K-C Chiu & L Xu, (2003), Proc. of 3rd International Workshop on Computational Intelligence in Economics and Finance (CIEF'2003), North Carolina, USA, September 26-30, 2003, pp 1096 -1099.

STOCK FORECASTING BY ARCH DRIVEN GAUSSIAN TFA AND ALTERNATIVE MIXTURE-OF-EXPERTS MODELS

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

In this paper, we introduce a new approach such that the recently developed arbitrage pricing theory (APT) based temporal factor analysis (TFA) coupled with the linear alternative mixture-of-experts model is used for stock price and index prediction. The original TFA model is extended such that the driving noise is capable of modelling any ARCH(p) process. Moreover, we use the linear alternative mixture-of-experts model rather than the traditional back-propagation networks in view of its automated model selection and better generalization ability. Comparisons with other traditional approaches are performed, with results indicating superior performance of the proposed approach.

1. INTRODUCTION

Financial time series prediction deals with the task of modelling the underlying data generation process using past observations and using the model to extrapolate the time series into the future. The application of backpropagation nets in the prediction of stock prices was initiated by [5] and subsequently developed in [3, 2]. To retain the benefit of better generalization of linear models yet exploit the advantage of wider scope of modelling power endowed with nonlinear modelling, the linear gaussian alternative mixture-of-experts model which is capable of modelling nonlinear relation via piecewise linear functions weighted by probability was adopted. For instance, stock price prediction based on learned extended normalized radial basis function (RBF) net as a special case of the alternative mixture-of-experts was realized in [6]. In the literature, Refenes [3] attempted to forecast stock prices within the framework of the arbitrage pricing theory (APT). However, the technique suffered from the constraint that factors could not be extracted from past time series data but had to be subjectively assumed to be some items on the balance sheets of companies in the universe of U.K. stocks.

THE WORK DESCRIBED IN THIS PAPER WAS FULLY SUPPORTED BY A GRANT FROM THE RESEARCH GRANT COUNCIL OF THE HONG KONG SAR (PROJECT NO: CUHK 4184/03E).

Recently, a new technique aiming at further analysis of the classical financial APT model termed temporal factor analysis (TFA) was proposed in [7]. In this paper, we consider how the APT-based gaussian TFA model can be integrated with the linear gaussian alternative mixture-of-experts model for stock price and index prediction. The proposed approach bases forecasting on a few APT factors recovered via TFA. Moreover, the TFA model is extended such that the driving noise is capable of modelling any ARCH(p) process. Comparisons with some similar, previously adopted techniques are also studied.

The rest of the paper is organized in the following way. Sections 2 briefly review the APT, the gaussian TFA model and the alternative mixture-of-experts model, respectively. Section 3 illustrates, via experimental comparisons, how gaussian TFA can be applied to stock index forecasting. Section 4 concludes the paper.

2. BRIEF REVIEW ON RELATED MODELS

2.1. The 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 [4]:

$$\mathbf{R_t} = \bar{\mathbf{R}} + \mathbf{A}\mathbf{f_t} + \mathbf{e_t} \tag{1}$$

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.

2.2. The Gaussian Temporal Factor Analysis Model

The gaussian TFA model assumes the relationship between a state $\mathbf{y}_t \in \mathbb{R}^k$ and an observation $\mathbf{x}_t \in \mathbb{R}^d$ is described by the first-order state-space equations as follows:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}\mathbf{y}_{t-1} + \varepsilon_t, \\ \mathbf{x}_t &= \mathbf{A}\mathbf{y}_t + \mathbf{e}_t, \qquad t = 1, 2, \ldots, N. \end{aligned} \tag{2}$$

where $\mathbf{e}_t$ and ε_t are further assumed to come from the following gaussian process:

$$G(\mathbf{e}_t|0,\mathbf{\Sigma_e}),\quad \mathrm{E}(\mathbf{e}_t)=0,\quad \mathrm{E}(\mathbf{y}_t\mathbf{e}_t^T)=0$$

$$G(\varepsilon_t|0,\mathbf{\Sigma}_\varepsilon),\quad \mathrm{E}(\varepsilon_t)=0,\quad E(\mathbf{y}_{t-1}\varepsilon_t^T)=0,\quad E(\mathbf{e}_t\varepsilon_t^T)=0,$$

ε_t and $\mathbf{e}_t$ are called driving and measurement noise respectively.

