

# Commonality in Codes of Ethics

Margaret Forster  
University of Notre Dame

Tim Loughran  
University of Notre Dame

Bill McDonald  
University of Notre Dame

January 30, 2009

**Abstract:** We create a database of company codes of ethics from firms listed on the Standard & Poor's 500 Index and, separately, a sample of small firms. The SEC believes that "ethics codes do, and should, vary from company to company." Using textual analysis techniques, we measure the extent of commonality across the documents. We find substantial levels of common sentences used by the firms, including a few cases where the codes of ethics are essentially identical. We consider these results in the context of legal statements versus value statements. While legal writing often mandates duplication, we argue that value-based statements should be held to a higher standard of originality. Our evidence is consistent with isomorphic pressures on smaller firms to conform.

Key words: Code of Ethics; Textual Analysis; Plagiarism; Regulation.

We thank Robert Battalio, Patrick Murphy, Ann Tenbrunsel (editor), Carolyn Woo, Jeanie Xamis, a referee, and participants at the 2008 University of Notre Dame Ethical Dimensions in Business Conference for helpful comments. Yuta Kambe provided research assistance.

One-sentence bio: University of Notre Dame Associate Professional Specialist Margaret Forster received a Bachelor of Science from the Escola Politecnica of the Universidade de Sao Paulo (Brazil), and an MBA, an MS, and a Ph.D. from Cornell University (Ithaca). University of Notre Dame Finance Professor Tim Loughran received a Bachelor of Arts and a Bachelor of Science from the University of Illinois (Urbana), an MBA from Indiana University, and a Ph.D. from the University of Illinois (Urbana). University of Notre Dame Finance Professor Bill McDonald received a Bachelor of Science from the Central Missouri State University and an MBA and Ph.D. from Arizona State University.

# Commonality in Codes of Ethics

## 1. Introduction

Reacting to public outrage over perceptions of unethical behavior by US managers, Congress passed and President George W. Bush signed into law the Sarbanes-Oxley Act of 2002 (SOX). Section 406 of SOX requires public firms to disclose whether they have adopted a code of ethics for management. The Securities and Exchange Commission (SEC) defines a code of ethics as written standards that are designed to deter wrongdoing and to promote the ethical conduct of firm managers and directors. The SEC notes that “ethics codes do, and should, vary from company to company.” The major US stock exchanges, following SOX, have adopted similar rules mandating codes of business conduct and ethics, extending their applicability to all employees. In this article, we consider the degree and nature of commonality across ethics codes.<sup>1</sup>

Legal practitioners tend to consider strict replication of language as a best practice in legal drafting; precedent—wording that has survived the tests of jurisprudence—overrides concerns of originality. This may serve as a basis for companies to borrow from other codes when developing the legal components of an ethics code. It should be noted that until a code of ethics has actually been vetted in a court of law, it cannot be properly used as precedent.

---

<sup>1</sup> As a result of the exchanges’ regulations, most of these documents are now titled “Code of Business Conduct and Ethics.” Some are labeled simply “Code of Ethics,” while others have more creative labels (e.g., General Electric’s “The Spirit & the Letter”). Throughout the paper we use the terms “code of business conduct and ethics,” “code of ethics,” “ethics code,” or simply “code” interchangeably.

Yet lifting language from other places can also be thought of as plagiarism, or the actual stealing of ideas. For ethics codes, this concern arises when value-based statements are made, in contrast to the legal boilerplate statements.

In academia, where original thought and writing are taken as the coin of the realm, plagiarism is a cardinal sin. It is interesting that, at a university, law schools are the unit most likely to adopt policy statements on plagiarism, while lawyers in the real world will look to the value of legal precedent in drawing up company policies or statements. Thus, depending on context, commonality in writing can be viewed from quite different moral perspectives.

The purpose of this paper is to document the extent to which codes of ethics vary across companies. We use textual analysis methods to examine ethics codes for firms included in the Standard & Poor's 500 Index (S&P 500) in addition to a sample of 100 of the smallest Nasdaq and New York Stock Exchange (NYSE) firms in the bottom decile of NYSE market capitalization.

Previous literature has used textual analysis to examine various aspects of document composition. For example, Li (2008) examined the readability of annual reports using the Fog index. The Fog index is a readability statistic created using only the number of words and proportion of words with three syllables or more per sentence. Thus, the higher the Fog index, the less readable is the document. Li (2008) finds that firms with lower earnings have less readable 10-Ks.

