THE DEMOGRAPHICS OF CHANGE: ENTERPRISE CHARACTERISTICS AND BEHAVIORS THAT INFLUENCE TRANSFORMATION

Zhongyuan Yu,1 Nicoleta Serban,2 and William B. Rouse1
1Center for Complex Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, USA
2School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Tolstoy asserts that happy families are all alike, but that every unhappy family is unhappy in its own way. Successful companies achieve strong performance based on many attributes—total revenue, current profitability, stock market valuation, and future potential profitability. Unsuccessful companies that experience decreases in these metrics may fail along any dimension, or combination of dimensions, for a wide range of reasons. We characterize the various combinations of changes, including explanations for how they arise, as the “demographics of change.” In this article, we elaborate the demographics of change by exploring companies’ changes in ranks and financial performance within two industry sectors: information technology and healthcare. We then pursue explanations of these changes using text mining of press releases, business news reports, and related materials to identify enterprise behaviors associated with changes in ranks and financial performance. The contribution of the article is in applying advanced yet easy-to-use techniques to study the demographics of change, understand the dynamics of industries as they change, explore reasons for enterprise changes, and provide potential insights into investment opportunities.

Keywords company ranking; enterprise transformation; financial performance; industry trends; ranking methodologies; text mining

INTRODUCTION

“If you are not Apple or Android, you better pack up and go home.” This is only a jest in the mobile device industry in light of Nokia’s falling...
market share and the ascendance of Apple and Samsung, but it reveals that in an openly competitive global market, changes are likely and perhaps inevitable.

Enterprise changes can be described in terms of company ranks, financial performance measures, or company-specific decisions and events. How should companies in an industry be ranked? How can we determine whether an industry has concentrated dominant market leaders or has many good performers? How should we interpret the volatility of rankings? Are there significant trends for an individual company’s financial performance? In which industry portfolios should we choose to invest? What happened to the companies that were undergoing these changes? This article introduces an integrated approach to analysis of companies and industries that enables addressing these questions using two analytic tools: data mining and text mining. While our methodology is new, the topic is not; numerous descriptive case studies have been conducted for individual company or specific events, such as mergers and acquisitions and market innovations (Buchanan and Simmons, 2007; Kanter, 2009; Sölvell and Porter, 2002; Stachowicz-Stanusch and Śląska, 2009; Yoffie and Kim, 2011). In the healthcare industry, most research has focused on the healthcare delivery system instead of companies in the broader healthcare space (Jacobson, Hall, and Swisher, 2006; Klassen and Rohleder, 1996; Lowery and Davis, 1999; Mahapatra et al., 2003). Few of these studies have analyzed companies and their relative positions at the industry level.

In both the information technology and healthcare sectors, research literature has largely moved away from employing regression to more sophisticated statistical analysis, but classic methods nevertheless remain popular. Such studies construct measures, such as the Herfindahl–Hirschmann Index, as a representation of market concentration and then regress it on the variable of interest controlling for observable confounding variables (Gaynor and Town, 2012). Mathematical modeling is another possible approach (Eibner et al., 2010; Liang and Sheng, 2011). Rouse (2005) proposed a theory of enterprise transformation that focuses on why and how transformation happens. Yu et al. (2011) created a computational version of this theory to understand companies’ best responses when changes happen. In using modeling tools, researchers necessarily have to make simplifications and assumptions that are less robust than financial and accounting data and that more or less distort interpretations of what happened in the real world.

In summary, most research has been restricted to a specific company with either descriptive studies that cannot be generalized or to outdated and/or complicated techniques that are not easy to implement. Studies of companies’ changes at the industry level and analyses of why changes happened with advanced but easy-to-use methodologies are scarce.

The key methodologies in this study are:
(1) a ranking methodology that, in effect, serves the same purpose as Fortune 500 and Russell 3000 rankings but includes multiple indicators and is inherently more comprehensive; and
(2) text mining, which enables inferences of deeper reasoning for changes in ranking.

These methodologies reveal enormous insights into industry structures as well as individual company’s positions and characteristics. Enterprise leaders could employ them to monitor future trends and anticipate potential crises from their own and other competitors’ changes in performance, rankings, and initiatives. Studies about ranking and changes in financial indicators can also suggest opportunities for investment.

In the next section, we discuss the industries selected, measures chosen, and data sources. Following that, ranking patterns derived using data mining techniques are illustrated based on the proposed novel ranking methodology. Next, trends in financial indicators for each company are determined, and we relate these trends to changes in ranks over time. In the penultimate section, text mining is employed to uncover insights that can explain the changes of company ranks in a richer way than possible with just financial metrics. The final section summarizes suggestions and implications for further research.

