Kai Chun Chiu, and Lei Xu, (2002), in J.R. Dorronsoro (Ed.), Lecture Notes in Computer Sciences, Vol. 2415, pp 1019-1024, Springer Verlag.

Financial APT-Based Gaussian TFA Learning for Adaptive Portfolio Management

Kai Chun Chiu and Lei Xu

Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, P.R. China
{kcchiu, lxu}@cse.cuhk.edu.hk

Abstract. Adaptive portfolio management has been studied in the literature of neural nets and machine learning. The recently developed Temporal Factor Analysis (TFA) model mainly targeted for further study of the Arbitrage Pricing Theory (APT) is found to have potential applications in portfolio management. In this paper, we aim to illustrate the superiority of APT-based portfolio management over return-based portfolio management.

1 Introduction

In view of the rapid expansion of today's capital markets, quantitative analysis of financial data has been studied in the context of neural networks and machine learning. Adaptive portfolio management [1–3] usually refers to the study of the traditional Markowitz's portfolio theory [8] in the context of artificial neural networks.

In literature, adaptive portfolio management via maximizing the well-known Sharpe ratio [4] was studied in [1, 2]. However, such approaches either treat the weights as constants or depend directly on the security returns.

Recently, a new technique called Temporal Factor Analysis (TFA) was proposed by [5] with an aim to provide an alternative way for implementing the classical financial APT model. In this paper, we consider how the APT-based Gaussian TFA model can be used for adaptive portfolio management. Comparisons with another similar, previously adopted technique is shown.

The rest of the paper is organized in the following way. Sections 2 and 3 briefly review the APT and the Gaussian TFA model respectively. Section 4 gives an algorithm for implementing the APT-based Gaussian TFA learning for adaptive portfolio management. Comparisons with the other approach by way of an empirical study is shown in section 5. Section 6 concludes the paper.

2 Review on Arbitrage Pricing Theory

APT begins with the assumption that the $n \times 1$ vector of asset returns, $\tilde{R}_t$, is generated by a linear stochastic process with k factors [6]:

$$\tilde{R}_t = \bar{R} + Af_t + e_t \qquad (1)$$

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 a $n \times 1$ 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.

3 Overview of Temporal Factor Analysis

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

$$y_t = By_{t-1} + \varepsilon_t, \tag{2}$$

$$x_t = Ay_t + e_t, \qquad t = 1, 2, \ldots, N. \tag{3}$$

where ε_t and e_t are mutually independent zero-mean white noises with $E(\varepsilon_i \varepsilon_j) = \Sigma_\varepsilon \delta_{ij}$, $E(e_i e_j) = \Sigma_e \delta_{ij}$, $E(\varepsilon_i e_j) = 0$, Σ_ε and Σ_e are diagonal matrices, and δ_{ij} is the Kronecker delta function. Specifically, it is assumed that ε_t is Gaussian distributed. The above model is generally referred to as the Gaussian TFA model. In the context of APT analysis, (1) can be obtained from (3) 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 (2) for modelling temporal relation of each factor. The added equation represents the factor series $y = \{y_t\}_{t=1}^T$ in a multi-channel auto-regressive 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. Details about the TFA model including its adaptive algorithms for implementation can be found in [7].

4 Using Gaussian TFA for Adaptive Portfolio Management

When APT-based Gaussian TFA learning is adopted for portfolio management, portfolio weights adjustment can be made under the control of independent hidden factors [7] behind securities in the portfolio. Specifically, we consider the return of a typical portfolio which is given by

$$R_t = (1 - \alpha_t) r_t^f + \alpha_t \sum_{j=1}^{m} \beta_t^{(j)} x_t^{(j)}, \text{ subject to} \tag{4}$$

$$\begin{cases} \alpha_t > 0, \\ 0 \le \beta_t \le 1, \\ \sum_{j=1}^m \beta_t^{(j)} = 1. \end{cases}$$

where r_t^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 j^{th} risky asset.

