T^{(r)}) - \exp(\xi_T^{(j)}) \right] \exp(\xi_T^{(j)})$$

$$\bigg/ \left[T\sqrt{[V(R_T)]^3} \left(\sum_{r=1}^{m}\exp(\xi_T^{(r)}) \right)^2 \right]$$

4.3.1. Simulation

For the purpose of simulation, we fix the rate of transaction cost at $r_c = 0.1\%$. Graphical comparison of profit gain between the two approaches using test data is shown in Fig. 3, while daily risk-return statistics of the portfolios are given in Table 4.

4.4. Scenario III: no transaction cost but short sale is permitted

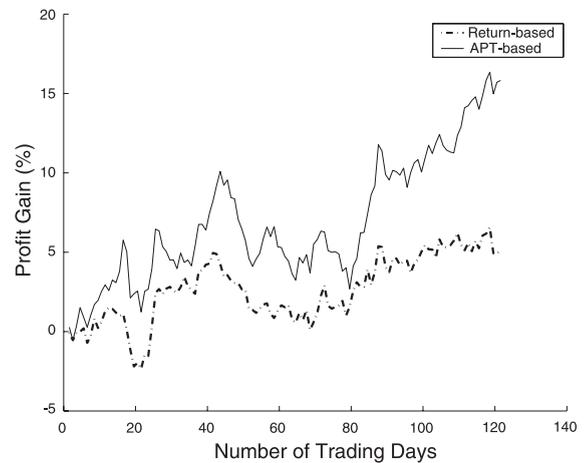Scenario III differs from Scenario I in that short sale is now permitted. By removing the

Fig. 3. Comparative profit gain of APT-based and return-based portfolios for Scenario II.

Table 4
Risk-return statistics of the portfolio for Scenario II

	Return-based portfolio	APT-based portfolio	Change in Sharpe ratio ΔS_p
Mean return	0.04%	0.12%	–
Risk	0.42%	0.73%	–
Sharpe ratio	0.0952	0.1644	↑ 72.69%

nonnegative constraints on α_t and β_t in Eq. (8), we get

$$R_t = (1 - \alpha_t)r^f + \alpha_t \sum_{j-1}^{m} \beta_t^{(j)} x_t^{(j)}$$
$$- r_c \, | \, \beta_t^{(j)} - \beta_{t-1}^{(j)} \, | \, (1 + x_t^{(j)})$$

subject to $\sum_{j=1}^{m} \beta_t^{(j)} = 1$ (17)

and the new objective function

$$\max_{\psi, \phi} S_p = \frac{M(R_T)}{\sqrt{V(R_T)}}$$

$$\text{subject to} \begin{cases} \alpha_t = \zeta_t = g(y_t, \psi), \\ \beta_t^{(j)} = \xi_t^{(j)} \Big/ \sum_{r=1}^{m} \xi_t^{(r)}, . \\ \xi_t = f(y_t, \phi) \end{cases} \qquad (18)$$

In implementation, the algorithm in Table 1 could be adopted, except the two terms $\nabla_{\zeta_T} S_p$ and $\nabla_{\xi_{T(j)}} S_p$ become, respectively,

$$\nabla_{\zeta_T} S_p = \frac{\left[V(R_T) - M(R_T)\left(R_T - M(R_T) - \frac{1}{T}\sum_{t=1}^{T}(R_t - M(R_t)) \right) \right]}{T\sqrt{[V(R_T)]^3}}$$

$$\times \left(\sum_{j=1}^{m} \left[\frac{\xi_T^{(j)} x_T^{(j)}}{\sum_{r=1}^{m} \xi_T^{(r)}} - r_c \left| \frac{\xi_T^{(j)}}{\sum_{r=1}^{m} \xi_T^{(r)}} - \frac{\xi_{T-1}^{(j)}}{\sum_{r=1}^{m} \xi_{T-1}^{(r)}} \right| (1 + x_T^{(j)}) \right] - r^f \right)$$

$$\nabla_{\xi_T^{(j)}} S_p = \frac{\left[V(R_T) - M(R_T)\left(R_T - M(R_T) - \frac{1}{T}\sum_{t=1}^{T}(R_t - M(R_t)) \right) \right] \zeta_T x_T^{(j)} \left(\sum_{r=1}^{m} \xi_T^{(r)} - \xi_T^{(j)} \right)}{T\sqrt{[V(R_T)]^3} \left(\sum_{r=1}^{m} \xi_T^{(r)} \right)^2}$$

4.4.1. Simulation

For the purpose of simulation, short selling is not applicable to the return-based approach. Graphical comparison of profit gain between the two approaches using test data is shown in Fig. 4, while daily risk-return statistics of the portfolios are given in Table 5.