Loughran, McDonald, and Yun (2008) find that firms using ethics-related phrases are more likely to be in the tobacco or alcohol industry, subject to a class action lawsuit, or have a low level of corporate governance. Lastly, the interaction between newspaper

columns and stock market prices is studied by Tetlock (2007). He finds that more pessimistic tone in a newspaper column leads to lower subsequent market returns.

To the best of our knowledge, we are the first paper to use textual analysis to examine commonality across codes of ethics. Our comparison of codes of ethics finds substantial levels of common sentences and content correlation, including some cases of virtually complete document duplication.

Commonalities among ethics codes of small firms are almost double those for firms listed in the S&P 500. The most frequently duplicated sentences are generally legalistic in tone. More than half our sample of publicly traded companies has codes of ethics correlations greater than 75%. In regression analysis, we find that firms of similar size, firms in the same industry, and firms listed on the same stock exchange have significantly higher pairwise correlations based on code content. Although these effects are statistically significant, their overall contribution in explaining the extent of correlation is insubstantial.

There may be several reasons as to why such mimicry occurs, including the need to conform to the external environment (i.e., regulation) and the role that competitive advantage plays in the process. Our empirical evidence that small firms are more likely to have common codes of ethics is consistent with institutional isomorphic pressures on managers to conform. Prior papers by DiMaggio and Powell (1983) and Carolan (2008) have argued that organizations can become isomorphic (that is, more similar with other institutions) due to government mandates.

External pressures to conform to governmental expectations are a possible explanation for the actions of small firms. Some managers desire similar codes of ethics

with other firms so that their company will not be seen as an outlier. Since small firms have fewer resources and less political power, they would be most likely to mimic the words of larger, more successful companies.

## **2. Literature on codes of ethics**

Whether a code of ethics is considered a legal document or a description of a company's value statements is key in determining whether one company is correct to copy or appropriate another's code in part or entirely. If a code of ethics is seen as a purely legal document, then precedent is valued, and copying another written code would be not only standard but also possibly a best practice. If a code of ethics is rather a document that sets forth a company's values or morals, however, it should arguably be a unique piece of work in that it establishes a code of behavior for its employees. Ironically, in this latter case, copying others' ethics codes would be unethical.

A normative view of a code of ethics is not a new concept. Schwartz (2002) argues that while a code may have some legal aspects, it goes beyond such a specific purpose as it sets behavioral guidelines. In fact, Schwartz argues that it should represent a set of universal moral standards, but should also allow for the fact that companies have unique characteristics and thus need different specific content. Murphy (2005), in a 2003 survey of 192 large US and non-US multinational companies, notes that many codes of ethics are "boilerplate," and suggests that companies should increase what he called "product differentiation."

One reason public companies may use "boilerplate" codes of ethics and a legal writing style could be that codes of ethics are mandated by the 2002 Sarbanes-Oxley Act,

the New York Stock Exchange, and the National Association of Securities Dealers, with slightly different content and disclosure requirements. While the SOX requirements are broader and less stringent, they are specific in that ethics codes are to be unique to each firm.

In its discussion of the final definition of “Code of Ethics,” the SEC notes

“We continue to believe that ethics codes do, and should, vary from company to company and that decisions as to the specific provisions of the code, compliance procedures and disciplinary measures for ethical breaches are best left for the company. ... we strongly encourage companies to adopt codes that are broader and more comprehensive than necessary to meet the new disclosure requirements.” (<http://sec.gov/rules/final/33-8177.htm>)

It would appear that the SEC itself is suggesting that codes of ethics should go beyond the letter of the law and describe the particular company’s code of behavior, a normative view. In fact, the SEC may be indirectly encouraging “boilerplate” codes by its broad definition of the code of ethics. The broader the code, the less specific they are to an individual company and thus the more likely they are to become generic or boilerplate in nature.

### **3. Legal and exchange requirements**

Section 406 of the Sarbanes-Oxley Act of 2002 requires all public companies to disclose whether they have adopted a code of ethics for principal executive, financial, and accounting officers. Firms that do not have such an ethics code must explain why not. SOX requirements are quite broad in nature and would allow a company to have different codes of ethics for officers, directors, and employees. Section 406 became effective March 3, 2003, and companies were required to comply for all fiscal years ended on or after July 15, 2003.

The NYSE adopted rules concerning codes of business conduct and ethics on November 4, 2003. Section 303A(10) of the NYSE's Listed Company Manual requires NYSE listed companies "to adopt and disclose a code of business conduct and ethics for directors, officers and employees." Unlike the SOX disclosure requirements, the NYSE rule is stricter and more specific. The NYSE not only requires a code of ethics for all company officials and employees to be reported on the company's website, but also is more specific in terms of the topics that must be addressed, such as statements regarding "compliance standards and procedures." Compliance was required by October 31, 2004.