According to Markowitz's portfolio theory [8], the optimal portfolio should lie on the efficient frontier. Alternatively, we can make use of the well-known Sharpe ratio [4] $S_p = M(R_t)/\sqrt{V(R_t)}$ to achieve the purpose [7].

$$\max_{\psi,\phi} S_p = \frac{M(R_T)}{\sqrt{V(R_T)}} \text{ subject to} \tag{5}$$

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

where $M(R_T) = \frac{1}{T}\sum_{t=1}^{T} R_t$ is a measure of 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 Sharpe ratio seeks to balance the tradeoff 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 [7]. Although the functions $g(y_t, \psi)$ and $f(y_t, \phi)$ are not known beforehand, it can be approximated via the adaptive Extended Normalized Radial Basis Function (ENRBF) algorithm in [9]. Specifically, $g(y_t, \psi)$ and $f(y_t, \phi)$ can be modelled by the ENRBF functions shown below.

$$g(y_t, \psi) = \sum_{p=1}^{k} (W_p^T y_t + c_p)\varphi(\mu, \Sigma, k) \tag{6}$$

$$f(y_t, \phi) = \sum_{p=1}^{\hat{k}} (\hat{W}_p^T y_t + \hat{c}_p)\varphi(\hat{\mu}, \hat{\Sigma}, \hat{k}) \tag{7}$$

where $\varphi(\mu, \Sigma, k) = \frac{e^{-0.5(y_t - \mu_p)^T \Sigma_p^{-1}(y_t - \mu_p)}}{\sum_{p=1}^{k} e^{-0.5(y_t - \mu_p)^T \Sigma_p^{-1}(y_t - \mu_p)}}$.

The set of parameters in (6) and (7) to 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 $\theta \in \Theta$, updating takes the following form:

$$\theta^{\text{new}} = \theta^{\text{old}} + \eta_0 \nabla_\theta S_p \tag{8}$$

where η_0 is the learning step size, $\nabla_\theta S_p$ 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.

5 Experimental Illustrations

5.1 Data Considerations

The analysis are based on past bank 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

Table 1. An adaptive algorithm for implementation of the APT-based portfolio management

Updating rules for the parameter set ψ

$$m_p^{\text{new}} = m_p^{\text{old}} + \eta(\nabla_{\zeta_T} S_p)\varphi(\mu, \Sigma, k)\tau(\mu, \Sigma, W_p, c, k)(y_T - \mu_p)$$
$$\Sigma_p^{\text{new}} = \Sigma_p^{\text{old}} + \eta(\nabla_{\zeta_T} S_p)\varphi(\mu, \Sigma, k)\tau(\mu, \Sigma, W - P, c, k)\kappa(\mu, \Sigma)$$
$$W_p^{\text{new}} = W_p^{\text{old}} + \eta(\nabla_{\zeta_T} S_p) y_T \varphi(\mu, \Sigma, k)$$
$$c_p^{\text{new}} = c_p^{\text{old}} + \eta(\nabla_{\zeta_T} S_p)\varphi(\mu, \Sigma, k)$$