4.5. Scenario IV: has transaction cost and short sale is permitted

Scenario IV differs from Scenario I in that the effects of both transaction cost and short sale on portfolio selection have to be treated appropriately. As a result, we have

$$R_t = (1 - \alpha_t)r^f + \alpha_t \sum_{j=1}^{m} [\beta_t^{(j)} x_t^{(j)} - r_c \, | \, \beta_t^{(j)}$$
$$- \beta_{t-1}^{(j)} \, | \, (1 + x_t^{(j)})],$$

subject to $\sum_{j=1}^{m} \beta_t^{(j)} = 1$ (19)

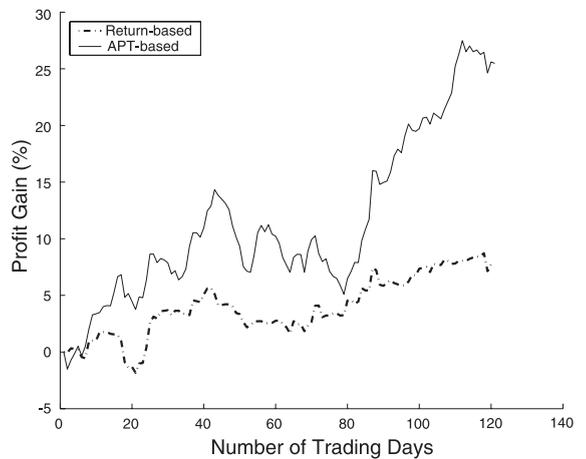Here, we have the objective function the same as Eq. (18). The APT-based algorithm in Table 1 could

Fig. 4. Comparative profit gain of APT-based and return-based portfolios for Scenario III.

Table 5
Risk-return statistics of the portfolio for Scenario III

	Return-based portfolio	APT-based portfolio	Change in Sharpe ratio ΔS_p
Mean return	0.06%	0.19%	–
Risk	0.48%	0.92%	–
Sharpe ratio	0.1250	0.2065	↑ 65.20%

still be adopted in this case, except that the two terms $\nabla_{\zeta_T} S_p$ and $\nabla_{\xi_T^{(j)}} S_p$ become, respectively,

$$
\nabla_{\zeta_T} S_p = \frac{\left[V(R_T) - M(R_T) \left(R_T - M(R_T) - \frac{1}{T} \sum_{t=1}^{T} (R_t - M(R_t)) \right) \right]}{T\sqrt{[V(R_T)]^3}}
$$
$$
\times \left(\frac{\sum_{j=1}^{m} \xi_T^{(j)} x_T^{(j)}}{\sum_{j=1}^{m} \xi_T^{(j)}} - r^f \right),
$$

$$
\nabla_{\xi_T^{(j)}} S_p = \left[V(R_T) - M(R_T) \left(R_T - M(R_T) \right. \right.
$$
$$
\left. \left. - \frac{1}{T} \sum_{t=1}^{T} (R_t - M(R_t)) \right) \right] \zeta_T
$$
$$
\times \left[x_T^{(j)} - r_c \mathrm{sign}\left(\xi_T^{(j)} - \xi_T^{(j-1)} \right) \right]
$$
$$
\times \left[\sum_{r=1}^{m} \xi_T^{(r)} - \xi_T^{(j)} \right]
$$
$$
\left/ \left[T\sqrt{[V(R_T)]^3} \left(\sum_{r=1}^{m} \xi_T^{(r)} \right)^2 \right] \right.
$$

4.5.1. Simulation

In simulation, we fix the rate of transaction cost at $r_c = 0.1\%$, and short selling is not applicable to the return-based approach. Graphical comparison of profit gain between the two approaches using test data is shown in Fig. 5, while daily risk-return statistics of the portfolios are given in Table 6.