**DATA SUMMARY**

This study highlights two different industry sectors as examples to explore the demographics of change. The two sectors are information technology, which experienced fierce competition over the past decade that will surely continue, and healthcare, in which competitive changes are now emerging due to both cost pressures and technological opportunities. The information technology sector, especially software and services, has a strong connection with the daily operations of nearly every business. This cushions the effect of the downside of business cycles. But the inexorable advance of high technology, e.g., recent cloud technology, prevents this sector from ever becoming fully mature. Increasing competition and pressures on profitability expunge players with obsolete offerings. Entries to this sector are quite common and not difficult, but it requires entrants to anticipate the right technology and appropriate (and perhaps niche) markets and to sustain these advantages over numerous rivals. Many entrants are eliminated in this competition or reduced to a survival state in which they cannot achieve scale and become acquisition targets.

The Global Industry Classification Standard (Morgan Stanley Capital International and Standard & Poor’s, 2008) further distinguishes each sector into several major industry groups. The information technology sector is divided into three industries. The first industry is semiconductors and semiconductor equipment (semiconductors) with a total of 123 companies, including Intel and Texas Instruments. The second industry is
software and services (software) with a total of 438 companies; examples of companies in this industry are Microsoft, IBM, and Google. The last industry is technology hardware and equipment (hardware) with a total of 344 companies, for instance, Apple, Hewlett-Packard, and Cisco Systems.

The healthcare sector was traditionally regarded as a stable industry. However, in recent years, things have changed due to an aging population, price pressures, and an altered industry landscape. Consequently, the healthcare sector has grown rapidly; this leaves much room for future expansion but also greater competition, including from non-traditional channels, as many companies are seeking to profit from these prospects. The healthcare sector contains two industries. One is the healthcare equipment and services industry (healthcare) with a total of 398 companies, e.g., Medtronic and United Health Group. The pharmaceuticals, biotechnology, and life sciences industry (pharmaceuticals) has a total of 408 companies, for instance, Pfizer and Johnson & Johnson.

Companies in these industries can be analyzed using numerous indicators. Hansen and Wernerfelt (1989) suggested that external market factors and organizational factors determine firm success. Teece, Pisano, and Shuen (1997) summarized the competitive advantages of firms. Most of the competitive advantages summarized above are difficult to evaluate from available data. Dess and Robinson (1984) focused on return on assets (ROA), growth in sales, and overall performance. Orlitzky et al. (2003) used accounting-based indicators, such as the firm’s ROA, return on equity (ROE), or earnings per share (EPS), and argued that these metrics capture a firm’s internal efficiency. Capon et al. (1990) discussed selected financial performance variables, such as growth and variability in profit, market value, assets, equity, cash flow, sales, and market/book value. In summary, returns, sales, revenues, enterprise values (EVs), and various ratios were the most commonly selected financial indicators to evaluate and rank companies.

Five indicators were selected for this article to track each company’s total revenue (TR), current profitability, stock market valuation, and future potential profitability. These five indicators are:

1. TR;
2. net operating profit after tax (NOPAT);
3. EV, i.e., market capitalization;
4. normalized enterprise profit, equal to the return on invested capital (ROIC) minus the weighted average cost of capital (WACC), which provides an indicator of current performance:

\[ EP_{Normalized} = \frac{EP}{IC} = \frac{(ROIC - WACC) \times IC}{IC} = ROIC - WACC; \]

5. normalized future value, equal to the difference of current value (CV) and the EV, divided by the invested capital (IC), provides an indicator
of future potential profitability, i.e., high stock prices compensate for a negative CV:

\[ F_{V_{\text{Normalized}}} = \frac{EV - CV}{IC} = \frac{EV}{IC} - \left( \frac{EP_{\text{WACC}} + IC}{IC} \right) = \frac{EV - \frac{ROIC}{\text{WACC}} \times IC}{IC} = EV - \frac{ROIC}{\text{WACC}}. \]

These indicators and data were derived from the Capital IQ database (2010). These five indicators are used and transformed in various ways in this article. These indicators were chosen to address the basics of enterprise performance based on the ground truth of their revenues, profits, stock prices, and returns. These metrics describe the overall conditions of companies and are not necessarily designed for investors who are interested only in returns, especially short-term returns.

**RANKING METHODOLOGY AND MINING**

Long-term investments in extremely dominant market leaders are regarded as safe but deliver lower upside returns; investment in rising stars is riskier but with potential higher returns. Investment in a company is risky when its industry is quite volatile in terms of companies’ ranking changes. Studying the demographics of change, for example, clusters and volatility of rankings could lead to a better understanding of the dynamics over time of these industries and, potentially, insights for investment. Common rankings of companies provided by Fortune 500 and Russell 3000 companies only include one performance metric; we propose a ranking methodology that combines the effects of a company’s current profitability, stock market valuation, and future potential profitability. Based on the ranking results, we analyze and compare patterns across industries.

