International Review of Economics and Finance 31 (2014) 46–58

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Evaluating multi-criteria ratings of financial investment options

Andrew N.K. Chen [a], Shin-Yun Wang [b,*], Po-Lung Yu [c]

[a] Accounting and Information Systems, University of Kansas, 1300 Sunnyside Ave, Lawrence, KS 66045-7585, United States
[b] Department of Finance, National Dong Hwa University, No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan
[c] Institute of Information Management, National Chiao Tung University, Hsin Chu, Taiwan

ARTICLE INFO

Article history:
Received 14 May 2013
Received in revised form 12 December 2013
Accepted 13 December 2013
Available online 31 December 2013

Keywords:
Multiple criterion ratings
Information quality
Implied ordered classes

ABSTRACT

In the financial market, most available rating information is based on multi-criteria and published by myriad agents or companies. Given a multi-criteria rating report on a finite number of assets (e.g., stocks, bonds, mutual funds), we can construct sets of ordered classes. If ratings from a published report have useful and valid information value as claimed, the average performances of assets within classes are expected to show some monotonic property. A set of hypotheses and empirical tests based on Value Line Mutual Fund Survey are provided to illustrate our proposed method. Implications and future research opportunities are also discussed.

1. Introduction

Investors (e.g., individuals, portfolio and fund managers, investment bank managers) who make investment decisions often encounter numerous ratings and reports of investment options including stocks, bonds, mutual funds, and derivatives. These investors often naively trust and use whatever available or familiar information to support their decision making without systematically judging the quality of information and the validity of intended interpretation of information. They simply assume that these ratings are of good quality and treat them as de facto legitimate. Paradoxically, these ratings often are estimated with different approaches but rarely proven or validated objectively in a systematic way.

After the 2008 financial crisis, the value and validity of published ratings (e.g., credit, bond) have been widely criticized and questioned. A growing voice of the need for more scrutiny of these ratings has gained stronger attention. Some contend that credit rating agencies have made errors of judgment in rating debt (e.g., particularly in assigning AAA ratings to structured debt where many cases have subsequently been downgraded or defaulted) (Carlo & Michelle, 2008). For example, credit rating agencies such as Moody's, Standard & Poor's, and Fitch Ratings still maintained at least “A” ratings on AIG and Lehman Brothers right before its bankruptcy and acceptance of government's bailouts respectively. Other people argue that credit rating agencies are plagued by conflicts of interest that might inhibit them from providing accurate and honest ratings. For example, many banks in the U.S. that were closed by the FDIC in 2010 actually achieved passing grades or even an A+ grade from the Better Business Bureau (BBB) beforehand. In the wake of recent financial crisis, which tarnishes the credibility of rating agencies, it is more important than ever to verify whether these ratings are valid and trustworthy and can provide quality information to their users. Under the notion that good decisions are based on good information, we conjecture that good investment decisions need to be supported by quality rating information.

If the use of these published ratings is intended for helping investors to make good investment decisions, objective evaluations and evidences should be put forward to justify that these ratings do provide meaningful and useful information (i.e., information quality) and are adequate and appropriate for investors' intended inferences and actions (i.e., validity). We conjecture that a

* Corresponding author. Tel.: +886 3 8633136; fax: +886 3 8633130.
E-mail address: gracew@mail.ndhu.edu.tw (S.-Y. Wang).

1059-0560/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.iref.2013.12.003

rating can be better perceived as useful and valid when there is an objective measurement shows what is rated by the rating actually has consistently good performance over a period of time. Therefore, in this study we propose a method to identify information quality and verify validity of these ratings with a sound mathematical foundation. We set to present our method as an effective tool for assisting investors to make more informed and better investments. A set of hypotheses and empirical tests based on Value Line Mutual Fund Survey are provided to illustrate our proposed method. Through the illustration, a sequential statistical series are also constructed to measure the extent of information quality provided by the Value Line Rating.

This study intends to address the quality of the information provided by the multi-criteria ratings by proposing a method to identify information quality and verify validity of these ratings with a sound mathematical foundation. With the use of our proposed method we can objectively evaluate these ratings and seek to answer the following questions:

> *Do these ratings provide useful information for investors to make investment decisions*?*Can we trust these ratings and their intended interpretation and use for investors in making investment decisions*?

We discuss the research background for this study in the following section. In Section 3, we demonstrate the rationale and methods to systematically construct the ordered classes of ratings. We then develop our hypotheses under suitable assumptions in Section 4. In Section 5, an empirical study based on Value Line Mutual Fund Survey was conducted to illustrate how to construct the ordered classes and was used to test our hypotheses. The information quality, depending on dominance relations and time, is also assessed. We conclude this study and offer remarks in Section 6.

2. Research background

2.1. Financial rating

The literature on the ratings of financial performance is abundant. Some scholars created Capital Asset Pricing Model (CAPM) and related models to rate the performance of a given set of financial assets/portfolios (Chiang, Kozhevnikov, & Wisen, 2005; Fama & French, 1993, 1996; Jensen, 1968, 1969; Lee, Wu, & Wei, 1990; Lehmann & Modest, 1987). Others studied the persistence of performance rating (Blake & Morey, 2000; Brown & Goetzmann, 1995; Goetzmann & Ibbotson, 1994; Grinblatt & Titman, 1992). Recently, stochastic dominance and DEA (Data Envelopment Analysis) have also been employed to study the rating, benchmarking and efficiency of the performance (Kopa & Post, 2009; Kuosmanen, 2004; Levy, 1998; Post, 2003). These studies tend to focus on how to rate the financial assets or portfolios over a set of known performance criteria. The derived results have been used to construct rating systems or methods, and to reward portfolio managers. Nonetheless, these studies rarely put effort on verifying the implied usefulness and validity of those ratings. Therefore, an objective measurement method should be developed to evaluate whether what is rated by these ratings actually has consistently good performance over a period of time. This method can be applied to almost all ratings supplied by rating industry and scholars in financial applications as well as other areas.

2.2. Information quality theory

Good information is needed for good decisions. One of the important functions of decision support is to gather information and uncover the quality ones for deriving good decisions. For example, leisure and business travelers who search ideal hotels often depend on certain rankings or "stars" of different aspects of hotels. Similarly, professional sports teams who try to identify prospective players count on rankings of a range of athletes' abilities from various scouting reports. Students choose and apply to colleges usually rely on some publications of rankings on various criteria of these colleges. However, a common problem for these scenarios is that there is no objective means or clear guidance on judging which rankings actually provide useful information and are trustworthy for decision making.

As O'Reilly (1982) stated that there is a direct relationship between the quality of information used by a decision maker and decision making performance. In general sense, there is support for the notion that good information leads to good decision making (e.g., Ge, 2009). In finance literature, only a handful of recent studies explicitly consider the impact of information quality (e.g., Ai, 2010; Brevik & d'Addona, 2010; Veronesi, 2000). We can consider information quality as the degree of usefulness or the "fitness of use" (Juran, 1992) for information users. Redman (2001) suggested that information is of high quality if it is fit for its intended uses in operations, decision-making, and planning. Wang (1998) also defined information quality as the characteristic of information to be of high value to its users. Eppler (2003) adopts both definitions of quality – meeting the customer expectations and meeting the activity requirements – acknowledging the important duality of quality: subjective (meeting the expectations) vs. objective (meeting the requirements). Therefore, for assisting investors to make good investment decisions, published ratings need to provide quality information. In the context of evaluating ratings, it is necessary to develop a mechanism which can assess and verify information quality of these ratings.

2.3. Theory of validity

According to Merriam-Webster, the term "validity" refers to the degree to which the conclusions (interpretations) derived from the results of any assessment are "well-grounded or justifiable; being at once relevant and meaningful." In the well-established field of educational measurement, validity is in general defined as an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment (Messick, 1989, p. 13). Cook and Beckman (2006) pointed out that validity is not a property of the instrument, but of the instrument's scores and their intended interpretations. What needs to be valid is the meaning or interpretation of the score; as well as any implications for action that this meaning entails (Cronbach, 1971). The term score is used generically in its broadest sense to mean any coding or summarization of observed consistencies or performance regularities of any assessment device including tests and portfolios (Messick, 1995). In this study we do not question how publishers derive their rating scores but focus on investigating the meaningfulness and usefulness of these ratings.