The National Association of Securities Dealers Rule 4350(n) mandates that Nasdaq listed companies must adopt a code of conduct that "complies with the definition of 'code of ethics' as set forth in Section 406(c) of the Sarbanes-Oxley Act" and that applies to all directors, officers, and employees. The NASD rules are also stricter and more specific than the SOX requirements. For example, they require inclusion of an

“enforcement mechanism” and a process for compliance and dealing with violations. Compliance was required by May 4, 2004.<sup>2</sup>

#### **4. Data and methods**

##### *Data*

We consider documents typically labeled Codes of Business Conduct and Ethics, which we refer to generically as “codes.” In most cases, the codes are available in the Corporate Governance section of the Investor Relations webpage for a given firm. We create two sample groups.

First, we download codes for firms listed on December 31, 2007 in the Standard & Poor’s 500 Index. The S&P 500 consists of mostly US large-capitalization firms representing a broad cross-section of industries. The downloading process took place during the first half of 2008, so it is possible that some of the codes have been updated post-2007.

Of the 500 firms, there are three where we could not readily identify a code and did not receive any e-mail response after querying investor relations. Thus our S&P 500 sample actually consists of 497 documents.<sup>3</sup>

Second, we create a sample of small firms for comparison with the S&P 500 group. We want to identify firms that are perennially small, and not new firms or fallen angels (i.e., formerly large companies that have fallen on hard times). To achieve this, we create a list of firms that have appeared in the bottom market capitalization decile for

---

<sup>2</sup> Although we do not include in our research any companies listed on the American Stock Exchange (AMEX), as of December 2003 the AMEX listing requirements (Sec. 807) require firms to have codes of conduct and ethics. AMEX essentially requires firms to comply with SOX.

<sup>3</sup> The three firms with missing codes are Harrah’s Entertainment, Tiffany’s, and Celgene Corp.

each of the past five years. The bottom decile market capitalization cutoff for each of the five years is based on all NYSE stocks listed at year end on the Center for Research in Security Prices database for the years 2003-2007.<sup>4</sup> From this list we begin with the smallest-capitalization firm and include cases where the firm's code is readily accessible on its web site. The final small-firm sample consists of the first 100 firms' codes downloaded using this procedure.

Most of the downloaded codes are in pdf (portable document format), but some are text within webpages, some are simple text documents, and one (Microsoft Corporation) is a Microsoft Word document. In all cases, we convert the documents to a pure text format so that we can parse the content. Words that appear within a graphic are lost in this transformation process, but this occurs in only a few instances, and should have little impact on our analysis. All of our data is available at <http://www.nd.edu/~ethicsdb>.

### *Comparing code documents by sentence*

Once we have each code in text format, our initial comparison is based on sentences. We first eliminate from the document all characters except for alphabetic characters and punctuation. We then use a regular expression method to parse the code into sentences based on spacing and punctuation. A sentence must have at least 50 characters to be included in the comparison. For sentence comparisons we refer to as "exact," we compare sentences across the various codes based on the string variable stored for each sentence. Comparisons are case-insensitive.

---

<sup>4</sup> Because there are a relatively small number of large companies, financial researchers typically use NYSE cutoffs to avoid small deciles dominated by extremely small and illiquid Nasdaq firms. The NYSE is used only for cutoff values; firms listed on either the Nasdaq or NYSE are included in our sample.

We also want to be able to identify essentially identical sentences; for example, different firm names might occur within the sentence, or only a few words may be different. Two sentences that are approximate matches must not be an exact match and must be within +/-20% in length (based on the number of characters). For approximate matches, we first parse the sentence into words of at least two characters.

We then create a dictionary for each sentence, tabulating each unique word and the number of occurrences. For two separate sentences to be counted as an approximate match, the sum of the differences in counts across all words occurring at least once in both sentences must be less than 20% of the total number of words in the sentence and the sentences must not be exact matches. Essentially, sentences must have an 80% overlap to be considered approximate matches.