**Ranking Methodology**

One of the most intuitive ways to explore the demographics of change in companies is through companies’ ranking changes within an industry. Many well-known company rankings are based on a single variables or assigning fixed weights to multiple uncorrelated variables. Fortune 500 companies are ranked by TR for their respective fiscal years (Fortune, 2012); Russell 3000 companies are ranked by market capitalization (Russell, 2012).

Russell also uses multiple financial indicators in the derivation of some market indices, e.g., the growth and value index. The Russell algorithm ranks different indicators separately and converts them to standardized units. This may raise some concerns. For example, assume that companies A and B are the first and tenth companies ranked by book-to-price and that their values are nearly the same. But company B may have doubled or tripled
by the Institutional Broker’s Estimate System forecast medium-term growth (2-year) value compared to company A. By ranking separately and assigning fixed weights, company A might have a better ranking, but this method ignores the difference across indicators.

Many rankings for companies simply assign fixed weights for multiple variables (Daily Beast, 2011). This is undoubtedly biased, as different people are likely to prefer different weightings. To reduce this inherent bias, the ranking system should be data driven. Theoretically, ranking methods can be divided into two categories: total and partial ranking methods, according to Pavan and Todeschini (2009). The main difference is that in partial ranking, not all objects are comparable, while in total ranking, every two members of the set are comparable and the set could generate a unique “linear” chain. The well-known total-order ranking methods for multi-criteria decision making (MCDM) include the Simple Additive Ranking (SAR), Pareto Optimal Ranking, and the Multi-Attribute Value Function method, which further includes utility functions. All methods require assigning weights for the criteria (financial indicators in this article), which may strongly affect the final results of the decision-making process. Most of the total-order ranking methods assume independence between indicators and thus cannot be applied to our dataset because the indicators or criteria are strongly correlated.

Partial-order ranking, on the other hand, is considered to be a non-parametric method, thus not a weighting-based approach. In brief, when we compare two elements in a set, only the mathematical relations of ≤, =, ≥, or “incomparable” status are needed to link all the elements. The graphical representation of the partial ordering is called the Hasse Diagram, introduced by Hasse and developed by Brüggemann et al. (1995). Pavan and Todeschini (2009) suggested that a small data sample (the example in their study was around 20 members) is more suitable for using the Hasse Diagram Technique (HDT). However, the enterprise financial data we consider in our analysis consist of multiple indicators observed across different industries with more than 300 members per industry in general. Another reason why the HDT might not be the best solution to our problem is that grouping top normalized future value companies (e.g., Constellation 3D or Brocade Communications Systems) and top EV companies (e.g., Hewlett-Packard) together is inconsistent with common knowledge of these companies.

This article proposes a ranking method combining the effects of a company’s current profitability, stock market valuation, and future potential using data-driven weights. The steps of ranking are as follows.

1. Transform correlated indicators into uncorrelated variables using principal component analysis (PCA).
2. Obtain final scores for companies; weightings are the explained percentage of the variance in each dimension of the uncorrelated variables.
3. Rank or cluster the scores.
 TABLE 1 Total Ranking of Hardware Industry Companies

<table>
<thead>
<tr>
<th>Ticker</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYSE:HPQ</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NasdaqGS:AAPL</td>
<td>257</td>
<td>47</td>
<td>50</td>
<td>50</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>NasdaqGS:CSCO</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>NasdaqGS:DELL</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>NasdaqGS:QCOM</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>TSX:RIM</td>
<td>30</td>
<td>87</td>
<td>139</td>
<td>19</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>HSE:NOKIV</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>NYSE:WDC</td>
<td>154</td>
<td>65</td>
<td>24</td>
<td>14</td>
<td>18</td>
<td>12</td>
<td>16</td>
<td>14</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>NYSE:XXR</td>
<td>26</td>
<td>48</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>NYSE:MSI</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>66</td>
<td>13</td>
</tr>
<tr>
<td>NasdaqGS:SYNA</td>
<td>190</td>
<td>153</td>
<td>35</td>
<td>32</td>
<td>45</td>
<td>51</td>
<td>64</td>
<td>31</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>NYSE:HRS</td>
<td>80</td>
<td>68</td>
<td>48</td>
<td>46</td>
<td>30</td>
<td>24</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>NYSE:NCR</td>
<td>38</td>
<td>21</td>
<td>25</td>
<td>53</td>
<td>20</td>
<td>17</td>
<td>34</td>
<td>32</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>NasdaqGM:MLAB</td>
<td>105</td>
<td>56</td>
<td>40</td>
<td>63</td>
<td>49</td>
<td>37</td>
<td>33</td>
<td>23</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>NasdaqCM:OVRL</td>
<td>95</td>
<td>73</td>
<td>41</td>
<td>49</td>
<td>104</td>
<td>181</td>
<td>276</td>
<td>274</td>
<td>273</td>
<td>249</td>
</tr>
</tbody>
</table>