Our goal of this study is not to test theory of information quality or theory of validity but to provide a mechanism which can practically and objectively verify the usefulness of information and validity of intended inference and use of these ratings. By doing so, our method can support investors alike to identify good ratings from the available ones and ultimately make better investment decisions.

3. Implied ordered classes by multi-criteria ratings

Given a value function, we can construct ordered isovalue curves. In the same vein, given a rating based on multi-criteria (e.g., expected return, risk, growth persistence) over a set of assets (e.g., mutual funds, stocks, bonds), we can construct a number of ordered classes over the set of assets. If the rating does provide quality information, the average performance of the ordered classes over a period of time should show certain monotonic property (increasing or decreasing) which is similar to that of the value for the isovalue curves for a value function. Otherwise, the rating can be assessed as lack of validity of intended interpretation and usage.

To facilitate our presentation, we use Value Line Mutual Fund ratings for illustration. Value Line offers rating on more than two thousand mutual funds according to three criteria: overall ranking, risk, and 5-year growth persistence. A rating ranges from 1 to 5 for each criterion, and a smaller number represents a better rating. Thus, a mutual fund with rating (2, 2, 3) is better than one with rating (3, 4, 5). We summarize our notations in Appendix A and present our method as follows.

We have n assets (stocks, bonds or funds) in a space $X = \{1, \ldots, n\}$. According to the criteria $C = \{1, \ldots, l\}$, we have rating information $y = (r_1, \ldots r_l)$ for an asset where $y_i = r_i$ is the rating of the ith criterion. Let $f(x)$, $x \in X$ be the rating vector for asset x, with $f_i(x)$, an integer, being the rating of ith criterion for asset x. Assume that for all i, $1 \le f_i(x) \le M$, and that a smaller number represents a better rating. In the case of Value Line's rating, $M = 5$. Many other ratings use number of stars or alphabetic ordering (e.g., AA, ABB) which can be easily transformed into numerical integer ordering. We have the *rating space* defined as $Y = \{f(x) \mid x \in X\}$. Note that Y is the collection of all possible rating combinations. There may be several assets which are associated with the same y vector and $y^x = f(x) \in Y$. Let us first define some dominance relations over Y.

3.1. Dominance ordered classes

We first introduce definitions to allow us to construct ordered classes containing rating vector(s) for a giving rating space as follows. Concerning these classes, we'd conjecture and later empirically test that certain amount of performance homogeneity over time of a class of assets can show some monotonic difference when compared with other classes.

Definition 1. Given y^{x_1}, y^{x_2} of Y, we say that y^{x_1} dominates y^{x_2}, denoted by $y^{x_1} \le y^{x_2}$, if $y^{x_1} \ne y^{x_2}$ and for each $i = 1, \ldots, l$, $y_i^{x_1} \le y_i^{x_2}$.

We denote $y_i^x = f_i(x)$ which is the rating of ith criterion for asset x. Intuitively, the definition y^{x_1} dominates y^{x_2} ($y^{x_1} \le y^{x_2}$), means that there is at least one rating criterion i of y^{x_1} is strictly better (smaller) than that of y^{x_2} while y^{x_1} is better than or equal to y^{x_2} for other rating criteria.

Clearly, if $y^{x_1} \le y^{x_2}$ and $y^{x_2} \le y^{x_3}$ then $y^{x_1} \le y^{x_3}$. Thus, the dominance relation "≤"over Y defined in Definition 1 is *transitive*. If rating information in vector y is valid and in good quality, and $y^{x_1} \le y^{x_2}$, we can expect that the performance of assets with rating y^{x_1} to be better than that of those with rating y^{x_2}.

Definition 2. Given $y^{x_1}, y^{x_2} \in Y$, we say that y^{x_1} strictly dominates y^{x_2}, denoted by $y^{x_1} < y^{x_2}$, if $y^{x_1} \ne y^{x_2}$ and $y_i^{x_1} < y_i^{x_2}$ for all $i = 1, \ldots, l$.

The strict dominance relation over Y in Definition 2, "<", is also transitive. Again, if rating information in vector y is valid and in good quality, and $y^{x_1} < y^{x_2}$, we can expect that the performance of assets with rating y^{x_1} to be better than that of those with rating y^{x_2}.

Example 1. Let a rating vector $y = (r_1, r_2)$ with r_1 for the expected return rating and r_2 for risk rating. There exists a rating space Y associated with 12 assets and $Y = \{(1,2),(1,3),(1,5),(2,5),(3,1),(3,4),(3,5),(4,1),(4,2),(5,2),(5,3),(5,5)\}$ (Fig. 1). We notice that $(1,2) \le (1,3) \le (3,4) \le (3,5)$, and $(1,2) < (3,4)$, $(1,3) < (3,5)$. It is not true that $(1,2) < (1,3)$ or $(3,4) < (3,5)$.

Definition 3. Given Y, a "non-dominated class" of rating vector(s) is defined as

$N(\le) = \{y^x \in Y \mid no\ y^* \in Y \text{ with } y^* \le y^x\}$; and a "dominated class" of rating vector(s) can be defined as $D(\le) = Y \backslash N(\le)$. Let $N_1(\le) = N(\le)$, $D_1(\le) = D(\le)$. For $k = 2, 3, \ldots$ where k denotes the kth "class", we define $N_k(\le) = \{y^x \in D_{k-1}(\le) \mid no\ y^* \in D_{k-1}(\le) \text{ with } y^* \le y^x\}$; $D_k(\le) = D_{k-1}(\le) \backslash N_k(\le)$.

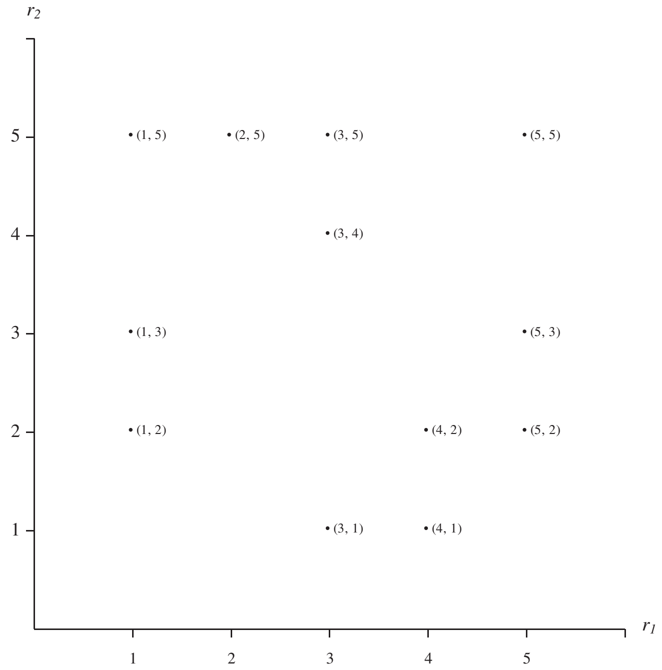

Fig. 1. The rating space of Example 1.

With respect to, $D_{k-1}(\leq)$, $N_k(\leq)$ is the set of all non-dominated rating vectors (i.e., none of $D_{k-1}(\leq)$ can dominate any point of $N_k(\leq)$), and $D_k(\leq)$ is the set of all dominated points (i.e., any rating vector of $D_k(\leq)$ is dominated by at least one rating vector of $N_k(\leq)$). Thus $N_k(\leq)$ and $D_k(\leq)$ form a partition of $D_{k-1}(\leq)$. That is, $D_{k-1}(\leq) = N_k(\leq) \cup D_k(\leq)$ and $N_k(\leq) \cap D_k(\leq) = \phi$. For this reason we call $N_k(\leq)$ the kth class of non-dominated set and $D_k(\leq)$ the kth class of dominated set.