In implementation, an adaptive algorithm suggested in [10] is adopted and shown below. Unlike its earlier counterpart [8] which uses first order approximation under the assumption of conditional independence between different states, this algorithm achieves marginal independence via the integral $p(y_t^{(j)}) = \int p(y_t^{(j)}|y_{t-1}^{(j)})p(y_{t-1}^{(j)})dy_{t-1}^{(j)}$.

Step 1 Fix $\mathbf{A},\mathbf{B},\mathbf{\Sigma}_{t-1}^{\mathbf{y}}$ and $\mathbf{\Sigma_e}$, estimate the hidden factors $\mathbf{y}_t$ by

$$\begin{aligned}
\mathbf{\Pi} &= \mathbf{B}^{\mathrm{old}}\mathbf{\Sigma}_{t-1}^{\mathbf{y}}\mathbf{B}^{\mathrm{old}}+\mathrm{I} && (3)\\
\hat{\mathbf{y}}_t &= [\mathbf{\Pi}^{-1}+\mathbf{A}^T{\mathbf{\Sigma_e}}^{-1}\mathbf{A}]^{-1}\mathbf{A}^T{\mathbf{\Sigma_e}}^{-1}\bar{\mathbf{x}}_t && (4)\\
\mathbf{e}_t &= \bar{\mathbf{x}}_t-\mathbf{A}\hat{\mathbf{y}}_t, && (5)
\end{aligned}$$

Step 2 Fix $\mathbf{y}_t$, update $\mathbf{A},\mathbf{B},\mathbf{\Sigma}_t^{\mathbf{y}}$ and $\mathbf{\Sigma_e}$ by the gradient ascent approach as follows:

$$\begin{aligned}
\mathbf{A}^{\mathrm{new}} &= \mathbf{A}^{\mathrm{old}}+\eta\mathbf{e}_t\hat{\mathbf{y}}_t^T, && (6)\\
{\mathbf{\Sigma_e}}^{\mathrm{new}} &= (1-\eta){\mathbf{\Sigma_e}}^{\mathrm{old}}+\eta\mathbf{e}_t\mathbf{e}_t^T && (7)\\
\mathbf{C}^{\mathrm{new}} &= \mathbf{C}^{\mathrm{old}}+\eta\mathrm{diag}\Big[(\mathrm{I}-{\mathbf{B}^{\mathrm{old}}}^2)\\
&\quad (\mathbf{\Pi}^{-1}\mathbf{y}_t\mathbf{y}_t^T\mathbf{\Pi}^{-1}-\mathbf{\Pi}^{-1})\mathbf{B}^{\mathrm{old}}\mathbf{\Sigma}_{t-1}^{\mathbf{y}}\Big] && (8)\\
\mathbf{\Sigma}_t^{\mathbf{y}} &= \mathbf{B}^{\mathrm{new}}\mathbf{\Sigma}_{t-1}^{\mathbf{y}}\mathbf{B}^{\mathrm{new}}+\mathrm{I} && (9)
\end{aligned}$$

where η denotes the learning rate, $\mathbf{B}$=diag$[b_1,\cdots,b_k]$, $b_j=\frac{\exp(c_j)-\exp(-c_j)}{\exp(c_j)+\exp(-c_j)}$, $\mathbf{C}=\mathrm{diag}[c_1,\cdots,c_k]$.