After choosing the indicators (discussed in the “Data Summary” section), our first step is to transform correlated indicators \( x_l (l = 1, \ldots, 5) \) into uncorrelated variables \( \xi_k (k = 1, 2, \ldots, r, r \leq 5) \) using PCA (Jolliffe, 2002). \( \xi_k \) denotes uncorrelated random variables called scores with zero mean and variances \( E(\xi_k^2) = \lambda_k \), where \( \sum_k \lambda_k < \infty \). For our data, on average, the first three principal components \( (r = 3) \) are sufficient to explain more than 85% of the variance across the indicators. Then we obtain the weighted scores for each company by weighting \( \xi_k \) with \( \frac{\lambda_k}{\sum_{j=1}^{r} \lambda_j} \). This step results in weighting less on indicators where companies have more similarities. The final score for each company \( i \) is derived as

\[
score_i = \sum_{k=1}^{r} \frac{\lambda_k}{\sum_{j=1}^{r} \lambda_j} \cdot \xi_{ik}, \quad i = 1, 2, \ldots, n.
\]

Having the final scores for each year within each industry, we could either rank them to reflect the direct ordering or cluster these final scores using, for example, the \( k \)-means method (Hartigan and Wong, 1979) to group companies with similar financial indicators. The \( k \)-means method aims to partition the \( n \) companies into \( k \) \( (k \leq n) \) groups of companies, defined as \( G = \{ G_1, G_2, \ldots, G_k \} \), so as to minimize the within-cluster sum of squares:

\[
\arg\min_G \sum_{i=1}^{k} \sum_{Score_j \in G_i} ||Score_j - \mu_i||
\]

Table 1 shows the total rankings of selected companies from the hardware industry using steps (1)–(3) as previously described. It is clear from this table that ranks do change over years, in some cases, dramatically.
Table 2 shows the group rankings of selected companies from the hardware industry using the clustering method. Group rankings appear to change less dramatically because of the aggregate nature of groups. Group ranking results in this article are data driven instead of being derived from subjective judgments. Boundaries among groups are determined by minimizing the within-cluster sum of squares only. For example, in the semiconductors industry, Intel ranks first and Texas Instruments ranks second. One could argue that both companies could be categorized in Group 1. However, the $k$-means method shows that the difference between Intel and Texas Instruments is large enough that should be categorized separately, as seen in Figure 1.

The methodology presented here was validated by the strong correlation of the resulting ranks with Fortune 500 and Russell 3000. The proposed method is considered to be a more comprehensive ranking method for two reasons. First, given correlated indicators, which are common for financial indicators, this methodology can transform correlated indicators into uncorrelated ones and obtain final scores for ranking. Second, the proposed ranking method does take multiple indicators into consideration and, thus, does not depend on any one of these indicators alone. Further, the weighting depends on the data rather than subjective judgments.

### Ranking Patterns

Based on the ranking methodology, we analyze and compare four patterns derived from the ranking results. The patterns are company clustering, which shows whether an industry has concentrated dominant market leaders or has many good performers. Consecutive changes of group ranking determines whether an investment in an industry should be short or long term. Volatility of ranks and trends of ranks are also discussed.
Company Clustering

Companies in each industry are grouped (clustered) according to the final scores in the proposed ranking method. The number of groups is based on industry size. The group index increases as companies’ performance decreases; i.e., Group 1 consists of the industry leaders. The difference between dominant market leaders and good performers depends on the number of companies in Group 1 or 2. Figure 2 shows the number of companies that ever belong to each cluster throughout 11 years. The number of best performers in the software industry is small—only 2 (0.7%) companies enter Group 1 and 13 (4.55%) companies enter Group 2. This indicates an industry with concentrated dominant market leaders. These leaders’ rankings remain quite stable throughout the 11 years and have clearly evident outstanding performance. There are more uncertainties in the healthcare industry. Many companies can be regarded as good performers, with 12 (4.88%) companies belonging to Group 1 and 25 (10.16%) companies belonging to Group 2.

The software and semiconductors industry has dominant market leaders; the hardware and healthcare industry has more good performers. The distribution of company clusters in Figure 2 reveals the structure of an industry. Industries have a heavy lower tail (Groups 9–11), because there are many new entrants, small-scale, or poorly performing companies. Top-performing groups represent the upper tail (Groups 1 and 2). The healthcare industry has a heavier upper tail than the software industry. It is harder for a company
to achieve top performance when the industry has a thin upper tail, because top performers dominate the industry, and the difference between top and good performers is large.

**Consecutive Changes of Group Ranking**

In general, the hardware and the software industries have more consecutive non-decreasing group rankings for short durations and non-increasing group rankings for longer durations. The healthcare industry displays the opposite phenomena, with more consecutive non-increasing group rankings for short durations and non-decreasing group rankings for longer durations. These observations suggest that although companies in the IT sector may achieve good performance for several years, in the long run, it is hard to remain competitive or be continuously successful. The healthcare sector, on the other hand, is more likely to experience short downturns currently, but they are also likely to have long-term stable ranks. Investors would prefer long-term investment in companies in the healthcare industry than the software industry during 1999–2009.

