Presentation Outline

• Role of Textual Data in Quantitative Analysis

• Sources of Data

• Selecting the Unit of Analysis

• Developing Concept Categories

• Coding the Data
Role of Text in Quantitative Analysis

• Narrative statements provide insights into the cognitive processes of individuals

• Published narratives are unobtrusive sources of information that can be collected post hoc, and are generally less subject to retrospective sensemaking
Role of Text in Quantitative Analysis

• Published narratives may be the primary way in which management communicates with certain stakeholder groups

• Concepts discussed in the narratives can be counted and used as either independent or dependent variables
Sources of Textual Data

• Interviews
• Open Ended Surveys
• Internal Documents (reports, memos, etc.)
• Required SEC Filings (proxies, annual reports)
• News Reports
• Speeches
• Blogs
Worth and Words Study

Examined factors influencing the frequency with which compensation committees used different justifications in explaining their executive pay practices

Sample: 266 companies included in the S&P 500 in 1992

Narrative Source: Compensation Committee Reports in 1993 Proxy statements

12,500 sentences were content analyzed

Four types of Justifications: External Validation, Shareholder Alignment, Accounting Perf. Measures, Market Perf. Measures
Selecting a Unit of Analysis

• Word
• Sentence
• Paragraph
• Story
Developing Concept Categories

Three Strategies for Developing Concept Categories

• Apply categories that have been previously defined using a particular theoretical framework or conceptual schema

• Begin with no predefined structure, and allow the categories to emerge from the data

• Combination Approach - start with a partially defined set of categories, and modify, add, or delete categories as the analysis proceeds
Coding the Data - CATA

• Concepts are identified based on the use of particular words and phrases

• The Hardest Part: Identifying the words and phrases that capture the concepts you want and avoid the concepts you don’t

• Sources of words and phrases
  – brainstorming general uses
  – thesaurus
  – vocabulary lists generated by your software
  – fuzzy searches
Coding the Data - CATA

• ZyIndex
  – Uses Boolean logic to identify phrases
  – Identifies and presents all hits in context
  – Provides a vocabulary list with frequency counts

• Example: Identify all variations of return to stockholders

  return* w/in 3 stock*

  returns to stockholders
  return to stockholders
  return to stockholder
  return on stockholder’s
  return on stockholders
  return on stockholder
  return on stockowner
  return on stockowner’s
  return on stockowners
  return to stockowners
Coding the Data - CATA

• VBPro
  – Analyzes text and provides concept counts based upon dictionaries of terms which you develop for each concept
  – Free!!! (http://www.yoshikoder.org/vbpro.html)
  – Provides both excerpted text with the concept marked and quantitative data which is easily imported into statistical analysis packages
  – Quantitative output can be either occurrence (1,0) or count
  – Counts provided at either the sentence, paragraph or story level of analysis
<table>
<thead>
<tr>
<th>CONCEPT</th>
<th>FREQUENCY</th>
<th>CONCEPT</th>
<th>FREQUENCY</th>
<th>CONCEPT</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXTERNAL VALIDATION</td>
<td>631</td>
<td>SHAREHOLDER</td>
<td>1399</td>
<td>ACCOUNTING PERFORMANCE</td>
<td>1531</td>
</tr>
<tr>
<td>advisors</td>
<td>5</td>
<td>shareholder</td>
<td>418</td>
<td>roa</td>
<td>19</td>
</tr>
<tr>
<td>consultant</td>
<td>92</td>
<td>shareholders</td>
<td>455</td>
<td>roe</td>
<td>97</td>
</tr>
<tr>
<td>consultants</td>
<td>137</td>
<td>shareowner</td>
<td>15</td>
<td>eps</td>
<td>82</td>
</tr>
<tr>
<td>consultant's</td>
<td>5</td>
<td>shareholders</td>
<td>19</td>
<td>ros</td>
<td>7</td>
</tr>
<tr>
<td>consulting</td>
<td>73</td>
<td>stockholder</td>
<td>180</td>
<td>rotc</td>
<td>9</td>
</tr>
<tr>
<td>hewitt</td>
<td>40</td>
<td>stockholders</td>
<td>312</td>
<td>return on assets</td>
<td>42</td>
</tr>
<tr>
<td>hewitt's</td>
<td>3</td>
<td></td>
<td></td>
<td>return on equity</td>
<td>164</td>
</tr>
<tr>
<td>Perrin</td>
<td>12</td>
<td></td>
<td></td>
<td>return on investment</td>
<td>21</td>
</tr>
<tr>
<td>tpf&amp;c</td>
<td>2</td>
<td></td>
<td></td>
<td>return on capital</td>
<td>29</td>
</tr>
<tr>
<td>ors</td>
<td>2</td>
<td></td>
<td></td>
<td>return on sales</td>
<td>17</td>
</tr>
<tr>
<td>hay</td>
<td>6</td>
<td></td>
<td></td>
<td>return on total capital</td>
<td>7</td>
</tr>
<tr>
<td>experts</td>
<td>5</td>
<td></td>
<td></td>
<td>return on average equity</td>
<td>3</td>
</tr>
<tr>
<td>survey</td>
<td>118</td>
<td></td>
<td></td>
<td>net income</td>
<td>172</td>
</tr>
<tr>
<td>surveys</td>
<td>111</td>
<td></td>
<td></td>
<td>earnings</td>
<td>558</td>
</tr>
<tr>
<td>surveyed</td>
<td>20</td>
<td></td>
<td></td>
<td>cash flow</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>credit rating</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ratio</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ratios</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on sales/revenues</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on targeted equity</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on expense</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on common equity</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on shareholders' equity</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on average common equity</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on opening equity</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on average equity</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>returns on the company's assets</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on beginning - of - year equity</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>return on average shareholder equity</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>operating profits</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>operating profit</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>operating income</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>operating margins</td>
<td>4</td>
</tr>
</tbody>
</table>
Example Output for Shareholder Alignment

Alza uses a shareholder alignment justification 4 times
Coding the Data - CATA

• The concepts coded must also be validated

• Difficult to get 100% inclusion of intended concepts and 100% exclusion of false hits

• Validation done manually against a sample of the machine coded text
  – 80% or better accuracy on concept hits
  – No more than 5% of coded sentences false hits
  – Misses assumed to be randomly distributed
External Validation

The corporation, in establishing base salaries, levels of incidental and/or supplemental compensation, and incentive compensation programs for its officers and key executives, assesses periodic compensation surveys and published data covering the electrical/electronics industry and industry in general.

Shareholder Alignment

In this way, long-term compensation is aligned to increased owners shareholder value and long-term service to the corporation.

Accounting Measures of Performance

In reviewing Mr. Bossidy's performance, the committee focused primarily on the company's performance in 1992, which exceeded goals with very substantial improvement in operating earnings per share and cash flow.
Our Findings – Worth & Words

• The frequency with which these justifications are used is affected by the level of CEO pay, company performance, and the presence of concentrated and active shareholders.

• Elements of pay, especially bonus compensation, and company performance interact with shareholder characteristics to influence the frequency with which different justifications are used.

• Companies used justifications both to explain away high levels of pay, and to highlight when executive compensation is not paid.
Coding the Data – Celebrity

• 42,000+ articles from Business Week on 291 companies (1991-2004)

• Used Linguistic Word Count software – provides validated dictionaries of positive emotion and negative emotion words

• Coded article as positive if >60% affective language positive; negative if >60% affective language negative; otherwise neutral

• Calculated Janis-Fadner Coefficient of Imbalance for each year

• Coded celebrity = 1 if in top 25% in both number of articles and JF Coefficient in the same year