### *Document correlations*

We use correlations to measure the degree of commonality between code documents. The correlations are based on word count dictionaries created for each document.<sup>5</sup>

To create word count dictionaries for each code, we first create a master dictionary with counts for each word occurring in the entire corpus of 597 codes. We eliminate all non-alphabetic characters from each document, and then parse the document into “tokens” (i.e., collections of letters separated by spaces). Each token is compared with a master word list to determine whether the token is a word. If so, then the word is

---

<sup>5</sup> The natural language processing literature provides a variety of ways to measure “semantic similarity.” See, for example, Manning and Schütze (2003), section 8.5. Most measures provide very similar results for our particular application, so we use simple correlations for their ease of interpretation.

added to the master dictionary, and each subsequent occurrence increases the count for that word.<sup>6</sup>

We then parse each individual code of ethics into tokens, and compare these with the master dictionary. In natural language processing, it is typical to eliminate words with very high and very low frequencies in order to focus on content. This is because words with high frequency (like “the,” “and,” or “that,” also called stopwords) typically do not differentiate one document from another. At the other end of the spectrum, words that occur only rarely are unique to a document and typically do not contribute to comparative content.<sup>7</sup>

Thus for a word to be tabulated in a firm’s document dictionary: (1) The token must be a word occurring in the master dictionary; (2) the word must have a count of 50 or more in the master dictionary, i.e., must appear in about 8% of the codes; and (3) the word cannot be a stopword.<sup>8</sup>

Once we have the document dictionary for each code of ethics, we calculate the 177,906 pairwise correlations for the word counts of each document. For each comparison, a word must occur at least once in one of the documents to be included in the calculation of the pairwise correlation.

---

<sup>6</sup> We use release 4.0 of the 2of12inf word list available at <http://wordlist.sourceforge.net/12dicts-readme.html>. The word list includes 81,520 words and word inflections. Abbreviations, acronyms, and names are not included.

<sup>7</sup> The most common word frequency is one, a phenomenon referred to as *hapax legomena*. The distribution of word counts in documents is known to have an extremely long tail of rare occurrences. These words are typically excluded in comparisons of documents.

<sup>8</sup> We use a stopword list consisting of 278 words taken from Van Rijsbergen (1979).

## 5. Results

### *Sample description*

Table I reports the mean, median, 5<sup>th</sup> percentile, and 95<sup>th</sup> percentile summary statistics for the sample. Panel A reports the results for 497 firms listed on the S&P 500, and panel B reports them for the 100 small-firm sample. The small firms, listed overwhelmingly on Nasdaq, have an average market value of only \$15 million compared to the average market capitalization of \$26.5 billion for companies in the S&P 500. Market value is stock price multiplied by shares outstanding as of the last trading day in December 2007. S&P 500 firm codes of ethics have an average of 226.7 sentences compared to 117.2 sentences for small firms (only sentences with at least 50 characters are included in the analysis).

### *Sentence comparisons*

To gauge overlap, we compare both exact and approximate sentence matches across all code documents. In the total sample of 597 firms, we make 177,906 document comparisons, producing more than 2.5 billion sentence comparisons.

For exact sentence matches, there is substantial skewness. As reported in Table I, the average number of exact sentence matches is 36.8 for S&P firms while the median is only 4. Thus, while for one-half of the sample there are only four or fewer exact sentences in their codes of ethics overlapping other S&P 500 codes, for the other half of the sample, there is a significantly higher number of sentences that match other codes. So, for the average firm, about 37 sentences in their code of ethics are exactly repeated in

other S&P 500 codes. At the 95<sup>th</sup> percentile, there is a huge overlap (222 sentence matches).

Since many sentences in codes are only slightly different, we create an approximate match measure. For a sentence to be defined as an approximate match, it cannot be an exact match and the difference in word count must represent less than 20% of the total number of words in the base comparison group. On average, 5.9 approximate sentence matches appear in the S&P 500 codes, compared to 7.2 approximate sentence matches for small firms.

One example of an approximate match between sentences occurs in the codes of ethics for AFLAC Inc. and Stanley Works:

*From the AFLAC Inc. code of ethics:*

You may not take unfair advantage of anyone through manipulation, concealment, abuse of privileged information, misrepresentation of material facts or any other unfair-dealing practice.

*From the Stanley Works code of ethics:*

None should take unfair advantage of anyone through manipulation, concealment, abuse of privileged information, misrepresentation of material facts, or any other unfair-dealing practice.

Table I also reports an average code of ethics commonality measure, defined as the (number of exact sentence matches + number of approximate sentence matches) divided by the total number of sentences in the code of ethics. In other words, commonality describes what proportion of a firm's code sentences are essentially repeated somewhere in other codes.

The average commonality measure for small firms (16.6%) is almost double the measure for firms in the S&P 500 (8.4%). Thus, low-market value firms repeat many more of the sentences in other companies' codes of ethics. This high level of

commonality is consistent with the role of isomorphic pressure to conform being larger for small firms.

For the median S&P firm, only about 3% of the sentences are common with other firms. At the 95<sup>th</sup> percentile, though, the overlap is enormous; over 44% of codes of ethics sentences are exact or approximate matches to other codes' sentences. For small firms, the 95<sup>th</sup> percentile observation shows substantial commonality (over 65% common sentences).

