New Approaches for financial prediction, portfolio management, and market modeling

Lei Xu
http://www.cse.cuhk.edu.hk/~lxu/

Department of Computer Science and Engineering,
The Chinese University of Hong Kong

1. Financial Prediction

- RPCL competitive learning based piecewise linear prediction
- Extended radial basis functions, Mixture of expert model and financial prediction
- Finite mixture of ARCH and GRACH models for prediction
2. Portfolio Management

- Portfolio Management by Learned Decisions
- Markowitz Portfolio, Sharpe's ratio and Downside risk
- Improved Portfolio Sharpe Ratio Maximization with Diversification
- Adaptive Portfolio Management based on Extended RBF nets and analyses of market factors

3. Arbitrage Pricing Theory

- Capital Asset Pricing Model vs. Arbitrage Pricing Theory
- Three Types of APT Implementation and Incapability of Factor analysis
- Temporal Factor Analysis (TFA) and APT
- TFA based APT for Prediction and Portfolio Management
4. Challenges and Advances of Statistical Learning

- Two types of Intelligent Ability: Learning from Samples
- Key Ingredients of Statistical Learning
- Two Key Challenges and Advances on Seeking Solutions
- A Unified Theory: Bayesian Ying-Yang Harmony Learning

1. Financial Prediction

- RPCL competitive learning based piecewise linear prediction
- Extended radial basis functions, Mixture of expert model and financial prediction
- Finite mixture of ARCH and GRACH models for prediction
- APT-TFA based prediction
The interface of HK stock analyzer and the graph shows the predicted stock price on 29th April and the historical stock price of HSBC Holding from 15th April to 25th April.
Autoregressive model \( z_{t+1} = \sum_{r=1}^{p} a_r z_{t+1-r} + \varepsilon_t \), \( P(\varepsilon_t) = G(\varepsilon_t \mid 0, \lambda^2) \)

\begin{align*}
\mathbf{z}_1 & \mathbf{z}_2 \cdots \mathbf{z}_{q_1} & \mathbf{y}_1 & \mathbf{y}_1 = \mathbf{z}_{q_1+1} \\
\mathbf{z}_2 & \mathbf{z}_3 \cdots \mathbf{z}_{q_1+1} & \mathbf{y}_2 & \mathbf{y}_2 = \mathbf{z}_{q_1+2} \\
\vdots & \vdots & \vdots & \vdots \\
\mathbf{z}_{N} & \mathbf{z}_{N+1} \cdots \mathbf{z}_{N+q_1} & \mathbf{y}_N & \mathbf{y}_N = \mathbf{z}_{q_1+N} \\
\end{align*}

\[ x_t = \begin{bmatrix} z_t & z_{t+1} & \cdots & z_{t+q_1} \end{bmatrix}^\intercal \quad \mathbf{a} = [a_1 \cdots a_{q_1}]^\intercal \]

\[ a^\intercal x_t - y_t = \varepsilon \quad \text{or} \quad y_t = a^\intercal x_t + \varepsilon \]

\[ P(\varepsilon) = G(\varepsilon \mid 0, \lambda^2) \]

\[ \min E \|\varepsilon\|^2 \]
\[ x_t = [z_t, z_{t+1}, \ldots, z_{t+q_t}]^t \quad x_t = [z_t, z_{t+1}, \ldots, z_{t+q_2}]^t \quad x_t = [z_t, z_{t+1}, \ldots, z_{t+q_k}]^t \]

\[ a_1 = [a_{11}, \ldots, a_{1q_1}]^t \quad a_2 = [a_{21}, \ldots, a_{2q_2}]^t \quad a_k = [a_{k1}, \ldots, a_{kq_k}]^t \]

\[ y_t = a_1^t x_t + c_1 + \varepsilon_1 \quad y_t = a_2^t x_t + c_2 + \varepsilon_2 \quad y_t = a_k^t x_t + c_k + \varepsilon_k \]

\[ z_{t+1} = \sum_{\tau=1}^{q_t} a_{j,\tau} z_{t+1-\tau} + c_j + \varepsilon_j, \quad P(\varepsilon_j) = G(\varepsilon_j \mid 0, \lambda_j^2), \quad j = 1, \ldots, k \]

Maximum Learning on \( P(\varepsilon_j \mid \theta) = \sum_{j=1}^{k} \alpha_j G(\varepsilon_i \mid 0, \lambda_j^2) \),

\[ \theta = \{ \alpha_j, \lambda_j^2, a_{j,\tau}, \tau = 1, \ldots, q_j \}_j \]

Implemented by the EM algorithm (Xu, 1995, Proc. of IEEE NNSP-1995, USA)

Another implementation: from Clustering to Gaussian mixture

\[
x_t = [z_t, z_{t+1}, \ldots, z_{t+q}]^t
\]

\[
a_1 = [a_{11}, \ldots, a_{1q}]^t \quad a_2 = [a_{21}, \ldots, a_{2q}]^t \quad a_k = [a_{k1}, \ldots, a_{kq}]^t
\]

\[
y_t = a_1^t x_t + c_1 + \varepsilon_1 \quad y_t = a_2^t x_t + c_2 + \varepsilon_2 \quad y_t = a_k^t x_t + c_k + \varepsilon_k
\]

\[
y_t - \mu_1 = a_1^t (x_t - m_1) + \varepsilon_1 \quad y_t - \mu_2 = a_2^t (x_t - m_2) + \varepsilon_2 \quad y_t - \mu_k = a_k^t (x_t - m_k) + \varepsilon_k
\]

\[
c_j = \mu_j - a_j^t m_j,
\]

\[
P(\varepsilon_j) = G(\varepsilon_j \mid 0, \lambda_j^2),
\]

\[
j = 1, \ldots, k
\]

\[
P(\varepsilon_j \mid \theta) = \sum_{j=1}^k \alpha_j G(\varepsilon_j \mid 0, \lambda_j^2)
\]

Phase I:
When \( m_1, m_2, m_3 \) are given, we can get \( a_j, \mu_j \) by the least square regression, i.e.

\[
\min \sum_t \|\varepsilon_t\|^2, \quad \varepsilon_t = y_t - \mu_2 - a_2^t (x_t - m_2)
\]
Phase II: solve $m_1, m_2, m_3$

Mean Square Error (MSE) Clustering

\[
\min \varepsilon(m) = \frac{1}{2} \sum_{k=1}^{M} \sum_{x_i \in S_k} \|x_i - m_k\|^2 \\
S_k = \left\{ x_i : \|x_i - m_k\| < \|x_i - m_j\|, j \neq k \right\}
\]

Voronoi division

K-means algorithm

Local search
Competitive Learning

(a) $m_1$ is the winner

(b) $m_2$ is the winner

(c) $m_1$ and $m_2$ are converged

Frequency Sensitive Competitive Learning

(a) $m_1$ is the winner

(c) $m_1$ and $m_2$ are converged

$$m_c^{new} = m_c^{old} + \eta (x_t - m_c^{old})$$

$$c = \arg \min_j d_j, \quad d_j = \|x_t - m_j^{old}\|^2$$

$\eta$ works

still a problem

How to decide $k$?

$$f_j \quad \text{frequency} \ m_j \ \text{wins}$$
Rival Penalized Competitive Learning (RPCL)

\[ d_3 > d_2 > d_1 \]

(a) \( m_1 \) is the winner
\( m_2 \) is the rival

\[ d_1 > d_2 > d_3 \]

(b) \( m_3 \) is the winner
\( m_2 \) is the rival

(c) \( m_1 \) and \( m_3 \) are converged
\( m_2 \) is driven far away

\[
m_c^{new} = m_c^{old} + \eta_c (x_t - m_c^{old}), \quad c = \arg \min_j d_j
\]

\[
m_r^{new} = m_r^{old} - \eta_r (x_t - m_r^{old}), \quad r = \arg \min_{j \neq c} d_j
\]

k is determined automatically during learning

See Xu, Oja & Krzyzak, Proc. ICPR02, then

Phase III:
When $a_j, \mu_j, m_j$ are given, we can predict
\[
\hat{y}_t = \mu_c + a_c^t (x_t - m_c), \quad c = \arg \min_j \|x_t - m_j\|^2
\]
Fig. 5. Trading system with two steps: prediction step followed by trading step. In prediction step, the prediction model outputs $\hat{y}(t)$ — the estimation of desired output $Y(t)$ when the input $Z(t) = [z_t, z_{t+1}, \ldots, z_{t+d-1}, z_{t+d}]$ is available. In trading step, the trading model will output a trading signal $I(t)$ based on $\hat{y}(t)$ and current information $z_{t+d}$ for the next trading activity.

$I(t) = \begin{cases} -1, & \text{if } \hat{z}_{t+1} - z_t > 0 \\ 0, & \text{if } \hat{z}_{t+1} - z_t = 0 \\ 1, & \text{otherwise} \end{cases}$

where $I(t) = 1, 0 \text{ and } -1$ stand for the “buy long,” “do nothing” and “buy short” signals respectively.
Fig. 12. Profits from different prediction models with trading strategy 1.
Gaussian Mixture

\[ l = \frac{1}{\alpha_1} \cdots \frac{1}{\alpha_k} \]

\[ G(x|\mu_1, \Sigma_1) \cdots G(x|\mu_k, \Sigma_k) \]

\[ x_1 \quad x_k \]

\[ \alpha_j \geq 0, \sum_{j=1}^{K} \alpha_j = 1 \]

\[ q(x) = \sum_{i=1}^{k} \alpha_i G(x|\mu_i, \Sigma_i) \]

\[ p(x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x-x_i) \]

The EM Algorithm

\[ l = \frac{1}{\alpha_1} \cdots \frac{1}{\alpha_k} \]

\[ G(x|\mu_1, \Sigma_1) \cdots G(x|\mu_k, \Sigma_k) \]

Bayesian Inversion

\[ p(l|x) = \frac{\alpha_i G(x|\mu_i, \Sigma_i)}{\sum_{i} \alpha_i G(x|\mu_i, \Sigma_i)} \]

\[ p(l|x) \text{ is free} \]

\[ \alpha_i^{\text{new}} = \frac{1}{N} \sum_{i=1}^{N} p(l|x_i) \]

\[ n_i^{\text{new}} = \alpha_i^{\text{new}} N \]

\[ \mu_i^{\text{new}} = \frac{1}{n_i^{\text{new}}} \sum_{i=1}^{N} p(l|x_i)x_i \]

\[ \Sigma_i^{\text{new}} = \frac{1}{n_i^{\text{new}}} \sum_{i=1}^{N} p(l|x_i) [x_i - \mu_i^{\text{new}}] [x_i - \mu_i^{\text{new}}]^T \]
Comparison of EM Algorithm And Gradient Approach

• For Gaussian mixtures, connection between EM and gradient is set up

\[ \Theta^{(k+1)} = \Theta^{(k)} + P(\Theta^{(k)}) \frac{\partial L}{\partial \Theta^{(k)}} \]

EM searches in a positive projection of \( \frac{\partial L}{\partial \Theta^{(k)}} \)

• EM converges from any initial conditions
  Automatic satisfaction of Constraints

• \( P(\Theta^{(k)}) \) varies adaptively: Quasi - Newton Speed.