**Volatility of Ranks**

Volatility of ranks in this article refers to a measure for variation of rankings throughout 11 years for each company. Top-ranked companies have very small volatility. In the software industry, volatility continuously increases as companies’ rankings decrease, and the highest volatility level occurs for the bottom-ranked companies. However, in the healthcare, pharmaceuticals, hardware, and semiconductors industries, volatility tends to decrease from
The Demographics of Change

FIGURE 3 Shape and scale of ranking volatility.

mid-ranked to bottom-ranked companies. Software, hardware, and healthcare industries include similar numbers of companies (around 250–280), but the ranges of volatility are different. The software industry has volatility from 0 to 80, hardware from 0 to 60, and healthcare from 0 to 40. Figure 3 shows the shape and scale of volatility in the software and healthcare industries.

The findings of ranking volatility can be summarized as follows.

(1) In all industries, top performers have the smallest volatility in the change of rankings.
(2) In the software industry, startup companies or companies with poor performance have more volatility and are thus more likely to change or strive to achieve better performances.
(3) In all other industries except software, bottom-ranked companies have less volatility than mid-ranked companies, thus companies with poor performance are more likely to remain around the bottom; mid-ranked companies may have many changes.
(4) Although the software and hardware industries are comparatively mature industries, they are experiencing changes of rankings. The healthcare industry is less mature with less volatility; this industry may be ripe for changes in ranks.

Trends of Ranks

We apply linear regression to each company’s ranking changes over 11 years by regressing time onto ranks and define the resulting regression slopes as ranking trends. In general, there are less evident ranking changes in the healthcare and pharmaceuticals industries compared with those in the
software and hardware industries. The healthcare and pharmaceuticals industries have more rapidly increasing trends than rapidly decreasing trends but less mildly increasing trends than mildly decreasing trends. These phenomena are opposite for the software and hardware industries, which have less rapidly increasing trends than rapidly decreasing trends but more mildly increasing trends than mildly decreasing trends. These observations suggest that the healthcare sector has no dominant directions, but many good companies have the potential to lead the market in the future. On the other hand, the IT sector is open for changes but more selectively. If a company cannot catch up with the latest technology or service, it will likely experience a rapid decrease in ranking.

However, the analysis of the above trends only identifies companies with significant changes and ignores rapid changes relative to the magnitude of changes. A company’s rank changing from 1 to 7 is more interesting than a company whose rank changes from 250 to 300. Therefore, the companies are sorted based on the ratio of ranking trend versus average ranking. A positive “trend/average” indicates that companies are increasing in ranks. Higher ratios suggest more obvious trends of increasing ranks in terms of its magnitude. Mastercard, Google, Apple, Western Digital, and Research in Motion are among the companies that have most rapid increase in ranks. In the hardware industry, two top ranked companies, Nokia and Motorola, display significant decreased rankings. In all other industries, we rarely find top performers with significant decreasing ranking. Previously, we stated that top performers have the smallest volatility, but we now also find that top companies in an industry do not necessarily retain unchanged ranks.

**TRENDS OF FINANCIAL INDICATORS IMPLY CHANGES IN RANKS**

In this section, our goal is to determine how different trends across indicators could imply changes in ranks for companies. Trends of TR, EV, NOPAT, normalized enterprise profit, and normalized future value for each company for the period 1999–2009 are approximated to constant (defined as C-shaped), linearly increasing (LI-shaped) or linearly decreasing (LD-shaped), and non-linear (NL-shaped) types. We link this trend analysis to changes in ranks by calculating some percentages. Some initial conjectures are that declining trends in financial performances may directly affect ranks negatively, while increasing trends in financial performance may not necessarily lead to increased ranks; increasing revenues are quite common even among decreasing rank companies, so if a company’s rank decreases, there must be some other factors than revenue alone. In addition, we summarize some shape–rank rules. For example, if any LD shape is detected, there is then an 80% chance that the ranking trend is non-increasing.
Shape Evaluation of Trends

We examine the data and find many companies that have missing values for the five financial indicators throughout the years 1999–2009. Since some of the companies are quite essential within their industry, the data of these companies should not be ignored or deleted. In addition, 11 years of annual data is insufficient to obtain accurate estimated confidence bands (CBs) for each company without accounting for the observed trends for other companies with similar business environments. This suggests using a statistical method that allows borrowing information in indicator patterns from other companies within the same industry. Ramsay and Silverman (1997) introduced the functional PCA (FPCA) technique, a way to analyze data so as to highlight various characteristics (called modalities in FPCA) and to identify important patterns and variations. Thus, we apply FPCA to our study of the demographics of change.