2.3. ARCH Driven TFA Model

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

$$\begin{aligned}
\varepsilon_t^{(j)} &= \nu_t^{(j)}\psi_t^{(j)},\quad \nu_t^{(j)}\sim N(0,1)\\
{\psi_t^{(j)}}^2 &= {a_0^{(j)}}^2+\sum_{\tau=1}^{p_j}{a_\tau^{(j)}}^2{\varepsilon_{t-\tau}^{(j)}}^2
\end{aligned}$$

To accommodate for the learning of ARCH effect, we do not fix the the covariance matrix of ε_t to be an identity matrix. Instead, it could be modelled to simulate the ARCH(p) process. As a result, learning of $\mathbf{\Pi}$ in step 1 and $\mathbf{\Sigma_t^y}$ in step 2 can be alternative done via updating $a_0^{(j)}$ and $\{a_\tau^{(j)}\}_{\tau=1}^{p_j}$ as shown below:

$$\begin{aligned}
a_0^{(j)\mathrm{new}} &= a_0^{(j)\mathrm{old}}+\frac{\eta a_0^{(j)}}{{a_0^{(j)}}^2+\sum_{\kappa=1}^{p_j}{a_\kappa^{(j)}}^2{\varepsilon_{t-\kappa}^{(j)}}^2}\\
&\quad\left(\frac{{\varepsilon_t^{(j)}}^2}{{a_0^{(j)}}^2+\sum_{\kappa=1}^{p_j}{a_\kappa^{(j)}}^2{\varepsilon_{t-\kappa}^{(j)}}^2}-1\right) && (10)\\
a_\tau^{(j)\mathrm{new}} &= a_\tau^{(j)\mathrm{old}}+\frac{\eta a_\tau^{(j)}{\varepsilon_{t-\tau}^{(j)}}^2}{{a_0^{(j)}}^2+\sum_{\kappa=1}^{p_j}{a_\kappa^{(j)}}^2{\varepsilon_{t-\kappa}^{(j)}}^2}\\
&\quad\left(\frac{{\varepsilon_t^{(j)}}^2}{{a_0^{(j)}}^2+\sum_{\kappa=1}^{p_j}{a_\kappa^{(j)}}^2{\varepsilon_{t-\kappa}^{(j)}}^2}-1\right) && (11)
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Pi} &= \mathrm{diag}\Big[{b_1^{\mathrm{old}}}^2{\sigma_y^{(1)}}^2+{\psi^{(1)}}^2,\cdots,{b_k^{\mathrm{old}}}^2{\sigma_y^{(k)}}^2+{\psi^{(k)}}^2\Big]\\
\mathbf{\Sigma}_t^{\mathbf{y}} &= \mathrm{diag}\Big[{b_1^{\mathrm{new}}}^2{\sigma_y^{(1)}}^2+{\psi^{(1)}}^2,\cdots,{b_k^{\mathrm{new}}}^2{\sigma_y^{(k)}}^2+{\psi^{(k)}}^2\Big]
\end{aligned}$$

where ${\psi^{(j)}}^2={a_0^{(j)}}^2+\sum_{\tau=1}^{p_j}{a_\tau^{(j)}}^2{\varepsilon_{t-\tau}^{(j)}}^2, j=1,2,\cdots,k, \varepsilon_t^{(j)}$ is the j-th component of ε_t, b_j and $\sigma_y^{(j)}$ denote the j-th diagonal elements of $\mathbf{B}$ and $\mathbf{\Sigma}_{t-1}^{\mathbf{y}}$ respectively.

2.4. The Linear Gaussian Alternative Mixture-of-Experts Model

Consider the task of learning the mapping $x\rightarrow y$ in regression. From the perspective of Bayesian Ying-Yang (BYY) harmony learning, the task can be solved with the linear gaussian alternative mixture-of-experts model [9]:

$$\begin{aligned}
p(y|\mathbf{x}) &= \sum_{j=1}^{k}G(y|\mathbf{W_j}\mathbf{x}+c_j,\varrho_j^2)p(j|\mathbf{x})\\
p(j|\mathbf{x}) &= \frac{G(\mathbf{x}|\mathbf{m_j},\mathbf{\Sigma_j})\alpha_j}{\sum_{i=1}^{k}G(\mathbf{x}|\mathbf{m_i},\mathbf{\Sigma_i})\alpha_i} && (12)
\end{aligned}$$

Basically, we have the regression $E(y|\mathbf{x})=\sum_{j=1}^{k}p(j|\mathbf{x})$ $(\mathbf{W_j}\mathbf{x}+c_j)$ weighted by the gate $p(j|\mathbf{x})$. Each $\mathbf{W_j^T}\mathbf{x}+\mathbf{c_j}$ represents a local linear segment. The linear gaussian alternative mixture-of-experts model approximates a globally nonlinear function by smoothly joining all piecewise linear segments. The set of parameters to be estimated is $\Theta=\{\mathbf{m_j},\mathbf{\Sigma_j},\mathbf{W_j},\mathbf{c_j}\}_{j=1}^{k}$.