Existing Efforts

- VC Dimension based SRM
- AIC
- BIC, SIC
- Cross Validation
- MML/MDL
- Bayesian Approach

A finite size \( N \) of samples

\[ k^* = \arg\min_k [\Delta(k) + F(p(x \mid \theta(k)), X)] \]

The existing efforts usually lead to a rough estimate \( \Delta(k) \)
Two Steps of Solving

Step 1  Enumerate $k$ for a set of candidate values, fixed at each candidate, make learning

$$\theta^*(k) = \arg \min_{\theta} F(p(x \mid \theta), X)$$

Step 2  Select the best one $k^*$ by

$$k^* = \arg \min_k [\Delta(k) + F(p(x \mid \theta(k)), X)]$$

Very computational extensive !!!

Gaussian mixture and clustering analyses
(Studied as a typical special case of BYY harmony learning)

- $J(k)$ for $k$, especially on a small size of samples,
- The smoothed EM algorithm,
  (Xu, Pattern Recognition Letter, 1997; ICONIP97)
- Adaptive algorithms with $k$ determined automatically during learning,
- Kernel density estimation via support vectors via BYY learning.
  (Xu, Int. J. Neural Systems, 2001; Neural Networks, 2002)

\[ H(\theta, k) \]
\[ \theta = \{\theta_1^*, \theta_2^*\}, \theta_2^* \text{ under constrain} \]

\[ H(\theta, k) \]
\[ \theta = \{\theta_1^*, \theta_2^*\}, \theta_2^* = 0 \]


Then, we can get $\mathbf{a}_j, \mu_j$ by the least square regression, i.e.

$$\min \sum_{t} \sum_{j=1}^{k} p(j \mid \mathbf{x}_t) \| y_t - \mu_j - \mathbf{a}_j^t (\mathbf{x}_t - m_j) \|^2$$

With $a_j, \mu_j, m_j, \Sigma_j$ obtained, we can predict

$$\hat{y}_t = \sum_{j=1}^{k} p(j \mid \mathbf{x}_t) [\mu_j + \mathbf{a}_j^t (\mathbf{x}_t - m_j)]$$

$$p(j \mid \mathbf{x}_t) = \frac{\alpha_j G(\mathbf{x}_t \mid m_j, \Sigma_j)}{\sum_{j=1}^{k} \alpha_j G(\mathbf{x}_t \mid m_j, \Sigma_j)}$$
**Prediction error based RPCL learning**

\[ d_j(\theta_j) = \| y_t - \mu_j - a'_j(x_t - m_j) \|^2 \]

\[ \theta_{c}^{\text{new}} = \theta_{c}^{\text{old}} - \eta_c \frac{\partial d_c(\theta_c)}{\partial \theta_c}, \quad c = \arg \min_j d_j \]

\[ \theta_{r}^{\text{new}} = \theta_{r}^{\text{old}} + \eta_r \frac{\partial d_r(\theta_r)}{\partial \theta_r}, \quad r = \arg \min_{j \neq c} d_j \]

\[ c_j = \mu_j - a'_j m_j, \]

\[ P(\varepsilon_j) = G(\varepsilon_j \mid 0, \lambda_j^2), \quad j = 1, \cdots, k \]

**Prediction error based Gaussian Mixture**

\[ c_j = \mu_j - a'_j w_j, \]

\[ x_t = [z_t, z_{t+1}, \cdots, z_{t+q_t}]^t \quad x_t = [z_t, z_{t+1}, \cdots, z_{t+q_2}]^t \quad x_t = [z_t, z_{t+1}, \cdots, z_{t+q_k}]^t \]

\[ a_1 = [a_{11}, \cdots, a_{1q_1}]^t \quad a_2 = [a_{21}, \cdots, a_{2q_2}]^t \quad a_k = [a_{k1}, \cdots, a_{kq_k}]^t \]

\[ y_t = a'_1 x_t + c_1 + \varepsilon_1 \quad y_t = a'_2 x_t + c_2 + \varepsilon_2 \quad y_t = a'_k x_t + c_k + \varepsilon_k \]

\[ z_{t+1} = \sum_{\tau=1}^{q_j} a_{j,\tau} z_{t+1-\tau} + c_j + \varepsilon_j, \quad P(\varepsilon_j) = G(\varepsilon_j \mid 0, \lambda_j^2), \quad j = 1, \cdots, k \]

Maximum Learning on \( P(\varepsilon_j \mid \theta) = \sum_{j=1}^{k} \alpha_j G(\varepsilon_j \mid 0, \lambda_j^2) \),

\[ \theta = \{\alpha_j, \lambda_j^2, a_{j,\tau}, \tau = 1, \cdots, q_j\} \]
E step:

\[ p(j \mid \epsilon_t) = \frac{\alpha_j G(\epsilon_t \mid 0, \lambda_j^2)}{\sum_{j=1}^{k} \alpha_j G(\epsilon_t \mid 0, \lambda_j^2)} \]

M step:

\[ \max \sum_{t} \sum_{j=1}^{k} p(j \mid \epsilon_t) \ln[\alpha_j G(\epsilon_t \mid 0, \lambda_j^2)] \]

\[ \max_{\alpha_j \geq 0, \lambda_j^2 \geq 0} \sum_{t} \sum_{j=1}^{k} p(j \mid \epsilon_t) \ln[\alpha_j G(\epsilon_t \mid 0, \lambda_j^2)] \]

\[ \min_{\mu_j, a_j, m_j} \sum_{t} \sum_{j=1}^{k} p(j \mid \epsilon_t) \| y_t - \mu_j - a_j^t (x_t - m_j) \|^2 \]


C.H. Wong, F. Yung, & L. Xu, Proc. of NCNN96, China
Summary

\[ \hat{y}_t = \sum_{j=1}^{k} p(j|x_t)[\mu_j + a'_j(x_t - m_j)] \]

\[ \hat{y}_t = \sum_{j=1}^{k} p(j|x_t)[\sum_{r=1}^{q_j} a_{j,r} z_{t+r-\tau} + c_j] \]

Examples of Mixture of experts
1. Financial Prediction

- RPCL competitive learning based piecewise linear prediction

- Extended radial basis functions, Mixture of expert model and financial prediction

- Finite mixture of ARCH and GRACH models for prediction

- APT-TFA based prediction
Mixture-of-experts

\[ E(y \mid x) = f(x, \theta) = \sum_j p(j \mid x) f(x, \theta_j) \]
\[ p(y \mid x) = \sum_j p(j \mid x) G(y \mid f(x, \theta_j), \sigma_j^2 I) \]

![Diagram of Mixture-of-experts model]

Linear Mixture-of-experts

\[ E(y \mid x) = f(x, \theta) = \sum_j p(j \mid x) f(x, \theta_j) \]
\[ p(y \mid x) = \sum_j p(j \mid x) G(y \mid f(x, \theta_j), \sigma_j^2 I) \]
\[ f(x, \theta_j) = \begin{cases} W_j x + c_j \\ c_j \end{cases} \]

![Diagram of Linear Mixture-of-experts model]
RBF nets and Extended RBF nets

\[ f_k(x) = \sum_{j=1}^{k} \frac{(x + c_j) \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}{\sum_{j=1}^{k} \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}, \]

\( \Sigma_j = \sigma^2 I, \) a RBF net

\[ f_k(x) = \sum_{j=1}^{k} \frac{0 + c_j) \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}{\sum_{j=1}^{k} \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}, \]

an elliptic RBF net

\[ f_k(x) = \sum_{j=1}^{k} \frac{(W^T x + c_j \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}{\sum_{j=1}^{k} \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}, \]

Extended RBF net

---

Step I of learning

Mean Square Error (MSE) Clustering

Gaussian Mixture

\[ G(x_i | m_1, \Sigma_1) \]
\[ G(x_i | m_2, \Sigma_2) \]
\[ G(x_i | m_3, \Sigma_3) \]

K-mean algorithm  EM algorithm
Step II of learning

\[ f_k(x) = \frac{\sum_{j=1}^{k} (W_j^T x + c_j) \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}{\sum_{j=1}^{k} \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])} = p(j \mid x) \]

We can get \( W_j, c_j \) by the least square regression, i.e.,

\[
\min \sum_{t} \sum_{j=1}^{k} p(j \mid x_t) \| y_t - W_j x - c_j \|^2
\]

Extended RBF nets: Specific cases of Linear ME

\[
f_k(x) = \frac{\sum_{j=1}^{k} (W_j^T x + c_j) \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}{\sum_{j=1}^{k} \phi([x - m_j]^T \Sigma_j^{-1} [x - m_j])}, \quad p(j \mid x) = \frac{\alpha_j G(x, \Sigma)}{\sum_{j=1}^{k} \alpha_j G(x, \Sigma)}, \quad \alpha_j = \frac{\| \Sigma_j \|}{\sum_{r} \| \Sigma_r \|},
\]

\[
p(j \mid x) = \frac{\exp(-0.5(x - m_j)^T \Sigma_j^{-1} (x - m_j))}{\sum_{j} \exp(-0.5(x - m_j)^T \Sigma_j^{-1} (x - m_j))}
\]
Mixture-of-experts

\[ p(y | x) = \sum_j p(j | x) G(y | f(x, \theta_j), \sigma_j^2 I) \]
\[ E(y | x) = f(x, \theta) = \sum_j p(j | x) f(x, \theta_j) \]