Because we are only interested in shapes, we first normalized the values of each indicator in each industry to a 0–1 scale. We then used the FPCA package in R to derive the estimated values, curves, and asymptotic simultaneous CBs for each company; detailed derivations are provided in the online supplementary information (SI). We deleted companies whose indicators in all years were close to zero. These companies account for about 10% of the total companies and are mainly in the category of over-the-counter stocks. Their changes can be ignored when compared with the industry average.

We propose a method to identify the trends of TR, EV, NOPAT, normalized enterprise profit, and normalized future value using (simultaneous) CBs. We approximated patterns across indicators to shapes as constant, linearly increasing or decreasing, and non-linear types. We defined a “constant” shape if there is a constant line that falls within the CBs. Similarly, we define a “linear” shape if there exists an increasing or decreasing line crossing within the CBs.

CBs play an important role in determining shapes in this article; therefore, accurate CB estimation is critical. For some companies and indicators, trends can be approximated to either constant or non-linear shapes depending on the width of bands. Thus, we include both the simultaneous CB calculated in the FPCA package and Krivobokova et al. (2010) derivation of CBs to ensure the shape evaluation process. Calculations are shown in the SI.

Figure 4 shows the CBs for Nokia’s TR, EV, and normalized future value from 1999 to 2009. The black circles are 11-year observations, the red triangular points are the fitted values, the smoothed curve is the fitted pattern using FPCA, the dotted line is the point-wise CB, and the dashed–dotted line is the simultaneous CB for shape evaluation. For example, Figure 4 shows that Nokia has a linearly increasing TR, linearly decreasing EV, and normalized future value from 1999 to 2009. Table 3 also lists other examples of the
shape evaluation results. Several companies, such as Nokia and Dell, have decreasing patterns across the five indicators; on the other hand, Western Digital has increasing patterns across the five indicators in 1999–2009. These shapes may imply rank changes for companies.

**Shapes Imply Changes in Ranking**

Do the shapes of a company’s TR, EV, NOPAT, normalized EV, and normalized future value shapes imply any rank changes for a company? Yes, they do. We examine the relationship between indicator shapes and changes in ranks of a company in the next several paragraphs.

By calculating percentages of companies with changes in ranks for each shape pattern, we find that if any LD shape of the five indicators is identified for a company, there is an 80% chance that this company has non-increasing ranking over time. If at least two LD shapes are detected, there is 100% likelihood that this company has non-increasing ranking. These suggest that declining trends in financial performances may directly affect ranks negatively. However, if any LI shape of the five indicators is detected for a company,

**TABLE 3** Examples of Shape Evaluation for Companies’ Five Indicators

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Total_Rev</th>
<th>EV</th>
<th>NOPAT</th>
<th>Normalized_EP</th>
<th>Normalized_FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLSE:NOK1V</td>
<td>LI</td>
<td>LD</td>
<td>C</td>
<td>C</td>
<td>LD</td>
</tr>
<tr>
<td>NYSE:HPQ</td>
<td>LI</td>
<td>LI</td>
<td>LI</td>
<td>C</td>
<td>NL</td>
</tr>
<tr>
<td>NasdaqGS:DELL</td>
<td>LI</td>
<td>LD</td>
<td>C</td>
<td>C</td>
<td>LD</td>
</tr>
<tr>
<td>NYSE:MSI</td>
<td>NL</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>NYSE:WDC</td>
<td>LI</td>
<td>LI</td>
<td>LI</td>
<td>LI</td>
<td>LI</td>
</tr>
<tr>
<td>NYSE:XXR</td>
<td>NL</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>TSX:RIM</td>
<td>NL</td>
<td>LI</td>
<td>LI</td>
<td>C</td>
<td>NL</td>
</tr>
<tr>
<td>NYSE:HRS</td>
<td>NL</td>
<td>LI</td>
<td>LI</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>NasdaqCM:OVRL</td>
<td>NL</td>
<td>LD</td>
<td>C</td>
<td>C</td>
<td>LD</td>
</tr>
</tbody>
</table>
there is only around a 40% chance that this company has non-decreasing ranking. Even if at least three out of five LI shapes are detected, it still cannot guarantee a non-decreasing ranking. These suggest that increasing trends in financial performance may not necessarily lead to increasing ranks. Particularly for the healthcare industry, a company with three or more LI shapes has only a 45% chance to be a non-decreasing ranking company, which means better financial performance is necessary but not sufficient for companies in the healthcare industry to increase their ranks.

We summarize shape–rank rules along with their corresponding percentages as follows.