4.6. Performance evaluation

To summarize the experimental results of the above four scenarios, we have noted the following two phenomena. First, the APT-based portfolio in general performs better than the return-based portfolio if the scope of comparison is limited to within each scenario, as evidenced by higher S_p attained in Tables 3–6. It should be noted that higher S_p may arise as a consequence of one of the following situations: (i) higher expected return, lower overall volatility; (ii) higher expected return, same overall volatility; (iii) same expected return, lower overall volatility; (iv) both expected return increase or decrease, with expected return increases (decreases) at a faster (lower) rate than overall volatility. Second, if we compare the performance of APT-based portfolios across all the four scenarios, especially the portfolio Sharpe ratio of scenario III against I (↑ 19.50%) and scenario IV against II (↑ 10.58%), we may conclude that performance may be further improved whenever short sale is permitted.

The first phenomenon reveals the fact that independent hidden factors may be more effective in controlling portfolio weights. Possible rationales include dimensionality reduction, as there are usually only a few hidden factors for a large number of securities. What seems to be a more important revelation is that the classical APT [16] model is still helpful here.
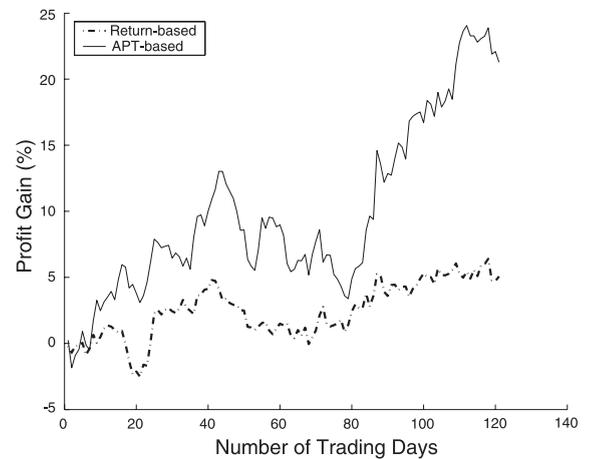

Fig. 5. Comparative profit gain of APT-based and return-based portfolios for Scenario IV.

Table 6
Risk-return statistics of the portfolio for Scenario IV

	Return-based portfolio	APT-based portfolio	Change in Sharpe ratio ΔS_p
Mean return	0.04%	0.16%	–
Risk	0.42%	0.88%	–
Sharpe ratio	0.0952	0.1818	↑ 90.97%

Although short selling is expensive for individual investors and not generally permissible for most institutional investors [2] in many markets, relevant experimental results reveal the hypothetical potential benefit such facility might add to the portfolio returns. The benefit mainly arises from the exploitation of downside trend in market price in addition to upward movement. This, in turn, reduces the chance that the fund is left idle due to declining stock prices for most stocks, which is more or less a phenomenon when the general market atmosphere is gloomy.

5. APT-based portfolio management by modified portfolio Sharpe ratio

In this section, we consider three variants of the portfolio Sharpe ratio. Specifically, we consider portfolio expected downside risk V_T^- which is represented by

$$V_T^- = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_t^{(i)} \beta_t^{(j)} \int_{-\infty}^{0} \int_{-\infty}^{0} x_t^{(i)} x_t^{(j)} p$$

$$(x_t^{(i)}, x_t^{(j)}) dx_t^{(i)} dx_t^{(j)}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_t^{(i)} \beta_t^{(j)} D_t^{(i,j)} \qquad (20)$$

where β_i and x_i denote the portfolio weight and return of the ith risky security, respectively, and $D_t^{(i,j)}$ is a constant.

In addition to considering the downside risk, the so-called portfolio expected upside volatility V_T^+ can be defined similarly as

$$V_T^+ = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_t^{(i)} \beta_t^{(j)} \int_{0}^{\infty} \int_{0}^{\infty} x_t^{(i)} x_t^{(j)} p$$

$$(x_t^{(i)}, x_t^{(j)}) dx_t^{(i)} dx_t^{(j)}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_t^{(i)} \beta_t^{(j)} U_t^{(i,j)} \qquad (21)$$

where β_i and x_i denote the portfolio weight and return of the ith risky security, respectively, and $U_t^{(i,j)}$ is a constant.

