22 Aug 2017 11:53:51 ET | 67 pages

Quantitative
Global

Searching for Alpha: Extracting Information from Earnings Calls

Part 1: Language Complexity, Saying Less Is More

- **Earnings Call Insights** — We conduct analysis of earnings conference call transcripts, focusing on the primary facets of calls ranging from call periodicity, timing, sector and country patterns to better understand the relationship these have with the variability of earnings calls.
- **Dealing with Big Data** — Big Data is gaining a lot of attention but the challenge of dealing with unstructured data sometimes gets over-looked. Using global earnings conference call transcripts as an example, we outline some of the more interesting and demanding problems faced with acquiring, parsing and normalising this unstructured source.
- **Measuring Disclosure and Readability** — We apply an innovative approach to measuring the levels of disclosure, by considering the relative amounts of time management spend on the discussion segments of a call and the amount of disclosure provided in response to analyst questions. Separately we investigate obfuscation through complexity measures applied to segments of the call, eventually building up to a metric that evaluates disclosure quality.
- **Complex Alpha** — We examine the relationship of our earnings call based signals to our standard style factors. In doing so we find these measures add incremental information and are most closely related to Earnings and Price Momentum styles.

Europe

Chris Montagu [AC]
+44-20-7986-3958
chris.montagu@citi.com

Helen H Krause, CFA
+44-20-7986-8653
helen.krause@citi.com

James J Murray, CFA
+44-20-7986-4579
james.murray@citi.com

David T Chew
+44-20-7986-7698
david.chew@citi.com

Zee Yusuf, CFA
+44-20-7986-4228
zee.yusuf@citi.com

Ying Zhu
+44-20-7986-8325
ying.zhu@citi.com

Asia-Pacific

Chris Ma
+852-2501-2404
chris.ma@citi.com

Liz Dinh, CFA
+61-2-8225-4896
liz.dinh@citi.com

Jason Kim
+852-2501-2735
jason.jh.kim@citi.com

Contents

Executive Summary

Earnings calls are widely followed by fundamental analysts as one of the main mechanisms corporates disclose information to the market. Improvements in technology have meant that many of these calls are being recorded and the contents of these events are being transcribed. Leveraging text mining techniques, this suite of Big Data and the resulting information, is now available to quantitative investors that traditionally would have been the domain of fundamental investors.

In part 1[1] of this series we start our analysis using relatively simple measures of complexity to earnings call transcripts, offering significant detail regarding data curation as well as signal construction. Despite earnings call transcripts being available in a standardized form we encountered a number of challenges with data collection. We discuss document parsing challenges, highlighting numerous issues, such as text encoding errors, and how we solved them. We can't stress enough the importance of data curation and the effort involved when dealing with (unstructured) Alternative or Big Data. This part of the analysis is an important consideration for those investors that want to explore text analysis or uses any other forms of unstructured alternative data.

For signal construction, our main hypothesis follows from prior financial research such as Li (2008) in believing simplicity of earnings calls is a positive predictor for future stock returns. Excessive obfuscation and lack of clarity from company insiders is indicative of management either hiding/withholding information, or alternatively, not being well informed enough about the businesses they run to offer well formulated insights.

To measure complexity we use readability measures commonly employed in the computational linguistics literature. Empirical results of selecting stocks (buy simple language, sell complex language) using these measures, adjusted for sector and county biases, was encouraging if not compelling.

Finally, we examine the relationship of these signals to our standard style factors. Doing so we find these measures add incremental information and are most closely related to Earnings and Price Momentum styles. Examining this further adding our earnings call signals into Earnings and Price Momentum composites improves risk adjusted performance through both enhancing returns and reducing risk.

As mentioned this is the first part of a research series that focuses on one relatively easily accessible Big Data source and go through the full process of data handling through to data analysis. From the data analysis side we start with using some basic techniques with the aim to further investigate more sophisticated data analysis methodologies.

Regards,

Citi Quant Research Team

[1] Actually we have done a few research reports on this topic. See: Searching for Alpha: Big Data: Navigating New Alternative Datasets and Machine Learning Algorithms in Sentiment Detection: Some Things Machines do Better

Introduction – Why Earnings Calls Matter

We begin our foray into the world of unstructured big data through examining analyst earnings conference calls – an information source which has been of paramount importance to our fundamental analyst colleagues but has increasingly become accessible for quantitative evaluation.

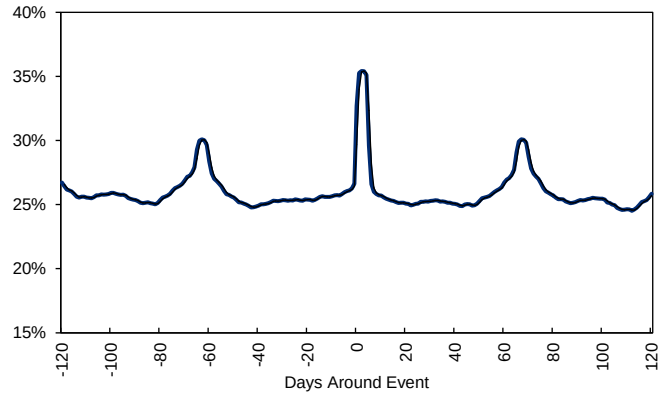
For many companies the earnings calls are one of the main mediums for management to present a summary and offer their views on the company's performance. This enables management to not only provide reassurance but also give guidance on the future prospects of the firm. From an investor's perspective, they provide a rare opportunity to hear from firm insiders in both a scripted (presentation) and unscripted (Q&A/discussion) setting.

There isn't a prescribed procedure surrounding the execution of earnings calls with the majority of factors ranging from timing to content, being at the discretion of management. Therefore, when and how earnings calls are conducted may be as informative as the information they contain. However, they typically occur on the day or following day of earnings announcements and start with an overview of business operations from management before opening for analyst questions.

The ability of analysts to question management is a distinguishing feature of earnings calls relative to other firm disclosure mechanisms. Analysts participating in the call play a crucial role in eliciting information from management, and the dynamic nature of calls prevents management from preparing disproportionately scripted responses. It's an opportune time for analysts to clarify issues that may be in contrary to their assessment of a firm and garner additional insight into company operations.

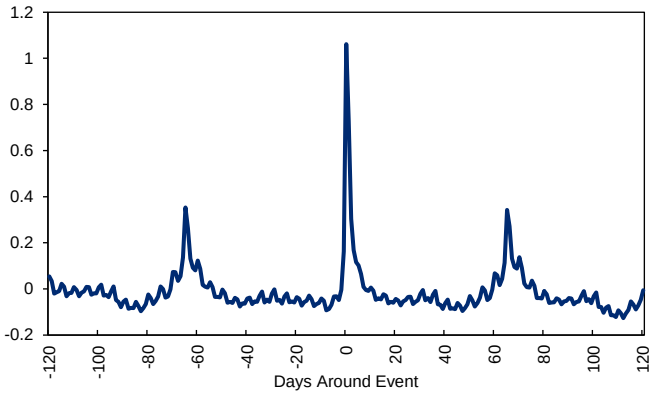
The numerical information in earnings announcements has been widely followed by market participants since the dawn of markets, with analyst forecasts of future earnings announcements also impacting price moves. In case of any doubt, the importance of earnings announcements can be seen clearly in Figure 1 and Figure 2 which shows a clear spike in both trading volume and stock volatility around the earnings announcement period.

Figure 1. 5-day Rolling Average Annualized Volatility Around Earnings Call



Source: Citi Research

Figure 2. Average Standardized Trading Volume Around Earnings Call



Source: Citi Research

Along with the 'hard' numerical information communicated through earnings announcements, relatively recent technological advances have allowed academics and practitioners to examine the circumstance around the delivery of earnings information. Earnings numbers are provided by those – meaning company management - with significant insights and information purporting to the current and future performance of a firm. Reported numbers are at the mercy of window dressing and are a result of historic operational performance which is not necessarily indicative of the future.

There has been a considerable amount of research surrounding earnings press releases in the academic literature, however slightly less so towards earnings conference calls themselves. We find this odd given the potential informational content of these events.

Scripted press releases are very easily controlled by the company enabling bias in language to be suppressed. Conversely earnings calls, listening to company management in a live, unscripted setting[2], are much harder to control and more likely to contain linguistic biases offering potential insights into future firm performance.

While this explains our interest in earnings calls over earnings press releases, it does little to explain why we have chosen to conduct our first piece of Natural Language Processing (NLP) research on earnings calls over other common sources such as the media (see Tetlock [2007]) or regulatory disclosures (Feldman et al [2009], Li [2010]). As will be shown, integrating new (particularly unstructured) information sources into an investment and research process is time consuming and challenging. As such the 'skill' in choosing which text sources to spend time analyzing is arguably as important as the analysis itself.

In terms of our motivation, the decision is in part due to our appreciation of the rationale behind the unpriced source of information. It does also come with some practical considerations which we anticipate our clients will also need to grapple with.

The size of the data is an important consideration. Earnings calls are at most quarterly giving four data points per year, per company making a relatively small dataset. Despite this, downloaded files from 2003-2017 are around 100Gb (around 10Gb after parsing) which is still manageable on desktop workstations (in acceptable computation times). Alternative textual sources such as media stories could be significantly larger perhaps requiring a different technology stack.

The dataset is also global in nature, making the research practically feasible to all our clients, unlike regulatory filings which are country specific. Clients running money across countries with different regulators (we suggest most) need to seriously consider this constraint.

Possibly, and most pragmatically, the dataset was available to us requiring less effort to collect (no web-scraping) and with a relatively long history. It therefore enables us to examine textual methods of analysis in an easy yet practically applicable setting: a good place to start. While these decisions are based on practicality rather than alpha generation or the 'cool factor' we surmises that we are not alone in having to trade-off idealistic approaches with plausible realities.

[2] While we acknowledge those speaking are discussing pre-prepared presentations and discussion points, it is unlikely management are simply reading pre-prepared presentations. As such the semantics around the key points – does the speaker use 'good' or 'great' – is where we suggest information may exist, and is hard to control by the firm.

In this research we first start with an overview of previous work in this area before spending considerable time on the collection and handling of earning call transcripts. While this is probably the least exciting part of big or alternative datasets, we argue it is the most important. The research then proceeds to examine the dataset gaining insights into both the earnings calls dataset, and features we may be able to exploit. Finally we conclude with an illustrative application of the data through examining disclosure and readability of calls in predicting firm performance.

Earnings Calls in Academia

Earnings information is used by investors in a variety of ways from valuation metrics to growth forecasts (e.g. see Basu [1977] and Little [1962] respectively) with Bradshaw [2011] offering a good review of the significant efforts which have gone into examining analysts forecasting ability of earnings.

In this research we are attempting to gain insight into company operations through leveraging the mechanism with which information are communicated to the public. Earnings announcements provide numerical information representing firm performance which is well covered by investors. However, they also provide an opportunity to hear from firm insiders through both the press release, or more interestingly to us, the earnings conference call following the press release. Such data provides an insight into the level of optimism or pessimism held internally at the company, or the extent to which their explanations are plausible or farfetched.

Although the methodologies and techniques for textual analysis have been applied in other fields for some time, its application in finance and accounting has only grown in significance over the last decade; much of this is due to the increasing availability or large corpuses of financial text coupled with advancement in technology and linguistic analysis. The increasing abundance of research papers on textual analysis across financial datasets, has also led to the emergence of literature surveys covering the application of text analysis in finance, Kearney [2014] and more recently Loughran Mcdonald [2014] providing a good summary of methods, pitfalls and literature highlights.

Research into earnings calls, sits closely alongside those of earnings press releases and filings due to their characteristic similarities. Most studies focus on applying tone/sentiment, readability and topic analysis to extract quantitative metrics which can be applied with firm fundamentals in a descriptive or predictive manner. There are also alternative streams that look at other interesting aspects such as hosting, periodicity and timings of calls.

In the context of press releases Davis et al [2011] analyse 23,000 press releases from 1998-2003. Through examining the frequency of optimistic versus pessimistic words in earnings releases as determined by the textual analysis program DICTION the authors find optimism leads future strong ROA. Demers and Vega [2008] follow a similar methodology showing positive tone increases stock prices with excess returns persisting for 60 days after the earnings release. Henry [2008] finds tone influences stock price reactions arguing framing results in positive terms makes investors more positive as suggested by Kahneman and Tversky [1981].

The focus of this research is earnings calls, which has also been examined in academia. Along with our prior assertion that earnings calls are more difficult to manipulate, there is also evidence that they contain distinct information not previously divulged. Sunder [2002] and Irani [2004] find post regulation Fair Disclosure[3] earnings calls have become more informative and increased in value relevant information. In a similar vein, Bowen, Davis and Matsumoto [2002] find earnings calls increase analysts' ability to forecast earnings accurately and lowers dispersion among analysts suggesting "…the calls increase the total information available about a firm..."

Earlier studies such as Frankel [1999] find that firms which hold earnings calls are larger and more profitable than those which do not. They also document elevated volatility at the time earnings calls are taking place highlighting the price relevant information such calls carry. Michaely, Rubin and Vedrashko [2016] also examine

[3] SEC ruling in 2000 which opened earnings conference calls to the public

the circumstance around the earnings call uncovering rational opportunistic behavior by management to reduce interaction with investors and analysts by scheduling earning calls when there are less active market participants.

In a similar vein, Matsumoto, Pronk and Roelofsen [2006] conduct a structured analysis of earnings calls content to study determinants which affect management presentation and discussion periods during earnings calls. They find that longer calls are associated with larger price reaction, larger trading volume, forecast revisions and improvements in forecast accuracy, indicating that longer conference calls are more informative to the market.

Other studies have used the bag of words approach of analyzing earnings calls. Price, Doran, Peterson and Bliss [2011] and Huang, Teoh and Zhang [2014] use the Loughran and McDonald [2011] dictionary to identify positive and negative tone in earnings calls finding, like Demers and Vega [2008], that positive (negative) tone leads to future prices persisting for 60 days. However, Huang, Teoh and Zhang [2014] also find a reversal effect whereby excessive positive tone is followed by 1-2 quarters of price underperformance. Druz, Wagner and Zeckhauser [2016] find disappointment predicts stock returns more strongly and has a longer decay than optimism.