In implementation, the following iterative algorithm [9] can be adopted for parameter learning.

Initialize $\alpha_j = \frac{1}{k}$ and $\tau = 1$.

♦ Step 1
$$j_t = \arg\min_j \Big[0.5(\ln \varrho_j^2 + \frac{\|e_{t,j}\|}{\varrho_j^2} + \ln|\mathbf{\Sigma_j}| + \mathbf{e}_t^{x\,T}\mathbf{\Sigma_j}^{-1}\mathbf{e}_t^x) - \ln\alpha_j\Big]$$
$$e_{t,j} = y_t - \mathbf{W_j x_t} - c_j,$$
$$\mathbf{e}_t^x = \mathbf{x_t} - \mathbf{m_j}^{\rm old}$$

♦ Step 2
$$\eta_{t,j} = \eta_0 \frac{p_t(j)}{\tau},$$
$$p_t(j) = \delta_{j,j_t}$$

♦ Step 3
$$\alpha_j^{\rm new} = \frac{\exp(\varsigma_j^{\rm new})}{\sum_{i=1}^k \exp(\varsigma_j^{\rm new})},$$
$$\varsigma_j^{\rm new} = \varsigma_j^{\rm old} + \eta_{t,j}(1 - \alpha_j^{\rm old}),$$
if $\alpha_j^{\rm new} \to 0$, we discard the corresponding cluster j.

♦ Step 4
$$\mathbf{m_j}^{\rm new} = \mathbf{m_j}^{\rm old} + \eta_{t,j}\mathbf{e}_t^x$$
$$\mathbf{\Sigma_j}^{\rm new} = \mathbf{\Gamma_j}^{\rm new}\mathbf{\Gamma_j}^{\rm new\,T},$$
$$\mathbf{\Gamma_j}^{\rm new} = \mathbf{\Gamma_j}^{\rm old} + \eta_{t,j} G_{\mathbf{\Sigma_j}}\mathbf{\Gamma_j}^{\rm old}$$
$$G_{\mathbf{\Sigma_j}} = \mathbf{\Sigma_j}^{-1}\mathbf{e}_t^x\mathbf{e}_t^{x\,T}\mathbf{\Sigma_j}^{-1} - \mathbf{\Sigma_j}^{-1}$$

♦ Step 5
$$\mathbf{W_j}^{\rm new} = \mathbf{W_j}^{\rm old} + \eta_{t,j} e_{t,j}\mathbf{x}_t^T$$
$$c_j^{\rm new} = c_j^{\rm old} + \eta_{t,j} e_{t,j}$$
$$\varrho_j^{\rm new} = \varrho_j^{\rm old} + \eta_{t,j}\varrho_j^{\rm old}(\|e_{t,j}\|^2 - \varrho_j^{\rm old})$$
$$\tau^{\rm new} = \tau^{\rm old} + 1$$

Model selection on an appropriate number of experts is made automatically during learning with k initialized at a large enough value. With the least complexity nature of BYY harmony learning, model selection is effected via discarding the corresponding cluster j with $\alpha_j^{\rm new} \to 0$.