5.1. Modified Sharpe ratio with minimum downside risk and maximum upside volatility

Given that portfolio variance can be broken down into portfolio downside risk and upside volatility, it is desirable to consider the maximization of the upside volatility and minimization of the downside risk at the same time in calculating the optimal portfolio. In other words, we can consider maximization of the following improved Sharpe ratio S_p'

$$\max_{\psi, \phi} S_p' = \frac{M(R_T) + V_T^+}{V_T^-}$$

$$\text{subject to} \begin{cases} \alpha_t = \exp(\zeta_t), \\ \zeta_t = g(y_t, \psi), \\ \beta_t^{(j)} = \exp(\xi_t^{(j)}) \Big/ \sum_{r=1}^{m} \exp(\xi_t^{(r)}), \\ \xi_t = f(y_t, \phi) \end{cases} \qquad (22)$$

In implementation, the algorithm in Table 1 could be adopted, except the two terms $\nabla_{\zeta_T} S_p$ and $\nabla_{\xi_T^{(j)}} S_p$ become, respectively,

$$\nabla_{\zeta_T} S_p' = \frac{\sum_{j=1}^m \exp(\xi_T^{(j)}) \left(\sum_{j=1}^m \exp(\xi_T^{(j)}) x_T^{(j)} - r^f \sum_{j=1}^m \exp(\xi_T^{(j)}) \right) \exp(\zeta_T)}{\sum_{t=1}^T \sum_{i=1}^m \sum_{j=1}^m \exp(\xi_T^{(i)}) \exp(\xi_T^{(j)}) D_t^{(i,j)}},$$

$$\nabla_{\xi_T^{(j)}} S_p' = \left[\sum_{t=1}^T \sum_{i=1}^m \sum_{j=1}^m \exp(\xi_T^{(i)}) \exp(\xi_T^{(j)}) D_t^{(i,j)} \right.$$

$$\times \left(\exp(\zeta_T) x_T^{(j)} + \frac{\sum_{i=1}^m \exp(\xi_T^{(i)}) U_T^{(i,j)}}{\sum_{r=1}^m \exp(\xi_T^{(r)})} \right)$$

$$- \left(\sum_{t=1}^T R_t + \sum_{t=1}^T \sum_{i=1}^m \sum_{j=1}^m \exp(\xi_T^{(i)}) \exp(\xi_T^{(j)}) \right.$$

$$\left. \times U_t^{(i,j)} \right) \frac{\sum_{i=1}^m \exp(\xi_T^{(i)}) D_T^{(i,j)}}{\sum_{r=1}^m \exp(\xi_T^{(r)})} \right]$$

$$\times \left[\left(\sum_{r=1}^m \exp(\xi_T^{(r)}) - \exp(\xi_T^{(j)}) \right) \exp(\xi_T^{(j)}) \right]$$

$$\Bigg/ \left(\sum_{t=1}^T \sum_{i=1}^m \sum_{j=1}^m \exp(\xi_T^{(i)}) \exp(\xi_T^{(j)}) D_t^{(i,j)} \right)$$

$$\left. \times \left(\sum_{r=1}^m \exp(\xi_T^{(r)}) \right) \right]^2$$

5.1.1. Simulation

We implement the modified Sharpe ratio simulation using the same set of data described before and the APT-based approach in Scenario I as benchmark for comparison. Graphical comparison of profit gain between the two approaches using test data is shown in Fig. 6, while daily risk-return statistics of the portfolios are given in Table 7.
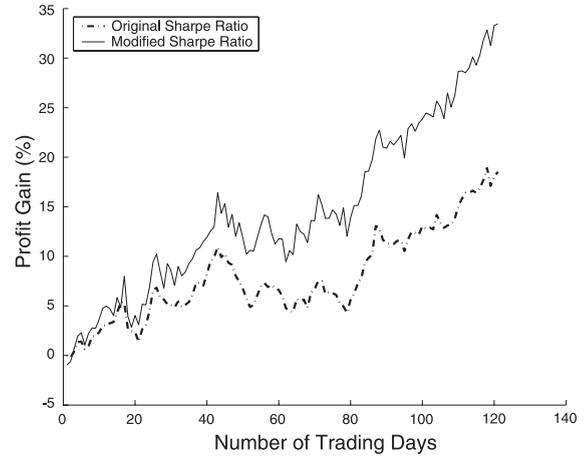

Fig. 6. Comparative profit gain under original and modified Sharpe ratio.