Are there identical sentences that appear in many codes of ethics? Table II reports the most common sentences for both S&P 500 and small firms. For 71 of 597 sample firms (about 12% of the sample), we read: "Theft, carelessness and waste have a direct impact on the company's profitability." Not surprisingly, this exact sentence is given in the NYSE regulatory document in a discussion of topics a company's codes should address.

The most common sentences appear to be boilerplate legal phrases and not value-based statements. In over 5% of the firms, these words appear: "This applies equally to email, internal memos and formal reports."

### *Correlations*

In Table III, the highest 15 correlations between codes of ethics are reported. The highest correlation (0.9840) indicates little difference between the ethics codes of Duke Energy and Spectra Energy. This is not completely unexpected, as Spectra Energy was part of Duke Energy until January 1, 2007. The reported correlation of 0.9199 between Altria (formerly Philip Morris) and Kraft Foods would also be expected, as Kraft was spun off from Altria in March 2007.

The high degree of overlap in a number of unrelated companies, however, is somewhat troubling if one expects codes of ethics to vary among firms. A measured correlation of greater than 0.9 for two documents means that the documents are essentially identical. Documents with moderately high correlations will frequently have duplicated lists of bullet-pointed items.

For example, the Abercrombie & Fitch and Home Depot codes contain a list consisting of approximately 480 words that differ by less than a handful of words. Qualcomm and TSR Inc. are not in the same industry, yet their codes are essentially identical. The same applies for two small firms listed on Nasdaq: Pyramid Breweries, and EDEN Bioscience. EOG Resources has unexpectedly high correlation with four different firms (Torchmark Corp., eOn Communications, Robert Half International, and General Growth Properties).

The evidence we find indicates that not all codes of ethics vary across firms. There is clearly a small subset of firms that have virtually identical codes of ethics.

### *Regression results*

Next we examine the relation between codes of ethics and three independent variables: the market value difference between the two firms, an exchange dummy, and an industry dummy. The difference variable is the absolute difference between the two market values (as of the last trading day in 2007) divided by the average of the two market values. If both firms are listed on the same exchange as of December 31, 2007, the Same Exchange Dummy takes the value of one (zero otherwise). If both firms have the same Fama and French (1997) 48-industry classification, the Same Industry Dummy takes the value of one (zero otherwise).

Table IV reports the regression results. In all four regressions, there are 177,906  $[(597 * 596) / 2]$  codes of ethics correlation observations. The t-statistics are reported in parentheses. In column 1, the coefficient on Market Value Difference is negative with a t-statistic of -18.98. This implies that greater differences in market values between the two firms lower the correlations. The second regression reports that if the two firms are listed on the same exchange, the higher the code of ethics correlation. The regression in column 3 reports that if the two firms are in the same Fama and French industry, the correlations are significantly higher, all else equal.

The last column of Table IV reports the results when all three of the explanatory variables are in the regressions together. There is little change in the coefficient values when the three variables are added together. Although all three of the coefficients are highly significant and have the expected sign, the  $R^2$  value is quite low (0.003). Thus, while the large sample provides for statistical significance, we conclude that the three variables explain little of the variation in the code correlations across the sample.

### *Who is without sin?*

A total of 64 S&P 500 firms have a commonality value of zero (that is, they have no sentence overlap with any other S&P firm). Among the 64 are such household names as Caterpillar, Costco, Exxon Mobil, Ford, and General Motors. Are these firms to be admired for their uprightness and integrity?

First, note that the interpretation of zero commonality is ambiguous. Zero commonality could represent the integrity of these firms' code creation, but it also could simply reflect the fact that these codes did not merit copying by others.

Second, for codes with duplication, we cannot determine which firm is the original. Only some of the codes include adoption dates and we would require the entire time series of code modifications in order to verify which firm in a common pair was the original. Thus a firm with a high measure of commonality and a high number of sentence matches could be a firm with a wholly original document—a document that was so outstanding that it served as a template for many other firms.

In addition, because our sample is limited to a small fraction of listed companies, zero commonality does not preclude duplication with organizations outside the sample.

### *Value statements versus legal statements*

All the most common sentences appearing in Table II are arguably legal statements, where duplication of exact wording is a positive attribute. Value statements—sentences expressing the core beliefs and values of a firm—cannot be objectively identified, but play an important role in identifying cases where duplication is not appropriate.

We use the words listed in Schwartz's (2002) set of universal and moral standards to search through all approximate and exact sentence matches and identify those we might classify as value statements.<sup>9</sup> Although this is a subjective process, in Table V we list ten sentences that we would categorize as value statements compared to the perfunctory legal statements in Table II.