- The EM algorithm (Jordan & Jacobs, 1994)
- Study on its convergence (Jordan & Xu, Neural Networks, 1995)
- J(k) for selecting k and automatic selection during learning
  (Xu, Neurocomputing, 98, Intl J Neural Systems, 01, Neural Networks, 02)

Extended RBF nets

- statistical consistency, convergence rates and receptive field size, among early major theoretical results in the literature of RBF nets.
  (Xu, Krzyzak, & Yuille, Harvard Robotic Lab, T.Rep, 1992, Neural Networks, 94)

- EM algorithm in place of the suboptimal clustering +LMS way
  (Xu, Neurocomputing, 98)

- J(k) for selecting k and automatic selection (either RPCL or BYY learning)
  (Xu, Neurocomputing, 98, Intl J Neural Systems, 01, Neural Networks, 02)

- applied to time series prediction, financial portfolio management.
**Mixture-of-experts**

\[ p(j | x_t) = \frac{e^{-h_j(x_t, \phi_j)}}{\sum_r e^{-h_r(x_t, \phi_r)}} \]

\[ \text{E step:} \]

\[ \text{M step:} \]

\[ \max_{\phi_j} \sum_t \sum_{j=1}^k p(j | x_t) \ln p(j | x_t) \]

\[ \min_{\theta_j} \sum_t \sum_{j=1}^k p(j | x_t) \left\| y_t - f(x_t, \theta) \right\|^2 \]

**Alternative Mixture-of-Experts (ME)**

\[ p(y | x) = \sum_j p(j | x) G(y | W_j x + c_j, \sigma^2_j) \]

\[ E(y | x) = f(x, \theta) = \sum_j p(j | x) (W_j x + c_j) \]

- Easy to be implemented by the EM algorithm
  (Xu, Jordan & Hinton, IJCNN, 1994; A. NIPS, 1995)
- J(k) for selecting k and automatic selection during learning
  (Xu, Neurocomputing, 98, Intl J Neural Systems, 01, Neural Networks, 02)
The EM algorithm for Alternative ME

E step:

\[ p(j | x_i) = \frac{\alpha_j G(x_i | m_j, \Sigma_j) G(y_i | W_j x_i + c_j, \sigma_j^2)}{\sum_{j=1}^{k} \alpha_j G(x_i | m_j, \Sigma_j) G(y_i | W_j x_i + c_j, \sigma_j^2)} \]

M step:

\[
\begin{align*}
&\max \sum_{t} \sum_{j=1}^{k} p(j | x_i) \ln[\alpha_j G(x_i | m_j, \Sigma_j)] \\
&\min \sum_{t} \sum_{j=1}^{k} p(j | x_i) \| y_i - (W_j x_i + c_j) \|^2
\end{align*}
\]


**RPCL Learning for Alternative ME**

\[ d_j(\theta_j) = -\ln[\alpha_j G(x_t | m_j, \Sigma_j) G(y_i | W_j x_t + c_j, \lambda_j^2 I)] = d_j^x(\theta_j^x) + d_j^y(\theta_j^y) \]

\[ c = \arg\min_j d_j, \quad r = \arg\min_j d_j \]

\[ d_j^x(\theta_j^x) = -\ln[\alpha_j G(x_t | m_j, \Sigma_j)], \quad d_j^y(\theta_j^y) = -\ln G(y_i | W_j x_t + c_j, \lambda_j^2 I) \]

\[ \theta_c^{\text{new}} = \theta_c^{\text{old}} - \eta_c \frac{\partial d_c^x(\theta_c^x)}{\partial \theta_c^x}, \quad \theta_r^{\text{new}} = \theta_r^{\text{old}} + \eta_r \frac{\partial d_r^x(\theta_r^x)}{\partial \theta_r^x} \]

\[ \theta_c^{\text{new}} = \theta_c^{\text{old}} - \eta_c \frac{\partial d_c^y(\theta_c^y)}{\partial \theta_c^y}, \quad \theta_r^{\text{new}} = \theta_r^{\text{old}} + \eta_r \frac{\partial d_r^y(\theta_r^y)}{\partial \theta_r^y} \]

**Versus Gaussian Mixture based RPCL learning**

\[ d_j^x(\theta_j^x) = -\ln[\alpha_j G(x_t | m_j, \Sigma_j)] \]

\[ \theta_c^{\text{new}} = \theta_c^{\text{old}} - \eta_c \frac{\partial d_c^x(\theta_c^x)}{\partial \theta_c^x}, \quad \theta_r^{\text{new}} = \theta_r^{\text{old}} + \eta_r \frac{\partial d_r^x(\theta_r^x)}{\partial \theta_r^x} \]

\[ m_c^{\text{new}} = m_c^{\text{old}} + \eta_c (x_t - m_c^{\text{old}}), \quad c = \arg\min_j d_j \]

\[ m_r^{\text{new}} = m_r^{\text{old}} - \eta_r (x_t - m_r^{\text{old}}), \quad r = \arg\min_j d_j \]

\[ \Sigma_j = S_j^T S_j, \quad \alpha_j = \gamma_j^2 \]
Prediction error based RPCL Learning

\[ d_3 > d_2 > d_1 \]

\[ m_1 \text{ is the winner} \]

\[ m_2 \text{ is the rival} \]

\[ d_j^y(\theta_j^y) = -\ln G(y_j | W_j^x t + c_j, \lambda_j^y I) \]

\[ \theta_{c_j}^{new} = \theta_{c_j}^{old} - \eta_c \frac{\partial d_j^y(\theta_j^y)}{\partial \theta_{c_j}^y}, \]

\[ \theta_{r_j}^{new} = \theta_{r_j}^{old} + \eta_r \frac{\partial d_j^y(\theta_j^y)}{\partial \theta_{r_j}^y}, \]

\[ \epsilon_{t,j} = y_t - W_j^x t - c_j \]

\[ W_{c_j}^{new} = W_{c_j}^{old} + \eta_c \epsilon_{t,c} x_t^T, \]

\[ W_{r_j}^{new} = W_{r_j}^{old} - \eta_r \epsilon_{t,r} x_t^T \]

\[ c_{c_j}^{new} = c_{c_j}^{old} + \eta_c \epsilon_{t,c}, \]

\[ c_{r_j}^{new} = c_{r_j}^{old} - \eta_r \epsilon_{t,r}, \]

\[ \lambda_{c_j}^{new} = (1 - \eta_c) \lambda_{c_j}^{old} + \eta_c \left( \left\| \epsilon_{t,c} \right\|_2^2 \right) / d, \]

\[ \lambda_{r_j}^{new} = \lambda_{r_j}^{old} + \eta_r \lambda_{r_j}^{old} \left( \lambda_c^{old} - \left\| \epsilon_{t,c} \right\|_2^2 \right) / d, \]

---

Table 1
The results of prediction on FOREX rate of USD-DEM-SET Type A (No. of units = 5)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>NRBF two-stage</th>
<th>EM-NRBF</th>
<th>ENRBF two-stage</th>
<th>EM-ENRBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training (NMSE)</td>
<td>0.553</td>
<td>0.394</td>
<td>0.143</td>
<td>0.152</td>
</tr>
<tr>
<td>Testing (NMSE)</td>
<td>2.92</td>
<td>0.774</td>
<td>0.452</td>
<td>0.448</td>
</tr>
</tbody>
</table>

Table 2
The results of prediction on FOREX rate of USD-DEM-SET Type A (by NRBF two-stage only)

<table>
<thead>
<tr>
<th>No. of units</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training (NMSE)</td>
<td>0.553</td>
<td>0.647</td>
<td>0.514</td>
<td>0.396</td>
</tr>
<tr>
<td>Testing (NMSE)</td>
<td>2.92</td>
<td>4.29</td>
<td>3.85</td>
<td>1.70</td>
</tr>
</tbody>
</table>

Table 3
The results of prediction on FOREX rate of USD-DEM-SET Type A (No. of units = 20)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Training flops *</th>
<th>Training (NMSE)</th>
<th>Testing (NMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRBF two-stage</td>
<td>(5.81 \times 10^5)</td>
<td>0.396</td>
<td>1.703</td>
</tr>
<tr>
<td>EM-NRBF (CCL) II</td>
<td>(5.94 \times 10^5)</td>
<td>0.238</td>
<td>0.768</td>
</tr>
<tr>
<td>ENRBF two-stage</td>
<td>(3.91 \times 10^5)</td>
<td>0.173</td>
<td>0.452</td>
</tr>
<tr>
<td>EM-ENRBF (CCL) II</td>
<td>(3.96 \times 10^6)</td>
<td>0.151</td>
<td>0.445</td>
</tr>
</tbody>
</table>

*Here one flop is counted by MATLAB as an addition or multiplication operation.
**Table 4**
The results of trading investment based on the prediction on USD-DEM-SET Type A

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Net profit point</th>
<th>Profit in US$ (in 112 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM-NRBF (CCL)</td>
<td>1425</td>
<td>9262.5</td>
</tr>
<tr>
<td>Adaptive EM-NRBF (CCL)</td>
<td>3966</td>
<td>25779.0</td>
</tr>
<tr>
<td>EM-ENRBF (CCL)</td>
<td>2063</td>
<td>13406.5</td>
</tr>
<tr>
<td>Adaptive EM-ENRBF (CCL)</td>
<td>2916</td>
<td>18954.0</td>
</tr>
</tbody>
</table>

**Table 5**
The results of trading investment by Supervised Decision Network on USD-DEM-SET Type A

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Net profit point</th>
<th>Profit in US$ (in 112 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM-NRBF (CCL)</td>
<td>1605</td>
<td>10432.5</td>
</tr>
<tr>
<td>Adaptive EM-NRBF (CCL)</td>
<td>4237</td>
<td>27540.5</td>
</tr>
<tr>
<td>EM-ENRBF (CCL)</td>
<td>2660</td>
<td>17290.0</td>
</tr>
<tr>
<td>Adaptive EM-ENRBF (CCL)</td>
<td>3207</td>
<td>20845.5</td>
</tr>
</tbody>
</table>

**Fig. 2.** The results of prediction on Forex data of USD-DEM-Set Type A (No. of units — 5), corresponding to Table 1, where the solid line is for data and the dashed line is for prediction result, and this convention is kept the same for all the figures in this paper: (a) by NRBF two-stage, (b) by EM-NRBF, (c) by ENRBF two-stage, (d) by EM-ENRBF.
Fig. 3. The results of prediction on Forex data of USD-DEM Set Type A (No. of units = 20), corresponding to Table 3: (a) by NRBF two-stage, (b) by EM-NRBF, (c) by ENRBF two-stage, (d) by EM-ENRBF Algorithm II.