3. THE TFA-ME APPROACH FOR STOCK PRICE PREDICTION

First, the gaussian TFA algorithm is used to recover independent hidden factors $\mathbf{y}_{t-1}$ at time $t-1$ from cross sectional stock returns $\mathbf{x_{t-1}}$. According to our previous work [1], the number of factors determined via the model selection ability of TFA is found to be 4 for HSI constituents. Then, the linear gaussian mixture-of-experts model is adopted for establishing the relationship between $\mathbf{y}_{t-1}, x_{t-1}^{(j)}$ and $x_t^{(j)}$. This approach is abbreviated as TFA-ME

To enable comparisons to be made, we implement two other variants of the methodology. They are respectively the ME and ICA-ME approach. Accounts of experiments using these two techniques can be found in [6, 8]

ME Approach The input vector at time t consisting of stationary returns $\tilde{R}_t$ is directly used as input to the linear gaussian mixture-of-experts algorithm. The index price at time t can be recovered from the predicted returns via $p_t = p_{t-1}(1 + \tilde{R}_t + \bar{R})$, where $\bar{R}$ denotes the mean return and $\tilde{R}$ denotes return with mean removed.

ICA-ME Approach This approach consists of two steps. First, the inverse mapping $\mathbf{y}_t = W\mathbf{x}_t$ is effected via independent component analysis (ICA) for higher-than-second order dependence reduction. For this step the stock returns of the corresponding index constituents at time $t-1$ are used as input to recover independent components $\mathbf{y}_{t-1}$. Then, the linear gaussian mixture-of-experts algorithm is adopted for establishing the relationship between $\mathbf{y}_{t-1}, x_{t-1}^{(j)}$ and $x_t^{(j)}$.

3.1. Data Considerations

The analysis are based on past Hong Kong stock and index data. Daily closing prices of three 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. Of the 86 equities, 30 of them are Hang Seng Index (HSI) constituents, 32 are Hang Seng China-Affiliated Corporations Index (HSCCI) constituents, and the remaining 24 are Hang Seng China Enterprizes Index (HSCEI) constituents.