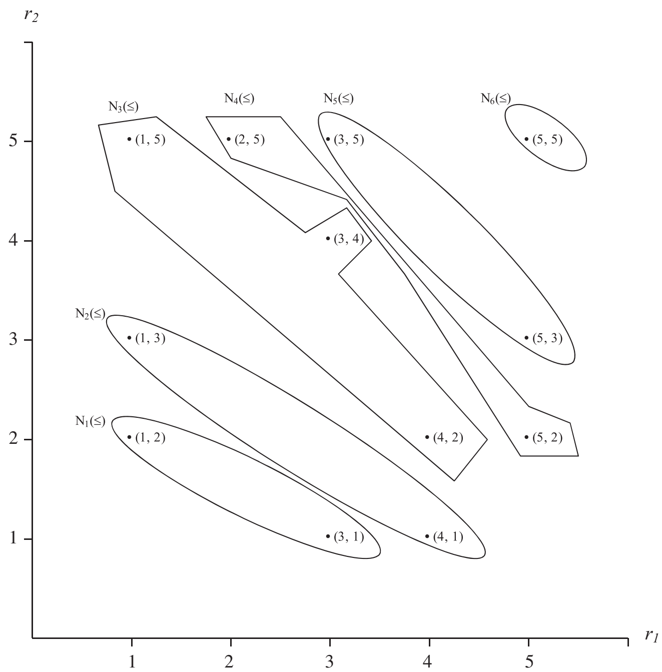

Fig. 2. The non-dominated classes (in Illustration 1) derived from Example 1.

Illustration 1. From Example 1 and Definition 3, we derive the following classes (see Fig. 2):

$$N_1(\le) = \{(3,1),(1,2)\}, N_2(\le) = \{(4,1),(1,3)\}, N_3(\le) = \{(4,2),(3,4),(1,5)\},$$

$$N_4(\le) = \{(5,2),(2,5)\}, N_5(\le) = \{(5,3),(3,5)\}, N_6(\le) = \{(5,5)\}$$

$$D_1(\le) = \{(1,3),(1,5),(2,5),(3,4),(3,5),(4,1),(4,2),(5,2),(5,3),(5,5)\},$$

$$D_2(\le) = \{(1,5),(2,5),(3,4),(3,5),(4,2),(5,2),(5,3),(5,5)\},$$

$$D_3(\le) = \{(2,5),(3,5),(5,2),(5,3),(5,5)\}, D_4(\le) = \{(3,5),(5,3),(5,5)\},$$

$$D_5\ (\le) = \{(5,5)\}, D_6(\le) = \phi.$$

Proposition 1.

- **(i)** *If $y^{x_2} \in N_{k'}(\le)$ and $1 \le k < k'$, then there exists $y^{x_1} \in N_k(\le)$ such that y^{x_1} dominates y^{x_2} (i.e., $y^{x_1} \le y^{x_2}$).*
- **(ii)** *If $k \ne k'$, then $N_{k'}(\le) \cap N_k(\le) =$.*
- **(iii)** *There exists $m \ge 1$, such that $Y = \bigcup_{k=1}^{m} N_k(\le)$.*
- **(iv)** *(iv) $D_k(\le) \supset D_{k+1}(\le)$ and $D_k(\le) \supset D_{k'}(\le)$, if $k < k'$ with $k, k' \in \{1,\ldots,m\}$.*

Proof.

- **(i)** By definition, as $k < k'$, $y^{x_2} \in D_k(\le)$. Thus y^{x_2} is dominated by some y^{x_1} in $N_k(\le)$.
- **(ii)** Suppose that $k < k'$, then $N_{k'}(\le) \subset D_k(\le)$. As $D_k(\le) \cap N_k(\le) =$,
- **(iii)** As Y consists of finite number of points, all points of Y will be contained by some $N_k(\le)$, $k = 1,\ldots,m$, and m is finite.
- **(iv)** By definition, for $k = 1, \ldots, m$, $D_k(\le) = N_{k+1}(\le) \cup D_{k+1}(\le)$. Thus $D_{k+1}(\le) \subset D_k(\le)$.

Theorem 1. *$\{N_1(\le), N_2(\le),\ldots,N_m(\le)\}$ forms a decomposition of Y and all non-dominated classes are mutually exclusive.*

Proof. From Proposition 1-(ii) and (iii).

Proposition 1-(i) essentially says that if $k < k'$, then $N_k(\le)$ collectively dominates $N_{k'}(\le)$ in the sense that for each $y^x \in N_{k'}(\le)$ there is $y^{x^*} \in N_k(\le)$ such that $y^{x^*} \le y^x$. Thus we could reasonably expect that the average performance of $N_k(\le)$ to be better than that of $N_{k'}(\le)$ if $k < k'$. We explore this further in the next section as testing hypothesis.

Similar to the dominance $(\le)$ and Definition 3, we define the case of "strict dominance $(<)$" in Appendix B with Proposition B and Theorem B which are comparable to Proposition 1 and Theorem 1 respectively.

Comparing Illustration 1 to Illustration B in Appendix B, we notice that the numbers of ordered classes in $\{N_k(\le)\}$ and $\{D_k(\le)\}$ are larger than those in $\{N_k(<)\}$ and $\{D_k(<)\}$ respectively. Thus it implies that the dominance relation of "$\le$" has a larger discrimination power than the strict dominance relation of "$<$".

4. Models and hypotheses

With the concepts and definitions introduced in Section 3 and Appendix B, we first build basic models and show derived properties of these models, and in turn introduce a sequence of hypotheses which will be tested with empirical data.

Recall that $y = (r_1,\ldots r_l)$ and $y_i^x = f_i(x)$, which is the rating of ith criterion for asset x. We define a value function which reflects the impact of rating y_i at time t as follow:

$$V(y_i, t) = \alpha_t + \sum_{i=1}^{l} \beta_{it} y_i + \varepsilon_t \tag{1}$$

α_t and ε_t are independent of y_i and reflect the general fluctuation (expected and unpredictable ones respectively) of the market. $\sum_{i=1}^{l} \beta_{it} y_i$ reflects the collective impact of given ratings.

If $\beta_{it} = 0$, the rating y_i has no impact. When rating y_i is "the smaller the better", we expect that $V(y^{x_1}, t) > V(y^{x_2}, t)$ t, if $y^{x_1} < y^{x_2}$. This property will hold if each $\beta_{it} < 0$, $i = 1,\cdots,l$. In addition, Eq. (1) can be used as a linear regression equation to statistically estimate α_t and β_{it} with suitable assumptions.

Since $V(y_i,t)$ can be nonlinear in y and t and we only know that a smaller y_i is better, we need to rely on the following underlying concepts for our use of the value function.

Definition 5. $V(y,t)$ is increasing in the "dominance (≤)" case; that is, $V(y^{x_1},t)>V(y^{x_2},t)$ wherever $y^{x_1}\le y^{x_2}$ and $y^{x_1}\ne y^{x_2}$ (i.e., y^{x_1} dominates y^{x_2} (Definition 1)).

$V(y,t)$ is increasing in "strictly dominance (<)" case; that is, $V(y^{x_1},t)>V(y^{x_2},t)$ wherever $y^{x_1}<y^{x_2}$ and $y^{x_1}\ne y^{x_2}$ (i.e., y^{x_1} strictly dominates y^{x_2} (Definition 2)).

We note that if $V(y,t)$ is increasing in "≤" then it is also increasing in "<"; however, the reverse may not hold true. For instance, let $y^{x_1}=(2,1)$, $y^{x_2}=(2,3)$, and $y^{x_3}=(3,3)$. If $V(y,t)$ is increasing in "≤", then $V(y^{x_1},t)>V(y^{x_2},t)>V(y^{x_3},t)$ is true. However, if $V(y,t)$ is increasing only in "<", $V(y^{x_1},t)>V(y^{x_3},t)$ is true but and it may not be true for $V(y^{x_1},t)>V(y^{x_2},t)$ or $V(y^{x_2},t)>V(y^{x_3},t)$. This inference echoes our earlier comment on that the strict dominance relation of "<" has a smaller discrimination power than the dominance relation of "≤".

Condition 1. $V(y,t)$ is increasing in "dominance (≤)".