Fig. 4. The results of prediction on Forex data of USD-DEM-SET Type B (No. of units = 20): (a) by EM-NRBF (CCL), (b) by Adaptive EM-NRBF (the prediction and the real data are almost overlapped).
1. Financial Prediction

- RPCL competitive learning based piecewise linear prediction
- Extended radial basis functions, Mixture of expert model and financial prediction
- Finite mixture of ARCH and GRACH models for prediction
- APT-TFA based prediction

Autoregressive model  \( z_{t+1} = \sum_{\tau=1}^{p} a_{\tau} z_{t+1-\tau} + \varepsilon_{i}, \quad P(\varepsilon_{i}) = G(\varepsilon_{i} \mid 0, \lambda^{2}) \)

- ARCH model :
  \[ \lambda^{2}(t) = \alpha_{0} + \sum_{i=1}^{q} \alpha_{i} \lambda^{2}(t - i) \]
- GARCH model :
  \[ \lambda^{2}(t) = \alpha_{0} + \sum_{i=1}^{q} \alpha_{i} \lambda^{2}(t - i) + \sum_{i=1}^{p} \beta_{i} \varepsilon_{t-i}^{2} \]

a conditional heteroscedastic variance
Mixture of ARCH models:

\[ \lambda_j^2(t) = \alpha_{0,j} + \sum_{i=1}^{q} \alpha_{i,j} \lambda_i^2(t-i) \]

Mixture of GARCH models:

\[ \lambda_j^2(t) = \alpha_{0,j} + \sum_{i=1}^{q} \alpha_{i,j} \lambda_i^2(t-i) + \sum_{i=1}^{p} \beta_{i,j} \varepsilon_i^2(t-i) \]

- a conditional heterodastic variance

outperforms GARCH model considerably

Best parametric model matching

参数模型最佳匹配

optimizing a matching cost

\[ F(p(x|\theta), X), \quad X = \{x_i\}_{i=1}^N \]

\[ p(x|\hat{\theta}(X)) \]

Maximum Likelihood (ML)

最大似然

One piece of evidence, take it by 100% 

Two pieces of evidence take each by 50% subject to the template

More pieces of evidence ...

The large number law

\[ p_n(x|\theta) \to p_n(x|\theta_0) \]

\[ \hat{\theta}(X) \to \theta_0 \quad \text{as} \quad N \to \infty \]
Maximum Likelihood Learning


Maximum Learning on \( P(\varepsilon_t \mid \theta) = \sum_{j=1}^{k} \pi_j G(\varepsilon_t \mid 0, \lambda^2_j(t)), \)

\( \theta = \{ \pi_j, c_j, a_{j,r}, \tau = 1, \ldots, q_j, \alpha_{i,j}, \beta_{i,j} \}_{j=1}^{k} \)

- **BHHH (Berndt, Hall, Hall and Hausman) algorithm:**

\[
\theta^{new} = \theta^{old} + \lambda_j [\hat{I}_{\theta \theta}]^{-1} \frac{1}{T} \sum_{t} \frac{\partial l_t}{\partial \theta} \\
\hat{I}_{\alpha,\alpha} = \sum - \frac{1}{2} a p(j \mid \varepsilon_t) \lambda^{-4}_j(t) \frac{\partial \lambda^3_j(t)}{\partial a_j} \frac{\partial \lambda^2_j(t)}{\partial a_j^T} \lambda^2_j(t)
\]

The EM algorithm

**E step:**

\[
p(j \mid \varepsilon_t) = \frac{\alpha_j G(\varepsilon_t \mid 0, \lambda^2_j(t))}{\sum_{j=1}^{k} \alpha_j G(\varepsilon_t \mid 0, \lambda^2_j(t))}
\]

**M step:**

\[
\max \sum_{t} \sum_{j=1}^{k} p(j \mid \varepsilon_t) \ln[\alpha_j G(\varepsilon_t \mid 0, \lambda^2_j(t))] \\
\alpha_j = \frac{1}{N} \sum_{t} p(j \mid \varepsilon_t) \\
\max_{\alpha_j > 0, a_{i,j}, \beta_{i,j}} \sum_{t} \sum_{j=1}^{k} p(j \mid \varepsilon_t) \ln[\alpha_j G(\varepsilon_t \mid 0, \lambda^2_j(t))] \\
\min_{\mu_j, a_j, m_j} \sum_{t} \sum_{j=1}^{k} p(j \mid \varepsilon_t) \| y_t - \mu_j - a_j^T (x_t - m_j) \|^2
\]
After the completion of parameter training, we have the prediction formula:

\[
\hat{y}_t = \sum_{j=1}^{k} p(j|\epsilon_t) \left[ \sum_{\tau=1}^{q_j} a_{j,\tau} z_{t+1-\tau} + c_j \right]
\]

\[
p(j|\epsilon_t) = \frac{\alpha_j G(\epsilon_t | 0, \lambda_j^2(t))}{\sum_{j=1}^{k} \alpha_j G(\epsilon_t | 0, \lambda_j^2(t))}
\]

- Four experiments conducted with real foreign exchange rate
  - USD vs DEM
  - USD vs GRP
  - USD vs SWF
  - USD vs FRN

W.C. Wong, F. Yip, & L. Xu, Proc. ICONIP98
Root-mean-square Error

<table>
<thead>
<tr>
<th></th>
<th>USD vs. GARCH model</th>
<th>Finite Mixture of GARCH model</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM</td>
<td>0.0114</td>
<td>0.0102</td>
<td>10.5</td>
</tr>
<tr>
<td>FRN</td>
<td>0.0405</td>
<td>0.0364</td>
<td>10.1</td>
</tr>
<tr>
<td>GRP</td>
<td>0.0028</td>
<td>0.0021</td>
<td>25.0</td>
</tr>
<tr>
<td>SWF</td>
<td>0.0107</td>
<td>0.0089</td>
<td>16.8</td>
</tr>
</tbody>
</table>

USD vs DEM(GARCH)
USE vs DEM (Finite Mixture of GARCH)

USD vs FRN (GARCH)
USD vs. FRN (Finite Mixture of GARCH)

USD vs. GRP (GARCH)
USD vs. GRP (Finite Mixture of GARCH)

USD vs. SWF(GARCH)
USD vs. SWF (Finite Mixture of GARCH)

Mixture of ARMA-GARCH models:

\[ z_{t+1} = \sum_{\tau=1}^{q_j} a_{j,\tau} z_{t+1-\tau} + c_j + \varepsilon_j(t) + \sum_{\tau=1}^{p_j} b_{j,\tau} \varepsilon_j(t-\tau), \]

\[ P(\varepsilon_j) = G(\varepsilon_j \mid 0, \lambda_j^2(t)), \quad j = 1, \cdots, k \]

\[ \lambda_j^2(t) = \alpha_{0,j} + \sum_{i=1}^{q} \alpha_{i,j} \varepsilon_j^2(t-i) + \sum_{i=1}^{p} \beta_{i,j} \lambda_{i,j}^2(t-i) \]

**E step:**

\[ p(j \mid \varepsilon_t) = \frac{\alpha_j G(\varepsilon_t \mid 0, \lambda_j^2(t))}{\sum_{j=1}^{k} \alpha_j G(\varepsilon_t \mid 0, \lambda_j^2(t))} \]

**M step:**

\[ \alpha_j = \frac{1}{N} \sum_{t} p(j \mid \varepsilon_t) \]

\[ \theta_{new} = \theta_{old} + \eta \sum_{t} \frac{\partial L_t}{\partial \theta} \]

Table 1: A summary of mean square errors for different approaches.

<table>
<thead>
<tr>
<th>Method</th>
<th>CK HDG</th>
<th>HSBC HDG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv. ARMA-GARCH</td>
<td>2.7550</td>
<td>2.3820</td>
</tr>
<tr>
<td>Mixture AR-GARCH</td>
<td>2.6025</td>
<td>2.2466</td>
</tr>
<tr>
<td>Mixture ARMA-GARCH</td>
<td>2.5524</td>
<td>2.1889</td>
</tr>
</tbody>
</table>

Figure 2: Prediction of CK prices with mixture of AR-GARCH.

Figure 4: Prediction of HSBC prices with conventional ARMA-GARCH.

Figure 5: Prediction of HSBC prices with mixture of AR-GARCH.

