**Stock portfolio performance Data Set**

|  |  |
| --- | --- |
| **Abstract**: The data set of **performances** of **weighted scoring** **stock** **portfolios** are obtained with mixture design from the **US stock market historical database**. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 315 | **Area:** | Business |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 12 | **Date Donated** | 2016-04-22 |
| **Associated Tasks:** | Regression | **Missing Values?** | N/A | **Number of Web Hits:** | 11127 |

**Source:**

Name: I-Cheng Yeh   
email addresses: (1) 140910 **'@'** mail.tku.edu.tw (2) icyeh **'@'** chu.edu.tw   
institutions: (1) Department of Information Management, Chung Hua University, Taiwan. (2) Department of Civil Engineering, Tamkang University, Taiwan.   
other contact information: 886-2-26215656 ext. 3181

**Data Set Information:**

There are three disadvantages of **weighted scoring stock selection models**:

* First, they **cannot identify** the **relations** between **weights of stock-picking concepts** and **performances of portfolios**.
* Second, they **cannot** systematically **discover** the **optimal** **combination** for **weights of concepts** to **optimize** the **performances**.
* Third, they are **unable** to **meet** various **investorsâ**€™ **preferences**.

This study **aims** to **more efficiently** **construct** **weighted scoring** **stock selection models** to overcome these disadvantages. Since the **weights of stock-picking concepts** in a weighted scoring stock selection model can be regarded **as** **components in a mixture**, we used **the simplex centroid mixture design** to **obtain** the **experimental sets of weights**.

These sets of **weights** are **simulated** with **US** **stock** **market** **historical** **data** to obtain their performances**. Performance prediction models** were built with **the simulated performance data set** and **artificial neural networks**. Furthermore, the **optimization models** to **reflect investorsâ€™** **preferences** were **built** up, and the **performance prediction models** were **employed** as the **kernel** **of the optimization models** so that the optimal solutions can now be solved with **optimization** **techniques**.

The empirical values of the performances of the optimal weighting combinations generated by the optimization models showed that they can meet various investorsâ€™ preferences and outperform those of S&Pâ€™s 500 not only during the training period but also during the testing period. [result]

**Attribute Information:**

The **inputs** are the **weights** **of the stock-picking concepts** as follows   
X1=the weight of the **Large B/P** concept   
X2=the weight of the **Large** **ROE** concept   
X3=the weight of the **Large S/P** concept   
X4=the weight of the **Large Return Rate in the last quarter** concept   
X5=the weight of the **Large Market Value** concept   
X6=the weight of the **Small systematic Risk** concept   
  
The **outputs** are the **investment performance** **indicators** (normalized) as follows   
Y1=**Annual** **Return**   
Y2=**Excess** **Return**   
Y3=**Systematic** **Risk**   
Y4=**Total** **Risk**   
Y5=**Abs**. **Win** **Rate**   
Y6=**Rel**. **Win** **Rate**

**Relevant Papers:**

[1] Liu, Y. C., & Yeh, I. C. Using mixture design and neural networks to build stock selection decision support systems. Neural Computing and Applications, 1-15. (Print ISSN 0941-0643, Online ISSN 1433-3058, First online: 16 November 2015, DOI 10.1007/s00521-015-2090-x)   
[2] Yeh, I. C., & Cheng, W. L. (2010). â€œFirst and second order sensitivity analysis of MLP,â€ Neurocomputing, Vol. 73, No. 10, pp. 2225-2233.   
[3] Yeh, I. C. and Hsu, T. K. (2011). â€œGrowth Value Two-Factor Model,â€ Journal of Asset Management, Vol. 11, No. 6, pp. 435-451.

problem statement,

description of dataset,

description of the statistical method(s) used,

presentation of the results.