Larcker and Zakolyukina [2012] take an alternative approach. They construct a dictionary based off financial restatements to identify the severity of accounting window dressing and subsequently label calls as deceptive or truthful. From this labeling they are able to identify words which are associated to being overly optimistic / pessimistic. However, their research was focusing on finding fraudulent CIO/CFO behavior rather than forecasting stock returns and so performance comparisons between the approaches are not available.

A related linguistic approach used by Li [2008] on annual reports, but not to our knowledge attempted on earnings calls, is examining the extent to which language complexity forecasts stock returns. Li uses the fog index to interpret the complexity of language in annual reports finding companies with lower earnings have financial reports which are harder to read with easier to read reports having more persistent positive earnings. Vocally the same scenario may apply. It is much harder to explain an event when trying to withhold information than when being completely transparent. As such, complexity of earnings call language may, as is found by Li [2008], be negatively related to future stock or firm performance.

In related research, Mayew and Venkatachalam [2012] use conference call audio files to analyze the emotional states of the conference call speaker finding positivity (negativity) is related to future positive (negative) earnings beats and stock performance. They also find that analysts do not incorporate this information into their forecasts. Borochin et al [2017] examine the effect of earnings call tone on the options market. Using the Loughran and McDonald [2011] dictionary while controlling for prefix and negation they find negative tone leads to increases in implied volatility.

The above studies provide us with some interesting takeaways. Firstly, the focus is on US data. While the US has a longer history than the rest of the world we have sourced earnings call transcripts from 2006 for other countries, enabling us to examine the robustness of author's findings in a different data sample. Secondly, the literature has focused as much on the circumstance around the call (characteristics of firms, when are calls performed, length of calls etc.) than on analysis of the call content (complexity of language, tone) with remarkably successful results.

Data Handling

One of the challenges of unstructured data is content handling and we dedicate a section in order to highlight some of the challenges and pitfalls that need to be overcome specifically when dealing with an unstructured data source. Although most of the issues highlighted here are specific to the conference call transcripts, they may be applicable to other sources utilising same file formats.

One of the reasons we emphasize the data handling section is that we believe that it's an aspect of previous 'Big Data' research reports that is often neglected or limited in discussion. Many reports talk about big data sources generally and then focus or discuss sophisticated methodologies to integrate the data but for most quantitative investors, the challenge lies in the data itself. Moving from structured data sets to un-structured data sets presents a new range of problems that investors need to deal with in order to get the data into usable form that analysis can then be conducted on. The following section of the research outlines how we source the data and identifies a number of problems along with our, at times imperfect, solutions. We hope that this highlights the importance of data curating in Big Data analysis as shown through our specific research.

Document Extraction

We source conference call transcripts from FactSet's Call Street, a data source within FactSet which provides access to a large range of corporate communication and related firm news events. FactSet captures earnings conference call transcripts which are conducted in English or where an English translation is present, and transcribes the content. While this is sufficient for North American investors, further in the research we examine the extent to which this limits non-English speaking regions.

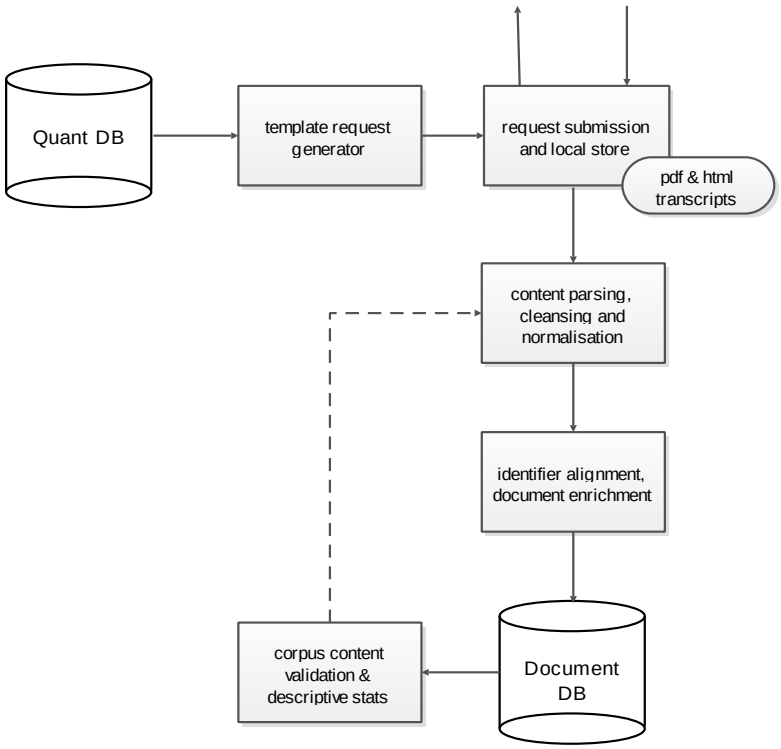
Call transcripts are typically loaded onto Factset Call Street within 3 days of the event. The initial transcript may be revised to correct for errors; and options are available to download the raw transcripts and corrected equivalent. However coverage for the raw transcripts is poor. We therefore focus our attention on corrected transcripts for this research and impose a minimum five day lag between the call date and their use so as to avoid any forward bias. While this is quite conservative for the current period of time, we have no way of assessing how realistic this is going back through time.

At the outset, we needed to adopt a process which enabled us to easily acquire, store and parse the content in a repetitive manner to handle regular updates, correct errors detected, and backfill historical transcripts.

In order to automate the task, we prepared a formula template using MSCI World constituents as a benchmark (~3000 unique companies) present within the index from Jan 2003 to June 2017. This benchmark enabled us to selectively filter for companies with sufficient breadth, size and liquidity to satisfy most institutional investors while avoiding survivorship bias issues.

The diagram below briefly outlines the steps which were undertaken to stage call transcripts in a local environment for further analysis.

Figure 3. Transcript acquisition and normalization process flow.



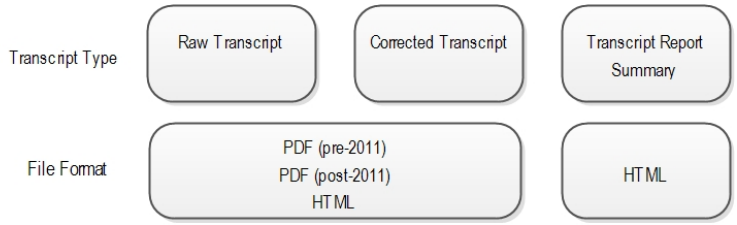
Source: Citi Research

The first step was to download the earnings call transcript from FactSet. For approximately 3000 stocks over 15 years of history this was not a job to be done manually. We programmatically prepared a template file which ran a macro to refresh FactSet paths, and then downloaded the data into a file server directory. In doing this we were quite meticulous with our error logging, enabling us to have multiple tries at failed downloads. Furthermore, along with the underlying file we extracted basic characteristics like date/time of the event, description and stock ID which we combined to form a unique download ID generating a primary reference key per document that could be linked to a specific company and point in time.

Transcript formats and types

Call transcripts are available in different report types and formats. Figure 4. Summary Call Transcript report types and formats provides a summary of call transcript report types and formats that are available to download.

Figure 4. Summary Call Transcript report types and formats

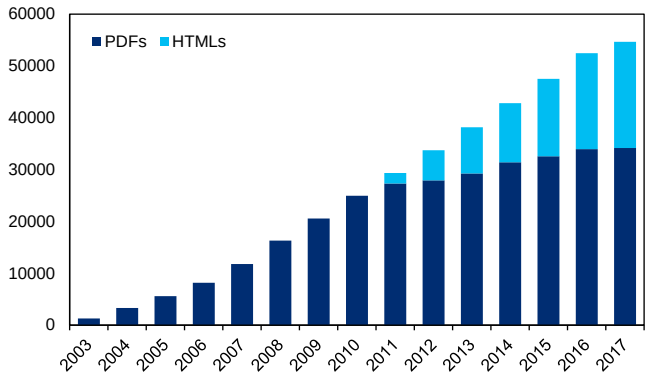


Source: Citi Research

While we downloaded 'Raw' and 'Corrected' transcript files, coverage is much better for the corrected files, with raw files seemingly being a legacy construct. For many calls FactSet also transcribe a 'Report' file which appears to be a summary of the call and its key themes. The coverage for these files is low going back through time but reasonably well covered currently. While we do not use these files in this report, the idea of applying topic analysis to the transcripts while using a human written summary to learn from, feels like an interesting way to obtain the factors which management, and those analysts which asked questions, consider to be most relevant for the stock.

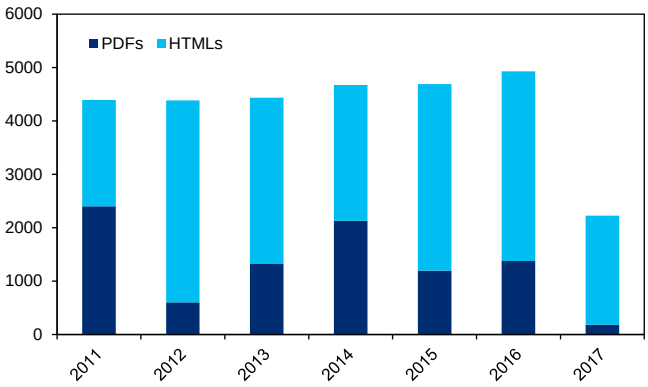
Along with the three types of reports, the files are available in either PDF or HTML format. Again we downloaded both and were pleased we did as coverage does vary, particularly back through time. As exhibited in Figure 6, all documents were available in PDF format only before 2011. After 2011 transcripts were made available in HTML and PDF formats, but the PDF document structure was changed significantly. To account for this we clean the two different types of PDF formatting separately. HTML formatted transcripts first appeared in 2011 and have had a fairly consistent markup representation ever since. When two sources were available for the same call event, we either combined or selected the one with greater detail for further processing. Figure 6 provides a breakdown of the transcript type selected for analysis by year.

Figure 5. Parsed transcripts selected for analysis by year



Source: Citi Research, Factset

Figure 6. Transcripts selection split since HTML formats introduced.



Source: Citi Research, Factset

Parsing the Documents

As expected, getting the data was relatively straight forward given the use of FactSet. To claim call transcripts from Factset are truly unstructured data is a stretch, we admit that some partial unification and cleansing is already handled by Factset, which makes our tasks less cumbersome. Parsing truly raw earnings calls either from a recording or from company provided transcripts which all have different document formats would be a significantly harder problem to resolve. Despite this preliminary cleansing, we still encountered numerous problems in processing the content which we discuss in the next section.

Initial errors in data handling and pre-processing have the potential to compound reducing the quality of analysis performed. As such we spent a significant amount of time cleaning and checking the data encountering different issues for different file types. Given the challenges in parsing, and robustly extracting information, where we have a file in different formats we attempt to parse both formats and use the two outcomes to validate the data we extract.

HTML Parsing

HTML parsing is easier and more robust than PDF parsing. This is because the documents carry information tags relating to text formatting and document sections which make it easier to identify key document features (Titles, speakers etc.). The HTML documents were also relatively easy to parse into analysable text in R using open-source packages[4]. Our general workflow was to open the HTML file, convert it to XML which can then be searched (as a transformed list or via XPath[5] queries of the object model) whereby different sections of the text can be identified and extracted.

Once files had been opened and converted to XML the file structure was examined to ensure that the document was as expected which held true for the vast majority of documents. Figure 7 and Figure 8 provide a much reduced code snippet employed in parsing the relevant sections of the document alongside its output.

Figure 7. Basic R code to parse a HTML Earnings Call

```
# LOAD USEFUL LIBRARY
library(xml2)
library(rvest)
# LOAD FILE
Doc<-read_html(paste(InPathHtml, i, sep=""))
# FIND BODY TEXT
Resp<-xml_find_first(Doc,"//*[@id='story_text']")
print(Resp)
# BODY TEXT SECTIONS
Sections<-xml_find_all(Resp,".//*[@id ='sections']")
print(xml_children(Sections))
# USING THE FIRST SECTION
Sec<-xml_children(Sections)[[1]]
# NAME OF FIRST SECTION
xml_text(xml_find_all(Sec, ".//div[@class = 'sectionHeader']"))
# SUBSECTION
SubSec<-xml_children(Sec)
# WHO IS SPEAKING
xml_text(xml_find_all(SubSec[5], ".//a[@class = 'ptLink ptContent']"))
# JOB ROLE
xml_text(xml_find_all(SubSec[5], ".//span[@class = 'ptInfo']"))
# WHAT DID THEY SAY
gsub("\\s+", " ",xml_text(xml_find_all(SubSec[5], ".//p[contains(@class,'pwrapper')]")))
```

Source: Citi Research

[4] We used rvest, and jsonlite, see Appendix – Terms and Definitions for details.

[5] XPath - XML Path Language, see Appendix – Terms and Definitions for details

We were able to extract the management discussion section and separate speaker by HTML markups. Similarly, we could split the question and answer section, and identify the participant asking the questions, responding and which firm they worked for. This was only possible when Factset themselves had captured this information; historical transcripts had less detail, which made it impossible to identify speaker/responder's detail.

