Journal of Banking & Finance 71 (2016) 50–61

Contents lists available at ScienceDirect

Journal of Banking & Finance

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

Stock returns and future tense language in 10-K reports

Rasa Karapandza

Department of Finance, Accounting and Taxation, European Business School (EBS), Gustav-Stresemann-Ring 3, 65189 Wiesbaden, Germany

ARTICLE INFO

Article history:
Received 3 November 2014
Accepted 23 April 2016
Available online 21 June 2016

JEL classification:
G11
G12
G14
D82
D83
M48
G3

Keywords:
Anomalies
Asset pricing
Text analysis, Information risk

ABSTRACT

This paper shows that firms talking less about the future in their annual reports generate positive abnormal returns of about 5% annually. I measure how much companies talk about the future in their annual 10-K reports by the frequency of the verbs *will*, *shall*, and *going to*. The evidence favors a risk-based interpretation: firms that use less future tense in their report offer higher returns since they are riskier. These results are consistent with finance theories stating that investors need to be rewarded for holding stocks of firms that put less information about the future in the marketplace.

1. Introduction

The finance literature has recognized that quantitative data in corporate reports may be important for pricing. For example, Haugen and Baker (1996) and Cohen et al. (2002) show that firms that are more profitable have unusually high average returns. Sloan (1996) finds that higher accruals predict lower stock returns. Fairfield et al. (2003) and Titman et al. (2004) show that more heavily invested firms exhibit lower stock returns. Ikenberry et al. (1995) show that firm returns are higher after stock repurchases, and Loughran and Ritter (1995) show that new issues underperform in the long run. The robustness of anomalous returns generated by strategies based on net stock issues, accruals, and momentum (Jegadeesh and Titman, 1993) has been demonstrated and confirmed by independent researchers (for a "dissection" of such anomalies, see Fama and French, 2008).

However, no links have been established between qualitative data contained in annual corporate reports and stock prices in the long run. I this paper I propose a new anomaly based on qualitative data contained in 10-K reports. I show that a trading strategy based on buying stocks of companies that talk less about the future in their 10-K annual reports generates significant positive abnormal returns of about 5% per year after controlling for common risk factors. I measure how much companies talk about the future in their annual 10-K reports by the frequency of the verbs *will*, *shall*, and *going to*. In terms of economic magnitude, these results are comparable to well-established anomalies as asset growth, profitability, size, book-to-market ratio, and accruals. These results are consistent with various finance theories (e.g., Easley and O'Hara, 2004) which predict that, given two otherwise identical stocks, the one for which there is less public information about the future will be riskier and thus will generate larger excess returns. Therefore, if talking about the future in 10-K reports does carry information then at least part of the variation in firms' excess returns should be explained by the frequency of auxiliary words *will*, *shall* and *going to* – which is what I find.

In support of these findings, I provide evidence that is based on a large sample of 10-K reports from 1993 to 2014. I demonstrate that the results are robust to the choice of both portfolio weighting methodology and the factor model of expected returns as well as to the exclusion or inclusion of microcap firms. In order to show that various alternative explanations are not consistent with my empirical findings, I also provide evidence that: (i) in a Fama–Macbeth (1973) framework, frequency of future tense sentences is neither subsumed by nor related to the previously reported proxies for

E-mail address: rasa.karapandza@ebs.edu
URL: http://www.rasakarapandza.com

http://dx.doi.org/10.1016/j.jbankfin.2016.04.025
0378-4266/

information disclosure, known anomalies, or other known measures of qualitative information contained in 10-K reports; (ii) frequency of future tense sentences is more than just a proxy for either industry-specific language or growth firms versus value firms; (iii) the frequency of future tense sentences is *unrelated* to level of debt, debt maturity, current profits, and other duration-related balance sheet items; and (iv) it is not possible to generate abnormal returns by sorting companies and forming portfolios based on any *other* word (or three-word combination) from the 100 most frequently used words in 10-K reports, so the frequency of future tense is probably not just a proxy for some general feature of English language that could predict returns just as successfully. I follow the convenction in the literature and focus my attention to 10-K annual reports since they are audited, unlike other types of reports that firms are required to fill with SEC. At the same time this also makes my results comparable to other papers in the field of textual analysis in finance (e.g. Loughran and McDonald, 2011).

This paper contributes to the literature that relates public information disclosure and returns. Linking information disclosure and returns is an old idea (see, e.g., the stream of papers by Easley and O'Hara). But to be more specific, I contribute to the literature on textual analysis and public disclosure in finance. In particular, this paper demonstrates that the qualitative information contained in 10-K reports is systematically related to long-term stock returns. I create a new measure for quantifying the qualitative information which is different than any of the dictionary methods proposed so far (cf. Tetlock, 2007; Tetlock et al., 2008; Loughran and McDonald, 2011) that measure the "sentiment" or "tone" of a body of text by counting words that have a negative or positive connotation. The technique proposed here is also different than naïve Bayes learning algorithms (Antweiler and Frank, 2004).

Also related is the work of Fang and Peress (2009), who show that higher returns are earned by stocks without than with media coverage. Despite the similarity of their findings to those reported here, authors show that their results are unrelated to the level of information disclosure and instead are rooted in Merton's (1987) "investor recognition hypothesis", which states that investors are not aware of all securities in informationally incomplete markets. As a consequence, stocks with lower investor recognition need to offer higher returns to compensate their holders for being imperfectly diversified.

Although the Securities and Exchange Commission encourages firms to discuss future plans in their 10-K reports—and offers firms the protection of "safe harbor" statements—they are not *required* to provide information beyond the events that shaped their previous fiscal year. According to the results presented in this paper, markets view firms that are more reticent (in their 10-K reports) about the future as being riskier. Awareness of this phenomenon could be an additional incentive for managers to change their corporate reporting in order to disclose more information about their firms' future.

2. Related literature

Linking information disclosure and returns is an idea that dates back at least to the stream of papers (starting with Easley and O'Hara, 2004) on information risk. In the model of Easley and O'Hara, private information is a priced risk factor and the distinct prediction is that information disclosure is associated with lower ex post returns. This notion has also been tested empirically. Fang and Peress (2009), for example, show that higher returns are earned by stocks without than with media coverage. Despite the similarity of their findings to those reported here, authors show that their results are unrelated to the *level* of information disclosure and instead are rooted in Merton's (1987) "investor recognition hypothesis", which states that investors are not aware of all securities in informationally incomplete markets. As a consequence, stocks with lower investor recognition need to offer higher returns to compensate their holders for being imperfectly diversified.

A growing body of finance and accounting literature has studied different techniques for mapping qualitative data into quantitative measures. The most prominent papers in this field—such as Tetlock (2007), Tetlock et al. (2008), and Loughran and McDonald (2011)—use external word lists to evaluate the "sentiment" or "tone" of a body of text. By looking at the frequencies of different word categories defined by standard dictionaries (such as the Harvard General Inquirer System; Stone and Hunt, 1963) or by creating their own, more appropriate dictionaries (Loughran and McDonald, 2011), authors seek to capture a text's sentiment. Tetlock et al. (2008) show that the fraction of negative words in company-specific news stories forecasts low earnings on the next day, and they also show that a small difference between the abnormal returns of firms with positive and negative news stories implies a simple trading strategy that could earn positive risk-adjusted profits. Specifically, at the close of each trading day, the authors form two equal-weighted portfolios based on the content of each firm's news stories during the prior trading day: they include all firms with positive (resp., negative) news stories on the prior trading day in the long (resp., short) portfolio. They hold both the long and short portfolios for one full trading day and then rebalance at the end of the next trading day. Although abnormal returns are detected, they are washed out in the long run by transaction costs entailed by the strategy's intensive daily trading.

Loughran and McDonald (2011) similarly capture the tone of 10-K reports. Instead of using the Harvard General Inquirer, these authors create a custom dictionary to define word categories that are more appropriate for the analysis of such reports. Although Loughran and McDonald demonstrate the importance of dictionary choice, they find no statistically significant relationship—between the tone of a firm's 10-K report text and the firm's long-run returns—that would warrant active trading by investors. Tetlock (2007) studies how the proportion of negative words in popular news columns on the stock market is incorporated into aggregate market valuation. A more tangentially related work is that of Garcia and Norly (2010), who measure a firm's geographic diversification by extracting state name counts from its 10-K reports.

An alternative and more complex approach for transforming qualitative into quantitative data relies on learning algorithms, such as naïve Bayes classification. This statistical technique is used to infer the tone of a document by detecting language patterns in the document being classified that are similar to patterns in a set of documents that have been pre-classified by humans into different tone categories. The most prominent work along these lines is that of Antweiler and Frank (2004), who train a naïve Bayes algorithm to assign a "bullish", "neutral", or "bearish" rating to more then 1.5 million messages posted at the Yahoo! Finance website about various companies. The authors find that these messages help predict market volatility but have a barely measurable effect on stock returns the day after the message board discussion took place.

This is the first paper to show that the qualitative information contained in 10-K reports has a systematic effect on long-term stock returns. Assessing the frequency of future-tense sentences is different than any of the dictionary methods proposed so far, and it is much simpler than naïve Bayes learning algorithms. Furthermore, the frequency of future tense is not related to measures that are designed to capture the tone of text.