Updating rules for the parameter set ϕ

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

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

$$M(R_T) = \frac{1}{T}\sum_{t=1}^{T} R_t, \quad V(R_T) = \frac{1}{T}\sum_{t=1}^{T}[R_t - M(R_T)]^2$$
$$\nabla_{\zeta_T} S_p = \frac{\{V(R_T) - M(R_T)[(R_T - M(R_T)](1 - \frac{1}{T})\}}{T\sqrt{[V(R_T)]^3}} \left(\frac{\sum_{j=1}^{m} e^{\xi_T^{(j)}} x_T^{(j)}}{\sum_{j=1}^{m} e^{\xi_T^{(j)}}} - r_t^f \right) e^{\zeta_T},$$
$$\nabla_{\xi_T^{(j)}} S_p = \frac{[V(R_T) - M(R_T)(R_T - M(R_T)(1 - \frac{1}{T})] e^{\zeta_T} x_T^{(j)} (\sum_{r=1}^{m} e^{\xi_t^{(r)}} - e^{\xi_T^{(j)}}) e^{\xi_T^{(j)}}}{T\sqrt{[V(R_T)]^3} \left(\sum_{r=1}^{m} e^{\xi_t^{(r)}} \right)^2},$$
$$\varphi(\mu, \Sigma, k) = \frac{e^{-0.5(y_t - \mu_p)^T \Sigma_p^{-1}(y_t - \mu_p)}}{\sum_{p=1}^{k} e^{-0.5(y_t - \mu_p)^T \Sigma_p^{-1}(y_t - \mu_p)}},$$
$$\kappa(\mu, \Sigma) = \Sigma^{-1}(y_T - \mu)(y_T - \mu)^T \Sigma^{-1} - 0.5\text{diag}[\Sigma^{-1}(y_T - \mu)(y_T - \mu)^T \Sigma^{-1}],$$
$$\tau(\mu, \Sigma, W_p, c, k) = \frac{(W_p^T y_T + c_p) - \sum_{p=1}^{k}(W_p^T y_T + c)\varphi(\mu, \Sigma, k)}{\sum_{p=1}^{k} e^{-0.5(y_t - \mu_p)^T \Sigma_p^{-1}(y_t - \mu_p)}},$$

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

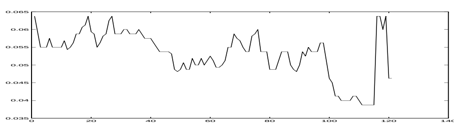
and the remaining 24 are HSCEI constituents. The index data are directly used for adaptive portfolio management while the stock prices are used by Gaussian TFA for recovering independent hidden factors y_t.

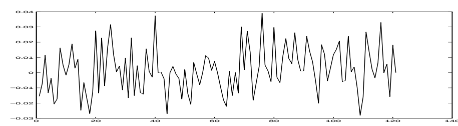
5.2 Methodology

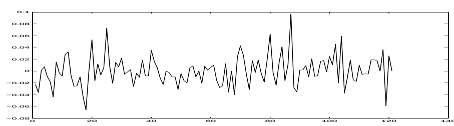
We consider the task of managing a portfolio which consists of four securities, the 1-week bank average bank interest rate and the three major stock indices in Hong Kong. The 1-week bank average bank interest rate is used as the proxy for the risk-free rate of return r_t^f. The first 400 samples are used for training and the last 121 samples for testing. Both training and test are carried out in an adaptive fashion. For the sake of comparison, three experiments are implemented. The first two, both based on the APT-based algorithm in Table 1 and uses hidden independent factors for controlling, are generally referred to as APT-based portfolio management. They differ only in the the situation where short sell of risky securities is allowed or disallowed. Short selling can be effected via modifying the algorithm in Table 1 and remembering that α and $\beta^{(j)}$

in (4) are not restricted to be non-negative. The third one, called return-based portfolio management, directly uses stock returns $\mathbf{x_t} = [\tilde{R}_{t-1}, \tilde{R}_{t-2}, \tilde{R}_{t-3}]^T$ instead of hidden factors $\mathbf{y}_t$. For each $\mathbf{y}_t$ 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 $\alpha_t = e^{\zeta_t}$ and $\beta_t^{(j)} = e^{\xi_t^{(j)}} / \sum_{r=1}^{m} e^{\xi_t^{(r)}}$. Finally, returns can be computed via (4). Fig. 2 shows the returns of individual securities that make up the portfolio during the testing period. It is clear that except for the risk-free security, investing in any single security in the portfolio cannot always guarantee positive returns.

Comparisons of the performance in terms of daily returns and cumulative profits between the APT-based portfolio and the return-based portfolio using the testing data are illustrated in Fig. 2(a) and (b) respectively. Risk-return statistics of the portfolio are given in Table 2. We may conclude that the APT-based portfolio generally outperforms the return-based portfolio.