The sentences must have at least one pairwise match to be included in the table, but many have multiple matches. For example, if we enter the key words in the first example from Table V, excluding "this company's," in a Google search, we get 569 matches. This also shows that a broader sample would document even more occurrences of overlap.

As much as adopting legal templates is considered a best practice, we would argue that code content that does not delineate or prescribe appropriate behaviors but instead makes subjective statements of company values should be held to a higher standard of originality. To the extent a code of ethics is something more than a statement required by law or regulation, and to the extent firms use these codes as a way to communicate their core values, only original prose is credible. However, if firms wish to follow another's set of values, it should be acknowledged and not merely copied.

## **6. Areas for future research**

Our sample includes only 597 firms. Thousands of companies are currently listed on the three main US trading venues. If the sample were expanded to include all listed firms, one would get a better sense of code commonality across firms. Certainly we would expect more firms with excessive commonality to be identified if the sample were expanded.

---

<sup>9</sup> The list includes words such as caring, honesty, trust, fairness, and citizenship.

Another approach might be to compare the codes of ethics of foreign-based firms listed in the US and US-based firms. Foreign-based firms listed on the NYSE or Nasdaq must also have a code of ethics. If codes have a significant value-based substance, for cultural reasons, it is reasonable to expect considerable differences between foreign and US-based companies.

Future research might also examine the stock market announcement reaction when companies disclose waivers of the code for directors or executives. The NYSE, for example, specifically notes that any waivers of the code “may be made only by the board or a board committee and must be promptly disclosed to shareholders.” The exchange believes this requirement will discourage questionable waivers and allow better shareholder monitoring of a board.

Another avenue would be to link financial performance of firms with codes of ethics. Do firms with similar codes have better stock or operating performance than companies with original codes? Future research could also study the links between copying a code of ethics and other bad management behavior at a corporation. Does copying a code of ethics make it more likely that a firm will engage in other unethical behavior?

Copying a code could be a signal of poor corporate governance. Future research could study the correlation between copying a code of ethics and items such as: excessive CEO pay, firm lawsuits, backdating of executive stock options, accounting restatements, and other various standard academic measures of corporate governance.

Our research also raises a broader concern about regulation. Regulation requiring codes of ethics may be problematic in that it removes good firms’ ability to signal their

type by creating a code of ethics. Now that every firm has to supply one, and can copy them at low cost from other firms, the benefits of writing your own unique code of ethics are minimal.

If left to their own devices, maybe managers would be more forthcoming with their ethical language. An interesting topic could be to examine the time trend of content for codes of ethics. Was there an abrupt change in the content after the passage of SOX? Loughran and McDonald (2008) report that following the passage of SOX, firms sharply decreased the proportion of strong modal verbs (i.e., always, best, highest) while increasing the usage of weak modal verbs (i.e., could, may, possibly) in their annual reports.

## **7. Conclusions**

The SEC states that codes of ethics for publicly traded firms “do, and should, vary from company to company.” Yet we document that for S&P 500 and small-market value firms, there is substantial overlap in the codes of ethics as written. Textual analysis reveals that commonality (exact and approximate sentence matches) averages 8.4% for S&P 500 firms and 16.6% for a sample of small firms. Given that smaller firms may be more likely to adopt sentences from other codes of ethics due to isomorphic pressures, the higher commonality value for small firms is not unexpected. Five percent of our sample has more than 50% overlap in sentences. Smaller firms, in their quest to conform to the external regulatory environment, appear more likely to mimic the words of larger, more successful firms.

In numerous cases codes of ethics are almost identical between two firms, including cases that cannot be explained by recent spinoffs. Although firm size, exchange listing, and industry have a statistically significant impact on the level of document correlation, the economic effect in terms of R-squared is trivial.

To the extent a code of business conduct and ethics is a legal document, duplication or copying could be good practice if the document has been vetted in a court of law. Many of the duplicate sentences we identify are, in fact, legalistic statements. Yet in other cases the duplicated sentences appear more normative. We argue that the standard for originality in such cases should be higher.

Under what conditions would copying another firm's code be unethical? What if a firm aspires to be an ethical firm and copies a code of ethics in order to do so? A bigger ethical issue may be less related to the ethics of copying and more related to how genuine efforts to increase ethical behavior do not work.

By requiring codes of ethics, the government may not achieve their desired results because of the lack of a consideration of the behavioral implications of such regulations. It may be that regulation attempts to force an outcome (i.e., ethical behavior by publicly-traded firms) which is really only a poor proxy for the more impactful value discussions which accompany them. The government should encourage firms to copy the good ethical actions of foremost firms, not simply their words.