Besides the relationship between public information disclosure and returns, the relationship between the trading based on private information and returns has been studied extensively. And these fields are related, since if more information is kept private it could be (not necessarily) that less information is released to the public. Easley et al. (2002) construct a proxy for PIN to capture information asymmetry between uninformed and informed traders. However, Duarte and Young (2009) show that liquidity effects unrelated to information disclosure explain the relation between PIN and a cross section of expected returns. Other measures—including idiosyncratic volatility (Ang et al., 2006) and dispersion in analyst coverage (Diether et al., 2002)—have been proposed as a proxy for asymmetric information or disagreement among investors. However, idiosyncratic volatility suffers from a problem similar to that observed with PIN. In particular, Ang et al. (2009) show that disseminated information does not (as previously hypothesized) explain the relationship between high idiosyncratic volatility and returns. On the other hand, the dispersion in analyst forecasts reported by Diether et al. (2002) is not related to risk. These authors show that stocks with higher dispersion in analysts' earnings forecasts earn significantly lower future returns than otherwise similar stocks. The authors suggest that the greater dispersion may be due to less public information. Yet their results clearly reject the notion that such dispersion is a good proxy for risk because the relation between dispersion and future returns is actually negative. In other words, their finding is not in line with that of Easley and O'Hara (2004), who predict that—all else being equal—a stock for which there is less public information will be riskier and will generate larger excess returns. In other words, Diether, Malloy, and Scherbina find not positive but rather negative excess returns for a portfolio of firms characterized by greater dispersion in analyst forecasts.

3. Frequency of future tense and auxiliary verbs *will*, *shall and going to*

How languages mark statements about the future is the defining characteristic when classifying languages in terms of weak versus strong future time reference (Thieroff, 2009; Chen, 2013). In fact, future time reference (FTR) was the main subject of the study conducted by the EUROTYP Theme Group on Tense and Aspect.[1] According to this study, English is a strong-FTR language because using future time markers is obligatory in nearly all circumstances (Chen, 2013).

Given that English is a strong-FTR language and that the future tense markers in English are *will*, *shall*, and *going to* (Szmrecsanyi, 2003), almost all sentences that talk about the future in English will be made using one of these three auxiliary verbs. In this study I exploit that characteristic of the English language. Observe that for German, it is not feasible to create an algorithm for identifying only sentences that talk about the future that easily; as noted by Chen (2013), German grammar does not require use of FTR markers in all statements. To demonstrate this rigidity of English language, Chen gives the following example. A German speaker predicting rain tomorrow can naturally do so in the present tense, saying: "Es regnet morgen" which literally translates to: "It rain tomorrow". In contrast, English would require the use of the future tense markers, "It *will* rain tomorrow" or "It is *going to* rain tomorrow".

Thus, it is clear from the literature in linguistics and economics that if a company wants to talk about the future in English language it will nearly always have to use one of the three auxiliary verbs *will*, *shall*, or *going to*. So, if talking less about the future does increase the informational risk, as predicted by various finance theories this increased risk should be compensated by higher return – and that is exactly what the data tells me.

4. Data, data treatment, and variables

4.1. Data and data treatment

I download all 10-K reports from the Securities and Exchange Commission (SEC) website using their ftp interface (ftp.sec.gov). The unique identifier for firms in the SEC database is Central Index Key (CIK) number. Returns come from the Center for Research in Security Prices (CRSP) database, where the unique identifier is permanent company number (PERMNO). In order to match CIKs with corresponding PERMNOs, I use the historical mapping provided by the Wharton Research Data Services. Because this is not a one-to-one mapping, my sample of 10-K reports is reduced from 157,143 to 98,046 to which I can link their corresponding returns in the CRSP database. Since I follow the same procedure as Loughran and McDonald (2011), I obtain similar results. These authors are not surprised by the extent of this reduction, given that many of the companies with missing PERMNOs are real estate, nonoperating, or asset-backed partnership/trusts that are not required to file with the SEC.

I follow Fama and French (2008) and include only ordinary common shares that were trading on the New York Stock Exchange (NYSE), the American Stock Exchange (Amex), and (after 1972) NASDAQ. Similarly, I follow Fama and French in excluding financial firms (Standard Industrial Classification codes between 6000 and 6999) and firms with negative book equity at $t-1$. The treatment of missing returns and of returns from delisted firms is as described by Beaver et al. (2007). Finally, I follow Fama and French in excluding firms until they have been in Compustat for two years; this reduces the survival bias inherent in the way Compustat adds firms to its sample (Banz and Breen, 1986). Following Loughran and McDonald (2011), I include only the first filing in a given year for each company. Applying these filters yields 46,078 10-K reports that are matched to their corresponding returns in the CRSP database. Similarly, Loughran and McDonald (2011) analyze 37,287 10-K reports. If I would restrict my observation period to the same one as in Loughran and McDonald (2011) my sample would have been 35,219. This difference in sample size is expected since they use slightly less restrictive filtering—for example, they disregard the Compustat bias and also retain financial companies—which are all standard procedures in the contemporary asset pricing literature (see Fama and French, 2008).

4.2. Definitions of variables

Timing is crucial for predicting returns and for testing whether some variable is an anomaly. Fama and French (1993, 2008) predict returns from July of year t through June of year $t+1$ using variables that correspond to the fiscal year ending in calendar year $t-1$. This ensures that any observed anomalous returns are persistent—in particular, either risk-related characteristics of expected returns or the result of behavioral biases that persist after the variable is observed (Fama and French, 2008). Yet even more important is that Fama and French have shown that to predict returns starting from a month earlier than July would introduce "look ahead" bias because the book equity values they use are those reported in the Moody's manuals. These manuals were published annually on 30 June of each year t to cover firms whose fiscal year ended in calendar year $t-1$.[2]

I follow the same convention and so predict returns of firms from July of year t through June of year $t+1$ using variables constructed from 10-K reports of firms whose fiscal year ends in calendar year $t-1$. Rules for submitting reports allow firms to submit

[1] The working group summarized their findings in an 846-page volume on tense and aspect (Dahl, 2000).

[2] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_historical_be_data.html.

their 10-Ks no more than 90 days after their fiscal year-end.[3] This requirement ensures that all information used to predict returns is from 3 to 27 months old, so that no look-ahead bias is introduced.

To make the timing clear, consider for example a 10-K report of a firm whose fiscal year ends on 31 December 2002. According to the SEC rules, this firm must submit its 10-K report by the end of March 2003. This report covers fiscal year 2002. and I use it to predict returns from July 2003 through June 2004. If a firm's fiscal year ends on 31 January 2003, then I use variables constructed from its 10-K report to predict returns from July 2004 through June 2005. Thus my timing structure complies with the convention used in the literature, and the SEC rule precludes any look-ahead bias.

Hence the following expression is used to predict returns from July of year t through June of year $t+1$:

$$\text{Frequency of future tense}_{it-1} = \begin{cases} \frac{1+\log(\text{Number of will, shall, going to}_{i,t-1})}{1+\log(\text{Number of Words}_{i,t-1})} & \text{if} \quad (\text{Number of will, shall, going to}_{i,t-1}) \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Here $\text{Number of will, shall, going to}_{i,t-1}$ is the number of appearances of the verbs *will*, *shall*, and *going to* in the 10-K report[4] of firm i whose fiscal year ends in calendar year $t-1$. Excluded from this number are appearances in sentences that are questions or in which the auxiliary verb is preceded by an article.[5] The variable $\text{Number of Words}_{i,t-1}$ is the total number of words counted in a 10-K report of firm i whose fiscal year ends in calendar year $t-1$ after XML tags and embedded binary data have been removed. In constructing this measure I closely follow Loughran and McDonald (2011).[6]

As a robustness check I use $Filesize_{it-1}$ (instead of $Number\ of\ Words_{it-1}$) as the normalization variable. The variable $Filesize_{it-1}$ is the size—in bytes, after removal of XML tags and embedded binary data—of the 10-K report filed by a firm i whose fiscal year ends in calendar year $t-1$.

[3] http://www.sec.gov/answers/form10k.htm.

[4] I examine the entire 10-K reports and not just particular subsections—for instance, MD&A (management discussion and analysis)—because Loughran and McDonald (2011) have shown that the informational content of MD&A and the complete 10-K do not differ.

[5] This filter is introduced because *will* is a noun (not an auxiliary verb) when it is preceded by an article.

[6] Loughran and McDonald (2011) calculate relative word frequencies as follows:

$$\text{Frequency of a Word k}_{jt-1} = \begin{cases} \frac{1+\log(\text{Count of Word } k_{jt-1})}{1+\log(\text{No. of Words}_{jt-1})} \log\left(\frac{\text{No. of Reports}}{\text{No. of Reports with Word } k}\right) & \text{if } (\text{Count of Word } k_{jt-1}) \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

I do exactly the same except for excluding the second log term, which Loughran and McDonald use to modify the relative impact if a word based on its commonality. As an example, they show that the word *loss* appears in most of the documents they examined but that *aggravates* appears in only a few of them. In this case, the term's relative scarcity increases its importance. That effect does not apply in this paper because the auxiliary verbs *will*, *shall*, and *going to* have the same meaning and thus are interchangeable (Szmrecsanyi, 2003). And since most of the 10-K reports contain all three auxiliary verbs, the results reported here would not be altered. The main results remain qualitatively the same also if I replace Loughran and McDonald's ratio of logarithms with a ratio based on raw word counts.