1. If any LD shape is detected for a company in any industry, there is an 80% chance of non-increasing rank for that company.
2. If at least two LD shapes are detected for a company in any industry, there is a 100% chance of non-increasing rank for that company. Consider Nokia’s CBs in Figure 4; even though Nokia has LI for TR, two LDs for EV and normalized future value indicate that Nokia’s ranking is surely non-increasing; indeed, it is among the companies that have most rapid decreases in ranks.
3. If at least two LI shapes are detected for a company in the software industry, there is a 70% chance of non-decreasing rank for that company.
4. If at least three LI shapes are detected for a company in the software industry, there is an 80% chance of non-decreasing rank for that company.
5. If an LI shape is detected for NOPAT for a company in the hardware industry, there is a 95% chance of non-decreasing rank for that company. Google has LI shapes for TR, EV, and NOPAT; having three LI shapes in the hardware industry suggests a 90% chance of non-decreasing ranking. In fact, the percentage of non-decreasing ranking is even higher, as an LI shape in NOPAT alone suggests a 95% likelihood of non-decreasing ranking.
6. If an LI shape is detected for normalized future value for a company in the hardware industry, there is a 90% chance of non-decreasing rank for that company.

We know that each of the five indicators belongs to at least one of the shapes (C shape, LI shape, LD shape, and NL shape), with 1,024 (or $4^5$) possible combinations. However, only less than 100 types exist for any industry. We order the types according to their frequency and apply shape–rank rules. The hardware and pharmaceuticals industries have more non-increasing rank rules that apply; the software and healthcare industries have more non-decreasing rank rules that apply; and the software industry has an even higher certainty of non-decreasing ranks than the healthcare industry. Thus, considering an investment in an industry portfolio, one would prefer investing in the software industry to other industries during 1999–2009.
TEXT MINING TO EXPLAIN CHANGES IN RANKS

We discovered many patterns of the demographics of changes in terms of ranking and financial performance in previous sections. We know that for a company to succeed and maintain its success, it needs to achieve good performance in many dimensions—current profitability, stock market valuation, and future potential profitability; i.e., “happy families are all alike.” But as every unhappy family is unhappy in its own way, if a company fails, failure in any dimension or combination of dimensions could lead to a decreasing rank. In this section, we explore deeper explanations of companies’ behaviors.

Top performers’ success could come from frequent releases of new products, strong strategic alliances, visionary acquisition of competitors, fledgling innovations, or perhaps something yet to be realized. We detected companies with evident trends in rankings in the “Ranking Methodology and Mining” section, and we now seek a deeper explanation than could be provided by financial metrics alone. For example, could Nokia’s announcement about its reduced expansion plans be related to its decreasing ranking? Text mining helps us link companies’ activities with ranking changes.

Feldman and Sanger (2006) defined text mining in the first page of the book as “a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. In a manner analogous to data mining, text mining seeks to extract useful information from data sources through identification and exploration of interesting patterns.” Feldman and Sanger (2006) noted that many domains have begun leveraging text-mining capabilities in business analytics.

In this article, we mine a large amount of corporate/news release textual data to assess enterprises’ characteristics, detect changes, identify correlations, and cluster similar behaviors. The goal is to unearth explanations for improvements or degradations of financial performance. Declining revenues, fading profits, and diminished rankings should be explainable in terms of enterprise behaviors that are reported in press releases, business news reports, and related materials. The addition of these types of explanations to the multiple-indicator model of financial performance enables a much richer view of the demographics of change.

To this end, we developed an enterprise transformation taxonomy library using Northern Light’s text-mining platform on their research portal Single Point (Seuss, 2011). First, we defined normalized concepts for business event types, for instance, new product announcement, business expansion, legal issues, and so on. Then, we defined terms and criteria that trigger these normalized concepts; for example, the term “layoff” triggers “workforce decrease”; the term “job” and “employee” trigger “workforce decrease” only if it meets the confidence query (e.g., “job” near 10 words around “loss” or “losing”; “employee” near 10 words of “cut” and “fire”). It is useful to note that
development, refinement, and validation of this taxonomy and its associated rules involved significant effort.

Validation of our taxonomy was accomplished by assessing the extent to which the text-mining results agreed with Capital IQ’s analysis of the same dataset. We continued to refine our taxonomy until agreement was quite high. A key factor in achieving this match was the design of rules that enabled correct interpretation of terms. Mergers, acquisitions, and legal issues are pretty straightforward. Business alliances, for example, are more subtle. Company X may mention Company Y in a press release because Y is a new alliance partner or just because Y is a large customer—they may do this without using the word alliance or customer or synonyms.

The textual data was from Capital IQ companies’ news reports (Capital IQ, 2011), which includes 583 sources. Some source examples are individual company websites, Reuters, Dow Jones News Service, and Capital IQ transaction database. The number of documents for each business event type provides rich insights into a company’s characteristics. For instance, Apple has more new product announcement, and much more business expansion and legal issues than other companies, while it has fewer announcements for merger and acquisition and alliance. Cisco, on the other hand, has significantly more merger and acquisition. The highest percentage of strategic alliance suggests that IBM emphasizes collaboration and industry alliances. A high percentage of discontinued operations or downsizing reflects the downtimes in the recent histories of HP and Motorola. These results can be observed in Figure 5.