5.2. Risk minimization with control of expected return

Some conservative investors are more concerned about risk than return. Therefore, a more appropriate investment strategy may be to minimize risk while controlling the expected return. Particularly, this can be achieved by setting the expected return in Eq. (30) to be a constant specified by the investor, and the optimization essentially becomes a minimization of downside risk and a maximization of upside volatility.

$$\max_{\psi,\phi} S_p' = \frac{r + V_T^+}{V_T^-}$$

$$\text{subject to} \begin{cases} \alpha_t = \exp(\zeta_t), \\ \zeta_t = g(y_t, \psi), \\ \beta_t^{(j)} = \exp(\xi_t^{(j)}) \Big/ \sum_{r=1}^m \exp(\xi_t^{(r)}), \cdot \\ \xi_t = f(y_t, \phi), \\ M(R_T) = r \end{cases} \quad (23)$$

To solve the above optimization problem with equality constraints, we adopt the augmented La-

Table 7
Daily Risk-return statistics of portfolio under original and modified Sharpe ratio

	Original Sharpe ratio	Modified Sharpe ratio
Mean return	0.14%	0.24%
Risk	0.81%	1.13%
Upside volatility	–	0.43%
Downside risk	–	0.35%
Sharpe ratio S_p	0.1728	1.9143

grangian method. Specifically, for the equality constrained problem,

$$\text{maximize} \quad f(x) \text{ with respect to } x$$
$$\text{subject to} \quad h(x) = 0, \tag{24}$$

the augmented Lagrangian function can be written as

$$L(x, \lambda) = f(x) - \lambda h(x) - \frac{1}{2} c[h(x)]^2 \tag{25}$$

where λ is the Lagrange multiplier, c is the penalty parameter. Then, a sequence of minimizations of the form

$$\text{maximize} \quad L_{c_k}(x, \lambda_k) \text{ with respect to } x$$
$$\text{subject to} \quad x \in \mathbb{R}^n \tag{26}$$

is performed, where $\{c_k\}$ is a sequence of positive penalty parameters sequence satisfying

$$0 < c_k < c_{k+1} \quad \forall k$$
$$c_k \to \infty \qquad \text{as } k \to \infty \tag{27}$$

The multiplier sequence $\{\lambda_k\}$ is generated by the iteration

$$\lambda_{k+1} = \lambda_k + c_k h(\hat{x}) \tag{28}$$

where $\hat{x}$ is the solution of Eq. (26).

Here, the augmented Lagrangian is given by

$$L = \frac{r + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_t^{(i)}\beta_t^{(j)}U_t^{(i,j)}}{\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_t^{(i)}\beta_t^{(j)}D_t^{(i,j)}}$$
$$- \lambda\left(\frac{1}{T}\sum_{t=1}^{T}R_t - r\right) - \frac{c}{2}\left(\frac{1}{T}\sum_{t=1}^{T}R_t - r\right)^2 \tag{29}$$

In implementation, the algorithm in Table 1 could be adopted, except the two terms $\nabla_{\zeta_T}S_p$ and $\nabla_{\xi_T^{(j)}}S_p$ are replaced by $\nabla_{\zeta_T}L$ and $\nabla_{\xi_T^{(j)}}L$, respectively, where

$$\nabla_{\zeta_T}L = \frac{\exp(\zeta_T)}{T}\left(r^f - \frac{\sum_{r=1}^{m}\exp(\xi_T^{(r)})x_T^{(r)}}{\sum_{r=1}^{m}\exp(\xi_T^{(r)})}\right)$$
$$\times (\lambda + c(M(R_T) - r))$$

$$\nabla_{\xi_T^{(j)}}L = \left[\left(\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\exp(\xi_T^{(i)})\exp(\xi_T^{(j)})D_t^{(i,j)}\right)\right.$$
$$\times \left(\frac{\sum_{i=1}^{m}\exp(\xi_T^{(i)})U_T^{(i,j)}}{\sum_{r=1}^{m}\exp(\xi_T^{(r)})}\right)$$
$$- \left(\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\exp(\xi_T^{(i)})\exp(\xi_T^{(j)})U_t^{(i,j)}\right)$$
$$\times \frac{\sum_{i=1}^{m}\exp(\xi_T^{(i)})D_T^{(i,j)}}{\sum_{r=1}^{m}\exp(\xi_T^{(r)})}\right)$$
$$\left./\left(\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\exp(\xi_T^{(i)})\exp(\xi_T^{(j)})D_t^{(i,j)}\right)^2\right.$$

$$-\frac{\exp(\zeta_T)x_T^{(j)}}{T}\left(\lambda + c(M(R_T)-r)\right)\Biggr]$$

$$\times \frac{\left(\sum_{r=1}^{m}\exp(\xi_T^{(r)}) - \exp(\xi_T^{(j)})\right)\exp(\xi_T^{(j)})}{\left[\sum_{r=1}^{m}\exp(\xi_T^{(r)})\right]^2}$$

