The Use of Content Analysis in Quantitative Research

Dr. Timothy G. Pollock

Smeal College of Business

The Pennsylvania State University

Academy of Management Annual Meeting

August 8, 2009

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 on stockholders**   **return on stockowner's**
  return to stockholders   **return on stockholder**   **return on stockowners**
  return to stockholder   **return to stockowner**   **return to stockowners**
  return on stockholder's   **return on stockowner**

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**

Frequency Counts of Dependent Variable Concept Words

CONCEPT	FREQUENCY	CONCEPT	FREQUENCY	CONCEPT	FREQUENCY
EXTERNAL VALIDATION	**631**	**SHAREHOLDER**	**1399**	**ACCOUNTING PERFORMANCE**	**1531**
advisors	5	shareholder	418	roa	19
consultant	92	shareholders	455	roe	97
consultants	137	shareowner	15	eps	82
consultant's	5	shareowners	19	ros	7
consulting	73	stockholder	180	rotc	9
hewitt	40	stockholders	312	return on assets	42
hewitt's	3			return on equity	164
Perrin	12	**ALIGN**	**1505**	return on investment	21
tpf&c	2			return on capital	29
ors	2	link	129	return on sales	17
hay	6	linking	39	return on total capital	7
experts	5	linked	82	return on average equity	3
survey	118	tie	42	net income	172
surveys	111	tying	12	earnings	558
surveyed	20	tied	109	cash flow	127
		align	523	credit rating	1
		aligned	26	ratio	36
		aligning	38	ratios	9
		interest	138	return on sales/revenues	1
		interests	367	return on targeted equity	1
				return on expense	1
				return on common equity	10
				return on shareholders ' equity	11
				return on average common equity	3
				return on opening equity	1
				return on average equity	3
				returns on the company's assets	1
				return on beginning - of - year equity	1
				return on average shareholder equity	3
				operating profits	7
				operating profit	19
				operating income	65
				operating margins	4

Coding the Data - CATA

Example Output for Shareholder Alignment

```
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  1,  1
#AZA ,  0,  1
#AZA ,  1,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  1,  3
#AZA ,  0,  0
#AZA ,  1,  1
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  1
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  0,  0
#AZA ,  1,  2
#AZA ,  0,  0
#AZA ,  0,  0
```

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**

Coding the Data - CATA

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 **SURVEY** **surveys** and published data covering the electrical/electronics industry and industry in general.

Shareholder Alignment

In this way, long-term compensation is **ALIGN** **tied** 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 **ACCTMEAS** **earnings per share** and **ACCTMEAS** **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