Figure 6: Prediction of HSBC prices with mixture of ARMA-GARCH.
Mixture-of-Experts of ARMA-GARCH models

\[ p(z_{t+1} \mid x_t) = \sum_j p(j \mid \Xi_t) G\left(z_{t+1} \mid \sum_{r=1}^{q_j} a_{j,r} z_{t+1-r} + c_j + \sum_{r=1}^{p_j} b_{j,r} \varepsilon_j(t-\tau), \lambda_j^2(t)\right) \]

\[ = \sum_j p(j \mid \Xi_t) G(\varepsilon_j(t) \mid 0, \lambda_j^2(t)) \]

\[ p(j \mid \Xi_t) = \frac{\alpha_j G(\Xi_t \mid m_j, \Sigma_j)}{\sum_{j=1}^k \alpha_j G(\Xi_t \mid m_j, \Sigma_j)} \]

\[ z_{t+1} = \sum_{r=1}^{q_j} a_{j,r} z_{t+1-r} + c_j + \varepsilon_j(t) + \sum_{r=1}^{p_j} b_{j,r} \varepsilon_j(t-\tau), \]

\[ P(\varepsilon_j) = G(\varepsilon_j \mid 0, \lambda_j^2(t)), \quad j = 1, \ldots, k \]

\[ \lambda_j^2(t) = \alpha_{0,j} + \sum_{i=1}^d \alpha_{i,j} \varepsilon_j^2(t-i) + \sum_{i=1}^p \beta_{i,j} \lambda_j^2(t-i) \]

Alternative Mixture-of-Experts

E step:

\[ h_j(t) = \frac{\alpha_j G(\Xi_t \mid m_j, \Sigma_j) G(\varepsilon_t \mid 0, \lambda_j^2(t))}{\sum_{j=1}^k \alpha_j G(\Xi_t \mid m_j, \Sigma_j) G(\varepsilon_t \mid 0, \lambda_j^2(t))} \]

M step:

\[ \max \sum_t \sum_{j=1}^k h_j(t) \ln[\alpha_j G(\Xi_t \mid m_j, \Sigma_j)] \]

\[ \max \sum_t \sum_{j=1}^k h_j(t) \ln G(\varepsilon_t \mid 0, \lambda_j^2(t)) \]

Figure 1: First step prediction of CK prices with conventional ARMA-GARCH.

Figure 2: First step prediction of HSBC prices with conventional ARMA-GARCH.

Table 1: Mean square errors of first step prediction using different ARMA-GARCH models.

<table>
<thead>
<tr>
<th></th>
<th>CK HDG</th>
<th>HSBC HDG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>2.3799</td>
<td>2.1651</td>
</tr>
<tr>
<td>Gaussian Mixture</td>
<td>2.1101</td>
<td>1.9899</td>
</tr>
<tr>
<td>Mixture-of-Expert</td>
<td>2.0030</td>
<td>1.9225</td>
</tr>
</tbody>
</table>

Figure 3: Prediction of CK prices with mixture of ARMA-GARCH.

Figure 4: First step prediction of HSBC price.

Figure 5: First step prediction of CK prices with mixture-of-expert ARMA-GARCH.

Figure 6: First step prediction of HSBC prices with mixture-of-expert ARMA-GARCH.
Figure 7: Second step prediction of CK prices with conventional ARMA-GARCH.

Figure 9: Second step prediction of CK prices with Gaussian mixture ARMA-GARCH.

Figure 11: Second step prediction of CK prices with mixture-of-expert ARMA-GARCH.

Figure 8: Second step prediction of HSBC prices with conventional ARMA-GARCH.

Figure 10: Second step prediction of HSBC prices with Gaussian mixture ARMA-GARCH.

Figure 12: Second step prediction of HSBC prices with mixture-of-expert ARMA-GARCH.

<table>
<thead>
<tr>
<th></th>
<th>CK HDG</th>
<th>HSBC HDG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>4.9627</td>
<td>4.2001</td>
</tr>
<tr>
<td>Gaussian Mixture</td>
<td>4.8216</td>
<td>3.9935</td>
</tr>
<tr>
<td>Mixture-of-Expert</td>
<td>4.4147</td>
<td>3.7120</td>
</tr>
</tbody>
</table>

Table 2: Mean square errors of second step prediction using different ARMA-GARCH models.
2. Portfolio Management

- Portfolio Management by Learned Decisions
- Markowitz Portfolio, Sharpe's ratio and Downside risk
- Improved Portfolio Sharpe Ratio Maximization with Diversification
- Adaptive Portfolio Management based on Extended RBF nets and analyses of market factors
Portfolio Management by learning decision signals


\[ I_t^i = I_t^a \cdot I_t^p \]
\[ I_t^a = 1 \]

the \( i \)th foreign currency of interest.

\[ \sum_{i=1}^{m} I_t^a = 1 \]

That is, we are limited to investing in, at most, only one currency each day. The second component is the position signal, \( I_t^p \), with

\[
I_t^p = \begin{cases} 
1, & \text{which means to take long position} \\
0, & \text{which means to take neutral position} \\
-1, & \text{which means to take short position} 
\end{cases}
\]

At the current day, \( t \), we can calculate yesterday’s return by

\[ r_t^i = -I_{t-1}^i \left( z_t^i - z_{t-1}^i \right) - |I_{t-1}^i - I_{t-2}^i| \gamma \]

with \( 1 \leq i \leq m \), where \( \gamma \) is a transaction cost rate. We know that the best investment decision for yesterday

\[ I_{t-1} = \left\{ I_{t-1}^i \right\}_{i=1}^{m} \]

should be the one that optimizes total returns

\[ \sum_{i=1}^{m} r_t^i \]

which results in

\[ I_{t-1} = [I_{t-1}^a, I_{t-1}^p] \]
\[ I_{t-1}^i = [I_{t-1}^i, I_{t-1}^p] \]

with

\[ I_{t-1}^i = \begin{cases} 
1, & \text{if } i = j \text{ with } j = \arg \max_{i \leq m} (T_i^t) \\
0, & \text{otherwise}
\end{cases} \]

and

\[ I_{t-1}^p = \begin{cases} 
1, & \text{if } z_i^t - z_{i+1}^t > 0 \\
0, & \text{if } z_i^t - z_{i+1}^t = 0 \\
-1, & \text{if } z_i^t - z_{i+1}^t < 0 
\end{cases} \quad (5) \]

We assume that there is some relationship as follows

\[ [I_i, I_p] = f(I_{t-1}, I_{t-1}^p, z_i^t, z_{i-1}^t, \ldots, z_{i-d+1}^t, i \leq m] \quad (6) \]

Use Extended RBF nets to learn it and then make a decision

\[ I_i^t = \begin{cases} 
\hat{I}_i^s \cdot \hat{I}_i^p, & \text{if } i = j \text{ with } j = \arg \max_{i \leq m} (\hat{T}_i^t) \\
0, & \text{otherwise}
\end{cases} \]

**Figure 1.** The USD-DEM rate series of 1096 data points. Each horizontal bar represents 10 data points.
Figure 2. The results by Existing ENRBF ASLD. Upper graph: the trading signal on the test data [-1,1]. Lower graph: the profit gained (%).

Figure 3. The results by Adaptive CCL-ENRBF ASLD. Upper graph: the trading signal on the test data [-1,1]. Lower graph: the profit gained (%).
Figure 4. The results by Existing ENRBF SASLD. Upper graph: the trading signal on the test data [-1,1]. Lower graph: the profit gained (%).

Figure 5. The results by Adaptive CCL-ENRBF SASLD. Upper graph: the trading signal on the test data [-1,1]. Lower graph: the profit gained (%).
2. Portfolio Management

- Portfolio Management by Learned Decisions
- Markowitz Portfolio, Sharpe's ratio and Downside risk
- Improved Portfolio Sharpe Ratio Maximization with Diversification
- Adaptive Portfolio Management based on Extended RBF nets and analyses of market factors
- TFA based Adaptive Portfolio Management

Figure 6. A plot of all results, along with results of an old prediction trading system by Cheung et al. [1996].
Existing Portfolio Selection Methods

- **Standard Portfolio Optimization (SMPO)**
  Markowitz’s 1952 landmark paper “Portfolio Selection”.
  - Choose portfolio weights that
  
  $$\max u w^T \bar{r} - w^T \Sigma w$$

  Acceptable level of risk set by user
  Expected portfolio returns
  Portfolio risk

  Weights $w_1, w_2, \ldots, w_i, \ldots, w_N$
  Asset $1, 2, \ldots, i, \ldots, N$
  Expected returns $\bar{r} = [\bar{r}_1, \ldots, \bar{r}_N]$

- **Sharpe’s Method**
  Sharpe’s ratio of asset $i$

  $$S_i = \frac{\bar{r}_i}{\sqrt{\text{var}(r_i)}}$$

  Expected return
  Risk (standard deviation)

  Appropriateness of investing in asset $i$

- **Single asset investment:**
  - Choose the asset with highest $S_i$ to invest in

- **Portfolio investment:**
  - Choey, Kang, Weigend (1997)
  - Moody & Wu (1997)
Downside risk

- **Traditional risk** (Markowitz 1952):
  - \[ \text{Upside fluctuation} \]
  - \[ \text{Downside fluctuation} \]

- **Downside risk** (Markowitz 1959, Fishburn 1977):
  - Only fluctuation below target counted as risk

\[
downV_\alpha(G) = \int_{-\infty}^{G} (G - r)^\alpha dF(r)
\]

Target returns

2. Portfolio Management

- Portfolio Management by Learned Decisions
- Markowian Portfolio, Sharpe's ratio and Downside risk
- Improved Portfolio Sharpe Ratio Maximization with Diversification
- Adaptive Portfolio Management based on Extended RBF nets and analyses of market factors
- TFA based Adaptive Portfolio Management

**Improved Portfolio Sharpe Ratio Maximization with Diversification** (IPSRM-D)

- Select portfolio weights according to:

\[
\max_w \frac{w^T \bar{r} + Hw^T Uw}{w^T Dw} + B w^T ([I] - w)
\]

\[
s.t. \quad \sum_{i=1}^{N} w_i = 1, \quad w_i \geq 0
\]
- **Portfolio Downside Risk**  \( w^T D w \)
  - measure fluctuation below target return \( G \)
  \[
  D = [d_{i,j}]
  \]
  \[
  d_{i,j} = \int_{-\infty}^{G} \int_{-\infty}^{G} \left( G - r_i \right)^\alpha \left( G - r_j \right)^\alpha p(r_i, r_j) \, d r_i \, d r_j
  \]

- **Portfolio Upside Volatility**  \( w^T U w \)
  - measure fluctuation above target return \( G \)
  \[
  U = [u_{i,j}]
  \]
  \[
  u_{i,j} = \int_{-\infty}^{G} \int_{-\infty}^{G} \left( r_i - G \right)^\alpha \left( r_j - G \right)^\alpha p(r_i, r_j) \, d r_i \, d r_j
  \]

  May be desired by active investor:
  
  check performance frequently, sell assets at high point

- **Diversification term**  \( w^T ([I] - w) \)
  - min. when one of \( \{w_i\} \) is 1 and others are 0
  - max. when all \( \{w_i\} \) are equal

  make the portfolio distribute to more assets
Experimental Demonstration (1)