Event types reveal how companies are involved in various activities, but do they explain ranking changes or why a ranking change happened? By

![Companies' Event Types](image.png)

**FIGURE 5** Companies’ event types.
counting the number of news documents that contain a normalized term throughout years, we were able to track what happened in companies. For example, business expansion in our library is defined as opening new factories, plants, and offices and hiring new employees. Downsizing is defined as closing factories, plants, and offices and cutting workforces. Apple had a rapid increase in the number of business expansion news and had a peak around 2007–2009, while almost no news was reported about Apple’s downsizing. On the other hand, Nokia had an increasing number of downsizing news reports, especially in 2009. The trends of text-mining events for Apple and Nokia match with their rankings.

Nokia and Google provide examples to find the connections between text mining and changes in ranks. From Figure 6, we find there were not many business expansion or downsizing news announcements around 2001–2002 for Nokia, as Nokia maintained its leading position for several years. The net income decrease in 2001 represented a downsizing effect due to the market bubble more than a company internal effect for ranking change. Between 2003 and 2008, Nokia kept expanding according to the text-mining results. However, Nokia’s ranking decreased from the market leader range. This might appear contradictory until one looks at what other companies were doing. For example, Apple’s expansion was more vigorous. Hewlett-Packard increased its scale by merging with Compaq. Starting from the end of 2008, the text-mining results show that there was a decline in Nokia’s business expansion while an increase in downsizing and discontinued operation. This aligns with the change of ranks, as well as the effects of the economic recession.

Google Inc., on the other hand, as seen from Figure 7, shows a rapid increase from a startup company through year 1999–2009. The discontinued operations and falling business expansion around 2008–2009 is more a sign
of the recession than internal transformation. Google’s trends from text-mining events match its increasing ranking. For both Nokia and Google, the text-mining-based explanations are richer than just the financial metrics can provide.

Space does not allow discussing the text-mining results for the X companies for which we had extensive textual materials. In general, we were able to connect enterprise behaviors to ranking changes. We hasten to note, however, that we see the enterprise transformation taxonomy library as a work in progress. We focused on the IT companies for which we had extensive textual datasets. We imagine that our ongoing research on healthcare companies will involve considerable refinement to clarify the nuances of terminology differences between IT and healthcare enterprises.

CONCLUDING REMARKS

Table 4 summarizes the demographics of change and implications for investments. Investment in one of many good performers is risky, as it has both possibilities to be dominating or dominated. Investment in the dominant market leader is safer but with less potential return. It is the same argument with volatility of rankings. A company with large volatility has both possibilities of rapid increasing or decreasing trends in ranks. Companies with small volatilities are safer but with less potential return if it turns out to be a rank-increasing company. Trends of financial indicators suggest that industries with more increasing trends and fewer decreasing trends are preferred for investment as a portfolio. Rules of ranking trends suggest that industries with more non-decreasing ranking rules detected are preferred for investments as a portfolio. Thus combining these two results together, we conclude that software and healthcare industries were ideal for portfolio investments during 1999–2009. Some of the “NAs” noted for the
<table>
<thead>
<tr>
<th>TABLE 4 Insights for Investment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology hardware and equipment</strong></td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>Market leaders</td>
</tr>
<tr>
<td>Notes about market leader</td>
</tr>
<tr>
<td>Invest in a market leader</td>
</tr>
<tr>
<td>Volatility of rankings</td>
</tr>
<tr>
<td>Invest in a company</td>
</tr>
<tr>
<td>Consecutive rank changes</td>
</tr>
<tr>
<td>Investment strategy</td>
</tr>
<tr>
<td>Trends of financial indicators</td>
</tr>
<tr>
<td>Invest in an industry as a portfolio</td>
</tr>
<tr>
<td>Rules of ranking trends</td>
</tr>
<tr>
<td>Notes about rules of ranking trends</td>
</tr>
<tr>
<td>Invest in an industry as a portfolio</td>
</tr>
</tbody>
</table>
semiconductor industry are due to its industry scale being incomparable with other industries or an insufficient number of companies to estimate accurate probabilities.

The demographics of change described in this article rely on a data-driven ranking methodology, which combines the effects of a company’s TR, current profitability, stock market valuation, and future potential profitability. Text mining was applied to unearth non-financial explanations of ranking changes. In general, the data mining presented in the earlier sections of this article enabled understanding of “what” happened to companies in the industries studied. The text mining discussed in the last section provided insights into “why” these changes of rankings happened. When combined, the two methodologies provide a rich understanding of the demographics of change.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge the help of David Seuss and his Northern Light team in structuring their Single Point platform to support this research.

REFERENCES