5. Main empirical findings

5.1. Raw returns

Fig. 1 plots the cumulative returns of the hedged Low–High Future Tense portfolio obtained from taking long and short positions in the Low and High Future Tense portfolios. The figure also plots the cumulative returns on three Fama–French factors. From this figure it is apparent that the spread of returns is present across the whole time period and is not driven by particular events. The hedged Low–High Future Tense portfolio outperforms all factors except the market factor over the sample and exhibits lower volatility. Actually, in terms of Sharpe ratio, Low–High Future Tense portfolio is the best performing one. Sharpe ratios are calculated from monthly returns and are as follows: $s(\text{Low} - \text{High Future Tense}) = 0.21$, $s(\text{Mkt} - \text{rf}) = 0.15$, $s(\text{Small} - \text{Big}) = 0.05$ and $s(\text{High} - \text{Low}) = 0.05$.

Table 1 shows averages and standard deviations of returns for the VW and equal-weighted (EW) High and Low Future Tense portfolios across different size groups and mimics Table I of Fama and French (2008). I define the size groups as in Fama and Franch (2008). Thus: Microcap stocks (Micro) are below the 20th percentile of the NYSE market cap at the end of June, Small stocks are between the 20th and 50th percentiles, and Big stocks are above the NYSE median. All portfolios except for those in the Micro size group include both Small and Big stocks. The Low Future Tense portfolio delivers higher returns than the High Future Tense portfolio across all size groups. These results are consistent with the theoretical predictions that firms that reveal less information about the future are riskier and thus should offer higher returns.

5.2. Risk-adjusted returns

In order to make sure that the results presented in previous section are not driven by known sources of risk in this paper my main results come from the methodology of Fama and French (1993) that dominates empirical research in asset pricing to this day (Cochrane, 2005). This methodology has been used in all of the previously cited studies that have successfully established a cross-sectional relationship between some firm specific characteristic and returns. To demonstrate the robustness of my results I also follow the non-parametric approach of Fama and French (2008).

In order to check whether a particular strategy would be profitable, a common approach is to form EW or VW portfolios by sorting stocks on the variable of interest covering period $t-1$ and then to focus on deciles, quintiles, terciles, or halves of that sort and on the hedge portfolio return obtained from taking long and short positions in the extremes for the period from July of year t till June of year $t+1$. Then excess returns of these portfolios are regressed on common risk factors, and the existence of a statistically significant intercept in the regression is considered to be proof of mispricing (relative to the risk factors employed) as in Fama and French (1993). Therefore, I study four different factor models: CAPM, the Fama–French (1993) three-factor model, the Carhart (1997) four-factor model, and a six-factor model. The six-factor extension of Carhart's model additionally includes the PIN factor of Easley et al. (2002) and the liquidity factor of Pastor and

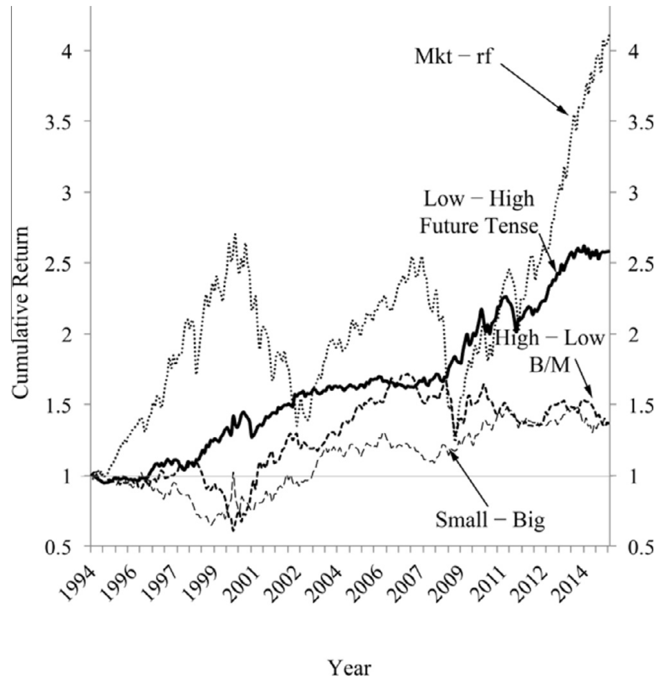

Fig. 1. Cumulative returns, July 1994 to June 2015. On 1 July of year *t* I remove firms that, at the end of March, had market capitalization smaller than the 20th percentile of the NYSE; the remaining firms are sorted according to *Frequency of Future* $Tense_{it-1}$. The High Future Tense portfolio consists of the top 50% of companies from the promises sort; the Low Fuure Tense portfolio consists of the bottom 50% of companies from this sort. The Low–High Future Tense portfolio takes a long position in the Low Future Tense portfolio and a short position in the High Future Tense portfolio. These portfolios are retained until the next July, when I rebalance them using the same rules. $\text{Frequency of Future Tense}_{it-1} = \begin{cases} \frac{1+\log(\text{Number of will, shall, going to}_{i,t-1})}{1+\log(\text{Number of Words}_{i,t-1})} & \text{if} \quad (\text{Number of will, shall, going to}_{i,t-1}) \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$ The variable *Number of will, shall, going to*$_{it-1}$ *is* the number of appearances of the verbs *will*, *shall*, and *going to* in a sentence that is not a question and that is not preceded by an article in a 10-K report of firm *i* whose fiscal year ends in calendar year $t-1$. *Number of Words*$_{it-1}$ is the total number of words counted in a 10-K report of firm *i* whose fiscal year ends in calendar year $t-1$ after XML tags and embedded binary data have been removed. Microcap stocks (Micro) are below the 20th percentile of NYSE market cap at the end of June. The figure reports cumulative returns of four value-weighted portfolios: the Low–High Future Tense portfolio; the Mkt − rf portfolio (i.e., the excess return value-weighted portfolio of all NYSE, Amex, and NASDAQ stocks from CRSP over the 1-month Treasury-bill rate); the High–Low B/M portfolio; and the Small − Big portfolio (i.e., the returns on the Fama–French factors).

Stambaugh (2003).[7] I also follow the non parametric approach of Fama and French (2008) where I measure abnormal returns of these sorted portfolios as returns net of the return on a matching portfolio formed on size and book-to-market equity (B/M) as proposed by Fama and French (2008). The matching portfolios are the 25 VW size-B/M portfolios of Fama and French (1993). The first approach of Fama and French (1993) has the advantage that it is easy to control for additional sources of risk like PIN or Liquidity, while the advantage of the non-parametric approach of Fama and French (2008) is that no assumption on the linearity between the risk and return is needed.

Table 2 shows abnormal returns as in Table II of Fama and French (2008). These abnormal returns are similar to the intercepts from the three-factor regression model of Fama and French (1993) estimated on the portfolio returns from the anomaly sorts which I report in Table 3.

The results indicate that both the Long-Short strategy and the strategy of going long in the stock of companies that use few future tense sentences in their 10-K reports generate positive and significant abnormal returns (at the 99% confidence level). This finding is robust in non-parametric setting as well as in all linear factor models, and the result is strongest in the six-factor model where it exceeds 7% annually.

The economic magnitude of these results is comparable to those (reported by Fama and French, 2008) of such well-established anomalies as those involving asset growth, profitability and accruals and is smaller only than the momentum anomaly. It is therefore probable that, in the long run, firms that use future tense less frequently are either riskier or mispriced.

These results are robust to alternative normalization measure, equal weighting (instead of value weighting) of the portfolio, and including microcaps—have a barely measurable effect on my findings. Furthermore, this claim is valid within all four of the asset pricing models listed in Table 3. The economic magnitude and statistical significance of my results remains virtually unchanged in response to these adjustments.

6. Risk versus mispricing

This paper shows that stocks of companies that use less future tense in their 10-K annual reports generate significant positive abnormal returns. This finding is consistent with the theories that companies that reveal less information about the future should offer larger returns due to increased risk and not due to mispricing. Fama and French (2008) discuss the difficulty of assessing how much the variation in expected returns reflects risk and how much it reflects mispricing.

In the spirit of Fama and French (1993) and as suggested by the literature summarized by Richardson et al. (2010), I provide evidence supporting the risk-based explanation via time-series regressions of zero-cost characteristic portfolios (deciles, quintiles, and terciles) on standard risk factors with the addition of a candidate mimicking factor portfolio returns: a base case Low–High Future Tense portfolio. The adjusted *R*-square in regressions of characteristic portfolios is increased across the spectrum of the sort, (ii) anomalous returns of the characteristic portfolios with low future tense frequency disappear, and (iii) the coefficient of the future tense-mimicking portfolio factor is always highly significant except within the second tercile and third quintile. The so-called GRS *F*-test of Gibbons et al. (1989) for the hypothesis that the four-factor model of Carhart (1997) explains the average returns of deciles, quintiles, and terciles generates the following respective *p*-values: 0.072, 0.081, and 0.035. Once the Low–High Future Tense factor is included, these *p*-values become (respectively) 0.218, 0.195, and 0.334. So far, the evidence has suggested that use less future tense in their reports generate positive abnormal returns; hence it would be appropriate to run the GRS *F*-test separately on portfolios made of firms that talk a less about the future and those that talk more about the future in their reports. In such a setting, a GRS test yields a *p*-value of 0.0068 for the hypothesis that the four-factor model of Carhart (1997) explains the average returns of the five deciles with lowest future tense usage. After inclusion of the Low–High Future Tense factor, the corresponding *p*-value is 0.199. A similar test on the top five deciles of the sort on future tense frequency produces insignificant results when applied to four-factor and five-factor models.