5.2.1. Simulation

We simulate the modified Sharpe ratio with control of expected return approach and use the modified Sharpe ratio approach in the previous subsection as benchmark. The predetermined expected return used for the simulation is $r = 0.15\%$. Graphical comparison of profit gain between the two approaches using test data is shown in Fig. 7, while daily risk-return statistics of the portfolios are given in Table 8.

5.3. Return maximization with control of expected downside risk

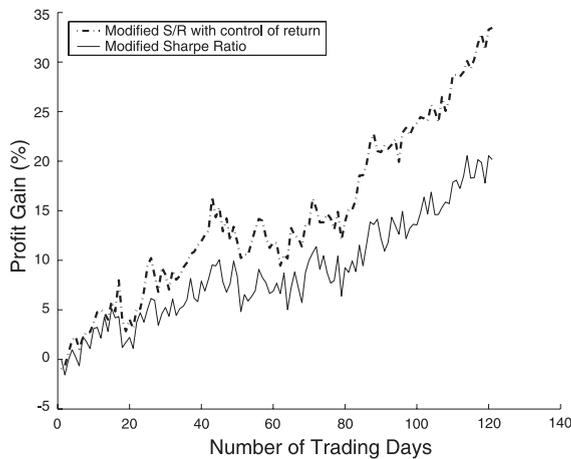Some aggressive investors are more concerned about return than risk. Therefore, a strategy that

Fig. 7. Comparative profit gain of portfolio with control of expected return.

Table 8
Risk-return statistics of portfolio with control of expected return

	Modified Sharpe ratio	Modified Sharpe ratio with control of expected return
Mean return	0.24%	0.17%
Risk	1.13%	0.79%
Upside volatility	0.43%	0.30%
Downside risk	0.35%	0.23%
Sharpe ratio S_p	1.9143	2.0435

may better serve them is to maximize the expected return while controlling the expected downside risk. In particular, this can be achieved by setting the expected downside risk in Eq. (30) to be a constant specified by the investor, and the optimization essentially becomes a maximization of expected return and upside volatility.

$$\max_{\psi,\phi} S_p' = \frac{M(R_T) + V_T^+}{v}$$

$$\text{subject to} \begin{cases} \alpha_t = \exp(\zeta_t), \\ \zeta_t = g(y_t, \psi), \\ \beta_t^{(j)} = \exp(\xi_t^{(j)})\Big/\sum_{r=1}^{m}\exp(\xi_t^{(r)}), \\ \xi_t = f(y_t, \phi), \\ V_T^- = v \end{cases} \quad (30)$$

Here, the augmented Lagrangian is given by

$$L = \frac{1}{Tv}\left(\sum_{t=1}^{T}R_t + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_t^{(i)}\beta_t^{(j)}U_t^{(i,j)}\right)$$
$$- \lambda\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_t^{(i)}\beta_t^{(j)}D_t^{(i,j)} - v\right)$$
$$- \frac{c}{2}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_t^{(i)}\beta_t^{(j)}D_t^{(i,j)} - v\right)^2 \quad (31)$$

In implementation, the algorithm in Table 1 could be adopted, except the two terms $\nabla_{\zeta_T}S_p$ and