- Six stocks:
  - S&P 500 Composite - Price Index (USA)
  - Hang Seng Index (Hong Kong)
  - NIKKEI 255 Stock Average (Japan)
  - Shanghai SE Composite - Price Index (China)
  - CAC 40 - Price Index (France)
  - Australia SE All Ordinary - Price Index (Australia)
- Transaction Cost 3%
- 1365 data points (1992 - 1997)
- In this experiment, \( a = 2, G = 0, H = B = 1 \)

Improved Portfolio Sharpe Ratio Maximization (IPSRM)
**IPSRM w/wo diversification**

<table>
<thead>
<tr>
<th></th>
<th>IPSRM</th>
<th>IPSRM-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of indexed involved</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Degree of diversification</td>
<td>0.000</td>
<td>0.559</td>
</tr>
<tr>
<td>Mean return</td>
<td>0.239</td>
<td>0.140</td>
</tr>
<tr>
<td>Variance of return</td>
<td>6.271</td>
<td>1.937</td>
</tr>
<tr>
<td>IPSR</td>
<td>1.692</td>
<td>1.676</td>
</tr>
<tr>
<td>Upside volatility</td>
<td>0.980</td>
<td>0.556</td>
</tr>
<tr>
<td>Downside risk</td>
<td>0.721</td>
<td>0.416</td>
</tr>
</tbody>
</table>

**Setting of $H$ and $B$**

Mean Return vs. $\{H,B\}$  
Downside Risk vs. $\{H,B\}$
Use of the parameters $H$, $B$

- Large $H$: more risky assets
- Small $H$: less risky assets

- Large $B$: distribute to more assets
- Small $B$: concentrate on few assets

High risk: high return
Low risk: low return

More convenient methods?

- I expect 10% return. Find $w$ with min. dn. risk.
- I can bare only $(8\%)^2$ dn. risk. Find $w$ with max. expected return.

In IPSRM-D
- How to set parameters $H$ & $B$ to meet specific expected return or risk?
- Difficult since the relationship is non-linear
The method with Control of Expected Portfolio Return

\[
\max_w \frac{r_{\text{spec}} + Hw^T Uw}{w^T Dw} + Bw^T (\mathbb{I} - w)
\]

subject to

\[
\begin{align*}
\sum_{i=1}^{N} w_i &= I, \quad w_i \geq 0 \\
\mathbf{w}^T \mathbf{r} &= r_{\text{spec}} \quad \text{fixed expected return}
\end{align*}
\]

Constrained Optimization:
by the Augmented Lagrange method

Experimental Demonstration (2)
Experimental demonstration (2)  
Control of Expected Portfolio Return

<table>
<thead>
<tr>
<th></th>
<th>Expected return fixed at</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.062</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>Mean return</td>
<td>0.117</td>
<td>0.137</td>
<td></td>
</tr>
<tr>
<td>Variance of return</td>
<td>0.384</td>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>Improved Portfolio Sharpe Ratio</td>
<td>2.228</td>
<td>2.175</td>
<td></td>
</tr>
<tr>
<td>Upside Volatility</td>
<td>0.307</td>
<td>0.371</td>
<td></td>
</tr>
<tr>
<td>Downside Risk</td>
<td>0.190</td>
<td>0.234</td>
<td></td>
</tr>
</tbody>
</table>

The method with  
Control of Portfolio Downside Risk

\[
\max_w w^T \bar{r} + Hw^T Uw \quad \left( \frac{v_{spec}}{\nu_{spec}} \right) + Bw^T ([I] - w)
\]

subject to

\[
\begin{cases}
\sum_{i=1}^{N} w_i = 1, \quad w_i \geq 0 \\
 w^T D w = v_{spec} \quad \text{fixed downside risk}
\end{cases}
\]

Constrained Optimization:  
by the Augmented Lagrange method
### Control of downside portfolio risk

<table>
<thead>
<tr>
<th></th>
<th>Expected downside risk fixed at</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.359</td>
</tr>
<tr>
<td></td>
<td>0.831</td>
</tr>
<tr>
<td>Mean return</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>0.132</td>
</tr>
<tr>
<td>Variance of return</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>0.741</td>
</tr>
<tr>
<td>Improved Portfolio Sharpe Ratio</td>
<td>2.193</td>
</tr>
<tr>
<td>Upside Volatility</td>
<td>0.282</td>
</tr>
<tr>
<td></td>
<td>0.388</td>
</tr>
<tr>
<td>Downside Risk</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>0.257</td>
</tr>
</tbody>
</table>
Summary

- Controlled Expected Return and Downside Risk
- Select $w$ in accordance to investor's preference
- Extension of the Sharpe Ratio to portfolio case
  New terms: Upside volatility, diversification term

2. Portfolio Management

- Portfolio Management by Learned Decisions

- Markowitz Portfolio, Sharpe's ratio and Downside risk

- Improved Portfolio Sharpe Ratio Maximization with Diversification

- Adaptive Portfolio Management based on Extended RBF nets and analyses of market factors
Adaptive Portfolio Management

Xu, L, "BYY harmony learning, independent state space and generalized APT financial analyses", IEEE Tr. on Neural Networks, 12 (4), 2001, 822-849.

The return of the portfolio on the $t^{th}$ day is defined by:

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

$\alpha_t$ : proportion of money spent on securities  
- borrowing from a risk-free bond is allowed  $\alpha_t > 0$

$r^f$ : return from the risk-free bond

$\beta_t^j$ : proportion of $\alpha_t$ spent on the $j^{th}$ security  
- short sale is not permitted  $1 \geq \beta_t^j \geq 0$

$x_t^{(j)}$ and $p_t^{(j)}$ are the return and closing price for the $j^{th}$ security on the $t^{th}$ day respectively.

$$S_p = \frac{M(R_T)}{\sqrt{V(R_T)}}, \quad : \text{the return obtained per unit of risk}$$

Observation based:

$$\alpha_t = e^{-\delta_t}, \quad \beta_t^j = \frac{e^{\xi_t^j}}{\sum_{r=1}^{m} e^{\xi_t^r}}$$

where $\delta_t$ and $\xi_t^j$ are controlled by observations

Hidden factors based:

$$\alpha_t = e^{-\delta_t}, \quad \beta_t^j = \frac{e^{\xi_t^j}}{\sum_{r=1}^{m} e^{\xi_t^r}}$$

where $\delta_t$ and $\xi_t^j$ are controlled by hidden market factors $y_t$

$$\delta_t = g(y_t, \psi), \quad \xi_t^j = f(y_t, \phi)$$
The exact functional form of both
\( g(y_t, \psi) \) and \( f(y_t, \phi) \) are unknown, it can be approximated by the adaptive Extended Normalized Radial Basis Function (ENRBF) algorithm [Xu, 1998]

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

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

where \( \varphi(\mu, \Sigma, k) = \frac{e^{-0.5(y_T-u_p)^T \Sigma_p^{-1} (y_T-u_p)}}{\sum_{p=1}^{k} e^{-0.5(y_T-u_p)^T \Sigma_p^{-1} (y_T-u_p)}} \)

---

An Adaptive Algorithm

- Use the gradient ascent approach

\[
\theta^{\text{new}} = \theta^{\text{old}} + \eta \nabla_{\theta} S_p
\]

\[ \theta = \psi \cup \phi \]

\[ \psi = \{ u_p, \Sigma_p, W_p, c_p \}_{p=1}^{k}, \quad \psi = \{ \hat{u}_p, \hat{\Sigma}_p, \hat{W}_p, \hat{c}_p \}_{p=1}^{\hat{k}} \]

### Detailed Updating Rules

#### Updating the parameter set $\psi$

\[
\begin{align*}
    u_p^{\text{new}} &= u_p^{\text{old}} + \eta \left( \nabla_{\xi_T} S_p \right) \varphi(\mu, \Sigma, k) \tau(\mu, \Sigma, W_p, c, k)(y_T - \mu_p) \\
    \Sigma_p^{\text{new}} &= \Sigma_p^{\text{old}} + \eta \left( \nabla_{\xi_T} S_p \right) \varphi(\mu, \Sigma, k) \tau(\mu, \Sigma, W_p, c, k) \kappa(\mu, \Sigma) \\
    W_p^{\text{new}} &= W_p^{\text{old}} + \eta \left( \nabla_{\xi_T} S_p \right) \varphi(\mu, \Sigma, k) y_T \\
    c_p^{\text{new}} &= c_p^{\text{old}} + \eta \left( \nabla_{\xi_T} S_p \right) \varphi(\mu, \Sigma, k)
\end{align*}
\]

#### Updating the parameter set $\phi$

\[
\begin{align*}
    \hat{u}_p^{\text{new}} &= \hat{u}_p^{\text{old}} + \hat{\eta} \left( \nabla_{\xi_T}^{(j)} S_p \right) \varphi(\hat{\mu}, \hat{\Sigma}, \hat{k}) \tau(\hat{\mu}, \hat{\Sigma}, \hat{W}_{p,r}, \hat{c}, \hat{k})(y_T - \hat{\mu}_p) \\
    \hat{\Sigma}_p^{\text{new}} &= \hat{\Sigma}_p^{\text{old}} + \eta \left( \nabla_{\xi_T}^{(j)} S_p \right) \varphi(\hat{\mu}, \hat{\Sigma}, \hat{k}) \tau(\hat{\mu}, \hat{\Sigma}, \hat{W}_{p,r}, \hat{c}, \hat{k}) \kappa(\hat{\mu}, \hat{\Sigma}) \\
    W_{p,r}^{\text{new}} &= W_{p,r}^{\text{old}} + \eta \left( \nabla_{\xi_T}^{(j)} S_p \right) \varphi(\hat{\mu}, \hat{\Sigma}, \hat{k}) y_T \\
    c_{p,r}^{\text{new}} &= c_{p,r}^{\text{old}} + \eta \left( \nabla_{\xi_T}^{(j)} S_p \right) \varphi(\hat{\mu}, \hat{\Sigma}, \hat{k})
\end{align*}
\]
where $\eta$ and $\hat{\eta}$ are learning rates,