All the results presented in this section are in line with each other, and they all support the risk-based explanation of the anomalous returns found for firms that talk less about the future in their 10-K reports. That means that investors consider firms that

[7] Using instead the liquidity factor of Sadka (2006) yields no qualitative changes and so, to save space, I do not report the results of those regressions.

Table 1
Summary statistics: average returns and standard deviations for Value-Weighted (VW) and Equally-Weighted (EW) portfolios formed using sorts on *Frequency of Future Tense*$_{it-1}$ in 10-K corporate reports, July 1994 to June 2015.

	Total Number of 10-K Reports	*Mean Frequency of Future Tense*$_{it-1}$			VW average return				EW average return			
					Low future tense		High future tense		Low future tense		High future tense	
		Low	High	*t*-Stat	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
Market	46,078	0.39	0.54	320	0.61	4.8	0.24	4.5	0.77	5.8	0.58	6.2
Micro	24,469	0.38	0.54	250	0.62	6.4	0.47	7.15	0.94	6.3	0.74	7.1
Small	10,498	0.39	0.55	170	0.48	6.1	0.32	6.29	0.57	6.2	0.44	6.5
Big	11,111	0.40	0.56	180	0.55	4.5	0.18	4.4	0.59	5.1	0.32	5.2
All but micro	16,449	0.39	0.55	200	0.55	4.6	0.20	4.5	0.64	5.7	0.38	5.4

The table shows averages of monthly value-weight (VW) and equal-weight (EW) average stock returns and their standard deviations for Low and High Future Tense portfolios among all stocks (Market) and among Micro, Small, Big, and All but Micro stocks in percent per month. It also shows the total number of reports belonging to the each size groups as well as the summary statistics for the *Frequency of Future Tense*$_{it-1}$ variable. On 1 July of year *t* the firms within each size group are sorted according to their *Frequency Future Tense*$_{it-1}$. The High Future Tense portfolio consists of the top 50% of companies from the Future Tense sort; the Low Future Tense portfolio consists of the bottom 50% of companies from this sort. I keep these portfolios until the next July, when I rebalance them using the same rules.

$$\text{Frequency of Future Tense}_{it-1} = \begin{cases} \frac{1+\log(\text{Number of will, shall, going to}_{i,t-1})}{1+\log(\text{Number of Words}_{i,t-1})} & \text{if} \quad (\text{Number of will, shall, going to}_{i,t-1}) \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

The variable *Number of will, shall, going to*$_{it-1}$ *is* the number of appearances of the verbs *will*, *shall*, and *going to* in a sentence that is not a question and that is not preceded by an article in a 10-K report of firm *i* whose fiscal year ends in calendar year $t-1$. *Number of Words*$_{it-1}$ is the total number of words counted in a 10-K report of firm *i* whose fiscal year ends in calendar year $t-1$ after XML tags and embedded binary data have been removed. Microcap stocks (Micro) are below the 20th percentile of NYSE market cap at the end of June, Small stocks are between the 20th and 50th percentiles, and Big stocks are above the NYSE median. All but Micro combines Small and Big stocks.

Table 2
Main findings: average abnormal returns Value-Weighted (VW) portfolios formed using sorts on *Frequency of Future Tense*$_{it-1}$ in 10-K corporate reports of all firms excluding microcaps, July 1994 to June 2015.

	All firms			All firms		
	Low	High	Low–High	Low	High	Low–High
	Average value-weigh returns			*t*-Statistics for average value-weight returns		
Market	0.42	−0.09	0.51	2.34	−0.68	3.01
Micro	0.25	−0.03	0.28	1.92	−0.23	2.71
Small	0.28	−0.05	0.32	2.05	−0.36	2.89
Big	0.47	−0.11	0.58	2.67	−1.11	4.34

This table reports the abnormal VW portfolio returns in percent per month and their corresponding *t*-statistics. The monthly abnormal return on a stock is as in Fama and French (2008) measured net of the VW return on a matching Fama and French portfolio formed on size and B/M. On 1 July of year *t* I remove firms that, at the end of March, had market capitalization smaller than the 20th percentile of the NYSE; the remaining firms are sorted according to *Frequency of Future Tense*$_{it-1}$. The High Future Tense portfolio consists of the top 50% of companies from the future tense sort; the Low Future Tense portfolio consists of the bottom 50% of companies from this sort. The Low–High Future Tense portfolio takes a long position in the Low Future Tense portfolio and a short position in the High Future Tense portfolio. I keep these portfolios until the next July, when I rebalance them using the same rules.

talk less about the future more risky and investors need to be rewarded for holding their stocks.

7. Alternative explanations and their deficiencies

7.1. Known anomalies, other hypothesized information disclosure-related variables, and the tone of text

As alternative explanations for observed empirical patterns, in related research the usual suspects include already known anomalies; hence it is crucial to rule these out. It might also be the case that frequency of future tense merely reflects other, hypothetical measures of information disclosure. Finally, the tone of the text in 10-K reports could be the driving force behind my empirical findings. But non of these explanations hold.

Thus I observe the following firm characteristics, as defined in Fama and French (2008): market capitalization (MC), book-to-market equity (B/M), net stock issues (NS), positive accruals (Ac/B Pos), negative accruals (Ac/B Neg), momentum (Mom), and momentum (Mom).

I observe three variables related to information disclosure: the PIN of Easley et al. (2002); idiosyncratic volatility as in Ang et al. (2006); and media coverage as in Fang and Peress (2009). I also add Amihud's (2002) illiquidity ratio.[8] Analyst coverage comes from I/B/E/S summary files. I follow Fang and Peress (2009) and measure analyst coverage by counting the number of analysis making fiscal year-end forecasts. I also follow Fang and Peress (2009) in estimating the fraction of individual ownership for each stock and year as 1 minus the fraction of total institutional ownership, obtained by aggregating 13-f filings.

The variables that I use to capture the *tone* of analyzed text are the frequencies of words that are considered negative, positive, litigation related, or uncertainty related as defined in Loghrand and McDonald (2008). Finally, the firm's age (from CRSP) is used to control for the information uncertainty effect of Zhang (2006).

I use all these variables in the cross-sectional regression approach of Fama and MacBeth (1973). As Fama and French (2008) argue, Fama–MacBeth regressions are able to disentangle the simultaneous effects of multiple anomalies on returns. This ability comes at the cost of assuming a specific functional relationship between returns and anomaly variables—a constraint that is not present in the nonparametric Fama and French (1993) framework.

The regression setup I use is the same employed by Fama and French (2008). I estimate the cross-sectional regressions of monthly returns on explanatory variables that are updated (for the most part) annually. Thus, I explain the cross section of monthly returns from July of year *t* to June of year $t+1$ using anomaly variables observed in June of year *t* or earlier. The exceptions to this rule are momentum and media coverage, which by construction can be updated monthly and thus reflect the month prior to the corresponding return. Following Fang and Peress (2009), I adjust the standard errors to correct for autocorrelation by using the Newey–West (1987) procedure with one lag.

Table 4 reports the results of the cross-sectional regressions and results show that the only consistently statistically significant variables (at the 5% level) are Ac/B Pos, B/M, NS, and *Frequency of Future Tense*$_{it-1}$ and also that each of these variables has the expected sign. The momentum variable is significant at 10%.

[8] Using other liquidity proxies like bid-ask spread, dollar trading volume does not influence my results qualitatively.

Table 3
Main findings: average abnormal returns and factor loadings for Value-Weighted (VW) portfolios formed using sorts on *Frequency of Future Tense*$_{it-1}$ in 10-K corporate reports of all firms excluding microcaps, July 1994 to June 2015.