Figure 8. Snippet sample extracting various segment of call transcripts

```
> # LOAD USEFUL LIBRARY
> library(xml2)
> library(rvest)
> # LOAD FILE
> Doc<-read_html(paste(InPathHtml, i, sep=""))
> # FIND BODY TEXT
> Resp<-xml_find_first(Doc,"//*[@id='story_text']")
> print(Resp)
{xml_node}
<div xmlns="http://www.w3.org/1999/xhtml" class="wrapper" id="story_text">
[1] <div class="content">\n  <div class="participants">\n    <div class="participantHeader"><a na
> # BODY TEXT SECTIONS
> Sections<-xml_find_all(Resp,".//*[@id ='sections']")
> print(xml_children(Sections))
{xml_nodeset (2)}
[1] <div><div class="line"> </div><div class="sectionHeader"><a name="MANAGEMENT DISCUSSION SECT
[2] <div><div class="line"> </div><div class="sectionHeader"><a name="QUESTION AND ANSWER SECTION
> # USING THE FIRST SECTION
> Sec<-xml_children(Sections)[[1]]
> # NAME OF FIRST SECTION
> xml_text(xml_find_all(Sec, ".//div[@class = 'sectionHeader']"))
[1] "Management Discussion Section"
> # SUBSECTION
> SubSec<-xml_children(Sec)
> # WHO IS SPEAKING
> xml_text(xml_find_all(SubSec[5], ".//a[@class = 'ptLink ptContent']"))
[1] "Peter Oppenheimer"
> # JOB ROLE
> xml_text(xml_find_all(SubSec[5], ".//span[@class = 'ptInfo']"))
[1] "Chief Financial Officer & Senior Vice President, Apple, Inc."
> # WHAT DID THEY SAY
> gsub("\\s+", " ",xml_text(xml_find_all(SubSec[5], ".//p[contains(@class,'pwrapper')]")))
 [1] " Thank you, Nancy. "
 [2] " We're very pleased to report the results of the first quarter of Apple's fiscal year 2014.
 [3] " Revenue for the quarter was $57.6 billion, up $3.1 billion or 6% from the year-ago quarter
 [4] " These three factors negatively impacted revenue by about $2.5 billion, and without them ou
```

Source: Citi Research

'New' PDF Parsing

Once again we relied upon open source packages[6] to read and parse the PDF content. This was a more involved process given the lack of meta data. Parsing the primary textual data into R was relatively straightforward with sample output in Figure 10 clearly showing the lack of meta data when compared to the HTML files. The challenge here was to extract the participant information, although present but far less structured.

Figure 9. Basic R code to parse a PDF Earnings Call

```
library(tm)
options(stringsAsFactors=FALSE)
options(encoding="UTF-8")
YearParse<-"2014"
pdf.func<-readPDF(engine = "xpdf",control=list(text="-layout"))
Doc<-pdf.func(elem=list(uri=paste(InPath, i, sep=""))
    , language="en",id = "id1")
Doc
head(Doc[["content"]], 25)
```

Source: Citi Research

[5] We used tm and xpdf, see Appendix – Terms and Definitions for details

Figure 10. Executed Code

```
> library(tm)
> options(stringsAsFactors=FALSE)
> options(encoding="UTF-8")
> YearParse<-"2014"
> pdf.func<-readPDF(engine = "xpdf",control=list(text="-layout"))
> Doc<-pdf.func(elem=list(uri=paste(InPath, i, sep="")), language="en",id = "id1")
> Doc
<<PlainTextDocument>>
Metadata:  7
Content:  chars: 98356
> head(Doc[["content"]], 25)
 [1] "                                                                Corrected Transcript"
 [2] ""
 [3] ""
 [4] ""
 [5] ""
 [6] "27-Jan-2014"
 [7] "Apple, Inc.                (AAPL)"
 [8] "Q1 2014 Earnings Call"
 [9] ""
[10] ""
[11] ""
[12] ""
[13] "                                                                Total Pages: 19"
[14] "1-877-FACTSET www.callstreet.com    Copyright © 2001-2014 FactSet CallStreet, LLC"
[15] "\fApple, Inc. (AAPL)
[16] "Q1 2014 Earnings Call
[17] ""
[18] ""
[19] ""
[20] "CORPORATE PARTICIPANTS"
[21] "Nancy Paxton
[22] "Senior Director, Investor Relations, Apple, Inc.
[23] ""
```

Source: Citi Research

The primary tool we utilised to extract data from PDF files is text pattern matching. This is less robust than the document structure tagging used in HTML files and also more at the mercy of typing and spelling mistakes. Again, this highlights the realities of handling unstructured data.

To remove the header and footer found on each page we performed a pattern match of the FactSet branding found in the headers and footers combined with the removal of surrounding lines. This worked well proving to be a reliable point to work with. In a similar vein a pattern match for the disclaimer section resulted in the main body of text.

The new PDF documents have each section separated with a line of dots as in line 53, 59 and 66 of Figure 11. Combining those dots with blank spaces and expected section headings enabled us to identify points of interest in the document. This was robust for the most part enabling us to identify who was speaking and what they were saying. However, we did find a number of ways to spell ‘Management Section’ along with variations on hyphenation. The general process was to try and parse, find failing documents and repeat.

Figure 11. Example New PDF Document

```
[47] ""
[48] "Nancy Paxton"
[49] "Senior Director, Investor Relations, Apple, Inc."
[50] ""
[51] "Thank you, Peter. We ask that you limit yourself to one question and one follow up. Operator, may we have the"
[52] "first question, please?"
[53] "..........................................................................................................."
[54] ""
[55] ""
[56] ""
[57] "QUESTION AND ANSWER SECTION"
[58] "Operator : [Operator Instructions] First we'll go to Katy Huberty with Morgan Stanley ."
[59] "..........................................................................................................."
[60] ""
[61] "Kathryn Huberty"
[62] "Analyst, Morgan Stanley & Co. LLC
[63] "Thanks. Good afternoon. Peter, revenue guidance embeds a sequential decline that is bigger than what you"
[64] "experienced in the March 2013 quarter and, I think, implies iPhone units may not grow in the March quarter year-"
[65] "on-year. So can you just talk about what's driving the cautious view on the top-line for March?"
[66] "..........................................................................................................."
[67] ""
[68] "Peter Oppenheimer"
[69] "Chief Financial Officer & Senior Vice President, Apple, Inc.
[70] "Sure. Katy , the biggest reason for the larger sequential decline in revenue this year relates to changes in channel"
```

Source: Citi Research

Along with metadata such as file size, creation data, tickers, call time etc. Data was then saved into a JSON file and uploaded to a NoSQL[7] database ready for analysis. Importantly, the order the text appeared was saved such that we could completely rebuild the key content of the HTML file from the JSON[8] file

While this approach worked for the majority of files, there were a small number for which this did not work.

'Old' PDF Parsing

Getting the data into R was the same process as above however extracting the relevant sections was less robust. The Old PDF documents did not list participants at the beginning of the document and do not have as distinct or consistent section markers. Regular expressions were used to find the 'Management and Discussion' and 'Question and Answer' sections which worked surprisingly well. One of the reasons it worked so well is that the section titles parsed into upper case letters making them easy to identify.

Finding the speaker in the Management and Discussion section was however harder as it is embedded in the text with little identifiers around it. To extract this information we noticed that the title is indented relative to everything else on the page. We therefore did a blank character count and found the indented lines. Checking there was a blank line above and below indented text validated the assertion with name and job role being separated with a comma. This was a less reliable process which was also difficult to validate resulting in lots of errors that needed to be checked manually.

[7] NoSQL – non SQL/non-relational databases

[8] JSON – Java Script Object Notation, see Appendix – Terms and Definitions for details

Figure 12. Speaker Indenting Example

```
> Doc[["content"]][1:100]
  [1] "                                    Apple Computer, Inc.                    AAPL          Q1 2006 Earnings Call          Jan. 18, 2006"
  [2] "                                                         Company          Ticker                  Event Type                  Date"
  [3] ""
  [4] ""
  [5] "                          MANAGEMENT DISCUSSION SECTION"
  [6] ""
  [7] "                          Operator: Good day and welcome to the Apple Computer First Quarter Financial Results"
  [8] "                          Conference Call. Today's call is being recorded. At this time, for opening remarks and"
  [9] "                          introductions, I would like to turn the call over to Ms. Nancy Paxton, Senior Director, Investor"
 [10] "                          Relations and Corporate Finance. Please go ahead, Ms. Paxton."
 [11] ""
 [12] ""
 [13] "                      Nancy Paxton, Senior Director of Investor Relations and Corporate Finance"
 [14] ""
 [15] "                          Thank you. Good afternoon and thanks to everyone for joining us. Speaking today is Apple's CFO,"
 [16] "                          Peter Oppenheimer, and he'll be joined by Apple's COO, Tim Cook, and Corporate Treasurer, Gary"
 [17] "                          Whipfler for the Q&A session with analysts."
 [18] ""
```

Source: Citi Research

Questions and Answers were relatively easy to find parsed in text with identifiers being <Q – NAME> or <A – NAME>. This likewise enabled the speaker to be identified. While this was robust there was no information about the speaker – company, job role etc. One possible way to get this information is to do a name match of covering analysts in the IBES database and get their company from that source.

Figure 13. Old PDF Question and Answer Section Example

```
 [35] "                          <Q - Benjamin Reitzes>: Okay. Then I'm going to cede the floor, one more. I see a situation"
 [36] "                          where you still may have some lower than expected channel inventory, but are you seeing anything"
 [37] "                          out there that is giving you any pause in the economy or in the consumer markets? Other than"
 [38] "                          some of these Apple-related events that you're talking about?"
 [39] ""
 [40] "                          <A - Timothy Cook>: What we are seeing is we had an extraordinary response last week to our"
 [41] "                          announcements at MacWorld. We are thrilled with the reaction that we've gotten, including the"
 [42] "                          bookings on the new iMac and the PowerBook, or MacBook Pro. The last quarter we did see a"
 [43] "                          pause as some customers began to speculate about the upcoming announcements at Macworld."
 [44] "                          However, the Mac beat our internal expectations as Peter had said. We obviously factored some"
 [45] "                          level of pause into our guidance and we were able to beat that number and we're very happy with it"
 [46] "                          and the overall momentum of the Mac business."
 [47] ""
 [48] "                          <Q - Benjamin Reitzes>: Thanks a lot."
 [49] ""
 [50] "                          <A - Nancy Paxton>: Thanks Ben. Could we have the next question please?"
 [51] ""
```

Source: Citi Research

We briefly outline some of the primary issues and challenges which we came across while parsing and normalizing call content.

Problems we encountered include;

- **Batch downloading of transcripts**

  FactSet manages access to transcripts through providing web links to documents which are dynamic and only valid per user, per connection. This means if your FactSet connection is dropped or times out, all of the paths are no longer valid. To mitigate such problems we downloaded data in batches, first obtaining paths, and then downloading transcript files for that batch. We found that downloading data year by year and in batches of 500 stocks at a time did not give us any connection problems.

- **Text encoding mismatch**

  Despite the relative simplicity in loading files, we did encounter a text encoding error in the FactSet documents. HTML transcripts contain a text encoding declaration informing the application how to interpret character symbols found within. UTF-8[9] is by far the most common type of text encoding, and all of our HTML files claimed to be in this format. However, in reality, there were a number of files which were encoded in Latin1[10] format. In basic cases this meant that certain characters; hyphens, commas accented characters etc. became corrupted. In extreme cases, we could not open the file at all.

  This was solved by opening the document under different text encodings with error handlers and then using the document with the least number of corrupt characters. While this simply solves the problem it does add significant processing time to complete the parsing and it took some time to identify that the file encoding was indeed the culprit.

- **The analyst question and answer section being in the management and discussion section.**

  This tended to happen if there was a missing tag or spelling error identifying the question and answer section. We incorporated validation checks (detailed in next section) to detect missing Q&A segments, these helped highlight the problem, however this was co-mingled with calls which genuinely contained no Q&A segments, and we combined these with checks of the number of question mark-ups within the management segment.

- **The question and answer section being in a different file all together.**

  This was rare but identifiable by the same company having a missing management discussion and question and answer section, and another equivalent transcript being found to only contain the missing segment.

- **Multiple calls on a single day.**

  This is common for companies with dual listings having calls at convenient times in different time zones. We handled this through comparing the different sections of the reports and taking the longest, rationalizing that this had the most information. In reality, this had little effect on the management and discussion sections which were very similar. However, the question and answer sections

[9] UTF-8 - character encoding format, see Appendix – Terms and Definitions for details

[10] Latin1 - character encoding format, see Appendix – Terms and Definitions for details

may lose information from questions asked in one region but not another. The flipside is questions being repeated in both calls.

- **The call not having a question and answer section.**

  Either no questions were asked or management did not open up for questions. In such circumstances we retain the management discussion section in our analysis.

- **Missing question symbolic markup.**

  It is quite common for analysts to raise multiple questions within a single recorded statement, however not all questions are terminated with question marks; this prevents us from obtaining an accurate representation of the number of questions asked.

- **PDF document word splits**

  During parsing, we also encountered some limitations with the open source solutions we adopted, one notable issue related to the pdf parsing engine, which resulted in words being incorrectly split, as shown in Figure 14.

Figure 14. Text Space Error Handling

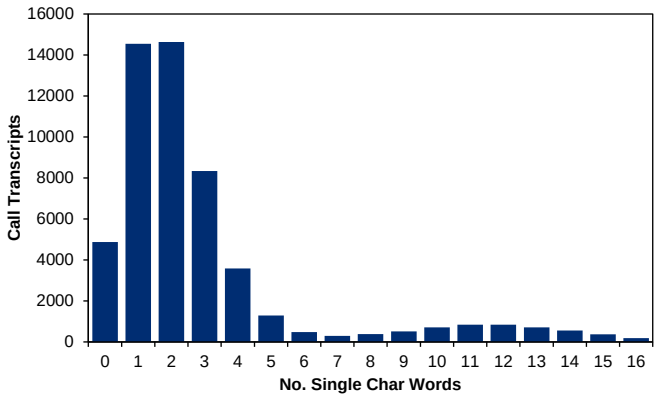
Thanks, Clay t. Starting with our Beauty business,* ***sale s* *grew 8%, led by double-digit growth in Prestige Fragrances, Skin Care, Feminine Care and Retail Hair Care. Cosmetics grew mid -singles, and Personal Care and cleansing products were roughly in line with the* ***y ear*** *ago. Retail hair color was down* ***v ersus***

Source: Citi Research, Factset CallStreet : Procter & Gamble, Q3 2007 Earnings Call, using xpdf 3.0.3

The presence of the above errors could be damaging for our analysis, if it were present over a large part of the corpus. To identify and mitigate this we examined the number of single character words in a corpus once common single character's 'a' and 'i' had been removed. Figure 15 and 16 presents a sample of the validation check conducted on our samples to detect these.