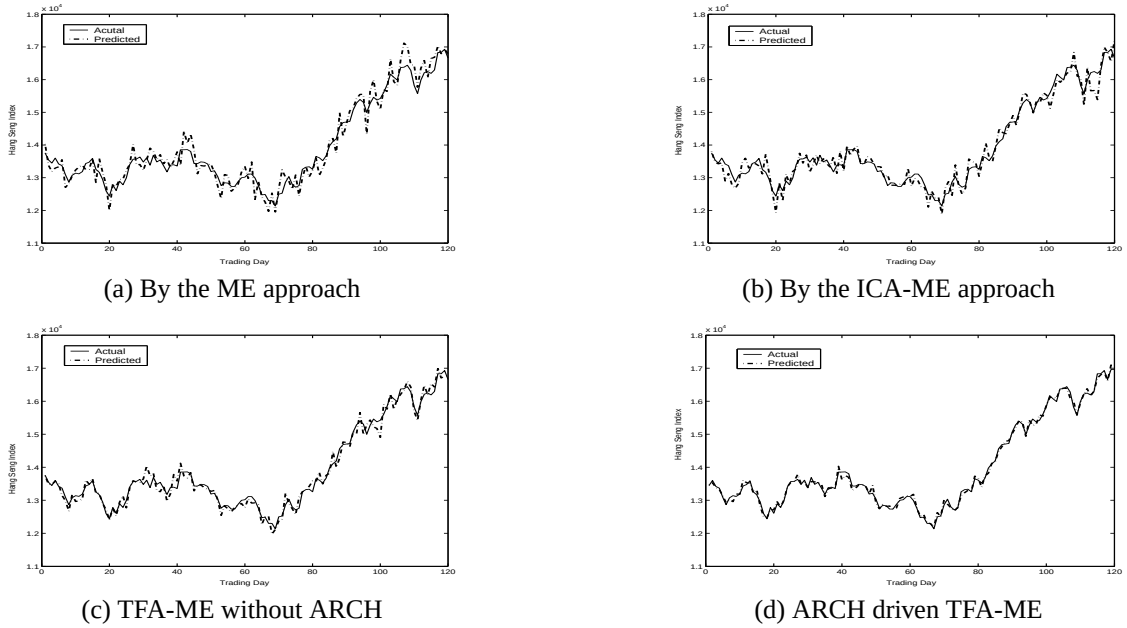
3.2. Experimental Results

Experimental investigation is based on the performance of prediction of the three stock indices, the HSI and one of the stocks, the HSBC Holding, which is also a HSI constituent. For the TFA-ME approach, we compile results based on two assumptions, one with ARCH considerations and another without ARCH. We use the first 400 data for training and the remaining 120 data for test. In the test phase, learning is carried out in an adaptive fashion such that the sample point at t is used to modify the network parameters once this point is known already (i.e., once the current time t is passed to $t+1$). We use static forecasts where actual values of lagged dependent variables are used. For the ARCH driven TFA-ME approach, we preset $p_j = 2$ for all j for simplicity. Typical results of HSI prediction using the ME, ICA-ME and TFA-ME approach are shown in Fig. 1(a)-(d) respectively.

As shown in Table 1, the ARCH driven TFA-ME approach consistently outperforms the other three approaches by having the least normalized MSE for all three indices and the stock HSBC Holding. The TFA-ME without ARCH approach comes second and and the ME approach the worst.

Table 1. normalized MSE using different approaches

Approach	HSI	HSCCI	HSCEI	HSBC
		($\times 10^{-2}$)		
ME	5.51	6.12	5.37	5.65
ICA-ME	3.85	4.12	3.10	3.71
TFA-ME	1.36	1.89	2.31	2.99
ARCH driven TFA-ME	1.15	1.63	2.05	2.53



(a) By the ME approach

(b) By the ICA-ME approach

(c) TFA-ME without ARCH

(d) ARCH driven TFA-ME

Fig. 1. Result of prediction on HSI prices.

4. CONCLUSION

In this paper, we investigate how the ARCH driven gaussian TFA model can be integrated with the mixture-of-experts model for application in stock forecasting. The mixture-of-experts model based on BYY harmony learning has the advantage of automatic model selection such that the number of experts or basis functions could be determined alongside parameter learning. Results by comparison reveal that the ARCH driven TFA-ME approach is superior to others in terms of the normalized MSE performance metrics.

5. REFERENCES

[1] K. C. Chiu and L. Xu, "A Comparative Study of Gaussian TFA Learning and Statistical Tests on the Factor Number in APT," **Proc. of International Joint Conference on Neural Networks (IJCNN'02)**, pp. 2243–2248, 2002.

[2] C. L. Giles, S. Lawrence and A. C. Tsoi, "Rule Inference for Financial Prediction using Recurrent Neural Networks," **Proc. of 1997 IEEE/IAFE Conf. of Comput. Intell. for Financial Engineering**, pp. 253–259, 1997.

[3] A. N. Refenes, M. Azema-Barac and A. D. Zapranis, "Stock Ranking: Neural Networks Vs Multiple Linear Regression," **IEEE International Conference on Neural Networks**, vol. 3, pp. 1419–1426, 1993.

[4] S. Ross, "The arbitrage theory of capital asset pricing," **Journal of Economic Theory**, vol. 13, pp. 341–360, 1976.

[5] H. White, "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," **IEEE International Conference on Neural Networks**, 1988.

[6] L. Xu, "RBF Nets, Mixture Experts, and Bayesian Ying-Yang Learning," **Neurocomputing**, vol. 19, pp. 223–257, 1998.

[7] L. Xu, "Temporal BYY Learning for State Space Approach, Hidden Markov Model and Blind Source Separation," **IEEE Trans. on Signal Processing**, vol. 48, pp. 2132–2144, 2000.

[8] L. Xu, "BYY Harmony Learning, Independent State Space and Generalized APT Financial Analyses," **IEEE Transactions on Neural Networks**, vol. 12, no. 4, pp. 822–849, 2001.

[9] L. Xu, "BYY Harmony Learning, Structural RPCL, and Topological Self-Organizing on Mixture Models," **Neural Networks**, vol. 15, pp. 1125–1151, 2002.

[10] L. Xu, "Mining Dependence Structures: (II) from An Independence Analysis Perspective," **Proceedings of IEEE ICDM 2003 Workshop on the Foundation of Data Mining and Discovery**, pp. 39–57, 2002.