## References

Carolan, B.: 2008, 'Institutional pressures and isomorphic change: The case of New York City's Department of Education', *Education and Urban Society* **40**, 428-451.

DiMaggio, P. J., and W. W. Powell: 1983, 'The Iron Cage Revisited: Institutional; Isomorphism and Collective Rationality in Organizational Fields', *American Sociological Review* **48**, 147-160.

Fama, E. and K. French: 1997, 'Industry costs of equity', *Journal of Financial Economics* **43**, 153-193.

Li, F.: 2008, 'Annual report readability, current earnings, and earnings persistence', *Journal of Accounting and Economics* **45**, 221-247.

Loughran, T. and B. McDonald: 2008, 'When is a liability not a liability?', University of Notre Dame.

Loughran, T., B. McDonald, and H. Yun: 2008, 'A sheep in wolf's clothing: The use of ethics-related terms in 10-K reports', forthcoming *Journal of Business Ethics*.

Manning, C. D. and H. Schütze: 2003, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA.

Murphy, P.: 2005, 'Developing, communicating and promoting corporate ethics statements: A longitudinal analysis', *Journal of Business Ethics* **62**, 183-189.

Schwartz, M.: 2002, 'A code of ethics for corporate code of ethics', *Journal of Business Ethics* **41**, 27-43.

Securities and Exchange Commission, Disclosure required by sections 406 and 407 of the Sarbanes-Oxley Act of 2002, 17 CFR parts 228, 229, and 249, January 24, 2003.

Securities and Exchange Commission, NASD and NYSE rulemaking: Relating to corporate governance, Release No. 34-48745, November 4, 2003.

Securities and Exchange Commission, Disclosure Required by Sections 406 and 407 of the Sarbanes-Oxley Act of 2002, Release No. 33-8177, January 24, 2003.

Tetlock, P.: 2007, 'Giving content to investor sentiment: The role of media in the stock market', *Journal of Finance* **62**, 1139-1168.

Van Rijsbergen, C.J.: 1979, *Information Retrieval*, Butterworth-Heinemann, Newton, MA.

**Table I**  
**Code of Ethics Summary Statistics for S&P 500 and Small-Firm Samples**

Panel A: S&P 500 Firms (N = 497)

	Mean	Median	5 <sup>th</sup>	95 <sup>th</sup>
Average market value (in millions)	\$26,502	\$12,669	\$2,503	\$122,176
Percentage listed on Nasdaq	15.1%	0.0%	0.0%	100.0%
Average number of sentences in code	226.7	201.0	43.0	494.0
Average number of exact sentence matches	36.8	4.0	0.0	222.0
Average number of approximate sentence matches	5.9	3.0	0.0	23.0
Average code commonality	8.4%	3.0%	0.0%	44.3%

Panel B: Small Firms (N = 100)

	Mean	Median	5 <sup>th</sup>	95 <sup>th</sup>
Average market value (in millions)	\$15	\$16	\$4	\$28
Percentage listed on Nasdaq	99.0%	100.0%	100.0%	100.0%
Average number of sentences in code	117.2	96.5	28.5	260.0
Average number of exact sentence matches	11.3	3.0	0.0	61.5
Average number of approximate sentence matches	7.2	4.0	0.0	24.0
Average code commonality	16.6%	7.5%	0.0%	65.3%

The sample includes 597 firms (497 firms in the S&P 500 Index and 100 small non-S&P firms) that had codes of ethics in 2008. Market value and listed exchange are as of the last trading day of December 2007. Only sentences with at least 50 characters are included in the analysis. For a sentence to be defined as an approximate match, it cannot be an exact match and the difference in word count must represent less than 20% of the total number of words in the base comparison group. Code commonality is defined as (exact sentence matches + approximate sentence matches) / total number of sentences.

**Table II**  
**Most Common Sentences in Codes of Ethics for S&P 500 and Small Firms**

Total Count	Percentage of Sample	Sentence
71	11.9%	Theft, carelessness, and waste have a direct impact on the company's profitability.
33	5.5%	This applies equally to email, internal memos, and formal reports.
29	4.9%	In most situations there is shared responsibility.
29	4.9%	Any suspected incident of fraud or theft should be immediately reported for investigation.
28	4.7%	We seek competitive advantages through superior performance, never through unethical or illegal business practices.
28	4.7%	Confidential information includes all non-public information that might be of use to competitors or harmful to the company, or its customers, if disclosed.
27	4.5%	It may help to get others involved and discuss the problem.
26	4.4%	This will enable you to focus on the specific question you are faced with and the alternatives you have.
25	4.2%	State and local governments as well as foreign governments may have similar rules.
25	4.2%	Remember that it is your supervisor's responsibility to help solve problems.
25	4.2%	Stealing proprietary information, possessing trade secret information that was obtained without the owner's consent, or inducing such disclosures by past or present employees of other companies is prohibited.