Condition 2. $V(y,t)$ is increasing in "strictly dominance (<)".

In the following we show that, under an additive assumption (Condition 3), if Condition 1 (Condition 2 respectively) is met, then a good investment portfolio should not contain any asset from $D(\le)$ ($D(<)$ respectively). That is, a good investment portfolio must be consisted of assets from $N(\le)$ ($N(<)$ respectively). Recall that we have n assets, $X=\{1, \ldots, n\}$ under consideration. A portfolio is a vector of n assets $P=(x_1, \ldots, x_n)$. x_p is the proportion of the investment put in the asset p; $p=1, \ldots, n$ assets. Note that $x_p\ge 0$ and $\sum_{p=1}^{n} x_p=1$.

Condition 3. *Collective rating* of a criterion i of all asset in portfolio P is calculated as $P_i=\sum_{p=1}^{n} x_p f_i(p)$

That is, the collective rating of a criterion i of all asset in portfolio P is the additive sum of the all rating $f_i(p)$ of its components (assets) $\{p|p=1, \ldots, n\}$ with weight x_p.

Theorem 3. *If Conditions 1 and 3 are met, then any portfolio $P=(x_1, \ldots, x_n)$ with $x_k>0$ and $k\in D(\le)$ cannot be the best portfolio. That is, we can find a portfolio $P'=(x'_1,\ldots,x'_n)$ where $P\ne P'$ and $V(f(P'),t)>V(f(P),t)$.*

Proof. As $k\in D(\le)$, we can find an asset r such that $f(r)=y^r\le y^k=f(k)$. Let P^0 be a portfolio such that $x_k^0=0$ (i.e., proportion of asset k is zero); $x_r^0=x_r+x_k$ (i.e., proportion of asset r is the sum of the proportion of asset r and proportion of asset k in P); and $x_i^0=x_i$, if $i\ne k, r$ (i.e., all other proportions of assets equal to those in P).

By additive assumption of Condition 3, we have $P_i^0=\sum_{p=1}^{n} x_p^0 f_i(p)=\sum_{p=1}^{n} x_p f_i(p)-x_k f_i(k)+x_k f_i(r)=P_i+x_k(f_i(r)-f_i(k))$.

As $x_k>0$ and $f(r)\le f(k)$, we have $y^0=f(x^0)\le f(x^P)=y^P$.

By Condition 1, we have $V(f(P^0),t)>V(f(P),t)$. Thus P cannot be the best portfolio.

Corollary 1. *If Conditions 1 and 3 are met, then a necessary condition for a portfolio $P=(x_1, \ldots, x_n)$ to be optimal is that whenever $x_k>0$ and $k\in N(\le)$.*

Proof. It is directly from $N(\le)=Y\backslash D(\le)$ in Definition 3.

Theorem 3 and Corollary 1 justify a reasonable process of good investment; that is, we can identify "not-dominated class" $N(\le)$ first and then form the investment portfolio of assets from $N(\le)$. On the other hand, the "dominated class" $D(\le)$ can be ignored if Conditions 1 and 3 are met. Similar to Theorem 3 and Corollary 1, we derive the followings.

Theorem 4. *If Conditions 2 and 3 are met, then any portfolio $P=(x_1, \ldots, x_n)$ with $x_k>0$ and $k\in D(<)$ cannot be the best portfolio. That is, we can find a portfolio $P'=(x'_1,\ldots,x'_n)$ where $P\ne P'$ and $V(f(P'),t)>V(f(P),t)$.*

Proof. Similar to that for Theorem 3.

Corollary 2. *If Conditions 2 and 3 are met, then a necessary condition for a portfolio $P=(x_1, \ldots, x_n)$ to be optimal is that whenever $x_k>0$ and $k\in N(<)$.*

Based on Theorems 3 and 4 and Proposition 1, we further intend to examine the usefulness of information and validity of intended inference and use of available ratings. That is, we'd like to verify whether a "rating" ($f(x)$, $x\in X$) is reliable and credible, or has any value for investment decision making. For instance, the Value Line Mutual Fund Survey ratings postulate that the *smaller a rating is, the better for each valuation criteria, which implies that these ratings satisfy Condition 1, or at least Condition 2. Hence, with the use of our techniques to generate ordered classes from these ratings, we would expect that the average performance of $N_k(\leq)$ to be better than that of $N_{k+1}(\leq)$, and that of $N_k(<)$ to be much better than that of $N_{k+1}(<)$, whenever these corresponding sets $\{N_k(\leq), N_{k+1}(\leq), N_k(<)$ and $N_{k+1}(<)\}$ are nonempty. Similarly, we would expect that the average performance of $D_k(\leq)$ to be better than that of $D_{k+1}(\leq)$, and that of $D_k(<)$ to be better than that of $D_{k+1}(<)$, whenever the corresponding sets are nonempty. If these expectations are proven wrong, we should be able to conclude that these ratings are not truly reliable or credible, and the rating information tend to have no value for investment decision making. Therefore, when Condition 1 is met, we hypothesize that:*

H1a: The average performance of those assets in $N_k(\leq)$ is better than that of those in $D_k(\leq)$, $k = 1,2,\ldots$, whenever the corresponding sets are not empty.

H1b: For $k = 1,2,\ldots$, the average performance of those assets in $N_k(\leq)$ is better than that of those in $N_{k+1}(\leq)$, if $N_{k+1}(\leq)$ is not empty.

H1c: For $k = 1,2,\ldots$, the average performance of those assets in $D_k(\leq)$ is better than that of those in $D_{k+1}(\leq)$, if $D_{k+1}(\leq)$ is not empty.

When Condition 2 is met, we hypothesize that:

H2a: The average performance of those assets in $N_k(<)$ is better than that of those in $D_k(<)$, $k = 1,2,\ldots$, whenever the corresponding sets are not empty.

H2b: For $k = 1,2,\ldots$, the average performance of those assets in $N_k(<)$ is better than that of those in $N_{k+1}(<)$, if $N_k(<)$ and $N_{k+1}(<)$ are not empty.

H2c: For $k = 1,2,\ldots$, the average performance of those assets in $D_k(<)$ is better than that of those in $D_{k+1}(<)$, if $D_{k+1}(<)$ is not empty.

5. Empirical study

5.1. Data

In order to illustrate the concepts introduced in the previous two sections and test our hypotheses, we used published data from the well known Value Line Mutual Fund Survey (henceforth called the "Survey"). For each mutual fund in the Survey, there was an integer rating ranging from 1 to 5 (i.e., $M = 5$) for each of the following three criteria (i.e., $l = 3$): (i) overall ranks, (ii) risk ranks, and (iii) 5-year growth persistence. For each criterion, the smaller the rating is, the better it is.

With the concept introduced in Section 3, we used rating data of the Survey published on August 22, 2006 to classify mutual funds into corresponding ordered classes: (i) $N(\leq)$ and $D(\leq)$ for testing Hypothesis 1, and (ii) $N(<)$ and $D(<)$ for testing Hypothesis 2. We also performed the same operations and analyses on Survey data published on two other dates (April 25, 2006 and December 26, 2006). In order to save space, we did not report these results which were similar to our major findings.

The year-to-date (YTD) rates of return provided by the Survey were used to measure the performance (corresponding to $V(y,t)$ of Section 4). For consistency of comparison, any mutual funds which were not consistently shown in the Survey from April 25, 2006 to April 24, 2007 (which is the time horizon of our study) were removed. It resulted in 2027 mutual funds for our study. The Survey published ratings once in every three weeks. We noted that sometimes two consecutive rating and rate of return could be identical (e.g., December 26, 2006 and January 9, 2007).

We are interested in the *posterior* (after August 22, 2006) *average rate of return* of each class $\{N_k(\leq)\}$, $\{D_k(\leq)\}$, $\{N_k(<)\}$ and $\{D_k(<)\}$ and derive them to test our Hypotheses. We report the corresponding results and discuss our findings as follows.