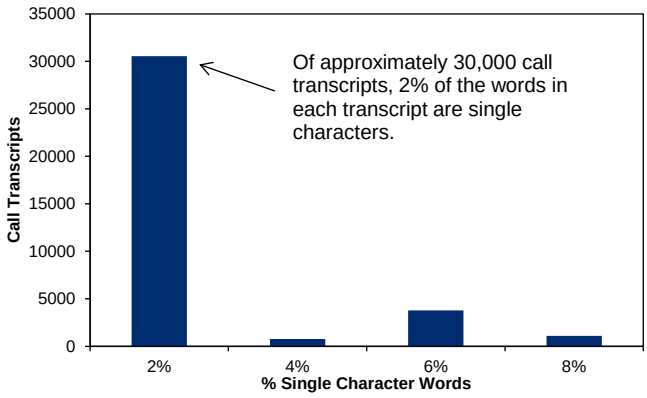
Fortunately, these issues although prevalent had little effect on our analysis as the problem primarily affected a small number of PDF transcripts for which we had HTML equivalents to substitute for.

Figure 15. Single Character Word Per Transcript



Source: Citi Research, Factset

Figure 16. Single Character Word Percentage of Total Transcript Word Count



Source: Citi Research, Factset

While we were ultimately able to fix a majority of these problems through data validation and writing a robust set of checks their existence makes us suspicious as to how many additional problems which could have gone unnoticed notice. It is clear that we are very reliant on the structure of the document being consistent. FactSet clearly have document templates for recording earnings calls which gives some credence to the document structure. However, as we noticed in 2011, this structure may change, at any point, potentially harming the efficacy of an alpha generative signal. As such, as with any unstructured data, significant quality and integrity checks are required to validate the data before it is used in any investment decision.

This is a challenge which is very easy to handle in a backtesting environment but much harder in a live setting. How sure are you that your unstructured data processing is capturing what you think it is, or what it was capturing when the signal was tested?

Processing these files to a high quality was an iterative approach with the vast majority being easy to process, followed by a long tail of spelling errors and missing tags which were being relied on. While all of these things can be handled in code, significant time and resource is required for error checking. Perhaps not obviously, we foresee this to be one of the significant challenges in using such alternative data in live strategies, and it is a topic we have to date heard very few people discuss.

Content Validation

Having multiple transcript formats of the same event enabled us to compare like for like transcripts to validate and capture formatting nuisances that may not be correctly accounted for in the custom parsing rules we had developed.

In order to minimise the manual effort in verifying the parsed content, we also ran the following pre-validations steps:

- Parsed transcript length checks (<100bytes).
- Documents which did not contain presentation segment.
- Documents without a discussion (question/answer) segment.

- Word counts, length of words within transcripts. (applied to presentation, question, answers sections).
- No of questions within management versus discussions.
- Calls containing answers without corresponding questions.
- Cross referencing speakers in discussion segments with participant's summary information.

This is a partial list of some of the hand curated rules which were applied to the corpus, prior to our analysis. This process was fairly iterative, and had to be repeated several times to refine the cleansing process for our content.

Persistence, Performance and Scalability

Throughout the parsing and data exploration stages of this project, we encountered a number of problematic issues, and although our analysis for this project had been rudimentary in nature, we had stumbled upon a number of challenges when attempting to conduct more thorough linguistic analysis of the corpus content. Some of these raise issues which clients should be aware of before embarking on textual analytics projects.

Firstly, although our corpus was relatively small in size, a fair amount of compute time was required to extract basic textual metrics. This research was conducted in R, and most of the 3rd party libraries/routines were written a single threaded usage. We set-up a basic parallel processing environment to optimise the processing. Maximising the capabilities from parallel processing requires some adjustment to design/coding approaches, which is best addressed early in the project to avoid having to conduct a retrofitting exercise.

Even with parallel processing, linguistic analysis, such as lemmatisation, part-of-speech (POS) tagging and dependency parsing (Figure 18) by their very nature are computationally intensive, and struggled on our workstations. These could potentially be addressed using a grid/cluster compute environment.

Figure 17. Sample Transcript Sentence Converted to Parse Tree.

**So to summarize, in 2015, the Group made good strategic and financial progress. The Group has a clear strategy and a differentiated …*

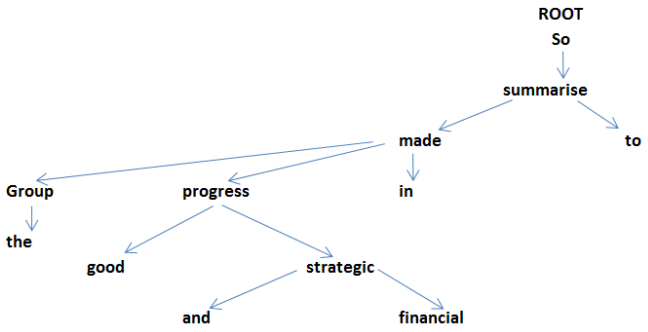
Source: Citi Research, Factset CallStreet : Lloyds Banking Group Plc Q4 2015

Figure 18. Output Parse Tree

Index	Governor	Dependent	Type	Gov. Index	Dep. Index
1	ROOT	so	root	0	1
2	summarize	to	mark	3	2
3	So	summarize	dep	1	3
4	made	in	mark	7	4
5	Group	the	det (predeterminer)	6	5
6	made	group	nsubj (nominal subject)	7	6
7	summarize	made	advcl (adverbial clause modifer)	3	7
8	progress	good	amod (adjectival modifier)	12	8
9	progress	strategic	amod (adjectival modifier)	12	9
10	strategic	and	cc (preconjunct)	9	10
11	strategic	financial	conj (conjunct)	9	11
12	made	progress	dobj (direct object)	7	12

Source: Citi Research

Figure 19. Tree Parse



Source: Citi Research

We refer to persistence in the context of storing our corpus data. In order to better understand how we would store earnings calls, we needed to have a firm idea of how it would be used. However, given the nature of exploratory analysis, especially in the context of textual data, there were many potential use cases. Modelling this in a relational database though possible, was rigid/inflexible. Structural modifications are cumbersome in a traditional database so is nesting, and the NoSQL solution seemed a natural fit.

Although a viable option, we did not incorporate the original PDF/HTML transcript source in our document data model as it would overly inflate our document database.

POS tagging and dependency parsing result in text parse trees and tags which add a significant amount of information to the base corpus (Figure 18). Given the computation time involved in generating them it's sensible to consider storing these.

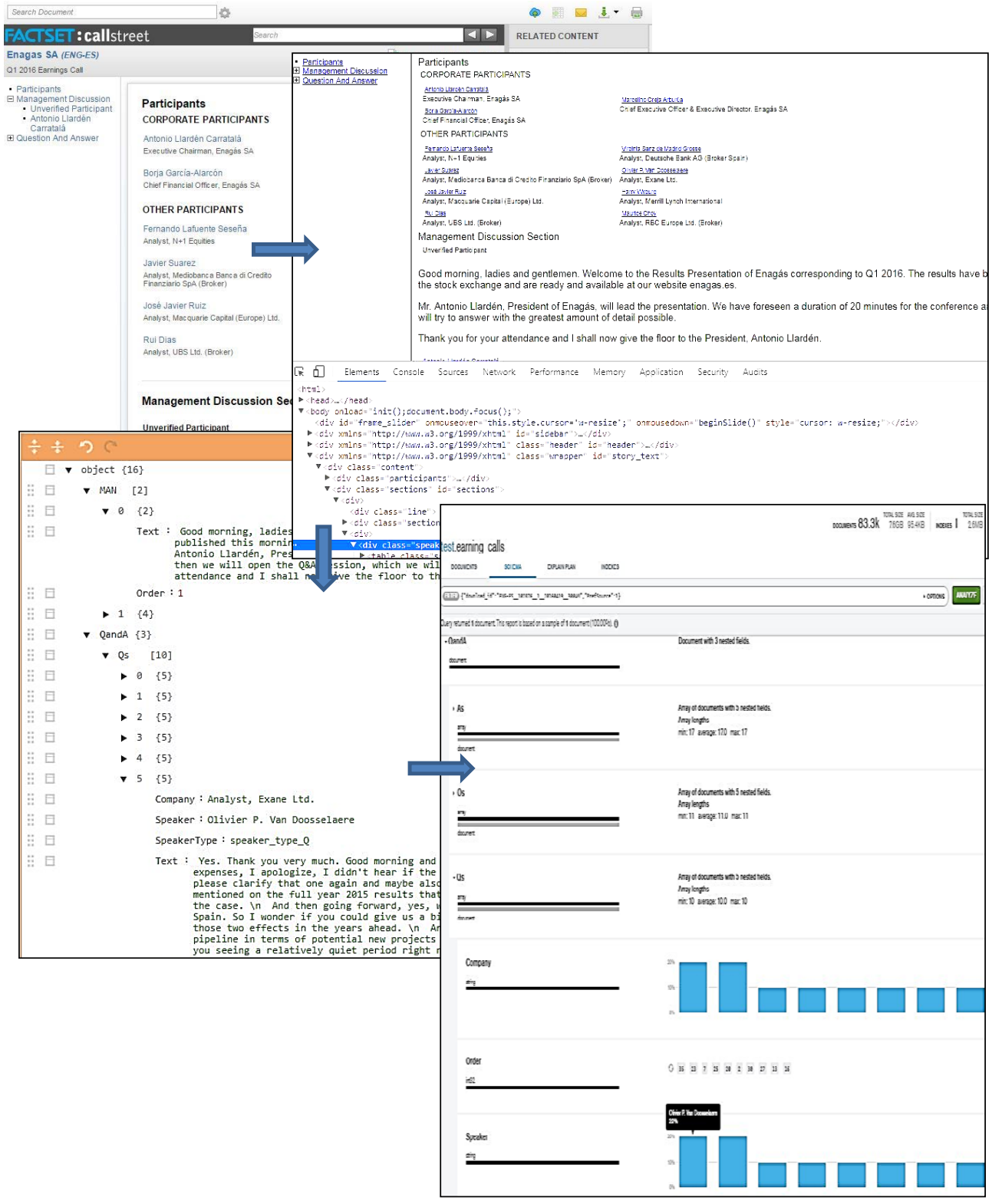
Depending on the analysis being undertaken, storing the original transcript and additional parsed tree information POS tags and dependent tree) may be of value; therefore the chosen persistence solution may need to also address these potential scaling issues.

Durability[11] of data is potentially an issue in a live setting, for our research tasks this was less of a problem, the NoSQL deployment we created did not guarantee this, and it's something that needs consideration when utilising in a production environment.

This briefly highlights some of the technological points to be taken into account for a textual analytics project and is by no means exhaustive, but serves as a guide based on the experienced we gained from this project.

[11] Durability refers to the concept that data transactions that have been committed to a data storage environment will survive permanently.

Figure 20. Principle transformation steps for call transcripts from origination to the NoSQL database.



Source: Citi Research, Factset, MongoDB Compass

Key Conclusions

The above description makes it sound much easier than it was. Again, it was an iterative process and exercise in document validation, identifying where the assumptions made about the structure of the document did not hold and in such cases, going to the individual document and examining how to rectify the problem. This iterative approach is acceptable when cleaning historical data. However, as mentioned, in a live environment unhandled instances are likely to occur requiring substantial data verification.

The more painful the process got, the more we realized that capturing such unstructured information may not be for every team, or even every organization.

For systematic investors, the use of such information is highly likely to require a data quality team (which we are aware already exists in many firms) or possibly an acceptance that (hopefully) a small proportion of positions will be impacted by erroneous data – which in diversified portfolios may matter little. Quantitative analysts maybe required to have a deeper understanding of individual data items and the possible errors contained in the data.

On the other side of the coin, fundamental investors arguably have a deeper understanding of company level fundamentals and the data meaning. Analysts interested in using unstructured data sources do however need to recognize the challenges of collecting such data and be patient through the learning curve, not simply losing confidence the first time an unintuitive number hits their desk.

It seems clear to us, that the more data that is collected from unstructured sources, the more time, effort and ultimately human-input are required to curate and validate its credentials. This is particularly true in the early years of collecting a new source of information. What is acutely important as well is a need to have a framework which enables comprehensive validation checks prior to any processing. As a result, gaining an information edge could become a more human intensive process resulting from mass automation of the collection of a myriad of data points rather than us all being replaced by robots. Should things transpire in this way, we postulate that the ‘alternative / unstructured data’ process will not be for every firm, with ‘off the shelf’ solutions that ultimately offer less unique information being deemed sufficient. Only time will tell.

Exploratory Data Analysis

The volume of information and metrics that can be derived from conference call content is significant and can be overwhelming at times. Our breakdown of the transcripts enables us to transverse conference calls by firm, participant, region/sector/country, time period, periodicity, call segments (presentation, question, answer) and textual features.

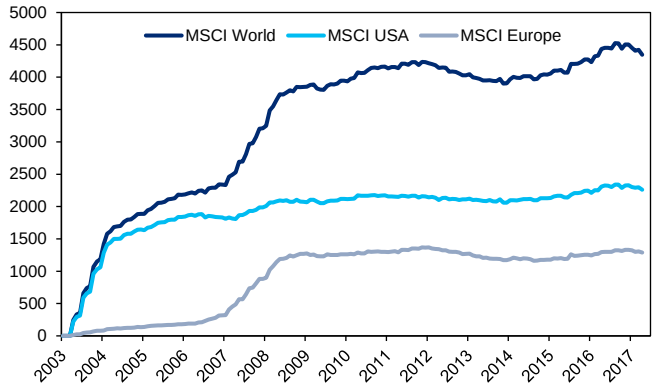
We begin by providing an overview of the content coverage through time by region, periodicity, country and sector before examining textual characteristics which can be used to make inferences about future stock performance.