	Low future tense		High future tense		Low–High future tense		Whole sample	
	Coefficient	*p*-Value	Coefficient	*p*-Value	Coefficient	*p*-Value	Coefficient	*p*-Value
CAPM								
Intercept (α)	0.28	0.01	−0.04	0.33	0.32	0.01	0.04	0.42
Mkt − rf	0.96	0.00	0.95	0.00	0.00	0.88	0.98	0.00
R-square		0.90		0.95		0.00		0.96
Three-factor model								
Intercept (α)	0.34	0.00	−0.05	0.61	0.38	0.01	0.06	0.21
Mkt − rf	0.96	0.00	0.95	0.00	0.00	0.91	0.93	0.00
Small − Big	−0.08	0.01	−0.03	0.08	−0.04	0.29	−0.04	0.08
High–Low B/M	−0.14	0.01	−0.05	0.03	−0.05	0.19	−0.08	0.01
R-square		0.91		0.94		0.01		0.97
Four-factor model								
Intercept (α)	0.36	0.01	−0.03	0.81	0.39	0.01	0.07	0.34
Mkt − rf	0.96	0.00	0.96	0.00	0.00	0.88	0.94	0.00
Small − Big	−0.06	0.05	−0.06	0.09	−0.04	0.21	−0.03	0.12
High–Low B/M	−0.09	0.02	−0.07	0.04	−0.03	0.38	−0.08	0.01
Momentum	0.00	0.89	−0.02	0.07	0.04	0.33	−0.1	0.39
R-square		0.92		0.92		0.02		0.97
Six-factor model								
Intercept (α)	0.46	0.01	−0.03	0.53	0.49	0.02	0.06	0.44
Mkt − rf	0.95	0.00	0.96	0.00	−0.04	0.67	0.98	0.00
Small − Big	−0.11	0.06	−0.01	0.67	−0.08	0.14	−0.03	0.07
High–Low B/M	−0.11	0.07	−0.03	0.75	−0.09	0.16	−0.11	0.03
Momentum	0.00	0.78	−0.04	0.04	0.04	0.33	−0.09	0.56
PIN	−0.11	0.21	0.04	0.62	−0.24	0.24	0.12	0.33
Pastor-Stambaugh	−0.22	0.94	−0.11	0.97	−0.14	0.85	−0.10	0.81
R-square		0.91		0.95		0.04		0.94

This table reports the intercepts (Jensen's alphas) in percent per month and their corresponding *p*-values as well as the factor loadings and *R*-squares from monthly time-series regressions of three different VW portfolio excess returns on the three- and four-factor Fama–French–Carhart models as well as a six-factor model. The six-factor extension of Carhart's model additionally includes the PIN factor of Easley et al. (2002) and the liquidity factor of Pastor and Stambaugh (2003). On 1 July of year *t* I remove firms that, at the end of March, had market capitalization smaller than the 20th percentile of the NYSE; the remaining firms are sorted according to *Frequency of Future Tense*$_{it-1}$. The High Future Tense portfolio consists of the top 50% of companies from the future tense sort; the Low Future Tense portfolio consists of the bottom 50% of companies from this sort. The Low–High Future Tense portfolio takes a long position in the Low Future Tense portfolio and a short position in the High Future Tense portfolio. I keep these portfolios until the next July, when I rebalance them using the same rules. I regress excess returns of these portfolios on three factors (excess return on market, High–Low B/M and Small − Big), on four factors (adding the momentum factor to the previous three) and six factors (adding the PIN and Pastor and Stambaugh liquidity factors to the previous four).

The significance of the media coverage variable has not been tested previously within the Fama–MacBeth framework. The non-significance of market capitalization is a result of the sample restriction imposed by including PIN and media coverage variables. Including the PIN variable truncates the sample at 2001 because this variable is unavailable after that year. Peress and Fang (2009) construct their media coverage variable for all NYSE firms but for only some (randomly selected) NASDAQ firms; this, too, restricts the sample. Nonetheless, I include these two variables in the main cross-sectional study because the aim of this section is to check whether related anomalies subsume the explanatory power of frequency of future tense in the cross section of stock returns.

I also present results from cross-sectional regressions that *exclude* the media coverage variable. The only change induced by the resulting increase in sample size is that the PIN and MC variables become significant and Ac/B Pos becomes insignificant. Excluding the PIN variable has the effect of increasing my sample in the time dimension, which renders Amihud's illiquidity variable statistically significant at 10%. The significance and sign of *Frequency of Future Tense*$_{it-1}$ are robust to these changes of sample, so interpretation of the results remains unchanged. The variable NS shows impressive robustness to the changes in specification.

Altogether, these considerations clearly indicate that none of the many previously reported anomalies subsume the explanatory power of frequency of future tense on the cross section of stock returns.

It would be interesting to know if some firm characteristic, such as size or B/M, explains the variation in the frequency of future tense. In order to explore this idea, I follow the methodology of Fang and Peress's (2009) examination of the variables that determine media coverage. As the dependent variables in the Fama–MacBeth (1973) regression method I use *Frequency of Future Tense*$_{it}$. As explanatory variables I use all the variables listed in Table 4. However, none of the variables is statistically significant. Results are reported in Table 5.

7.2. Growth firms versus value firms

It is well known that companies with high ratio of book value to market value (B/M) of equity, known as *value* firms, have higher returns than low B/M companies, which are known as *growth* companies. It could be that my measure is but another proxy that captures this firm characteristic, since growth (resp. value) firms will naturally spend more (resp. less) time talking about the future. Yet even though I control for the B/M ratio in the standard Fama and French (1993) setting, it may be that—because of imperfections in the multifactor models—I am actually proxying for growth versus value firms but that the model is unable to distinguish between them. If so, then there should be some correlation between my measure of frequency of future tense and the B/M ratio; however, none exists. In Section 7.1 I show that: (i) the predictive power of my frequency of future tense variable is not subsumed by B/M in a Fama–MacBeth type of regression; and (ii) there is no correlation between B/M and the variable capturing the frequency of future tense as reported in Table 5.

Table 4
Average slopes, p-values, and R-square values from monthly cross-section regressions.

	All variables		Without media coverage		Without media coverage and PIN		Fama and French (2008) sample of firms	
	Average	p-Value	Average	p-Value	Average	p-Value	Average	p-Value
Intercept	1.76	0.02	1.57	0.06	−0.27	0.42	1.34^{**}	0.04
Market capitalization (MC)	0.06	0.68	−0.18	0.13	−0.16	0.08	-0.14^{*}	0.06
Book-to-market ratio (B/M)	0.18	0.04	0.26	0.06	0.32	0.06	0.22^{**}	0.03
Net stock issue (NS)	−0.8	0.04	−1.33	0.01	−1.89	0.003	-1.84^{***}	0.006
Positive accruals (Ac/B Pos)	−1.54	0.05	−0.32	0.23	−0.68	0.08	-0.4^{**}	0.02
Negative accruals (Ac/B Neg)	2.18	0.22	0.08	0.74	−0.07	0.73	−0.03	0.66
Momentum (Mom)	0.54	0.07	3.23	0.07	2.30	0.23	0.62^{*}	0.09
Frequency of Future Tense$_{t-1}$	−0.42	0.05	−0.28	0.04	−0.33	0.05		
Fraction of negative words	−10.40	0.43	32.23	0.47	44.22	0.56		
Fraction of positive words	−12.38	0.40	17.33	0.11	83.41	0.22		
Fraction of uncertainty related words	20.53	0.45	18.43	0.22	74.31	0.23		
Fraction of litigation related words	13.87	0.69	6.12	0.45	−22.71	0.42		
CRSP age	−0.03	0.54	0.02	0.18	0.00	0.56		
Idiosyncratic volatility	−0.08	0.41	−0.03	0.65	0.05	0.22		
Amihud's illiquidity ratio	−0.01	0.21	0.01	0.25	−0.001	0.08		
Analyst coverage	0.03	0.43	0.04	0.53	0.13	0.48		
Fraction of individual ownership	−0.07	0.72	−0.05	0.33	−0.14	0.37		
PIN	0.85	0.54	−2.35	0.04				
Media coverage	−0.02	0.34						
Adjusted R-square		0.16		0.06		0.06		0.03

The table shows average slopes and their p-values from monthly Fama–Macbeth cross-section regressions to predict stock returns. The variables used to predict returns for July of year t to June of year $t+1$ are: MC, natural log of market cap (in millions) in June of year t; B/M, natural log of the ratio of book equity (for the last fiscal year-end in $t-1$) divided by market equity (in December of $t-1$); NS, net stock issues or the change in the natural log of the split-adjusted share from $t-2$ to $t-1$ divided by book equity per split-adjusted share; Ac/B, accruals or the change in operating working capital per split-adjusted share from $t-2$ to $t-1$ divided by book equity per split-adjusted share in $t-1$ (Ac/B Neg is Ac/B for firms with negative accruals, and Ac/B Pos is Ac/B for firms with positive accruals); Mom, momentum or cumulative for month j, the cumulative return from month $j-12$ to $j-2$; Media Coverage, the number of newspaper articles about a stock that appeared in a previous month in New York Times, USA Today, Wall Street Journal and Washington Post; PIN, Probability of Informed Trading of Easley et al. (2002); Amihud's illiquidity ratio, Stock's absolute return divided by its daily dollar trading volume, scaled by 10^6; fraction of negative, positive, uncertainty and litigation words, the fraction of these word categories in a 10-K report where these categories are as defined in Loghran and McDonald (2008); CRSP age, rounded number of years that a firm was present in CRSP database prior to its inclusion in a portfolio; analyst coverage, natural log of 1 plus the number of analysts issuing earnings forecasts on the stock in the past year; fraction of individual ownership, percentage of the stock's shares outstanding owned by individuals; *Frequency of Future Tense*$_{it-1}$ as in Table 1. All variables are timed as in Fama and French (2008) or equivalent. The full data sample spans from July 1994 to June 2015. The regressions that include the PIN variable have sample truncated at 2001 because this variable is unavailable after that year. The regressions that include the media coverage of Fang and Peress (2009) are truncated in the time dimension in 2002 but also in the cross-section, since authors do not construct their media coverage variable for all NYSE firms but for only some (randomly selected) NASDAQ firms; this, too, restricts the sample.