$\nabla_{\xi_T^{(j)}} S_{\mathrm{p}}$ are replaced by $\nabla_{\zeta_T} L$ and $\nabla_{\xi_T^{(j)}} L$, respectively, where

$$\nabla_{\zeta_T} L = \frac{\exp(\zeta_T)}{Tv} \left(\frac{\displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)}) x_T^{(r)}}{\displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)})} - r_f \right)$$

$$\nabla_{\xi_T^{(j)}} L = \left[\frac{\exp(\zeta_T) Tv \displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)}) x_T^{(j)} + \displaystyle\sum_{i=1}^{m} \exp(\xi_T^{(i)}) U_T^{(i,j)}}{Tv \displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)})} \right.$$
$$\left. - \frac{\displaystyle\sum_{i=1}^{m} \exp(\xi_T^{(i)}) D_T^{(i,j)}}{T \displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)})} (\lambda + c(V_T^- - v)) \right]$$
$$\times \frac{\left(\displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)}) - \exp(\xi_T^{(j)}) \right) \exp(\xi_T^{(j)})}{\left[\displaystyle\sum_{r=1}^{m} \exp(\xi_T^{(r)}) \right]^2}$$

5.3.1. Simulation

We simulate the modified Sharpe ratio with control of expected downside risk approach and use the modified Sharpe ratio approach in the previous subsection as benchmark. The predetermined expected downside risk used for the simulation is $v = 0.20\%$.
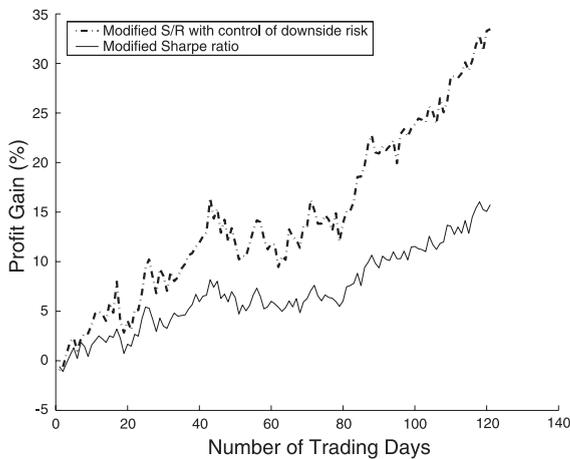

Fig. 8. Comparative profit gain of portfolio with control of expected downside risk.

Table 9
Risk-return statistics of portfolio with control of expected downside risk

	Modified Sharpe ratio	Modified Sharpe ratio with control of downside risk
Mean return	0.24%	0.15%
Risk	1.13%	0.71%
Upside volatility	0.43%	0.23%
Downside risk	0.35%	0.19%
Sharpe ratio S_{p}	1.9143	0.2000

Graphical comparison of profit gain between the two approaches using test data is shown in Fig. 8, while daily risk-return statistics of the portfolios are given in Table 9.

5.4. Performance evaluation

The investment strategy with control of expected return is well suited for risk-averse investors. By comparing the statistics shown in Table 8 with that of Table 7, we can see that not only is the expected return under control, but also is risk lowered. As a result, S_{p} remains more or less constant. This observation agrees with the tenet in finance that risk and return go hand in hand with each other. Similar reasoning could also be extended to include the case of aggressive profit-seeking investors by comparing the statistics shown in Table 9 with that of Table 7.

6. Conclusion

In this paper, we introduce how to utilize the APT-based Gaussian TFA model for adaptive portfolio management. Since TFA is based on the classical APT model and has the benefit of removing rotation indeterminacy via temporal modelling, using TFA for portfolio management would allow portfolio weights to be indirectly controlled by several hidden factors. Moreover, the approach is extended to tailor for investors according to their risk and return objectives. Simulation results reveal that APT-based portfolio management in general excels return-based portfolio management and portfolio returns may be somehow enhanced by short selling, especially when the general market climate is not that favorable.

Acknowledgements

References

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

[2] L. Chan, J. Karceski, J. Lakonishok, On portfolio optimization: forecasting covariances and choosing the risk model, The Review of Financial Studies 12 (5) (1999) 937–974.

[3] M. Choey, A.S. Weigend, Nonlinear trading models through Sharpe ratio optimization, International Journal of Neural Systems 8 (3) (1997) 417–431.

[4] M. Dungey, V. Martin, A. Pagan, A multivariate latent factor decomposition of international bond yield spreads, Journal of Applied Econometrics 15 (2000) 697–715.

[5] P.C. Fishburn, Mean-risk analysis with risk associated with below target returns, The American Economic Review 67 (2) (1977) 116–126.

[6] A. Harvey, E. Ruiz, E. Sentana, Unobserved component time series models with ARCH disturbances, Journal of Econometrics 52 (1992) 129–157.