Coverage

Figure 21 shows the rolling last twelve months cumulated earnings calls by region; we initiate our coverage from 2003. Factset primarily focus on capturing transcripts for large cap companies, small cap company transcripts are available, however this is captured by client request possibly generating a selection bias.

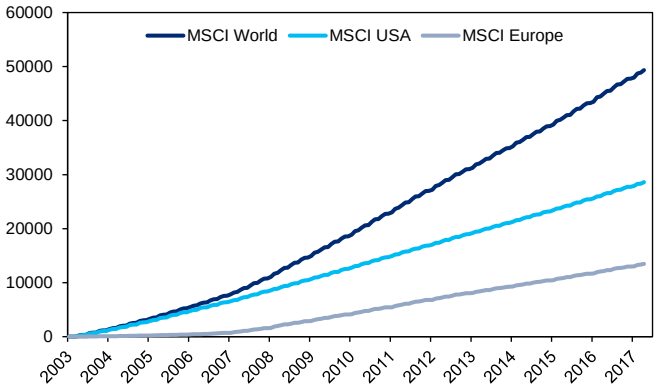
Figure 22 outlines the cumulative number of historical earnings calls we've captured from FactSet CallStreet. These represent conference calls for all constituents that have been part of the MSCI World benchmark, in English, which have been successfully parsed. The majority of prior academic work in the field has focused on the US using varying amounts of transcripts over different date ranges.

Figure 21. Last Twelve Months Rolling Conference Calls



Source: Citi Research, Factset, MSCI

Figure 22. Cumulative Conference Calls By Region



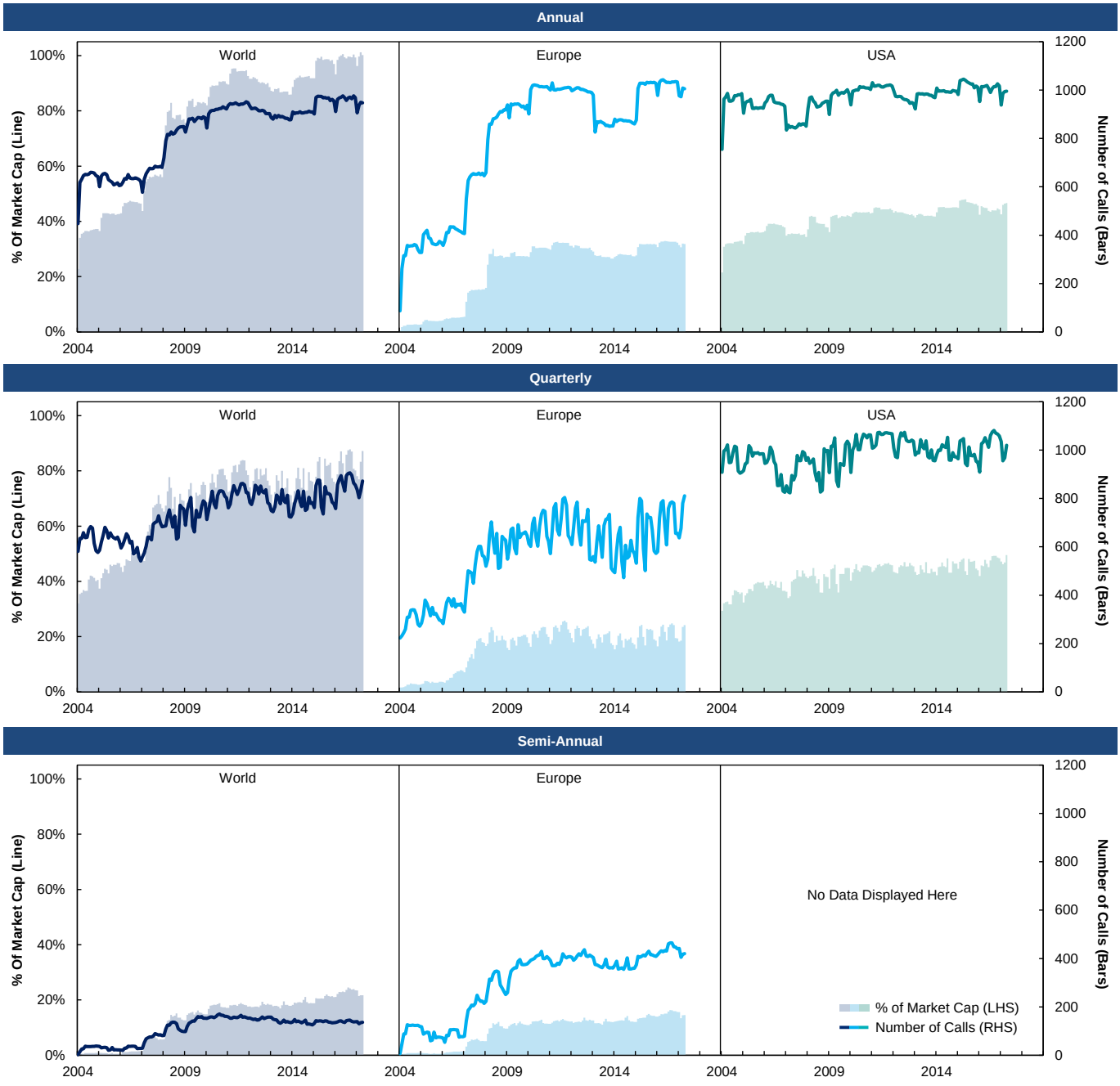
Source: Citi Research, Factset, MSCI

Periodicity

The number of earnings calls a company makes is clearly dependent on the number of times they announce earnings with different regions having different reporting frequencies. This presents a challenge to expanding our analysis on global equities but also offers insight into cross regional differences.

Figure 23 shows call coverage for annual and interim earnings calls using the counts of transcripts downloaded aligned to WorldScope reporting periods over the previous 12 months averaged by the number of constituents within each universe. Coverage is over 80% for MSCI World, US and Europe after 2008 which is clearly broad enough to be an interesting source of data. Before 2008, US coverage is strong whereas Europe and Global coverage begins to dwindle.

Figure 23. Summary Earnings Call Coverage by Periodicity and Region



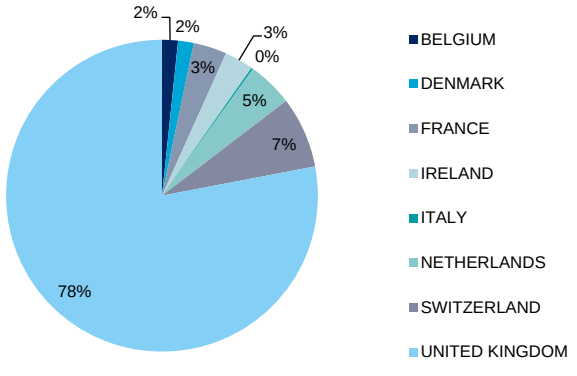
Source: Citi Research, Factset, MSCI

Note: We compute market cap coverage of conference calls using the number of unique call transcripts captured and parsed over the past <ndays> for universe constituents in the current period, where ndays is defined to be 91, 182, 365 for quarterly, semi and annual periods.

When examining interim calls there are clear geographic differences. In North America, quarterly earnings releases and conference calls are common with financial analysts forecasting quarterly interim numbers in IBES.

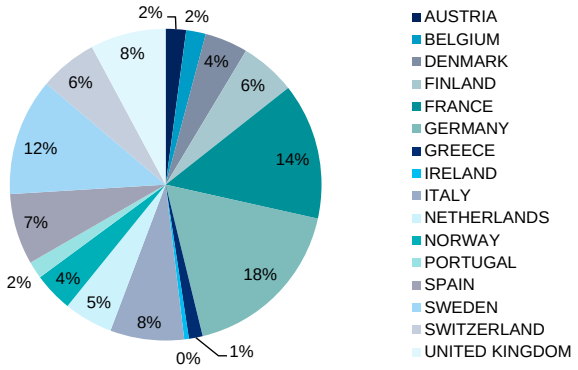
For much of continental Europe, quarterly updates are also common however, financial analyst coverage for interim earnings estimates is low limiting the level of analysis that can be performed on interim numbers. Finally some countries, such as the UK and Australia, follow a semiannual earnings update calendar which naturally feeds into semi-annual earnings calls.

Figure 24. MSCI Europe Semi-Annual Calls by Country



Source: Citi Research, Factset, MSCI

Figure 25. MSCI Europe Quarterly Calls by Country



Source: Citi Research, Factset, MSCI

As earnings calls are generally dictated by reporting frequency, it's unsurprising, that we find the majority of UK-listed firms hosting semiannual earnings calls. The requirement for reporting by the FCA has varied over time, and in 2007[12], firms were required to start issuing quarterly reports. However, these obligations were subsequently revised in 2014. As a consequence we observe a large discrepancy of reporting and earnings call hosting in the UK.

The differing frequency of earnings calls has significant implications for the persistence in alpha that earnings call signals need to have. As previously discussed, the earnings transcripts only become available 2-3 days after the call takes place. They are therefore unlikely to help in any 'post earnings announcement drift' based analysis which has a short duration. Instead, we focus on uncovering a longer duration signal, trying to find signals which persist over interim reporting periods which are suitable for large institutional investors. Regions that report quarterly can have a shorter alpha decay than those which report less frequently.

Interim Information

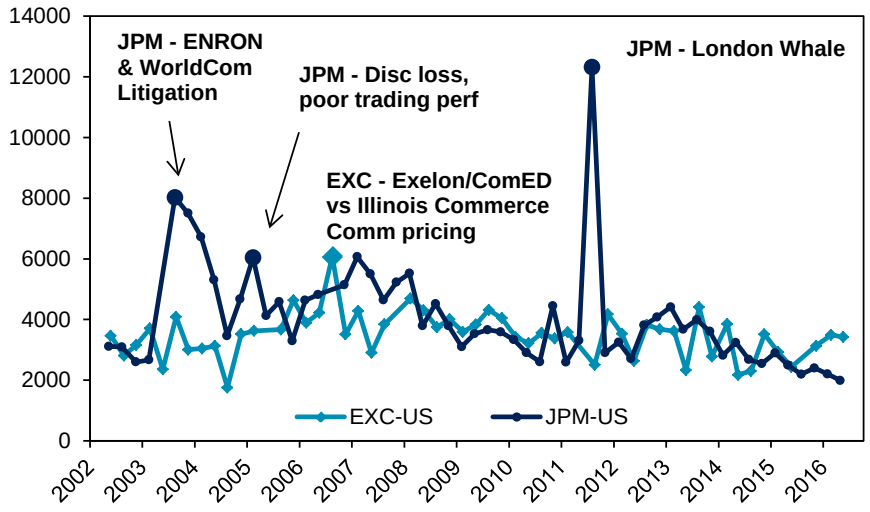
Along with the data coverage being affected by earnings reporting frequency the informational content of interim calls in comparison to full year calls is interesting. Comparing words and sentence counts in interim calls, we find that both of these metrics are lower in interim vs annual calls. In the US the mean and median word counts are marginally lower (-10%), In Europe, interims are on average 37% percent less than annuals. Seasonality is more pronounced in European earning calls with much greater call and sentence length in Q4, presumably due to the greater information overlay with year-end calls. We outline a summary of the call word/sentence counts in the descriptive statistics summary in the Appendix.

[12] The Disclosure and Transparency Rules (DTR;FSA 2006), required all UK public companies to issue interim statements for first and third quarters, in addition to annual and semi-annual financial statements.

Disclosure

We utilise word counts from transcripts as a measure of disclosure. Call disclosure appears higher in Europe relative to the US, across the primary segments of calls (presentation, answers). This bias is likely to be due to the greater number of firms in our corpus within Europe which only report annually. Although aggregate call lengths appear stable through time, there is considerable variation across firm periods. We include as an illustration quarterly presentation word counts of JP Morgan Chase & Co (JPM-US) contrasted with Exelon Corporation (EXC-US).

Figure 26. JPM vs Exelon Total Transcript Length Comparison



Source: Citi Research, Factset

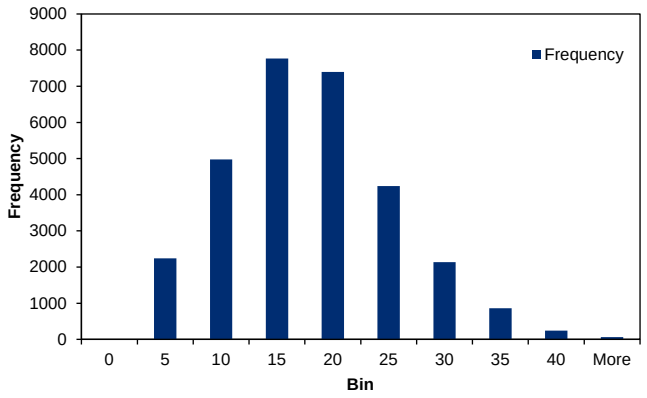
Disclosure lengths can vary quite significantly from quarter to quarter. Certainly for most financial firms, the (07-08) macro-economic environment warranted greater disclosure from banks, whereas utilities were fairly insulated from this. As can be seen from Figure 26, the macro-economic environment is just one of many factors to could determine call length, and isn't always as substantial as company specific events which potentially necessitate greater disclosure.

Other Information

Another issue we wanted to underline was the lack of analyst coverage in Europe for interim estimate measures. Figure 27 and Figure 28 highlights the disparity between the numbers of analyst providing interim estimates for Europe versus US. Not only are the coverage numbers small, they have remained static over the past few years. One possible explanation for this could be the IBES 'freshness' policy, which excludes estimates from analysts if they have not been revised over a period of time.