7.3. Industry-specific language

Loughran and McDonald (2011) suggest that certain words—such as *cancer*, *capital*, and *mine*—are strongly linked to the language of specific industry segments and so could actually be proxies for industries. That is to say, it could be that managers in different industries write in different styles when constructing their 10-K reports. As a result, these authors suggest that such words be omitted from any analysis of company documents. Loughran and McDonald to not explicitly identify the auxiliary verbs *will*, *shall*, and *going to* as being industry specific. Nevertheless, in order to ensure that these verbs are not merely industry proxies, Table 6 reports the time-series means, medians, and standard deviations of the proportion of firms in Low and High Future Tense portfolios belonging to different industry sectors (as classified under the Global Industry Classification Standard developed by Standard & Poor's and MCSI Barra). The differences are small and are never statistically significant.

7.4. Sapir–Whorf hypothesis

The notion that language can influence thought has become known as the Sapir–Whorf hypothesis (SWH; Whorf, 1956), though the idea has a rich history in linguistics, philosophy, and psychology. In a recent study, Chen (2013) is the first to look at the effects of language structure—in particular, future time reference—on economic behavior. Chen shows that speakers of weak-FTR languages save more, retire with more wealth, smoke less, practice safer sex, and are less obese. Chen explains these findings in terms of his "linguistic savings" hypothesis (LSH): that being required to speak in a distinct way about future events leads speakers to take *fewer* future-oriented actions. According to this hypothesis, grammatically separating the future and the present leads speakers to dissociate the future from the present; doing so makes the future seem more distant and, since saving involves current costs for future rewards, makes it hard to save. In short, Chen posits that a habit of speech (feature of language) that dissociates future from present can lead speakers to devalue future rewards.[9]

The original LSH of Chen (2013) cannot explain my findings. In this study, all reports are written in English, which is a strong-FTR language. Nonetheless, different reports discuss the future to a greater or lesser extent. Might one create a form of the SWH that is similar to the LSH? If this were possible then we should observe a difference in debt rates, current profits, debt maturity, and cash held between the set of firms that talk a lot about the future in

[9] Relating behavior to language structure is not new. Summarizing this literature, MacLaury et al. (1992) note that languages around the world possess anywhere from two to eleven basic color terms. So, for example, whereas nearly all languages distinguish between black, white, and red, there are several languages that use a single term for yellow, green, and blue as well as many that have no basic word for purple, pink, orange, or gray. Brown and Lenneberg (1954) find that Zuñi speakers (a language that does not distinguish between yellow and orange) have trouble distinguishing orange and yellow colors. Similarly, Winawer et al. (2007) find that Russian speakers do better than English speakers in distinguishing shades of blues since Russian makes an obligatory distinction between light blue (goluboy) and dark blue (siniy). They also show that these differences are eliminated when subjects must simultaneously perform a verbal (but not a spatial) distractor task. Further implicating the importance of language in perception, Franklin et al. (2008) find that the difference is observed in adults but not in pre-linguistic infants of different nations.

Table 5
Determinants of Future Tense Frequency: average slopes, *p*-values, and *R*-square values from monthly cross-section regressions.

	All variables		Without media coverage		Without media coverage and PIN	
	Average	*p*-value	Average	*p*-value	Average	*p*-value
Market capitalization (MC)	−0.06	0.18	−0.02	0.37	−0.13	0.12
Book-to-market ratio (B/M)	0.43	0.24	0.11	0.56	0.21	0.37
Net stock issue (NS)	0.81	0.14	0.43	0.29	0.28	0.45
Positive accruals (Ac/B Pos)	3.21	0.23	−1.34	0.53	1.43	0.38
Negative accruals (Ac/B Neg)	1.11	0.34	0.31	0.44	−0.26	0.41
Momentum (Mom)	−2.53	0.22	−1.23	0.17	1.07	0.64
Fraction of negative words	2.12	0.26	0.04	0.78	0.72	0.31
Fraction of positive words	1.27	0.18	1.01	0.33	−0.88	0.32
Fraction of uncertainty related words	2.32	0.32	−0.56	0.41	1.31	0.41
Fraction of litigation related words	0.31	0.43	0.73	0.21	0.87	0.58
CRSP age	−3.21	0.13	−1.38	0.19	−1.18	0.15
Idiosyncratic volatility	0.14	0.28	0.02	0.46	−0.28	0.52
Amihud's illiquidity ratio	0.04	0.54	0.10	0.47	−0.01	0.44
Analyst coverage	−0.17	0.18	0.23	0.76	−0.82	0.28
Fraction of individual ownership	0.38	0.27	0.03	0.88	−0.11	0.73
PIN	1.23	0.61	0.68	0.33		
Media coverage	−1.11	0.38				
Adjusted *R*-square		0.07		0.05		0.03

The table shows average slopes and their *p*-values from monthly Fama–Macbeth cross-section regressions to explain *Frequency of Future Tense*. The variables used to explain *Frequency of Future Tense* are the same as in Table 4 and the *Frequency of Future Tense$_{it}$* is as in Table 1. The timing of variables is as in Fang and Peress (2009) or equivalent. The full data sample spans from July 1994 to June 2015. The regressions that include include the PIN variable have sample truncated at 2001 because this variable is unavailable after that year. The regressions that include the Media Coverage variable of Fang and Peress (2009) are truncated in the time dimension in 2002 but also in the cross-section, since authors do not construct their media coverage variable for all NYSE firms but for only some (randomly selected) NASDAQ firms; this, too, restricts the sample.

Table 6
Descriptive statistics: time-series averages of cross-sectional means, medians, and standard deviations of the proportion of sample firms belonging to different industry sectors.

Sector name (GSECTOR number)	Low future tense			High future tense		
	Mean	Median	SD	Mean	Median	SD
Energy (10)	0.06	0.07	0.02	0.07	0.07	0.01
Materials (15)	0.08	0.07	0.04	0.07	0.07	0.03
Industrials (20)	0.16	0.17	0.04	0.15	0.15	0.03
Consumer discretionary (25)	0.29	0.21	0.03	0.20	0.21	0.04
Consumer staples (30)	0.08	0.07	0.02	0.07	0.06	0.04
Health care (35)	0.11	0.12	0.03	0.13	0.14	0.03
Financials (40)	0.01	0.01	0.02	0.01	0.01	0.02
Information technology (45)	0.22	0.21	0.06	0.20	0.19	0.04
Telecommunication services (50)	0.02	0.01	0.01	0.02	0.02	0.02
Utilities (55)	0.04	0.04	0.04	0.06	0.07	0.03

This table reports the time-series means, medians and standard deviations (SD) of the proportion of firms in Low and High Future Tense portfolios belonging to different industry sectors as classified under the Global Industry Classification Standard (GICS®) developed by Standard & Poor's and MSCI Barra (Compustat variable GSECTOR). Low Future Tense and High Future Tense portfolios are formed as in Table 1. The sample covers the period from 1994 to 2015. For the proportion of firms belonging to any specific industry sector, the spread in means between Low Future Tense and High Future Tense firms is never significantly different from 0 (at the 90% level).

their 10-K reports and the set of those that talk less about the future. For instance, firms with higher-duration debt would talk more about the future and firms with higher current profits would talk more about the present. However, this is not the case: there is no correlation between these variables and the frequency of future tense in 10-K reports.

I follow the same approach as in Fang and Peress (2009), who examined variables that determine media coverage. As the dependent variable in a Fama–MacBeth (1973) style regression, I use the *Frequency of Future Tense$_{it-1}$*. As explanatory variables I use the following variables (Compustat names in parentheses): total assets (AT), cash (CH), debt due in year 1 (DD1), debt due in years 2 and 3 (DD2 + DD3), debt due in years 4 and 5 (DD4 + DD5), total debt in current liabilities (DLC), earnings per share (EPSFI), and gross profit (GP). However, none of the variables is statistically significant. Because some of these variables are highly correlated, it could be that their statistical insignificance is due to multicollinearity. Yet when only one of the correlated variables is retained, I still obtain the same—statistically insignificant—results. Since none of these variables is statistically significant, these results are omitted for the sake of brevity.

In sum, it is therefore unlikely that my results can be explained by a derivative of the Sapir–Whorf hypothesis that is similar in spirit to the linguistic savings hypothesis of Chen (2013).

7.5. Future tense versus other words

It might be argued that the frequency of future tense is simply a proxy for some general feature of the English language that could be captured just as well by some other combination of words. However, I can demonstrate that this is not the case. In particular, I show that no investment strategy based on any *other* of the 100 most frequently used words (or their possible word triplets) is able to pass a small subset of robustness checks that the strategy based on frequency of future tense does pass: using an alternative normalization measure, equal weighting (instead of value weighting) of the portfolio, and including microcaps.

7.5.1. Most Frequently Used Words in 10-K Reports and Returns

In Loughran and McDonald (2011) are listed the frequencies of all words in 10-K reports. I check whether it is possible to generate abnormal returns by forming portfolios based on frequencies of these 100 most frequent words *taken separately* (the choice of exactly 100 words is arbitrary and is constrained by the availability of computer resources). I make three remarks as follows about concentrating on the most frequent words.