(a) Risk-free Returns


(b) Returns of HSI


(c) Returns of HSCCI

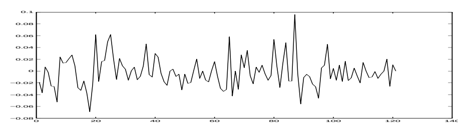
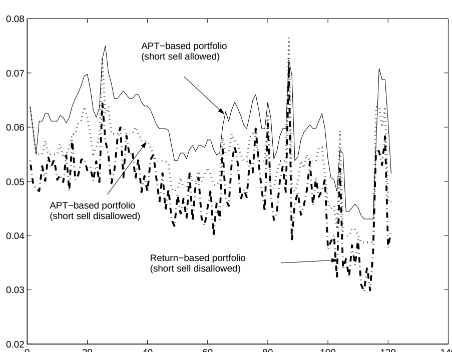
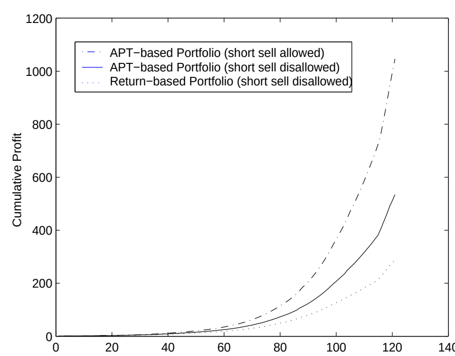

(d) Returns of HSCEI

Fig. 1. Returns of individual securities in the portfolio



(a) Daily returns of portfolios



(b) Cumulative profits of portfolios

Fig. 2. Returns and profits of APT-based and return-based portfolio: a comparison

Table 2. Risk-return statistics of the portfolio under test

Component Name	Expected Return (Mean)	Risk (Standard Deviation)
Risk-free Security	0.0528	0.0064
HSI	0.0018	0.0148
HSCCI	0.0003	0.0251
HSCEI	-0.002	0.0255
Return-based Portfolio (short sell disallowed)	0.0471	0.0110
APT-based Portfolio (short sell disallowed)	0.0538	0.0075
APT-based Portfolio (short sell allowed)	0.0709	0.0065

5.3 Performance Evaluation

The better performance of APT-based portfolio over return-based portfolio may be attributed to the contribution of independent hidden factors in controlling portfolio weights. Two advantages are evident. The first benefit arises from dimensionality reduction as there are usually only a few hidden factors for even a large number of securities. Second, the portfolio weights may be controlled more appropriately by hidden factors rather than security returns, considering the proposition of classical APT [6] that returns are generated by several hidden factors.

6 Conclusion

In this paper, we introduce how the Gaussian TFA model can be appropriately applied to adaptive portfolio management. We find that the APT-based portfolio management demonstrates superior performance over return-based portfolio management.

References

1. Choey, M., Weigend, A.S.: Nonlinear trading models through Sharpe ratio optimization. International Journal of Neural Systems **8** (1997) 417–431
2. Moody, J., Wu, L.: Optimization of trading systems and portfolios. Proc. of Computational Intelligence for Financial Engineering (CIFEr'97) (1997) 300–307
3. Xu, L., Cheung, Y.M.: Adaptive supervised learning decision networks for traders and portfolios. Proc. of Computational Intelligence for Fin. Eng. (CIFEr'97) (1997) 206–212
4. Sharpe, W.F.: The Sharpe ratio - properly used, it can improve investment. Journal of Portfolio Management (1994) 49–58
5. Xu, L.: Temporal BYY learning for state space approach, hidden Markov model and blind source separation. IEEE Trans. on Signal Processing **48** (2000) 2132–2144
6. Ross, S.: The arbitrage theory of capital asset pricing. J. Econ. Theory **13** (1976) 341–360
7. Xu, L.: BYY harmony learning, independent state space and generalized APT financial analyses. IEEE Trans. on Neural Networks **12** (2001) 822–849
8. Markowitz, H.: Portfolio Selection: Efficient Diversification of Investments. NY Wiley (1959)
9. Xu, L.: RBF nets, mixture experts, and Bayesian Ying-Yang learning. Neurocomputing **19** (1998) 223–257