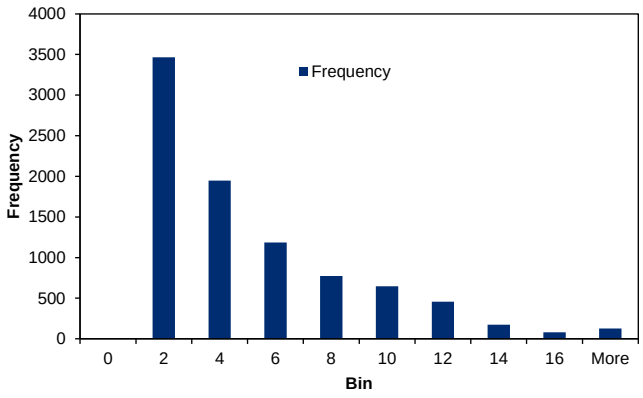
Following on from the above, it seems intuitive to us to use quarterly earnings calls for US analysis. Interim calls are of similar length to annual calls (circa 10% shorter) and our source of analyst estimates is well covered for interim numbers.

Figure 27. No of Contributing Analyst to IBES Interim Consensus Estimates – MSCI US



Source: Citi Research

Figure 28. No of Contributing Analyst to IBES Interim Consensus Estimates – MSCI Europe



Source: Citi Research

For Europe, it seems most intuitive to only examine annual earnings calls. Interim calls carry significantly less information and our analyst estimates data source has scant coverage for interim numbers. Furthermore, focusing on full year numbers resolves the challenges associated to different countries within Europe having different reporting frequencies.

Sector Coverage

The following tables highlight the sector coverage for annual conference calls globally and in US and Europe, these metrics are constructed using the counts of transcripts downloaded over the previous 12 month period divided by the average number of constituents within each sector/universe.

Figure 29. MSCI World (Annual) conference call coverage

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Cons. Disc	26%	28%	32%	28%	45%	60%	63%	68%	66%	59%	62%	69%	67%	65%
Cons. Staples	22%	27%	26%	31%	49%	59%	65%	71%	71%	62%	67%	70%	68%	74%
Energy	43%	51%	49%	51%	60%	71%	71%	76%	78%	70%	77%	77%	77%	80%
Financials	23%	28%	27%	31%	43%	53%	59%	63%	66%	65%	65%	72%	73%	100%
Health Care	40%	43%	44%	48%	64%	74%	82%	79%	76%	71%	76%	82%	81%	73%
Industrials	17%	20%	16%	23%	40%	53%	56%	61%	59%	61%	60%	64%	65%	65%
Info Tech	41%	46%	42%	43%	55%	57%	66%	74%	75%	69%	71%	73%	72%	70%
Materials	16%	19%	21%	34%	49%	62%	61%	65%	66%	67%	62%	72%	74%	71%
Telco Serv	27%	34%	38%	53%	82%	79%	73%	84%	85%	69%	77%	73%	90%	78%
Utilities	31%	29%	26%	30%	46%	58%	60%	63%	56%	59%	59%	63%	63%	71%

Source: Citi Research, Factset, MSCI

Figure 30. MSCI USA (Annual) Conference Call Coverage

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Cons. Disc	79%	72%	86%	59%	75%	90%	89%	96%	88%	82%	90%	92%	87%	88%
Cons. Staples	82%	85%	76%	71%	74%	72%	81%	91%	89%	79%	89%	86%	81%	88%
Energy	85%	80%	77%	79%	76%	76%	76%	84%	83%	75%	85%	91%	79%	84%
Financials	68%	78%	73%	63%	73%	81%	83%	85%	91%	85%	81%	88%	83%	100%
Health Care	83%	81%	77%	68%	77%	87%	91%	89%	87%	83%	87%	90%	83%	75%
Industrials	81%	87%	78%	66%	77%	82%	79%	83%	77%	82%	85%	87%	81%	87%
Info Tech	82%	89%	82%	71%	82%	76%	87%	94%	91%	86%	88%	90%	84%	84%
Materials	87%	77%	89%	62%	84%	84%	85%	91%	88%	79%	79%	85%	85%	88%
Telco Serv	89%	88%	73%	58%	69%	69%	54%	54%	58%	60%	80%	67%	88%	78%
Utilities	79%	73%	61%	56%	75%	77%	78%	81%	69%	82%	74%	76%	84%	84%

Source: Citi Research, Factset, MSCI

Figure 31. MSCI Europe (Annual) Conference Call Coverage

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Cons. Disc	4%	6%	7%	19%	50%	74%	84%	89%	86%	70%	67%	85%	86%	73%
Cons. Staples	7%	14%	13%	30%	55%	75%	82%	82%	75%	70%	67%	74%	79%	88%
Energy	13%	28%	18%	43%	60%	88%	100%	100%	89%	75%	76%	91%	79%	76%
Financials	3%	6%	7%	27%	46%	53%	69%	79%	82%	78%	80%	85%	88%	100%
Health Care	16%	19%	22%	43%	79%	93%	100%	90%	89%	77%	77%	92%	100%	82%
Industrials	1%	4%	3%	21%	53%	72%	82%	84%	84%	85%	78%	84%	92%	82%
Info Tech	13%	14%	19%	38%	68%	75%	75%	100%	100%	80%	88%	88%	79%	72%
Materials	2%	6%	8%	40%	66%	92%	80%	82%	79%	86%	82%	98%	95%	88%
Telco Serv	9%	21%	26%	62%	91%	81%	82%	100%	100%	61%	78%	75%	95%	85%
Utilities				15%	41%	72%	80%	76%	78%	67%	76%	86%	82%	95%

Source: Citi Research, Factset, MSCI

These charts show broad coverage across sectors from 2008 making our results widely applicable, particularly after 2008. Before 2008, when coverage gets more challenging, there is still a broad representation across sectors with European utilities being the only unrepresented sector at the beginning of the sample.

Sector Biases

As previously highlighted sector coverage varies through time, but are there other sector characteristics which we should account for when analysing transcript content? Figure 32 and Figure 33 compares presentation length ranked by sector on a yearly basis. Presentation length (word count) is fairly stable in both regions, although there are variations by sector, for instance in Europe presentation length for Industrials is the lowest across all sectors for almost all years, whereas in the US it one of the highest across sectors.

Figure 32. MSCI USA Sector – Word Count Rank

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Avg
Financials	4	7	9	10	10	9	10	10	7	9	10	8	10
Industrials	10	9	8	6	8	8	7	7	9	10	4	6	9
Info Tech	9	8	2	4	9	10	9	9	6	8	7	9	8
Health Care	7	6	7	8	4	7	8	8	8	7	8	7	7
Cons Staples	6	4	4	9	5	5	6	6	10	5	9	10	6
Utilities		2	10	5	7	6	3	1	2	6	3	5	5
Cons Disc	5	5	6	7	3	2	4	2	5	2	5	4	4
Energy	2	10	3	2	6	4	5	3	4	3	1	2	3
Telco Srvs	3	3	5	3	2	3	2	5	1	4	2	1	2
Materials	1	1	1	1	1	1	1	4	3	1	6	3	1

Source: Citi Research, Factset, MSCI

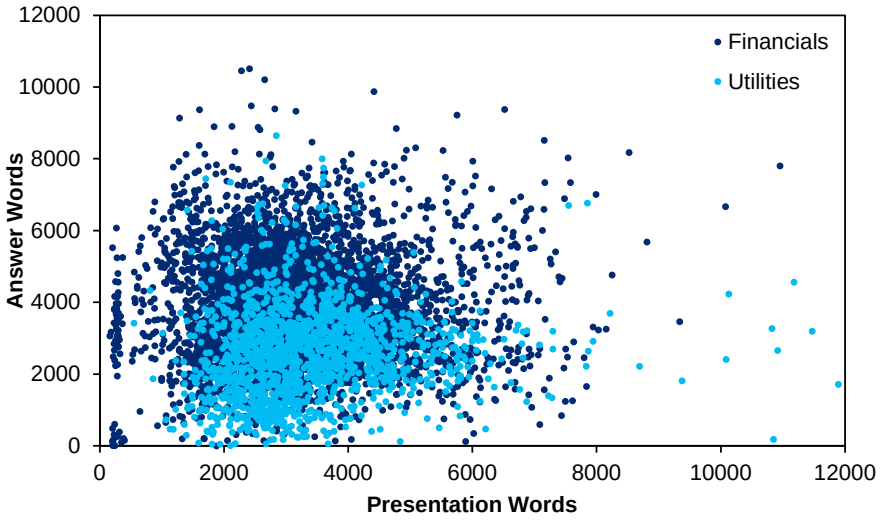
Figure 33. MSCI Europe Sector – Word Count Rank

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Avg
Cons Disc	9	9	9	9	9	8	9	10	8	10	10	10	10
Financials	6	7	10	10	10	10	10	9	7	8	9	8	9
Cons Staples	8	8	6	5	8	9	6	7	2	2	2	3	8
Info Tech	7	6	8	8	6	6	3	6	5	5	5	9	7
Utilities		10	2	6	2	1	8	5	3	4	6	4	6
Telco Srvs	5	5	3	3	1	4	4	8	6	6	4	1	5
Materials	1	1	7	7	7	7	7	4	4	7	8	7	4
Health Care	4	2	4	2	4	5	5	3	9	9	7	6	3
Energy	3	3	1	4	5	3	1	2	10	3	1	5	2
Industrials	2	4	5	1	3	2	2	1	1	1	3	2	1

Source: Citi Research, Factset, MSCI

Figure 34 provides an alternative view of this; the scatter incorporates presentation and answer segments between Financials and Utility firms in the US. We do not observe this pattern across all sectors and it's more pronounced within the US than in Europe. Logically we would expect some sector/industry related bias simply as a consequence of the difference in business operations.

Figure 34. MSCI USA Quarterly Call Disclosures – Financials compared to Utilities



Source: Citi Research, Factset, MSCI

To demonstrate the need to control for sector biases, we combine the findings from Matsumoto [2006] on call disclosure with the primary determinants utilised by Li [2008]'s study on readability and earnings persistence and overlay this incorporating sector fixed effects across on US (Quarters) and Europe (Annual) earning calls as determinants for call disclosure.

In our combined regression setting we include the variables, MISS, an indication a firm has missed consensus targets, Size the log market capitalization, Book/Market as a valuation measure, Price Momentum (3 months/12 months), the stock price performance since last interim/annual call, year and sector fixed effects. In addition year fixed effects are introduced for handle macro-economic conditions, e.g. boom and recessionary years which may influence the language and length of calls. All variables are winsorised between 1 and 99% to control for outliers; we ran our analysis using raw values of word counts in each segment, the regressions were also run using rank percentile word counts to handle for non-linearity but there was negligible change. Figure 35 and Figure 36 outline the results from running the panel regression in Europe (annual) and US (quarter) earnings calls.

Figure 35. MSCI USA (Quarterly) Presentation and Answer Word Count Determinants

	Presentation Word Count						Answer Word Count					
	No Dummy Year Added		Dummy Year Added		Including Sectors and Dummy Year		No Dummy Year Added		Dummy Year Added		Including Sectors and Dummy Year	
	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
MISS	48.95**	2.57	48.78**	2.57	59.72***	3.14	44.20**	2.4	41.22**	2.24	92.64***	5.1
Size	163.50***	20.18	198.19***	23.83	189.90***	22.66	129.25***	16.43	105.70***	13.04	95.29***	11.86
BookMarket	5.03*	1.76	5.28*	1.86	7.16**	2.54	2.15	0.79	1.93	0.71	2.88	1.08
Mom (3M)	-4.64***	-5.77	-4.89***	-5.9	-5.06***	-6.15	-3.05***	-3.93	-3.31***	-4.13	-2.66***	-3.39
Year	No		Yes		Yes		No		Yes		Yes	
Consumer Staples					54.58	1.5					50.2	1.44
Energy					-237.15***	-6.6					-234.78***	-6.87
Financials					-256.05***	-9.23					85.07***	3.21
Health Care					254.42***	7.88					104.52***	3.4
Industrials					-260.96***	-8.33					129.58***	4.33
Information Technology					-185.77***	-6.39					105.42***	3.81
Materials					-255.85***	-6.4					-20.68	-0.54
Telecommunication Services					-99.73	-1.42					-247.91***	-3.69
Utilities					-8.75	-0.22					-1,063.59***	-28.03
Constant	-460.63**	-2.43	-993.09***	-5.05	-700.00***	-3.54	727.85***	3.96	1,172.66***	6.12	1,429.04***	7.55
Observations	26685		26685		26685		26358		26358		26358	
R2	0.02		0.03		0.04		0.01		0.02		0.06	
Adjusted R2	0.02		0.03		0.04		0.01		0.02		0.06	
F Statistic	106.25***		42.66***		44.18***		69.60***		26.54***		63.26***	

Note: p*, p**, p*** < 0.01

Source: Citi Research, Factset, MSCI

Figure 36. MSCI Europe (Annual) Presentation and Answer Word Count Determinants

	Presentation Word Count						Answer Word Count					
	No Dummy Year Added		Dummy Year Added		Including Sectors and Dummy Year		No Dummy Year Added		Dummy Year Added		Including Sectors and Dummy Year	
	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
MISS	-123.78	-1.55	-129.59	-1.62	-90.89	-1.14	-18.36	-0.36	-30.2	-0.6	-41.58	-0.82
Size	376.21***	10.21	385.88***	10.25	353.64***	9.12	453.97***	19.36	457.64***	19.09	467.67***	18.9
BookMarket	0.01	0.09	0.03	0.17	0.04	0.25	0.15	1.6	0.14	1.52	0.15	1.58
Mom (3M)	-7.61***	-4.37	-8.93***	-4.36	-9.28***	-4.53	-4.84***	-4.38	-5.10***	-3.9	-5.47***	-4.17
Year	No		Yes		Yes		No		Yes		Yes	
Consumer Staples					68.33	0.4					-382.03***	-3.49
Energy					-310.81	-1.57					26.29	0.21
Financials					-355.19***	-2.63					-289.46***	-3.36
Health Care					420.57**	2.33					-81.04	-0.7
Industrials					-146.08	-1.07					-143.09	-1.63
Information Technology					-487.68**	-2.32					-435.29***	-3.26
Materials					-921.96***	-5.8					-354.65***	-3.49
Telecommunication Services					184.66	0.92					-239.87*	-1.87
Utilities					109.5	0.54					-682.36***	-5.3
Constant	-3,610.06***	-4.2	-4,223.67***	-3.98	-3,358.58***	-3.12	-6,625.82***	-12.11	-7,147.51***	-10.61	-7,165.02***	-10.46
Observations	3961		3961		3961		3877		3877		3877	
R2	0.03		0.04		0.05		0.09		0.1		0.11	
Adjusted R2	0.03		0.03		0.05		0.09		0.09		0.1	
F Statistic	27.80***		8.29***		8.41***		94.26***		23.04***		17.37***	

Note: p*, p**, p*** < 0.01

Source: Citi Research, Factset, MSCI

Adding sector fixed effects improved our fit for both presentation and answers for both regions. Surprisingly this impact was more dominant than fiscal year controls.