(i) The verb *will* is ranked number 49, and the verb *shall* is ranked number 520. Because they consider *going to* as two separate words, this combination does not appear on their list as a verb. Therefore, since *will*, *shall*, and *going to* are quite frequent, I want to use words of comparable frequency.
(ii) Using words that appear in fewer than 50% of the reports for a given year (and this characterizes most of the words) would lead to an unreliable bottom portfolio. If, say, 70% of reports have a value of 0 for a given word's frequency, then how should I split the sample into two portfolios?
(iii) One downside of the high-frequency approach is that selecting the top 100 words could place the focus on words with little content (a.k.a. stop words). Loughran and McDonald provide a list of stop words,[10] according to which 42 of my 100 most frequent words are classified as "low content". Hence the analysis here might benefit if such words were removed from consideration. However, the stop words list marks *will* and *shall* as low-content words, which runs counter to this paper's conclusions (and would render invisible the very subject of my inquiry). To avoid this or any other error arising from a filtered word set, I take a conservative approach and do not remove the stop words. Indeed, I take as many words as possible given the constraints on computational resources. This rationale preserves my top 100 words. Although some of these words are low content, at least 60 of them are not.

Just as in the case of future tense (cf. Table 3), I sort companies according to log(*Number of Specific Word*$_{it-1}$)/log(*Number of Words*$_{it-1}$) and then, for each of the 100 most frequently used words, I form equal-weighted and value-weighted hedged portfolios. The results indicate that none of the 100 strategies is able to generate abnormal returns that are statistically significant (at the 95% confidence interval) in terms of the Fama–French three- or four-factor models of EW and VW hedged portfolios—much less in terms of any additional robustness tests.[11]

7.5.2. Word triplets made of the 100 most frequently used words in 10-K reports and returns

Our "future tense frequency" is the combined frequency of the verbs *will*, *shall*, and *going to*. I establish a benchmark for this robustness check by looking at other strategies based on a combined frequency of word triplets whose components are selected from the 100 most frequently used words. In this case the number of possible word triplets is "100 choose 3", or $\binom{100}{3} = 161,700$. I try all of them.

I start with the regression that served as the paper's main example (see Table 3). But now, instead of sorting by log(*Number of Future Tense*$_{it-1}$)/log(*Number of Words*$_{it-1}$), I sort firms by log(*word1*$_{it-1}$ + *word2*$_{it-1}$ + *word3*$_{it-1}$)/log(*Number of Words*$_{it-1}$) and then filter the results as follows.

Step 1: Of the 161,700 possible word combinations, only 104 yield alphas that are significant at the 95% confidence level for both VW and EW portfolios in both three- and four-factor models.
Step 2: Normalizing by file size instead of word count leaves 18 word combinations that yield consistently significant alphas.[12]
Step 3: Inclusion of microcaps leaves 10 significant alphas.
Step 4: Using CAPM as a model for expected returns renders these remaining 10 alphas insignificant.

None of the 161,700 strategies generates abnormal returns that are robust to this four-step procedure. In marked contrast, the future tense anomaly survives all four steps and more.

7.5.3. Future tense versus other words: concluding remarks

The results just presented demonstrate that—unlike strategies based on frequency of future tense—trading strategies based on the frequency of the 100 most often used words (or of any three-word combination) in 10-K reports do *not* yield robust abnormal returns. These tests indicate that the frequency of future tense captures a property of text in 10-K reports that is unlikely to be captured by the frequency of other words. But these tests show more. They show that the robustness checks described in Section 5.2—using an alternative normalization measure, equal weighting (instead of value weighting) of the portfolio, and including microcaps—are tests that actually make sense and are not just a form imposed by convention. These three robustness checks are standard and reported frequently in the literature, and it can hardly be asking too much for them to be satisfied—but in fact they are not. For each robustness test, approximately 5% of 161,800 trading strategies based on frequencies of 100 words (and their combinations) do, as expected, generate alphas that are statistically significant at 5%; however, no such strategy produces statistically significant alphas in all three of the tests. The sole exception is a trading strategy based on frequency of future tense, which passes all three of these robust checks (in addition to others presented in previous sections of the paper). Finally, these tests tell us that the ability of frequency of future tense to predict returns so robustly is a rare event in statistical terms, occurring only once in 161,800 chances (p-value equivalent of 0.000006).

In addition to supporting the hypothesis that future tense frequency is important for pricing, the approach used here—in which the detected anomaly is contrasted with other, similar trading strategies in a series of robustness tests—is novel.[13] This type of approach can also be used to check the uniqueness of other known or suspected anomalies. For example, Sloan (1996) shows that there is a negative relationship between accounting accruals and subsequent stock returns. This work launched a slew of papers dissecting the "accruals anomaly": testing whether operational accruals should also be taken into account, whether asset and liabilities accruals should be assessed separately, and so forth.[14] By testing whether

[10] The list of low-content words is available at the website of Bill McDonald (http://www.nd.edu/~mcdonald/Word_Lists.html).

[11] Note that, for each combination of weighting methodology and factor model of expected returns, 5–9 different words based strategy yield statistically significant abnormal returns at 95% significance level. We do expect to find some significant Jensen's alphas by pure chance. But note that there is not a single word that would serve as a basis for a strategy that is robustly significant in all 4 regressions as the future tense strategy.

[12] Because word count and net report size are highly correlated, this result may seem unimportant; however, some word combinations fail to survive even this

[13] Numerous technical trading rules have previously been compared on a large scale, but not in the context of asset pricing or risk-adjusted investment strategies (Hsu et al., 2010; Sullivan et al., 1999) and were not faced with a series of robustness tests. Karapandza and Marin (2011) use supercomputers to track the risk-adjusted performance of many randomly selected strategies, but this research assesses market efficiency in the aggregate and does not focus on the performance of any single trading strategy.

combinations of accounting variables similar to those used by Sloan (1996) yielded abnormal returns, we could have learned more from accruals in 1996 about the multiplicity of these anomalies.

8. Discussion

8.1. On data mining and origin of the counting idea

In this paper I test 161,801 investment strategies. I test: 100 strategies based on the occurrence of the 100 most frequent words within the 10-K reports, as documented by Loughran and McDonald (2011); 161,700 strategies based on all the possible triplets of those words; and the future tense strategy based on frequencies of the verbs *will*, *shall*, and *going to*. Of these 161,801 strategies, only one survives the full battery of robustness checks: the strategy based on frequency of the future tense. One reasonable doubt could be that this strategy is merely the product of an extensive data-mining exercise. Yet this is not the case because, in the list of words provided by Loughran and McDonald (2011), *shall* is ranked number 520 and *going to* does not even appear (because it is treated as two separate words). Therefore, our proposed strategy could not possibly arise from any data-mining technique based on the frequency of words reported to appear in 10-K reports.

8.2. Policy implications

There is a fundamental disagreement between managers and shareholders with regard to the firm's disclosure policy (Almazan et al., 2008). Managers may be reluctant to disclose information because, for example, doing so may reflect prior poor managerial decisions and thereby bring their performance into question. Of course, it is for this reason that inducing managers to disclose information is essential (Almazan et al., 2008). Dissatisfied with the disclosures that firms were providing about future plans in their annual 10-K reports, the Securities and Exchange Commission (SEC, 1980) issued a revised requirement that granted protection under "safe harbor" rules and that is still in force. The rules stipulate that a company may not be held liable under federal securities laws for financial projections and other forward-looking statements if they are accompanied by meaningful cautionary statements. Despite these safe harbor provisions, companies avoid disseminating forward-looking information because of uncertainty about judicial interpretations and fear of state court litigation (Grundfest and Perino, 1997). This paper might give managers additional incentives to disclose more information, since the findings indicate that markets view firms that talk less about the future as being more risky.

9. Conclusion

Using a large sample of more than 46,000 reports, I show that investment strategies based on buying stocks of companies that talk less about the future in their 10-K annual reports generate significant positive abnormal returns of more than 5% per year. These anomalous returns survive a set of standard robustness checks. As a measure of how much companies talk about the future I use the frequency of the verbs *will*, *shall*, and *going to*. The evidence presented here favors an interpretation under which a firm that talk less about the future provide less information to the marketplace, so investors must be compensated for holding its inherently riskier stock. Several alternative explanations for the observed anomalous returns are rejected by the data. The reported results hold in both the Fama and French (1993) and Fama and Macbeth (1973) frameworks.

My results are consistent with theoretical work predicting that stocks with less public information about the future will, on average, offer larger returns. Yet my results are neither explained nor subsumed by previously hypothesized proxies for information disclosure.

Several recent papers have mapped the qualitative data of annual reports to quantitative data, seeking thereby to learn how the market interprets such information; these papers have concentrated mainly on studying whether the overall tone of the report text is negative versus positive or, equivalently, bearish versus bullish. I take a different approach and compare firms that talk less to those that talk more about the future. Although the measure I employ is different, my goal is the same. As shown by Loughran and McDonald (2011), the analysis of text can contribute to our modest ability to understand the impact of information on stock returns. To paraphrase these authors: even if the text's tone (or, in this paper, its frequency of future tense) is not itself causally related to returns, it could be an efficient way for analysts and investors to capture other sources of information. Similarly, I do not claim that my crude measure of language subsumes or transcends traditional accounting measures of a firm's fundamentals. Rather, I investigate whether the frequency of future tenses can improve our ability to explain the cross section of stock returns. Frequency of future tense is a noisy measure of qualitative information, and to the extent that it errs it does so by *understating* the true importance of qualitative information (cf. Tetlock et al., 2008).