5.2. Dominance ordered classes and Hypothesis 1

With Definition 3, we classified the 2027 mutual funds into 24 classes: $\{(N_k(\leq), D_k(\leq))|k = 1, 2, \ldots, 12\}$. The distribution of the number of assets in each class is presented below.

k	1	2	3	4	5	6	7	8	9	10	11	12
$N_k(\leq)$	5	96	203	330	222	194	342	203	127	116	102	84
$D_k(\leq)$	2022	1926	1723	1393	1171	977	635	432	305	189	87	3

The corresponding average YTD rates of return for each class from August 22, 2006 to April 24, 2007 (denoted by R(8/22/06), …, R(04/24/07) with 13 data points) are listed in Table 1. Note that the columns of R(8/22/06) and R(9/12/06) are identical because the Survey gave identical YTD rate of return. It is same for columns of R(12/26/06) and R(1/9/07). The indices of the last three rows show the frequency of falsification of H1a, H1b, and H1c respectively (i.e., the inequality of hypothesis is not supported) on corresponding dates. We further illustrate data from Table 1 in Fig. 3 which shows a "rough monotonic property" stated in Hypothesis 1 from August 22, 2006 to February 20, 2007 (i.e., the first 10 data points).

Table 1
The average YTD rates of returns of "dominance ordered classes (≤)".

	R(8/22/06)	R(9/12/06)	R(10/3/06)	R(10/24/06)	R(11/14/06)	R(12/5/06)	R(12/26/06)	R(1/9/07)	R(1/30/07)	R(/2/20/07)	R(3/13/07)	R(/4/3/07)	R(4/24/07)
N_1	0.0724	0.0724	0.0904	0.0998	0.0998	0.1188	0.1434	0.1434	0.1614	0.1614	0.0096	0.0150	0.0150
D_1	0.0292	0.0292	0.0497	0.0617	0.0618	0.0916	0.1185	0.1185	0.1277	0.1278	0.0128	0.0103	0.0103
N_2	0.0637	0.0637	0.0854	0.0933	0.0933	0.1205	0.1454	0.1454	0.1650	0.1660	0.0090	0.0119	0.0123
D_2	0.0275	0.0275	0.0479	0.0602	0.0602	0.0902	0.1172	0.1172	0.1258	0.1259	0.0130	0.0102	0.0102
N_3	0.0663	0.0663	0.0885	0.0973	0.0973	0.1271	0.1550	0.1550	0.1736	0.1735	0.0130	0.0147	0.0148
D_3	0.0230	0.0230	0.0431	0.0558	0.0558	0.0858	0.1127	0.1127	0.1202	0.1203	0.0130	0.0097	0.0097
N_4	0.0479	0.0479	0.0677	0.0783	0.0783	0.1078	0.1356	0.1356	0.1455	0.1476	0.0130	0.0122	0.0123
D_4	0.0170	0.0170	0.0373	0.0504	0.0505	0.0806	0.1073	0.1073	0.1142	0.1138	0.0130	0.0091	0.0091
N_5	0.0552	0.0552	0.0741	0.0827	0.0825	0.1150	0.1479	0.1479	0.1680	0.1666	0.0092	0.0075	0.0076
D_5	0.0098	0.0098	0.0304	0.0443	0.0444	0.0741	0.0996	0.0996	0.1040	0.1038	0.0137	0.0094	0.0093
N_6	0.0351	0.0351	0.0516	0.0635	0.0639	0.0922	0.1153	0.1153	0.1259	0.1234	0.0104	0.0082	0.0082
D_6	0.0048	0.0048	0.0261	0.0405	0.0405	0.0705	0.0965	0.0965	0.0996	0.0999	0.0144	0.0096	0.0096
N_7	0.0230	0.0230	0.0420	0.0589	0.0589	0.0874	0.1094	0.1094	0.1169	0.1179	0.0125	0.0077	0.0077
D_7	−0.0050	−0.0050	0.0176	0.0307	0.0306	0.0614	0.0896	0.0896	0.0903	0.0902	0.0153	0.0107	0.0106
N_8	0.0346	0.0346	0.0547	0.0556	0.0556	0.0894	0.1251	0.1251	0.1247	0.1244	0.0123	0.0128	0.0126
D_8	−0.0237	−0.0237	0.0001	0.0190	0.0189	0.0483	0.0729	0.0729	0.0741	0.0741	0.0168	0.0097	0.0096
N_9	−0.0025	−0.0025	0.0200	0.0366	0.0354	0.0674	0.0897	0.0897	0.0942	0.0944	0.0165	0.0082	0.0077
D_9	−0.0325	−0.0325	−0.0081	0.0116	0.0120	0.0403	0.0658	0.0658	0.0657	0.0656	0.0168	0.0103	0.0104
N_{10}	−0.0177	−0.0177	0.0046	0.0219	0.0228	0.0492	0.0741	0.0741	0.0733	0.0753	0.0158	0.0144	0.0144
D_{10}	−0.0415	−0.0415	−0.0160	0.0054	0.0054	0.0349	0.0608	0.0608	0.0611	0.0597	0.0174	0.0078	0.0080
N_{11}	−0.0427	−0.0427	−0.0193	0.0003	0.0003	0.0317	0.0599	0.0599	0.0606	0.0581	0.0180	0.0118	0.0120
D_{11}	−0.0401	−0.0401	−0.0120	0.0113	0.0113	0.0386	0.0618	0.0618	0.0616	0.0616	0.0168	0.0031	0.0033
N_{12}	−0.0389	−0.0389	−0.0111	0.0123	0.0123	0.0397	0.0627	0.0627	0.0629	0.0629	0.0168	0.0036	0.0038
D_{12}	−0.0747	−0.0747	−0.0380	−0.0167	−0.0167	0.0087	0.0360	0.0360	0.0247	0.0247	0.0167	−0.0113	−0.0113
(i) # of times falsified H1a	1	1	1	1	1	1	1	1	1	1	8	4	4
(ii) # of times falsified H1b	4	4	4	3	3	5	5	5	5	5	5	4	4
(iii) # of times falsified H1c	1	1	1	1	1	1	1	1	1	1	6	4	4

In this subsection, we denote $\overline{N_{kt}}$ and $\overline{D_{kt}}$ as the average YTD rates of return at time t for the mutual funds of $N_k(\leq)$ and $D_k(\leq)$ respectively.

For H1a, it is expected that $\overline{N_{kt}} > \overline{D_{kt}}$ for all t (i.e., 13 dates from August 22, 2006 to April 24, 2007). In addition, in our study we derive 12 classes (i.e., $k = 1, 2, \ldots, 12$) so that there are 12 comparisons for each time t and each set ($N(\leq)$ and $D(\leq)$). For the data points from August 22, 2006 to February 20, 2007 (i.e., the first 10 columns) listed in Table 1, we can see that H1a holds for all comparisons but one. Therefore, during this time horizon, the hypothesis is largely supported where the data shows a support of 11/12 (11 out of 12 comparisons) which is significantly better than a random selection would show (6/12). It is also interesting to note that after $t =$ February 20, 2007, the data do not support this hypothesis.

For H1b, it is expected that $\overline{N_{(k-1)t}} > \overline{N_{kt}}$ for all t. There are 11 comparisons ($k = 2, 3, \ldots, 12$) between $N(\leq)$ classes. For the data points from August 22, 2006 to February 20, 2007 (i.e., the first 10 columns) listed in Table 1, we can see that H1b is falsified between 3 to 5 times. Therefore, during this time horizon, the hypothesis is not supported where the data shows a support of only from 6/11 to 7/11which is not significantly better than a random selection 1/2.

For H1c, it is expected that $\overline{D}_{(k-1)t} > \overline{D_{kt}}$ for all t. Similarly, there are 11 comparisons ($k = 2, 3, \ldots, 12$) between $D(\leq)$ classes. From the first 10 data points in Table 1, we can see that H1c is falsified only once. Therefore, during this time horizon, the hypothesis is largely supported where the data shows a support of 10/11 which is significantly better than a random selection would show (1/2). Again, it is interesting to note that after $t =$ February 20, 2007, the data do not support this hypothesis.