[7] M. Haugh, A. Lo, Computational challenges in portfolio management, Computing in Science and Engineering 3 (3) (2001) 54–59.

[8] K.K. Hung, C.C. Cheung, L. Xu, New Sharpe-ratio-related methods for portfolio selection, Proceedings of Computational Intelligence for Financial Engineering (CIFEr 2000) (2000) 34–37.

[9] K.K. Hung, L. Xu, Further improvement of adaptive supervised learning decision (ASLD) network in stock market, Proceedings of the International Joint Conference on Neural Networks (IJCNN'99) 6 (1999) 3860–3865.

[10] K.G. Jöreskog, A.S. Goldberger, Factor analysis by generalized least squares, Psychometrika 37 (1972) 243–260.

[11] H. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959.

[12] J. Moody, L. Wu, Y. Liao, M. Saffell, Performance functions and reinforcement learning for trading systems and portfolios, Journal of Forecasting 17 (1998) 441–470.

[13] R. Neuneier, Optimal asset allocation using adaptive dynamic programming, Advances in Neural Information Processing Systems 8 (1996) 952–958.

[14] R. Roll, S. Ross, An empirical investigation of the arbitrage pricing theory, Journal of Finance 35 (1980) 1073–1103.

[15] R. Roll, S. Ross, The arbitrage pricing theory approach to strategic portfolio planning, Financial Analysts Journal 40 (1984) 14–26.

[16] S. Ross, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13 (1976) 341–360.

[17] W.F. Sharpe, Mutual fund performance, Journal of Business 39 (1966) 119–138.

[18] L. Xu, Temporal BYY learning for state space approach, hidden markov model and blind source separation, IEEE Transactions on Signal Processing 48 (2000) 2132–2144.

[19] L. Xu, BYY harmony learning, independent state space and generalized APT financial analyses, IEEE Transactions on Neural Networks 12 (4) (2001) 822–849.

[20] L. Xu, RBF nets, mixture experts, and Bayesian Ying–Yang learning, Neurocomputing 19 (1998) 223–257.

[21] L. Xu, Y.M. Cheung, Adaptive supervised learning decision networks for traders and portfolios, Journal of Computational Intelligence in Finance 5 (6) (1997) 11–15.

[22] F. Yip, L. Xu, An application of independent component analysis in the arbitrage pricing theory, Proceedings of the International Joint Conference on Neural Networks (IJCNN'2000) 5 (2000) 279–284.

Kai-Chun Chiu received his BBA (Hon.) in accountancy from the Chinese University of Hong Kong in 2000. He is a member of Beta Gamma Sigma. He is currently a PhD student of the Department of Computer Science and Engineering, the Chinese University of Hong Kong.

Lei Xu is a chair professor in the Department of Computer Science and Engineering of the Chinese University of Hong Kong. He has been concurrently a full professor at Peking University since 1992, and guest professor at three other universities in China and the United Kingdom. After completing his PhD thesis in Tsinghua University by the end of 1986, he joined Peking University in 1987 first as a postdoc and then became one of a few exceptionally promoted young associate professors of the university in 1988. During 1989–1993, he worked as postdoc or senior research associate in several universities in Europe and North America, including Harvard and MIT. Prof. Xu is a governor of the International Neural Network Society, the chair of the Computational Finance Technical Committee of the IEEE Neural Networks Society, a past president of Asia–Pacific Neural Networks Assembly, and an associate editor for six international journals on neural networks, including *Neural Networks* and *IEEE Transactions on Neural Networks*. He has published more than more than 240 papers in refereed journals, edited books and international conferences, with a number of them being well-cited contributions. He has received several Chinese national prestigious academic awards, including 1994 Chinese National Nature Science Award, and 1988 Chinese State Education Council Fok Ying Tung Award, and international awards (including an 1995 INNS Leadership Award). In addition, he has given over 40 keynote/plenary/invited/tutorial talks in international major neural networks conferences, such as ICONIP, WCNN, IEEE-ICNN, IJCNN, etc. He was an ICONIP'96 Program Committee Chair, a Joint-ICANN-ICONIP03 Program Committee co-chair and a general chair of IDEAL'98, IDEAL'00, IEEE CIFER'03 and IEEE FE04. Prof. Xu is an IEEE fellow, a fellow of the International Association for Pattern Recognition and a member of the European Academy of Sciences.