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

$$\nabla_{\varepsilon_{ij}} S_p = \frac{\left\{ V(R_T) - M(R_T)[R_T - M(R_T)] \left[ 1 - \frac{1}{T} \right] \right\} \sum_{j=1}^{m} e^{\varepsilon_{ij}} x_T^{(j)} \left( \sum_{j=1}^{m} e^{\varepsilon_{ij}} - e^{\varepsilon_{ij}} \right)}{T \sqrt{[V(R_T)]^3} \sum_{\varepsilon_{ij}} (x_T^{(j)})^2} e^{\varepsilon_T},$$

$$\nabla_{\varepsilon kj} S_p = \frac{\left\{ V(R_T) - M(R_T)[R_T - M(R_T)] \left[ 1 - \frac{1}{T} \right] \right\} e^{\varepsilon_T} x_T^{(j)} \left( \sum_{j=1}^{m} e^{\varepsilon(r)} - e^{\varepsilon_{ij}} \right)}{T \sqrt{[V(R_T)]^3} \sum_{\varepsilon_{ij}} (x_T^{(j)})^2} e^{\varepsilon_{ij}} ,$$

$$\varphi(\mu, \Sigma, k) = \sum_{p=1}^{k} e^{-0.5(y_T - u_p)^T \Sigma^{-1}_p (y_T - u_p)} ,$$

$$\kappa(\mu, \Sigma) = \Sigma^{-1}_p (y_T - u_p)(y_T - u_p)^T \Sigma^{-1}_p - 0.5 \text{diag} \left( \Sigma^{-1}_p (y_T - u_p)(y_T - u_p)^T \Sigma^{-1}_p \right) ,$$

$$\tau(\mu, \Sigma, W_p, c, k) = \frac{\left( W_p^T y_T + c_p \right) - \sum_{p=1}^{k} \left( W_p^T y_T + c_p \right) \varphi(\mu, \Sigma, k)}{\sum_{p=1}^{k} e^{-0.5(y_T - u_p)^T \Sigma^{-1}_p (y_T - u_p)}} .$$

**Simulation Result and Performance Evaluation**

**Data Considerations**

- Experiments based on the interest rate and stock prices of Hong Kong stock market
  - Data obtained from 1 Nov 2001 to 11/11/2002
  - The performance was studied for 180 trading days.
  - Eight stocks were selected to form the portfolios:

<table>
<thead>
<tr>
<th>Code</th>
<th>Stock Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>CHECUNG KONG</td>
</tr>
<tr>
<td>0002</td>
<td>CLP HOLDINGS</td>
</tr>
<tr>
<td>0003</td>
<td>HK &amp; CHINA GAS</td>
</tr>
<tr>
<td>0004</td>
<td>WHARF</td>
</tr>
<tr>
<td>0005</td>
<td>HSBC HOLDINGS</td>
</tr>
<tr>
<td>0008</td>
<td>PCCW</td>
</tr>
<tr>
<td>0992</td>
<td>LEGEND GROUP</td>
</tr>
<tr>
<td>1038</td>
<td>CKI HOLDINGS</td>
</tr>
</tbody>
</table>

Evaluate the performance of the portfolio management system in 4 different scenarios, where transactions were not in lots

The four scenarios are:

- **Experiment 1**: No Transaction Cost and Short Sale Not Permitted
- **Experiment 2**: Has Transaction Cost, Short Sale Not Permitted
- **Experiment 3**: No Transaction Cost, Short Sale Permitted
- **Experiment 4**: Has Transaction Cost, Short Sale Permitted

No Transaction Cost, Short Sale Not Permitted

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Portfolio</td>
<td>1.0541</td>
<td>0.0235</td>
<td>1.1020</td>
<td>0.9910</td>
<td>44.8553</td>
</tr>
<tr>
<td>Hang Seng Index</td>
<td>0.9771</td>
<td>0.0727</td>
<td>1.1099</td>
<td>0.8211</td>
<td>13.4402</td>
</tr>
</tbody>
</table>
No Transaction Cost, Short Sale Not Permitted

Compare Portfolio Return with its Constituents

<table>
<thead>
<tr>
<th>0001 CHECUNG KONG</th>
<th>0002 CLP HOLDINGS</th>
<th>0003 HK &amp; CHINA GAS</th>
<th>0004 WHARF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0005 HSBC HOLDINGS</td>
<td>0008 PCCW</td>
<td>0992 LEGEND GROUP</td>
<td>1038 CKI HOLDINGS</td>
</tr>
</tbody>
</table>

No Transaction Cost, Short Sale Not Permitted

Compare Portfolio Return with its Constituents

<table>
<thead>
<tr>
<th>The Portfolio</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Portfolio</strong></td>
<td>1.0541</td>
<td>0.0235</td>
<td>1.1020</td>
<td>0.9910</td>
<td>44.8553</td>
</tr>
<tr>
<td>0001</td>
<td>0.8712</td>
<td>0.1213</td>
<td>1.0621</td>
<td>0.6262</td>
<td>7.1822</td>
</tr>
<tr>
<td>0002</td>
<td>1.0188</td>
<td>0.0222</td>
<td>1.0680</td>
<td>0.9660</td>
<td>45.8919</td>
</tr>
<tr>
<td>0003</td>
<td>1.0572</td>
<td>0.0361</td>
<td>1.1413</td>
<td>0.9859</td>
<td>29.2853</td>
</tr>
<tr>
<td>0004</td>
<td>0.9938</td>
<td>0.1109</td>
<td>1.2764</td>
<td>0.8205</td>
<td>8.9612</td>
</tr>
<tr>
<td>0005</td>
<td>1.0257</td>
<td>0.0464</td>
<td>1.1239</td>
<td>0.9078</td>
<td>22.1056</td>
</tr>
<tr>
<td>0008</td>
<td>0.8678</td>
<td>0.1855</td>
<td>1.0864</td>
<td>0.4764</td>
<td>4.6782</td>
</tr>
<tr>
<td>0992</td>
<td>0.8985</td>
<td>0.0854</td>
<td>1.0815</td>
<td>0.6593</td>
<td>10.5211</td>
</tr>
<tr>
<td>1038</td>
<td>1.0591</td>
<td>0.0450</td>
<td>1.1545</td>
<td>0.9959</td>
<td>23.5356</td>
</tr>
</tbody>
</table>
With Transaction Cost, Short Sale Not Permitted

every change in $\beta^j$ involves a transaction of which a transaction cost

$$c_t = -\alpha_t \sum_{j=1}^{m} r_c |\beta^j_t - \beta^j_{t-1}|(1 + x^j_t)$$

where $r_c$ is the rate of transaction cost

$$R_t = (1 - \alpha_t) r^f + \alpha_t \sum_{j=1}^{m} \left( \beta^j_t x^j_t - r_c \sum_{j=1}^{m} r_c |\beta^j_t - \beta^j_{t-1}|(1 + x^j_t) \right)$$

$$\frac{dSp}{d\xi^j_t} = \left( V(R_T) - M(R_T) \right) e^{\xi^j_t} \left[ x^j_t - r_c \text{sign}(e^{\xi^j_t} - e^{\xi^j_{t-1}}) \right]$$

$$\cdot \left( \sum_{r=1}^{m} e^{\xi^j_t} - e^{\xi^j_{t-1}} \right) \frac{e^{\xi^j_t}}{T \sqrt{V(R_T) \left( \sum_{r=1}^{m} e^{\xi^j_r} \right)^2}}$$

Compare Portfolio Return with Hang Seng Index

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Portfolio</td>
<td>1.0452</td>
<td>0.0220</td>
<td>1.0911</td>
<td>0.9886</td>
<td>47.5091</td>
</tr>
<tr>
<td>Hang Seng Index</td>
<td>0.9771</td>
<td>0.0727</td>
<td>1.1099</td>
<td>0.8211</td>
<td>13.4402</td>
</tr>
</tbody>
</table>
Compare Portfolio Return with its Constituents

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Portfolio</td>
<td>1.0452</td>
<td>0.0220</td>
<td>1.0911</td>
<td>0.9886</td>
<td>47.5091</td>
</tr>
<tr>
<td>0001</td>
<td>0.8712</td>
<td>0.1213</td>
<td>1.0621</td>
<td>0.6262</td>
<td>7.1822</td>
</tr>
<tr>
<td>0002</td>
<td>1.0188</td>
<td>0.0222</td>
<td>1.0680</td>
<td>0.9660</td>
<td>45.8919</td>
</tr>
<tr>
<td>0003</td>
<td>1.0572</td>
<td>0.0361</td>
<td>1.1413</td>
<td>0.9859</td>
<td>29.2853</td>
</tr>
<tr>
<td>0004</td>
<td>0.9938</td>
<td>0.1109</td>
<td>1.2764</td>
<td>0.8205</td>
<td>8.9612</td>
</tr>
<tr>
<td>0005</td>
<td>1.0257</td>
<td>0.0464</td>
<td>1.1239</td>
<td>0.9078</td>
<td>22.1056</td>
</tr>
<tr>
<td>0008</td>
<td>0.8678</td>
<td>0.1855</td>
<td>1.0864</td>
<td>0.4764</td>
<td>4.6782</td>
</tr>
<tr>
<td>0992</td>
<td>0.8985</td>
<td>0.0854</td>
<td>1.0815</td>
<td>0.6593</td>
<td>10.5211</td>
</tr>
<tr>
<td>1038</td>
<td>1.0591</td>
<td>0.0450</td>
<td>1.1545</td>
<td>0.9959</td>
<td>23.5356</td>
</tr>
</tbody>
</table>
With Transaction Cost, Short Sale Not Permitted

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>1.0541</td>
<td>0.0235</td>
<td>1.1020</td>
<td>0.9910</td>
<td>44.8553</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>1.0452</td>
<td>0.0220</td>
<td>1.0911</td>
<td>0.9886</td>
<td>47.5091</td>
</tr>
</tbody>
</table>

- all the return related attributes decreased
- transaction cost was charged for every transaction
- The standard deviation and Sharpe Ratio better
- Still better than the performance of Hang Seng Index