5.3. Strictly dominance ordered classes and Hypothesis 2

With the Definition B in Appendix B, we classified the 2027 mutual funds into four classes: $\{(N_k(<), D_k(<)) | k = 1, 2\}$. The number of distribution of each class is as follows.

k	1	2
$N_k(<)$	1346	653
$D_k(<)$	681	28

There are only four classes because it is much harder to have "*strict dominance*" than to have "*dominance*". The corresponding average YTD rates of return for each class from August 22, 2006 to April 24, 2007 (denoted by R(8/22/06), …, R(04/24/07) with 13 data points) are listed in Table 2. The indices of the last three rows show the frequency of falsification of H2a, H2b, and H2c respectively (i.e., the inequality of hypothesis is not supported) on corresponding dates. We illustrate data from Table 2 in Fig. 4 which again shows a "rough monotonic property" stated in Hypothesis 2 from the first 10 data points.

In this subsection, we denote $\overline{N}_{kt}$ and $\overline{D_{kt}}$ as the average YTD rates of return at time t for the mutual funds of $N_k(<)$ and $D_k(<)$ respectively.

For H2a, it is expected that $\overline{N}_{kt} > \overline{D_{kt}}$ for all t. For the data points from August 22, 2006 to February 20, 2007 (i.e., the first 10 columns) listed in Table 2, we can see that H2a holds for all comparisons. However, after $t =$ February 20, 2007, the data do not support this hypothesis.

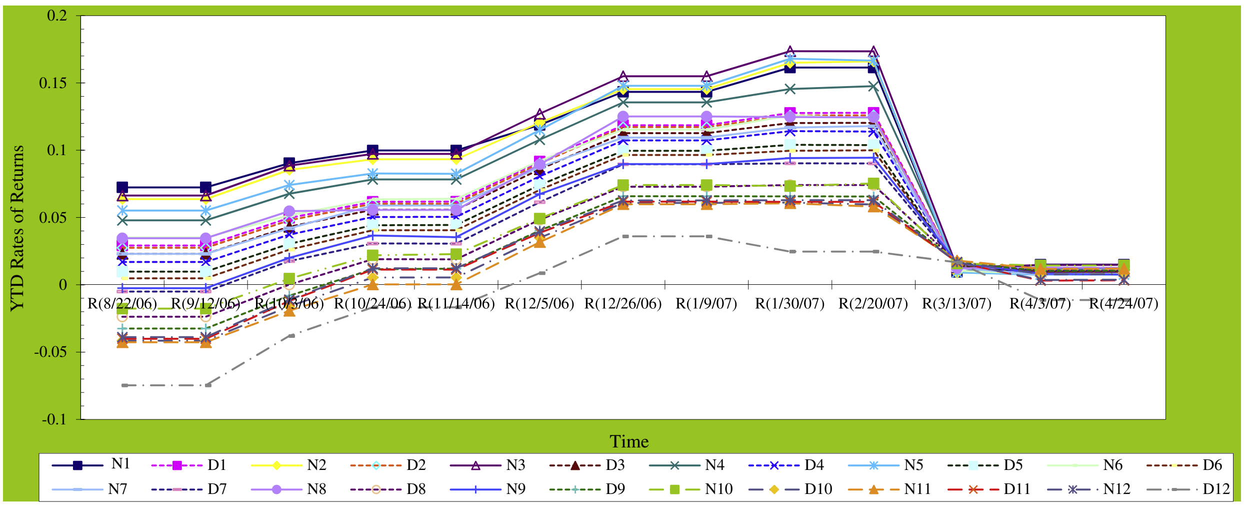

Fig. 3. The YTD rates of returns of non-dominated classes and dominated classes (≤).

Table 2
The average YTD rates of returns of "strictly dominance ordered classes (<)".

	R(8/22/06)	R(9/12/06)	R(10/3/06)	R(10/24/06)	R(11/14/06)
N_1	0.0450	0.0450	0.0648	0.0746	0.0746
D_1	−0.0016	−0.0016	0.0202	0.0367	0.0366
N2	−0.0007	−0.0007	0.0209	0.0375	0.0374
D2	−0.0205	−0.0205	0.0042	0.0180	0.0180
(i) # of times falsified H2a	0	0	0	0	0
(ii) # of times falsified H2b	0	0	0	0	0
(iii) # of times falsified H2c	0	0	0	0	0

(*continued on next page*)

For H2b, it is expected that $\overline{N_{1t}} > \overline{N_{2t}}$ for all t. This hypothesis is fully supported with the data points from August 22, 2006 to February 20, 2007 but is not supported with the data points after February 20, 2007.

For H2c, it is expected that $\overline{D_{1t}} > \overline{D_{2t}}$ for all t. Similarly, there are 11 comparisons ($k = 2, 3,\ldots, 12$) between $D(\leq)$ classes. This hypothesis is fully supported with the data points throughout August 22, 2006 and February 20, 2007.

5.4. Discussions and implications

With respect to YTD rates of return, we summarize the results of hypothesis testing in Table 3. In addition, we also performed the same operations and analyses on Survey data published on two other dates (April 25, 2006 and December 26, 2006). In order to save space, we did not report these results which were similar to our major findings.

Base on the dominance ordered class "≤", H1a and H1c (but not H1b) are largely supported for a certain period of time (about one half year). After that, Hypothesis 1 is not supported. This suggests that ratings (e.g., published by the Survey on August 22, 2006) can offer useful and valid information value for interested investors by considering specific circumstances. With our introduced technique which classifies assets into different non-dominated classes and dominated classes based on specific published ratings, investors can expect to make better decisions with respect to rate of return. For example, our results from testing H1a suggest that investors should invest in assets belonging to a non-dominated class instead of those in corresponding dominated class. Our results from testing H1c suggest that among dominated classes investors should invest in assets belonging to a lower-ordered class (i.e., the smaller k of kth class) instead of those in higher-ordered class. It is also interesting to point out that investors should be cautious with the use of rating information which is not timely and current (e.g., more than one half year based on our results).

Considering the strictly dominance ordered class "<", Hypothesis 2 is fully supported for a certain period of time (about one half year). After that, only H2c is supported. This suggests that with our introduced technique which classifies assets into different strictly non-dominated classes and strictly dominated classes based on specific published ratings, investors can expect to make better decisions with respect to rate of return. Our results from testing H2a suggest that investors can invest in assets belonging to a strictly non-dominated class instead of those in corresponding strictly dominated class. Our results from testing H2b (and H2c) suggest that among strictly non-dominated classes (and strictly dominated classes) investors can invest in assets belonging to a lower-ordered class instead of those in higher-ordered class.

Strict dominance requires that the rating in each criterion be strictly dominated and thus fewer mutual funds could be strictly dominated. As a result, the number of classes of $\{N_k(<), D_k(<)\}$ are much less than that of the corresponding $\{N_k(\leq), D_k(\leq)\}$. In other words, the classification based on "<" cannot be as fine as that based on "≤"; in turn, we conjecture earlier that the strict dominance relation of "<" has a smaller discrimination power than the dominance relation of "≤". However, the flip side of this reasoning is that strict dominance classes (once got classified) are better differentiated between each other since it is harder to derive "strict dominance" than just "dominance". Our results show that for dominance based on "≤", the Survey are only *largely consistent* with some corresponding hypotheses, and that for strict dominance based on "<", the Survey are *fully consistent* with all corresponding hypotheses.

6. Conclusions

In this study, we consider information quality as the characteristic of information to be of high value to its users. From the financial investment decision-making perspective, investors need to rely on the support of quality information to make good investment decisions. We propose a method to identify information quality and verify validity of published rating reports with a sound mathematical foundation.