No matter what explanation is offered for the phenomenon that I document—and regardless of whether the information captured by future tense frequency is directly processed by the markets or instead represents a proxy for some other, unknown source—an awareness of this finding should affect how corporate reports are constructed.

Acknowledgments

I am especially grateful to Timothy Loughran and Bill McDonald for detailed comments on the previous draft. I also thank to Yakov Amihud, Miloš Božović, Martin Dufwenberg, Diego Garcia, Benjamin Golez, Reint Gropp, André Güttler, Boyan Jovanovic, Gueorgui I. Kolev, Jose Marin, Mlađan Mrđan, Joel Peress, Paul Tetlock, Hans Joachim Voth, Michael Wolf, and members of equity investment group of Barclays Global Investors as well as to seminar participants at University Pompeu Fabra, University of British Columbia, Catolic University at Lisbon, IE Business School, and European Business School for helpful comments and suggestions. I am also grateful to anonymous referees on their fair and constructive comments. As usual, all the remaining errors are mine.

References

Almazan, Andres, Banerji, Sanjay, de Motta, Adolfo, 2008. Attracting attention: cheap managerial talk and costly market monitoring. The Journal of Finance 63, 1399–1436.
Amihud, Yakov, 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.
Ang, Andrew, Hodrick, Robert J., Xing, Yuhang, Zhang, Xiaoyan, 2006. The cross-section of volatility and expected returns. The Journal of Finance 61, 259–299.
Ang, Andrew, Hodrick, Robert J., Xing, Yuhang, Zhang, Xiaoyan, 2009. High idiosyncratic volatility and low returns: international and further U.S. evidence. Journal of Financial Economics 91, 1–23.
Antweiler, Werner, Frank, Murray Z., 2004. Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance 59, 1259–1294.
Banz, Rolf W., Breen, William J., 1986. Sample-dependent results using accounting and market data: some evidence. The Journal of Finance 41, 779–793.

[14] Richardson et al. (2005) is a useful review of the literature dealing with accruals.

Beaver, William, McNichols, Maureen, Price, Richard, 2007. Delisting returns and their effect on accounting-based, market anomalies. Journal of Accounting and Economics 43, 341–368.
Brown, Roger H., Lenneberg, Eric H., 1954. A study in language and cognition. The Journal of Abnormal and Social Psychology 49, 454–462.
Carhart, Mark M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.
Chen, Keith M., 2013. The effect of language on economic behavior: evidence from savings rates, health behaviors, and retirement assets. American Economic Review 103, 690–731.
Cochrane, John H., 2005. Asset Pricing, 2nd ed. Princeton University Press, Princeton, NJ.
Cohen, Randolph B., Gompers, Paul A., Vuolteenaho, Tuomo, 2002. Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. Journal of Financial Economics 66, 409–462.
Dahl, Östen (Ed.), 2000. Tense and Aspect in the Languages of Europe. Mouton de Gruyter, Berlin.
Diether, Karl B., Malloy, Christopher J., Scherbina, Anna, 2002. Differences of opinion and the cross section of stock returns. The Journal of Finance 57, 2113–2141.
Duarte, Jefferson, Young, Lance, 2009. Why is PIN priced? Journal of Financial Economics 91, 119–138.
Easley, David, O'Hara, Maureen, 2004. Information and the cost of capital. The Journal of Finance 59, 1553–1583.
Easley, David, Hvidkjaer, Soeren, O'Hara, Maureen, 2002. Is information risk a determinant of asset returns? The Journal of Finance 47, 2185–2221.
Fairfield, Patricia M., Whisenant, Scott, Yohn, Terry L., 2003. Accrued earnings and growth: implications for future profitability and market mispricing. The Accounting Review 78, 353–371.
Fama, Eugen F., French, Kenneth R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.
Fama, Eugen F., French, Kenneth R., 2008. Dissecting anomalies. The Journal of Finance 63, 1653–1678.
Fama, Eugen F., MacBeth, James, 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.
Fang, Lily, Peress, Joel, 2009. Media coverage and the cross-section of stock returns. The Journal of Finance 84, 2023–2052.
Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T., 2008. Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. Proceedings of the National Academy of Sciences 105, 3221–3225.
Garcia, Diego, Norly, Oyvind, 2010. Geographic dispersion and stock returns. Working Paper, Kenan-Flagler Business School The University of North Carolina at Chapel Hill.
Gibbons, Michael, Ross, Stephen, Shanken, Jay, 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.
Grundfest, Joseph A., Perino, Michael A., 1997. Testimony before the Subcommittee on Securities of the Committee on Banking, Housing, and Urban Affairs of the United States Senate, <http://securities.stanford.edu/research/articles/19970723sen1.html>.
Haugen, Robert A., Baker, Nardin L., 1996. Commonality in the determinants of expected stock returns. Journal of Financial Economics 41, 401–439.
Ikenberry, David, Lakonishok, Josef, Vermaelen, Theo, 1995. Market underreaction to open market share repurchases. Journal of Financial Economics 39, 181–208.
Jegadeesh, Narasimhan, Titman, Sheridan, 1993. Returns to buying winners and selling losers: implications for stock market efficiency. The Journal of Finance 48, 65–91.
Karapandza, Rasa, Marin, Jose, 2011. Dissecting Market Efficiency. In: Working Paper, Madrid Institute for Advanced Studies.
Loughran, Tim, McDonald, Bill, 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance 66, 35–65.
Loughran, Tim, Ritter, Jay R., 1995. The new issues puzzle. The Journal of Finance 50, 23–51.
MacLaury, Robert E., Hewes, Gordon W., Kinnear, Paul R., Deregowski, J.B., Merrifield, William R., Saunders, B.A.C., Stanlaw, James, Toren, Christina, Van Brakel, J., Wescott, Roger W., 1992. From brightness to hue: an explanatory model of color-category evolution. Current Anthropology 33, 137–186.
Merton, Robert C., 1987. A simple model of capital market equilibrium with incomplete information. The Journal of Finance 42, 483–510.
Newey, Whitney, West, Kenneth, 1987. A simple, positive semi-definite, heteroscedastic and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Pastor, Lubos, Stambaugh, Robert F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.
Richardson, Scott A., Sloan, Richard G., Soliman, Mark T., Tuna, Irem, 2005. Accrual reliability, earnings persistence and stock prices. Journal of Accounting & Economics 39, 437–485.
Richardson, Scott A., Tuna, Irem, Wysocki, Peter, 2010. Accounting anomalies and fundamental analysis: a review of recent research advances. Journal of Accounting & Economics 50, 410–454.
Sadka, Ronnie, 2006. Momentum and post-earnings-announcement drift anomalies: the role of liquidity risk. Journal of Financial Economics 80, 309–349.
SEC, 1980. Securities Act Release No. 6231.
Sloan, Richard G., 1996. Do stock prices fully reflect the information in accruals and cash flows about future earnings? Accounting Review 71, 289–315.
Stone, Philip J., Earl, B. Hunt, 1999. A computer approach to content analysis: studies using the General Inquirer system. In: Proceedings of the May 21–23, 1963, Spring Joint Computer Conference, May 21–23, Detroit, Michigan.
Sullivan, Ryan, Timmerman, Allan, White, Halbert, 1999. Data-snooping, technical trading rule performance, and the bootstrap. The Journal of Finance 54, 1647–1691.
Szmrecsanyi, Benedikt, 2003. Be going to versus will/shall: does syntax matter? Journal of English Linguistics 31, 295–323.
Tetlock, Paul C., 2007. Giving content to investor sentiment: the role of media in the stock market. The Journal of Finance 62, 1139–1168.
Tetlock, Paul C., Saar-Tsechansky, Maytal, Macskassy, Sofus, 2008. More than words: quantifying language to measure firms' fundamentals. The Journal of Finance 63, 1427–1467.
Thieroff, Rolf, 2009. On the areal distribution of tense-aspect categories in Europe. In: Dahl, Östen (Ed.), Tense and Aspect in the Languages of Europe. Mouton de Gruyter, Berlin, pp. 309–328.
Titman, Sheridan, Wei, K.C. John, Xie, Feixue, 2004. Capital investments and stock returns. Journal of Financial and Quantitative Analysis 39, 677–700.
Whorf, Benjamin Lee, 1956. Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. In: Carroll, John B. (Ed.). MIT Press, Cambridge, MA.
Winawer, Jonathan, Witthoft, Nathan, Frank, Michael, Wu, Lisa, Wade, Alex, Boroditsky, Lera, 2007. Russian blues reveal effects of language on color discrimination. Proceedings of the National Academy of Sciences 104, 7780–7785.
Zhang, X. Frank, 2006. Information Uncertainty and Stock Returns. The Journal of Finance 61, 105–137.