No Transaction Cost, Short Sale Permitted

\[ \alpha_i = \delta_i \]

\[ \beta_i^j = \frac{\xi_i^j}{\sum_{r=1}^{m} \xi_i^r} \]

\[
\frac{\partial S_p}{\partial \delta_T} = \frac{[V(R_T) - M(R_T)(R_T - M(R_T))]}{T \sqrt{V(R_T)^3}} \left( \frac{\sum_{r=1}^{m} e_i^{r} x_r}{\sum_{r=1}^{m} e_i^{r}} - r^f \right)
\]

\[
\frac{\partial S_p}{\partial \xi_T^i} = \frac{[V(R_T) - M(R_T)(R_T - M(R_T))]}{T \sqrt{V(R_T)^3}} \left( \frac{\sum_{r=1}^{m} e_i^{r}}{\sum_{r=1}^{m} e_i^{r}} \right)
\]
No Transaction Cost, Short Sale Permitted

Compare Portfolio Return with Hang Seng Index

<table>
<thead>
<tr>
<th>The Portfolio</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hang Seng Index</td>
<td>0.9771</td>
<td>0.0727</td>
<td>1.1099</td>
<td>0.8211</td>
<td>13.4402</td>
</tr>
<tr>
<td>The Portfolio</td>
<td>1.0700</td>
<td>0.0336</td>
<td>1.1376</td>
<td>0.9953</td>
<td>31.8452</td>
</tr>
</tbody>
</table>

No Transaction Cost, Short Sale Permitted

Compare Portfolio Return with its Constituents

<table>
<thead>
<tr>
<th>0001 CHECUNG KONG</th>
<th>0002 CLP HOLDINGS</th>
<th>0003 HK &amp; CHINA GAS</th>
<th>0004 WHARF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0005 HSBC HOLDINGS</td>
<td>0008 PCCW</td>
<td>0992 LEGEND GROUP</td>
<td>1038 CKI HOLDINGS</td>
</tr>
</tbody>
</table>
No Transaction Cost, Short Sale Permitted

Compare Portfolio Return with its Constituents

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Portfolio</td>
<td>1.0700</td>
<td>0.0336</td>
<td>1.1376</td>
<td>0.9953</td>
<td>31.8452</td>
</tr>
<tr>
<td>0001</td>
<td>0.8712</td>
<td>0.1213</td>
<td>1.0621</td>
<td>0.6262</td>
<td>7.1822</td>
</tr>
<tr>
<td>0002</td>
<td>1.0188</td>
<td>0.0222</td>
<td>1.0680</td>
<td>0.9660</td>
<td>45.8919</td>
</tr>
<tr>
<td>0003</td>
<td>1.0572</td>
<td>0.0361</td>
<td>1.1413</td>
<td>0.9859</td>
<td>29.2853</td>
</tr>
<tr>
<td>0004</td>
<td>0.9938</td>
<td>0.1109</td>
<td>1.2764</td>
<td>0.8205</td>
<td>8.9612</td>
</tr>
<tr>
<td>0005</td>
<td>1.0257</td>
<td>0.0464</td>
<td>1.1239</td>
<td>0.9078</td>
<td>22.1056</td>
</tr>
<tr>
<td>0008</td>
<td>0.8678</td>
<td>0.1855</td>
<td>1.0864</td>
<td>0.4764</td>
<td>4.6782</td>
</tr>
<tr>
<td>0992</td>
<td>0.8985</td>
<td>0.0854</td>
<td>1.0815</td>
<td>0.6593</td>
<td>10.5211</td>
</tr>
<tr>
<td>1038</td>
<td>1.0591</td>
<td>0.0450</td>
<td>1.1545</td>
<td>0.9959</td>
<td>23.5356</td>
</tr>
</tbody>
</table>

No Transaction Cost, Short Sale Permitted

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2.1</td>
<td>1.0541</td>
<td>0.0235</td>
<td>1.1020</td>
<td>0.9910</td>
<td>44.8553</td>
</tr>
<tr>
<td>Experiment 2.2</td>
<td>1.0452</td>
<td>0.0220</td>
<td>1.0911</td>
<td>0.9886</td>
<td>47.5091</td>
</tr>
<tr>
<td>Experiment 2.3</td>
<td>1.0700</td>
<td>0.0336</td>
<td>1.1376</td>
<td>0.9953</td>
<td>31.8452</td>
</tr>
</tbody>
</table>

• Short sales were permitted → the portfolio was still able to generate money from stock market even though the stock prices declined
  The portfolio was doing very well

• Hang Seng Index decreased for more than 10%
  However, the portfolio increased for about 14%

• Short sales brought return to the portfolio, but also brought risk
  The risk was still low
With Transaction Cost, Short Sale Permitted

\[ R_t = (1 - \alpha_t)r^f + \alpha_t \sum_{j=1}^{m} \left( \beta_t^j x_t^j - r_c \sum_{c=1}^{r_c} \beta_t^j - \beta_t^{j-1} \right) \left( 1 + x_t^j \right) \]

\[ \frac{\partial Sp}{\partial \delta_T} = \frac{[V(R_T) - M(R_T)(R_T - M(R_T))]}{T \sqrt{V(R_T)^3}} \left( \frac{\sum_{r=1}^{m} e^{\xi_T^r} x_T^r}{\sum_{r=1}^{m} e^{\xi_T^r}} - r^f \right) \]

\[ \frac{dSp}{d\xi_{jT}} = [V(R_T) - M(R_T)(R_T - M(R_T))] \xi_{jT} \left[ x_{jT} - r_c \text{sign} \left( \xi_{jT}^l - \xi_{jT}^{l-1} \right) \right] \frac{\sum_{r=1}^{m} \xi_T^r - \xi_{jT}^l}{T \sqrt{V(R_T)^3} \left( \sum_{r=1}^{m} \xi_T^r \right)^2} \]

With Transaction Cost, Short Sale Permitted

Compare Portfolio Return with Hang Seng Index

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Portfolio</td>
<td>1.0629</td>
<td>0.0296</td>
<td>1.1231</td>
<td>0.9952</td>
<td>35.9088</td>
</tr>
<tr>
<td>Hang Seng Index</td>
<td>0.9771</td>
<td>0.0727</td>
<td>1.1099</td>
<td>0.8211</td>
<td>13.4402</td>
</tr>
</tbody>
</table>
With Transaction Cost, Short Sale Permitted

Compare Portfolio Return with its Constituents

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Portfolio</td>
<td>1.0629</td>
<td>0.0296</td>
<td>1.1231</td>
<td>0.9952</td>
<td>35.9088</td>
</tr>
<tr>
<td>0001</td>
<td>0.8712</td>
<td>0.1213</td>
<td>1.0621</td>
<td>0.6262</td>
<td>7.1822</td>
</tr>
<tr>
<td>0002</td>
<td>1.0188</td>
<td>0.0222</td>
<td>1.0680</td>
<td>0.9660</td>
<td>45.8919</td>
</tr>
<tr>
<td>0003</td>
<td>1.0572</td>
<td>0.0361</td>
<td>1.1413</td>
<td>0.9859</td>
<td>29.2853</td>
</tr>
<tr>
<td>0004</td>
<td>0.9938</td>
<td>0.1109</td>
<td>1.2764</td>
<td>0.8205</td>
<td>8.9612</td>
</tr>
<tr>
<td>0005</td>
<td>1.0257</td>
<td>0.0464</td>
<td>1.1239</td>
<td>0.9078</td>
<td>22.1056</td>
</tr>
<tr>
<td>0008</td>
<td>0.8678</td>
<td>0.1855</td>
<td>1.0864</td>
<td>0.4764</td>
<td>4.6782</td>
</tr>
<tr>
<td>0992</td>
<td>0.8985</td>
<td>0.0854</td>
<td>1.0815</td>
<td>0.6593</td>
<td>10.5211</td>
</tr>
<tr>
<td>1038</td>
<td>1.0591</td>
<td>0.0450</td>
<td>1.1545</td>
<td>0.9959</td>
<td>23.5356</td>
</tr>
</tbody>
</table>
Evaluate the performance of the portfolio management system in 4 different scenarios, where transactions were in lots.

<table>
<thead>
<tr>
<th>Stock Code</th>
<th>Company Name</th>
<th>No. of Shares per Board Lot</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>CHEUNG KONG</td>
<td>1000</td>
</tr>
<tr>
<td>0002</td>
<td>CLP HOLDINGS</td>
<td>500</td>
</tr>
<tr>
<td>0003</td>
<td>HK &amp; CHINA GAS</td>
<td>1000</td>
</tr>
<tr>
<td>0004</td>
<td>WHARF HOLDINGS</td>
<td>1000</td>
</tr>
<tr>
<td>0005</td>
<td>HSBC HOLDINGS</td>
<td>400</td>
</tr>
<tr>
<td>0008</td>
<td>PCCW</td>
<td>1000</td>
</tr>
<tr>
<td>0992</td>
<td>LEGEND GROUP</td>
<td>2000</td>
</tr>
<tr>
<td>1038</td>
<td>CKI HOLDINGS</td>
<td>1000</td>
</tr>
</tbody>
</table>

The four scenarios are:

- **Experiment 1**: No Transaction Cost and Short Sale Not Permitted
- **Experiment 2**: Has Transaction Cost, Short Sale Not Permitted
- **Experiment 3**: No Transaction Cost, Short Sale Permitted
- **Experiment 4**: Has Transaction Cost, Short Sale Permitted
No Transaction Cost and Short Sale Not Permitted

No Transaction Cost, Short Sale Permitted

With Transaction Cost, Short Sale Permitted

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>1.0541</td>
<td>0.0235</td>
<td>1.102</td>
<td>0.9910</td>
<td>44.8553</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>1.0493</td>
<td>0.0214</td>
<td>1.0936</td>
<td>0.9928</td>
<td>49.0327</td>
</tr>
</tbody>
</table>

• The following values decreased:
  - Mean
  - Standard deviation
  - Maximum return

• Because the transactions were now in lots
  → Fewer stocks could be purchased
  → More money was placed in risk-free bond

• Return from stock market is usually greater than the return from risk-free bond
  → Lowered the return of the portfolio