In the financial market, most available rating information is based on multi-criteria and published by myriad agents or companies. Many investors who naively rely on these ratings and are unaware of their quality may result in poor investment decisions and be prone to mistakes. Using our method to systematically analyze published ratings and follow the implications from this study to objectively evaluate these ratings would be not only computationally feasible but also a useful task. By doing so, investors can be informed whether these ratings actually provide useful information and whether they can trust these ratings and intended interpretation for the use of making investment decisions. In turn, informed investors (e.g., individuals, portfolio and

Table 2 (*continued*)

R(12/5/06)	R(12/26/06)	R(1/9/07)	R(1/30/07)	R(/2/20/07)	R(3/13/07)	R(/4/3/07)	R(4/24/07)
0.1045	0.1324	0.1324	0.1441	0.1444	0.0112	0.0100	0.0100
0.0663	0.0914	0.0914	0.0954	0.0952	0.0159	0.0110	0.0109
0.0673	0.0923	0.0923	0.0968	0.0966	0.0160	0.0110	0.0110
0.0445	0.0686	0.0686	0.0619	0.0619	0.0145	0.0097	0.0097
0	0	0	0	0	1	1	1
0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0

fund managers, investment bank managers) can subsequently rely on identified "trustworthy" ratings and target asset(s) in non-dominated or good classes in general to invest and expect relatively better returns.

Following our method and given a multi-criteria rating report (e.g., Value Line, Standard & Poor's) on a finite number of assets (e.g., stocks, bonds, mutual funds), we can construct four sets of ordered classes class: $\{N_k(\leq)\}$, $\{D_k(\leq)\}$, $\{N_k(<)\}$ and $\{D_k(<)\}$. If ratings from a published report have useful and valid information value as claimed, the average performances of assets within classes (e.g., $N_k(\leq)$ within $\{N_k(\leq)|k = 1,2,\ldots\},$) are expected to show some monotonic property in k. We systematically describe the rationale and methods to construct the above-mentioned four sets of ordered classes (Section 3) and construct hypotheses (Section 4) to describe the expected monotonic properties.

For illustrations, we use Value Line Mutual Fund Survey's rating on August 22, 2006 (also April 25, 2006 and December 26, 2006, though not reported) to derive the corresponding four sets of ordered classes and test the corresponding hypotheses. We found that the usefulness and validity of published ratings depend on (1) the degree of consistency with the expected monotonic property and the length of time duration, and (2) the type of dominance ("≤" or "<"). As the dominance of "<" is more strict than that of "≤", the classification based on "≤" is sharper than that based on "<". Thus given the same length of time duration, the consistence in "≤" is more difficult but more valuable than that in "<".

It is worth noting that in this study one of the rating criteria is risk rank and we use realized average rate of returns to represent the performance of a class of assets. It is commonly recognized that assets with lower risk (volatility) usually realize lower returns on average. However, one of the reasons that we chose Value Line data for our empirical testing is that its safety criterion is the risk-adjusted performance. It is calculated by dividing an asset's three-year total return by the standard deviation of its return and is commonly known as the Sharpe ratio. In addition, it is not always true that lower risk of an asset leads to lower performance since the judgment of risk itself can depend on the risk attitude of investors. Therefore, we believe that using Value Line data for current study is appropriate to derive our findings. At the same time, further study with using other empirical data and finding relationships between volatility and expected payoff of assets can be pursued.

Another type of ordered classes (i.e., negative dominance classes) can also be constructed by using our technique. Especially, when the size of a rating space becomes large (i.e., many assets to be considered), the classification outcomes of dominance ordered classes and negative dominance ordered classes can be different. Therefore, it can be useful to construct and examine both types of ordered classes with our technique and it might be interesting to further study whether there exist any relationships of these classes as well as the implications derived from these classes.

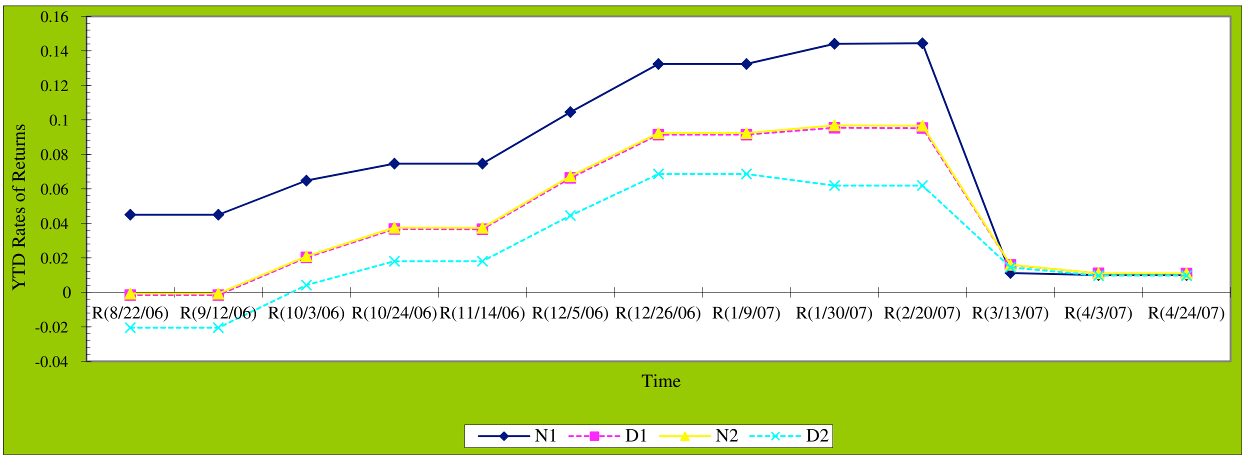

Fig. 4. The YTD rates of returns of strictly non-dominated classes and strictly dominated classes (<).

Table 3
Summary of hypothesis testing.

Data points (Survey dates)	8/22/2006–2/20/2007 (First 10 dates)	3/13/2007–4/24/2007 (Last 3 dates)
Hypothesis 1 (≤)	H1a is largely supported	Not supported
	H1b is not supported	Not supported
	H1c is largely supported	Not supported
Hypothesis 2 (<)	H2a is fully supported	Not supported
	H2b is fully supported	Not supported
	H2c is fully supported	Supported

To further extend our classification method, we can explore many other dominance cones, varying from "<" to a half space. For example, $V(y,t)$, as defined in Formula (1), with $\beta_{it} < 0$, $i = 1,\ldots,l$ is increasing in a cone $\Lambda = \left\{(h_1, \ldots, h_l) \middle| \sum_i h_i\beta_{it} > 0\right\}$ which is a half space and much larger than that of "≤" or "<". A careful study of information contents of ratings in terms of varying dominance cones could further refine our evaluation of these ratings. For further discussion of this extension, please see of Yu (Chapters 7 and 10, 1985).

With current study as a base and a first step, future testing of our hypotheses can be conducted with (1) ratings published by many agencies other than the Value Line, (2) ratings of various financial assets other than mutual funds, (3) ratings published in shorter and/or longer time intervals, (4) ratings published longer time ago or more recently, and (6) ratings collected in longer time horizon. We believe that these extended empirical investigations can potentially offer more detailed and insightful results for financial investors.

Future studies can arrange and examine different combination and/or number of criteria of ratings. With the use of our proposed method, we can then discover the "right" criteria to be included in ratings for an asset to be evaluated so that more useful and reliable investment aids can be extracted. In addition, by doing so, our method not only can identify "poor" existing assets that investor should avoid but also can timely detect inferior upcoming or new assets and weep them out for investors.

Our method can also be applied with other multi-criteria ranking practices. For example, it can be used for evaluating ratings from hotel rating agents (e.g., AAA, hotels.com, and tripadvisor) by tracking hotel performance (e.g., occupancy rate, customer satisfaction). In this fashion, evidence can be provided that specific hotel rating agents consistently provide useful information for travelers to choose ideal places to stay by relying on their ratings of hotels. Other practices can be considered are sports scouting, university admission, professor evaluation, and college choices with corresponding performance measurements.