The sample includes 597 firms (497 firms in the S&P 500 Index and 100 small non-S&P firms) that had codes of ethics in 2008. All sentences must be exact matches and have at least 50 characters to be included in the analysis.

**Table III**  
**Highest 15 Code of Ethics Correlations**

	Firm Name	Firm Name	Correlation
1.	Duke Energy	Spectra Energy	0.9840
2.	Qualcomm	TSR Inc.	0.9840
3.	Pyramid Breweries	EDEN Bioscience	0.9677
4.	EOG Resources	Torchmark Corp.	0.9330
5.	Wegener Corp.	MOD-PAC Corp.	0.9323
6.	UnitedHealth Group	Consulier Engineering	0.9310
7.	LSI	Novellus Systems	0.9297
8.	Robert Half Intl	eOn Communications	0.9294
9.	EOG Resources	eOn Communications	0.9291
10.	EOG Resources	Robert Half Intl	0.9268
11.	Neurobiological Tech	Schmitt Industries	0.9203
12.	Altria	Kraft Foods	0.9199
13.	Forest Labs	Pativ Corp.	0.9190
14.	EOG Resources	General Growth Properties	0.9186
15.	Cognizant Technology	Distributed Energy Systems	0.9166

A sentence must have at least 50 characters and 20 words to be included in the correlation analysis. All words must have at least 2 characters.

**Table IV**  
**Regressions with Code Correlations as the Dependent Variable**

	(1)	(2)	(3)	(4)
Intercept	0.53 (704.30)	0.50 (881.80)	0.51 (1383.36)	0.51 (492.51)
Market Value Difference	-0.01 (-18.98)			-0.01 (-11.94)
Same Exchange Dummy		0.02 (20.77)		0.01 (14.46)
Same Industry Dummy			0.01 (5.88)	0.01 (5.21)
R <sup>2</sup>	0.002	0.002	0.000	0.003
N	177,906	177,906	177,906	177,906

The sample includes 597 firms (497 firms in the S&P 500 Index and 100 small non-S&P firms) that had codes of ethics in 2008. Among the 597 firms, there are a total of 177,906 code correlations. The dependent variable is the paired code correlations for our sample firms. The average correlation is 0.513. The market value difference variable is the absolute difference between the two market values (as of the last trading day in 2007) divided by the average of the two market values. Market value is defined as share price multiplied by shares outstanding. If both firms are listed on the same exchange as of December 31, 2007, the same exchange dummy takes the value of one (zero otherwise). If both firms are in the same Fama and French (1997) industry, the same industry dummy takes the value of one (zero otherwise). T-statistics (in parentheses) use errors that are heteroskedasticity-robust.

**Table V**  
**A Sample of Ten Duplicate Sentences Categorized as Value Statements**

1.	Obeying the law, both in letter and in spirit, is the foundation on which this company's ethical standards are built.
2.	By acting with integrity we earn the trust of our customers, consumers, shareholders, coworkers, regulators, suppliers, and the communities where we live and work—those whose trust we must have to be successful.
3.	Due to the complexity of our business and the competitive realities of the marketplace, we must be clear about our corporate values and maintain a vigorous commitment to acting with integrity.
4.	The company is committed to promoting the values of honesty, integrity, and fairness in the conduct of its business and sustaining a work environment that fosters mutual respect, openness, and individual integrity.
5.	We are honest and treat people with respect, courtesy, fairness, and dignity, and we work hard to make XXXX a great place to work and a company that is respected for the quality of its people and products.
6.	The company encourages employees to exercise their rights of citizenship by voting, making personal political contributions as permitted by law if they wish to do so with their own funds, and being otherwise politically active in support of candidates, or parties, of the employee's own personal selection.
7.	In order to sustain that success for future generations, it is our responsibility to preserve and strengthen our core ethical beliefs and the values upon which our company was founded.
8.	There are no shortcut formulas or automatic answers to the choices we have to make in business today, however we should decide these questions in ways that are consistent with XXXX's values.
9.	We must work zealously, honestly, and in good faith with our hospital and physician partners on behalf of the millions of patients who entrust themselves to that partnership.
10.	We have a responsibility to ourselves—integrity is at the heart of who we are.

Sentences must have at least one pairwise exact or approximate match.