Other factors affecting the number of words include the size of the firm, and book/market and missing consensus; the latter was more applicable in the US than Europe.

Country

Figure 37 outlines conference call coverage by country. We observe increasing numbers of conference calls across all countries consistent with FactSet downloaded data however there appears to be persistently fewer calls in Asia (Singapore, Hong Kong, Japan) in comparison to the rest of developed World.

In one regard this is reassuring, as large developed non-English speaking nations such as Germany and France are well covered by the data. However, the lack of coverage in Asia does warrant further investigation.

Our Head of Quant Research in Asia confirmed that earning calls occurred less often in that region, and weren't always conducted in English; equally important was Factset's poor collection of transcripts in Asia. This was re-affirmed by Factset, who indicate their call coverage is driven primarily by market cap and client demand. As their client interest/base diversifies, this could lead to improving coverage in this region in the future.

Figure 37. MSCI World Conference Call Transcript Coverage By Country

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Trend
Australia	3%	0%	0%	5%	17%	49%	62%	82%	87%	87%	86%	87%	92%	97%	
Austria	0%	0%	8%	18%	25%	44%	63%	63%	50%	33%	50%	57%	14%	60%	
Belgium	0%	0%	5%	20%	18%	43%	62%	57%	50%	45%	36%	45%	55%	80%	
Canada	13%	26%	21%	30%	55%	66%	62%	58%	64%	63%	65%	64%	68%	77%	
Denmark	0%	0%	5%	23%	52%	87%	92%	100%	92%	91%	91%	100%	100%	100%	
Finland	5%	5%	10%	22%	70%	94%	94%	88%	100%	79%	92%	100%	100%	92%	
France	2%	11%	10%	41%	59%	53%	76%	87%	89%	67%	79%	77%	84%	75%	
Germany	11%	19%	20%	29%	75%	76%	84%	81%	84%	67%	78%	89%	76%	86%	
Greece	0%	0%	0%	22%	33%	36%	50%	63%	100%	50%					
Hong Kong	0%	0%	0%	0%	7%	9%	22%	15%	17%	24%	28%	33%	40%	29%	
Ireland	8%	7%	0%	13%	43%	100%	100%	100%	100%	75%	100%	100%	100%	80%	
Israel								57%	31%	50%	44%	11%	25%	23%	
Italy	0%	0%	0%	28%	43%	51%	76%	72%	69%	64%	74%	73%	84%	65%	
Japan	0%	1%	0%	1%	3%	8%	10%	14%	16%	14%	16%	19%	22%	23%	
Netherlands	16%	35%	35%	58%	96%	85%	70%	90%	86%	82%	70%	87%	92%	79%	
New Zealand	0%	0%	0%	0%	10%	20%	0%	20%	60%	60%	60%	57%	71%	43%	
Norway	0%	7%	0%	26%	50%	100%	100%	100%	80%	100%	100%	100%	100%	100%	
Portugal	0%	0%	0%	9%	27%	70%	56%	67%	57%	80%	80%	100%	100%	33%	
Singapore	0%	0%	2%	8%	11%	26%	31%	29%	28%	35%	50%	52%	50%	48%	
Spain	0%	0%	3%	13%	55%	48%	59%	64%	52%	61%	59%	78%	84%	92%	
Sweden	2%	7%	4%	26%	63%	90%	100%	91%	97%	84%	87%	94%	100%	81%	
Switzerland	16%	11%	16%	24%	56%	66%	73%	65%	69%	79%	42%	76%	80%	71%	
United Kingdom	5%	9%	10%	31%	52%	93%	95%	100%	99%	92%	88%	91%	98%	89%	
USA	80%	82%	80%	68%	79%	83%	85%	91%	88%	84%	87%	90%	87%	88%	

Source: Citi Research, Factset, MSCI

Country Biases

We perform a similar analysis to that undertaken within sectors. It would not be surprising to see some kind of country-specific nuance with earnings calls. However, the level of significance and whether it overshadows sector effects is intriguing. Figure 38 exhibits the word count by presentation segment, ranked on a yearly basis.

Figure 38. MSCI Europe Country – Word Count Rank

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Avg
Finland	10	15	16	16	16	16	16	15	15	15	9	14	16
Greece		12	14	15	14	15	15	16	16				15
Austria	11	16	15	13	10	9	12	14	11	13	15	12	14
Sweden	2	13	13	14	15	14	14	13	14	12	11	11	13
Denmark	7	11	12	11	11	11	11	12	12	10	10	9	12
Portugal		14	10	12	12	10	9	8	9	7	8	4	11
Norway		9	11	10	13	13	6	11	10	6	4	7	10
Italy		3	4	7	9	12	13	10	6	14	14	10	9
Belgium	9	1	8	8	7	7	10	6	12	9	12	15	8
Netherlands	8	7	7	2	5	4	8	7	8	11	13	13	7
Spain	3	10	9	9	6	3	4	5	4	4	7	6	6
Ireland		2	5	4	8	8	7	3	5	8	6	3	5
Germany	4	5	3	5	4	5	5	9	7	5	5	8	4
Switzerland	5	6	2	1	2	2	3	4	1	3	2	5	3
United Kingdom	6	4	6	6	3	6	2	2	3	1	1	1	2
France	1	8	1	3	1	1	1	1	2	2	3	2	1

Source: Citi Research, Factset, MSCI

The highest rank value reveals the shortest call presentation word length. Word counts are shortest in Finland, and longest in France. The key message here is that they seem consistent and fairly stable over the years, which suggest some kind of country specific distinction. Applying a country fixed effect onto our model results in the following.

Figure 39. MSCI Europe (Annual) Presentation and Word Count with Country Dummies

	Presentation Word Count				Answer Word Count			
	No Country Dummy		Country Dummy		No Country Dummy		Country Dummy	
	Coef	t-stats	Coef	t-stats	Coef	t-stats	Coef	t-stats
MISS	-90.89	-1.14	35.9	0.48	-41.58	-0.82	-4.17	-0.09
Size	353.64***	9.12	184.63***	4.99	467.67***	18.9	406.39***	16.65
BookMarket	0.04	0.25	0.08	0.55	0.15	1.58	0.11	1.2
Mom (12M)	-9.28***	-4.53	-8.72***	-4.58	-5.47***	-4.17	-5.23***	-4.16
Year	Yes		Yes		Yes		Yes	
Sector	Yes		Yes		Yes		Yes	
Country			Yes				Yes	
Constant	-3,358.58***	-3.12	-904.04	-0.86	-7,165.02***	-10.46	-6,537.92***	-9.47
Observations	3,961		3,957		3,877		3,874	
R2	0.05		0.19		0.11		0.19	
Adjusted R2	0.05		0.19		0.1		0.18	
F Statistic	8.41***		22.02***		17.37***		21.03***	

Note: p*, p**, p*** < 0.01

Source: Citi Research, Factset, MSCI

Figure 39 presents the output of regressing presentation and answer segments against our base determinants, regressions (2) and (4) include country effect, result in a considerable increase in model fit, and do not reduce the influence of size or momentum on both segments of the transcripts.

Given the aforementioned differences in earnings calls across both sectors and countries we perform all analysis using signals which are sector and country neutral.

Time of Call

In addition to extracting the textual and participant details of conference calls, another interesting facet available to us from earnings call transcript data is the date and time of conference calls. Of primary significance is the timeliness of the calls relative to the earnings number. Having large misaligned calls relative to earnings announcements, could potentially add another aspect in complicating analysis around price action on the back of earnings. Fortunately, this is less of a concern, as highlighted below; most earnings calls take place on announcement date with over 99% of our calls taking place by the end of the second working day.

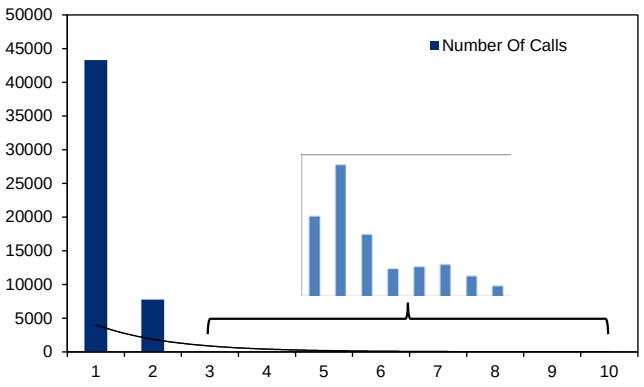
However, it does raise some interesting questions, what dictates the timings of earnings calls, and is there a hidden agenda behind this?

Figure 40. Table - Days between earnings announcement and conference calls

Days From Announcement	Number Of Calls	Cumulative Percentage
0	43299	84.0%
1	7762	99.1%
2	79	99.2%
3	130	99.5%
4	61	99.6%
5	27	99.6%
6	29	99.7%
7	31	99.7%
8	20	99.8%
9	10	99.8%

Source: Citi Research, Factset, MSCI

Figure 41. Chart - Days between earnings announcement and conference calls



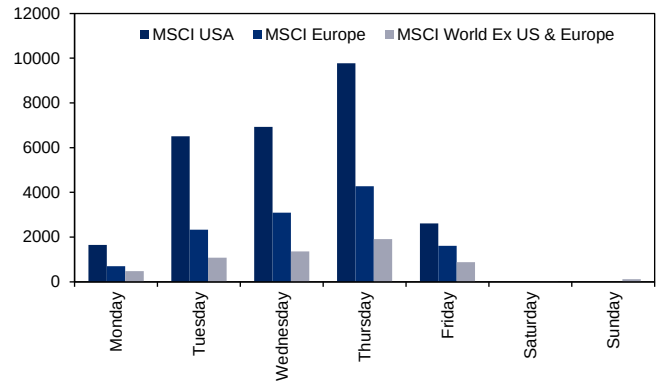
Source: Citi Research, Factset, MSCI

Across all regions, there appears to be a preference to host earnings calls towards the middle of the week. Most firms appear to avoid Friday and Mondays.

The average earnings surprise (SUE)[13] score is lowest on Fridays and the number of companies missing consensus is highest relative to the numbers reporting. Our results are similar to a prior study conducted by Michaely et al (2016), which analyses the strategic timing of earnings calls and Druz et al (2016). We also observe similarities in earnings surprises aggregates, with poorer more concentrated scores reported on Fridays, in both regions.

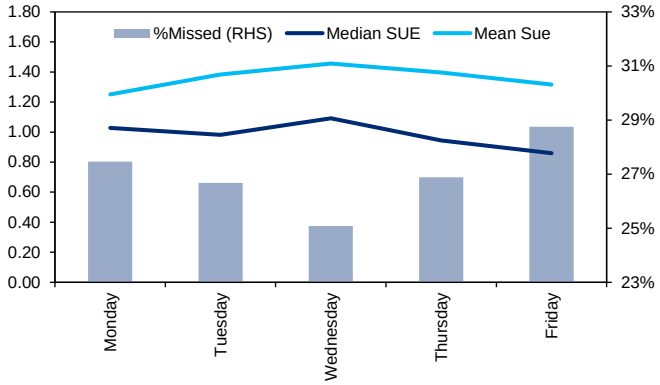
[13] SUE - We compute SUE by taking the difference between reported earnings and IBES consensus mean earnings estimates for the period, standardized by the deviation in consensus earnings.

Figure 42. Conference call day



Source: Citi Research, Factset, MSCI

Figure 43. Conference call day vs SUE



Source: Citi Research, Factset, MSCI, Thomson Reuters

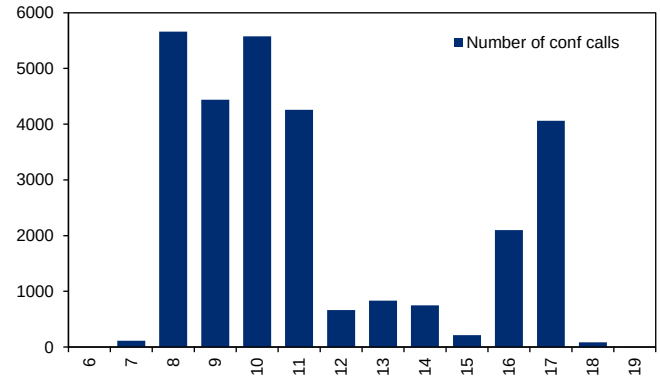
It is also worth highlighting that one of the challenges with collecting and using time-related artefacts using Factset CallStreet is that the time elements retrieved are specific to the timezone on the requesting workstation; these have to be adjusted to local timezones, depending on the host conference call location.

After accounting for this we find that the majority of calls take place either in the morning or post market close. However unlike the choice of conference call day, there is less of a pattern to the scheduling time for earnings calls.

That said, grouping calls by day and time slots, then sorting by earnings surprise scores (SUE) is quite revealing.