Appendix A. Summary of notations

Notation	Meaning
X	Asset space, $x \in X$, $x = 1, \ldots, n$ assets
C	Criteria space, $i \in C$, $i = 1, \ldots, l$ criteria
y	Rating vector, $y = (r_1, \ldots r_l)$
r_i	The rating of ith criterion in C
y_i	$y_i = r_i$
Y	Rating space, $Y = \{f(x)\|x \in X\}$
$f(x)$	Rating vector for asset x
$f_i(x)$	The rating of ith criterion for asset x; $f_i(x)$ is an integer and $1 \le f_i(x) \le M$; M is the largest number of a ranking for a criterion; 1 is the best rating and M is the worst rating.
y^x	$y^x = f(x)$
y_i^x	$y_i^x = f_i(x)$
$N(\le)$	Non-dominated class — a set (class) of rating vectors that do not have all rating criteria that are worse than or equal to those of compared rating vectors
$D(\le)$	Dominated class — a set (class) of rating vectors that are dominated by $N(\le)$; i.e., all rating criteria in any vector in $D(\le)$ that are always worse than or equal to those of *at least one* vector in $N(\le)$
$N_k(\le)$	The kth class of $N(\le)$; $N_1(\le)$ is the class that is not dominated by any other class
$D_k(\le)$	The kth class of $D(\le)$; $D_1(\le)$ is the class that is dominated by $N_1(\le)$; $D_2(\le)$ is the class that is dominated by both $N_1(\le)$ and $N_2(\le)$; and so forth
$N(<)$	Strictly non-dominated class — a set (class) of rating vectors that do not have all rating criteria that are strictly worse than those of compared rating vectors
$D(<)$	Strictly dominated class — a set (class) of rating vectors that are strictly dominated by $N(<)$; i.e., all rating criteria in any vector in $D(<)$ that are always strictly worse than those of *at least one* vector in $N(<)$
$N_k(<)$	The kth class of $N(<)$; $N_1(<)$ is the class that is not strictly dominated by any other class
$D_k(<)$	The kth class of $D(<)$; $D_1(<)$ is the class that is strictly dominated by $N_1(<)$; $D_2(<)$ is the class that is strictly dominated by both $N_1(<)$ and $N_2(<)$; and so forth
$V(y_i,t)$	Value function reflecting rating vector y at time t, $V(y_i, t) = \alpha_t + \sum_{i=1}^{l} \beta_{it} y_i + \varepsilon_t$
P	Portfolio vector of n assets; $P = (x_1, \ldots, x_n)$; $\sum_{p=1}^{n} x_p = 1$ where x_p is the portion of the investment put in the asset p; $p = 1, \ldots, n$ assets
P_i	Collective rating of a criterion i of all asset in portfolio P; $P_i = \sum_{p=1}^{n} x_p f_i(p)$

Appendix B. Strict dominance ordered classes

Definition B. Given Y, a "strictly non-dominated class" of rating vector(s) is defined as $N(<) = \{y^x \in Y | no\, y^* \in Y \text{ with } y^* < y^x\}$; and a "strictly dominated class" of rating vector(s) can be defined as $D(<) = Y \backslash N(<)$. Let $N_1(<) = N(<)$, $D_1(<) = D(<)$. For $k = 2, 3, \ldots$ where k denotes the kth "class", we define $N_k(<) = \{y^x \in D_{k-1}(\leq) | \, no\, y^* \in D_{k-1}(<) \text{ with } y^* < y^x\}$; $D_k(<) = D_{k-1}(<) \backslash N_k(<)$.

Illustration B. Given Y as in Example 1, we have $N_1(<) = \{(1,2), (1,3), (1,5), (3,1), (4,1)\}$, $N_2(<) = \{(2,5), (3,4), (3,5), (4,2), (5,2)\}$, $N_3(<) = \{(5,5), (5,3)\}$ (see Fig. B1 below).

Proposition B.

(i) If $y^{x_2} \in N_{k'}(<)$ and $1 \leq k < k'$, then there exists $y^{x_1} \in N_k(<)$ such that y^{x_1} strictly dominates y^{x_2} (i.e., $y^{x_1} < y^{x_2}$).

(ii) If $k \neq k'$, then $N_{k'}(<) \cap N_k(<) =$.

(iii) There exists $m \geq 1$, such that $Y = \bigcup_{k=1}^{m} N_k(<)$.

(iv) $D_k(<) \supset D_{k+1}(<)$ and $D_k(<) \supset D_{k'}(<)$, if $k < k'$ with $k, k' \in \{1,\ldots,m\}$.

Theorem B. *$\{N_1(<), N_2(<), \ldots, N_m(<)\}$ forms a decomposition of Y and all strictly non-dominated classes are mutually exclusive (proof is from Proposition B-(ii) and (iii)).*

From (i), it essentially says that if $k < k'$, then $N_k(<)$ collectively dominates $N_{k'}(<)$ in the sense that for each $y^x \in N_{k'}(<)$ there is $y^{x^*} \in N_{k'}(<)$ such that y^{x^*} strictly dominates y^x (i.e., $y^{x^*} < y^x$). Thus we could reasonably expect that the average performance of $N_k(<)$ to be better than that of $N_{k'}(<)$ if $k < k'$.

Fig. B1. The strictly non-dominated classes derived from Example 1.

References

Ai, H. (2010). Information quality and long-run risk: Asset pricing implications. *Journal of Finance*, *65*(4), 1333–1367.
Blake, C. R., & Morey, M. R. (2000). Morningstar ratings and mutual fund performance. *Journal of Financial and Quantitative Analysis*, *35*(3), 451–483.
Brevik, F., & d'Addona, S. (2010). Information quality and stock returns revisited. *Journal of Financial and Quantitative Analysis*, *45*(6), 1419–1446.
Brown, S. J., & Goetzmann, W. N. (1995). Performance persistence. *Journal of Finance*, *50*(2), 679–698.
Carlo, R. W., & Michelle, H. W. (2008). The credit crunch and credit rating agencies: Are they really striving towards more transparency? *Journal of Securities Compliance*, *1*(4), 322–336.
Chiang, K. C. H., Kozhevnikov, K., & Wisen, C. H. (2005). Ranking properties of Morningstar risk-adjusted ratings. *Journal of Investing*, *14*(1), 90–98.
Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*, 166.e7–166.e16.
Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 443–507) (2nd ed.). Washington, DC: American Council on Education.
Eppler, M. (2003). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes.* Berlin, Germany: Springer-Verlag.
Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56.
Fama, E., & French, K. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*, 55–84.
Ge, Mouzhi (2009). *Information quality assessment and effects on inventory decision-making.* (PhD thesis). Dublin City University.
Goetzmann, W. N., & Ibbotson, R. G. (1994). Do winners repeat? *Journal of Portfolio Management*, *20*(2), 9–18.
Grinblatt, M., & Titman, S. (1992). The persistence of mutual fund performance. *Journal of Finance*, *47*(5), 1977–1984.
Jensen, M. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, *23*, 389–416.
Jensen, M. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business*, *42*, 167–247.
Juran, J. (1992). *Juran on quality by design.* New York: The Free Press.
Kopa, M., & Post, T. (2009). A portfolio optimality test based on the first-order stochastic dominance criterion. *Journal of Financial and Quantitative Analysis*, *44*(5), 1103–1124.
Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, *50*(10), 1390–1406.
Lee, C. F., Wu, C., & Wei, K. C. J. (1990). The heterogeneous investment horizon and the capital asset pricing model: Theory and implications. *Journal of Financial and Quantitative Analysis*, *25*(3), 361–376.
Lehmann, B. N., & Modest, D. (1987). Mutual fund performance evaluation: A comparison of benchmarks and a benchmark of comparisons. *Journal of Finance*, *21*, 233–265.
Levy, H. (1998). *Stochastic dominance: Investment decision making under uncertainty. Studies in Risk and Uncertainty.* Boston, Dordrecht and London: Kluwer Academic.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103) (3rd ed.). New York American Council on Education and Macmillan.
Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50*(9), 741–749.
O'Reilly, C. A. (1982). Variations in decision makers' use of information sources: The impact of quality and accessibility of information. *The Academy of Management Journal*, *25*(4), 756–771.
Post, T. (2003). Empirical tests for stochastic dominance efficiency. *Journal of Finance*, *58*(5), 1905–1931.
Redman, T. (2001). *Data quality: The field guide, 2001.* Newton, Massachusetts: Digital Press.
Veronesi, P. (2000). How does information quality affect stock returns? *Journal of Finance*, *55*(2), 807–837.
Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–65.
Yu, P. L. (1985). *Multiple-criteria decision making, concepts, techniques, and extensions.* New York: Plenam Press.