Results indicate that the lowest reported earnings surprise scores occur on Friday afternoons. Calls on Friday afternoons have the largest proportion of missed earnings relative to other time periods, suggesting that this may be strategically chosen as a good time to bury bad news.

Figure 44. Number of conference calls by time slots (MSCI USA) EST



Source: Citi Research, Factset, MSCI

Figure 45. Double sort timing of call and with day of the week

	Time (00-24)	MedianSUE	MeanSUE	IQR
Friday	1300-1700	0.56	0.92	2.62
Wednesday	1300-1700	0.68	1.00	2.74
Thursday	1300-1700	0.72	1.02	2.68
Monday	400-800	0.86	1.30	1.86
Tuesday	1300-1700	0.89	1.30	3.59
Friday	800-1200	0.89	1.16	2.42
⋮				
Thursday	400-800	1.12	1.62	3.04
Tuesday	1700-2100	1.13	1.46	2.30
Friday	400-800	1.27	1.20	2.84
Monday	1700-2100	1.31	1.75	2.99
Thursday	1700-2100	1.38	1.96	3.09
Wednesday	1700-2100	1.56	1.71	2.61

Source: Citi Research, Factset, MSCI, Thomson Reuters

Constructing a Call Transparency Signal

The remainder of the report focuses on analyzing earnings call transcripts with a particular focus on the clarity with which management communicate their story. This section follows the academic literature in using relatively rudimentary signals to infer information about both firm operations and management.

Going forward we focus on using MSCI USA and Europe benchmark constituents, as coverage remains stubbornly poor in other sub-regions. As discussed in the previous section, we use full year transcripts for European listed companies and quarterly earnings calls for US firms with our US data starting 2004 and our European data having sufficient coverage after 2008. We follow our standard backtesting procedure of splitting stocks into 5 portfolios rebalancing at month end. To avoid possible forward looking bias transcripts from earnings calls after the 25th of each month are not used until the following month.

All analysis is performed on signals which have been made both sector and country neutral to avoid structural biases as outlined in the previous section.

Readability

There is an extensive amount of literature that compares the complexities of written firm communication by looking at various characteristics within text. In widely cited research, Li [2008], focuses on obfuscation by management of poor results using the Fog readability measure in corporate filings.

Li [2008] finds that firms with lower reported earnings tend to have annual reports that are harder to read. However a contradictory finding by Bloomfield [2008], explains that this may be caused by poorly performing firms needing to have more text and longer sentences to fully explain their situation to investors.

Guay et al. [2015] find that companies with less readable annual reports (based on six different readability measures including the Fog Index) tend to mitigate this negative readability effect by issuing more managerial forecasts of earnings per share, sales, and cash flows.

Miller [2010] finds that small investors trade significantly fewer shares of firms with high Fog Index values and word counts (i.e., less readable annual reports) around the 10-K filing date.

De Franco, Hope, Vyas, and Zhou [2015] analyze a sample of over 365,000 broker annual reports issued during 2002-2009 for readability characteristics. One of their readability measures is an aggregation of three different readability indexes (Fog, Flesch, and Flesch-Kincaid) finding higher ability analysts write more readable reports generating a larger volume reaction in the market.

Although readability stats have been primarily aimed at written media in the past, it has also been applied to spoken communication, and most recently in the analysis of political statements where Schumacher [2016] et al create a readability model looking at grammatical constructs at grade level to analyse spoken language.

The motivations for management to obfuscate in conference calls are similar to those of annual filings, the use of complex language in calls is aimed primarily to increase the processing time and reduce transparency, in the hope that obfuscation results in a lower price impact and impairs analysts' ability to provide a true valuation of a firm. Alternatively, complicated language could be indicative of management struggling to cast operational performance in a positive light.

Conversely simple and concise management communication implies operational performance requires less gloss, as well as offering analysts clarity, being positive for stock performance.

To measure complexity of language we use the below measures taken from the computational linguistics literature inverted where required such that larger numbers represent easier text.

Flesch-Kincaid

Flesch-Kincaid is a test that measures and grades the readability of text developed in the 70s, and has found proliferated usage across a number of industries. Flesch-Kincaid metrics consists of two component formulas.

The Flesch reading-ease test calculates a score based upon the ratio of total words to sentences and syllables to words.

Flesch reading-ease is defined as follow,

$$206.835 - 1.015\left(\frac{total\ words}{total\ sentences}\right) - 84.6\left(\frac{total\ syllables}{total\ words}\right)$$

By definition higher scores indicate that the text segment is easier to read.

Flesch-Kincaid grade provides an alternative metric which measures readability in a manner that presents a grade score, it is a proxy to the years of education generally required to understand a text. With this measure higher the grade level, the harder a text is to understand.

Flesch-Kincaid grade is defined as follow,

$$0.39\left(\frac{total\ words}{total\ sentences}\right) + 11.8\left(\frac{total\ syllables}{total\ words}\right) - 15.59$$

Notice that both the read-ease and grade versions use the same underlying metrics of total words to total sentences and total syllables to total words but with different weightings.

Fry

The Fry readability score is an estimation of how difficult a text is to read based on randomly sampling three lengthy sections within a segment of text. (> 100 words).

Once selected, two metrics are computed, the average number of sentences per 100 words (fry_sent) and the average number of syllables per 100 words (fry_syl) over the three sections. Intuitively, more sentences per 100 words imply sentences are shorter, and the text is simpler. Conversely, more syllables per 100 words suggest more complexity.

Gunning Fog

We include the Gunning-Fog index, which is amongst the more widely cited measures utilised in complexity scoring financial disclosures. Li [2008], Biddle, Hillary and Verdi [2009] and Franco,Hope,Vyas and Zhou [2015] all use this benchmark

Fog is defined as:

$$0.4\left[\left(\frac{total\ words}{sentences}\right)+100\left(\frac{complex\ words}{total\ words}\right)\right]$$

With complex words being defined as words containing three or more syllables. Larger scores represent harder to read text.

Coleman-Liau

To complete our list of readability measures, we incorporate Coleman-Liau's metric of readability. Unlike the other indices Coleman-Liau avoids relying on syllable counts returning a grade level score, from 1 to 12. At the time of its creation syllable detection was rather primitive based on a set of functional rules and as such, not to be relied on.

Coleman-Liau is defined as follow

$$0.0588\left(\frac{total\ letters}{total\ words}*100\right)-0.296\left(\frac{total\ sentences}{total\ words}*100\right)$$

Following our standard backtesting framework, the following tables show the performance of investing in firms which use simple language in earnings call applied to both the presentation and management answer section of the earnings call.

Results are broadly positive with the majority of measures providing positive risk adjusted performance. It is noticeable that performance appears to be stronger in the presentation section rather than the answer section. This is a little surprising as the presentation section is pre-prepared whereas the answer section is not.

One explanation is that this work uses linguistic measures designed for written text. Prepared presentations are intuitively going to more closely resemble written text than off the cuff answers to questions. Another is simply that linguistic biases are not (yet) a focus of management meaning they transcend rehearsed and unrehearsed answers.

Figure 46. EU Summary Statistics

	Presentation						Answers					
	FK_ReadEase	FK_Grade	FRY_Sent	FRY_Syll	FOG	COLE	FK_ReadEase	FK_Grade	FRY_Sent	FRY_Syll	FOG	COLE
Return	2.8%	3.8%	1.8%	3.4%	-1.3%	2.1%	2.7%	3.7%	0.6%	1.2%	-3.8%	1.3%
Risk	4.9%	4.9%	4.7%	5.4%	5.0%	4.8%	6.4%	5.5%	3.7%	4.8%	5.3%	6.0%
IR	**0.58**	**0.78**	**0.39**	**0.63**	**-0.27**	**0.43**	**0.43**	**0.69**	**0.15**	**0.25**	**-0.71**	**0.22**
Max	6.1%	3.6%	3.8%	7.1%	2.9%	7.2%	7.1%	4.7%	3.3%	5.1%	4.0%	8.5%
Min	-2.7%	-2.8%	-2.3%	-5.5%	-4.1%	-3.2%	-4.5%	-3.7%	-3.1%	-3.9%	-6.1%	-3.5%
Hit Rate	59%	59%	50%	59%	48%	50%	55%	57%	56%	47%	43%	48%

Source: Citi Research, Factset, MSCI

Figure 47. US Summary Statistics

	Presentation						Answers					
	FK_ReadEase	FK_Grade	FRY_Sent	FRY_Syll	FOG	COLE	FK_ReadEase	FK_Grade	FRY_Sent	FRY_Syll	FOG	COLE
Return	1.4%	2.2%	0.9%	0.9%	-0.1%	0.3%	0.5%	1.5%	-0.2%	0.7%	-1.5%	0.2%
Risk	5.6%	4.9%	4.1%	4.3%	4.6%	5.8%	5.9%	5.2%	3.8%	5.1%	4.3%	6.1%
IR	**0.24**	**0.44**	**0.23**	**0.20**	**-0.02**	**0.04**	**0.08**	**0.30**	**-0.06**	**0.13**	**-0.36**	**0.03**
Max	5.1%	3.3%	5.2%	2.6%	3.5%	7.9%	5.0%	4.4%	3.7%	4.0%	4.4%	5.8%
Min	-9.3%	-9.3%	-3.0%	-3.9%	-4.0%	-7.2%	-9.3%	-7.1%	-3.3%	-8.5%	-3.3%	-10.6%
Hit Rate	55%	57%	52%	55%	47%	53%	50%	50%	45%	55%	50%	52%

Source: Citi Research, Factset, MSCI

Looking across measures, the Flesch-Kincaid measures appear to offer the best, and most robust, performance having the highest information ratios over both sections of the text and both regions in our study. This is attractive as Flesch-Kincaid is a very simple measure which is easy to implement and takes little computation time. Along with the Flesch-Kincaid measure, Coleman-Liau also offers positive information ratios in all sections and for both regions. Interestingly, Coleman-Liau and Flesch-Kincaid both use the intuitive words per sentence measure with the difference being Flesch-Kincaid also using syllables per word rather than letters per word.

The Fry measures also generate positive performance. However, we dislike this approach as it is based on samples of the text rather than the entire text which feels unnecessary with modern computational power.

The widely used Fog measure generates interesting results, with very negative returns implying complexity is preferred to simplicity. Fog uses the popular words to sentences metric as leveraged in both Flesch-Kincaid and Coleman-Liau. However, it also considers the proportion of words which are complex with complex words being defined as those with greater than or equal to three syllables.

This metric is often criticized as there are a number of long words, such as 'interesting', which cannot really be considered as complex despite fulfilling the criteria. Along with this criticism, we have another which is rather use case specific.

Earnings calls are comprised of company management discussing operations with well-informed analysts. Both parties are, or should be, well versed in company operations and terms which are perhaps linguistically complex but widely known in a specific industry. As a simple financial illustrative example, try explaining what a collateralized debt obligation is only using words less than three syllables. It is highly likely that complex terms are necessary to communicate performance and outlook in a concise, clear, manner. We construct signals to try and adjust for this later in the research[14]

Performance differentials are mirrored in return correlations with Flesch-Kincaid and Coleman-Liau measures being correlated and FOG and FRY measures having less reliable relationships.

Figure 48. Return Correlation (EU Lower diagonal, US upper diagonal)

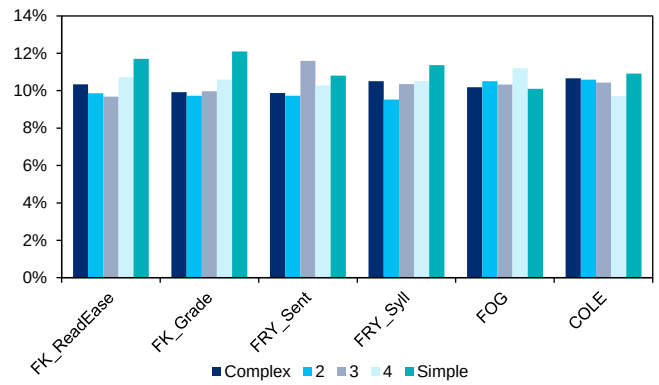
	FK_ReadEase	FK_Grade	FRY_Sent	FRY_Syll	FOG	COLE
FK_ReadEase		0.90	0.34	0.70	-0.21	0.88
FK_Grade	0.85		0.40	0.59	-0.15	0.77
FRY_Sent	0.40	0.39		0.07	-0.28	0.23
FRY_Syll	0.76	0.63	0.36		-0.05	0.79
FOG	-0.44	-0.29	-0.03	-0.32		-0.17
COLE	0.85	0.73	0.28	0.77	-0.47	

Source: Citi Research, Factset, MSCI

Examining returns across each basket, Figure 48 shows a relatively monotonic relationship for Flesch-Kincaid offering credence to our backtest results. With this measure, returns are particularly high in the simple (Quintile 5 – ease of readability), and almost simple (Q4), portfolios with complexity not necessarily implying poor future performance. This suggests low complexity is a better predictor of future strong performance, than high complexity is for future poor performance.

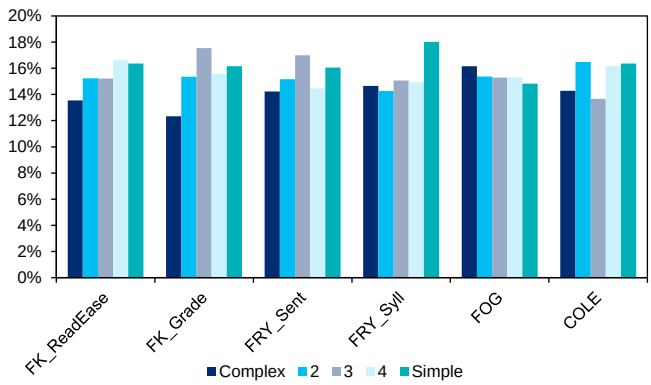
[14] See 'Waffle' section

Figure 49. US Returns by Basket – Management Section



Source: Citi Research, Factset, MSCI

Figure 50. EU Returns by Basket – Management Section



Source: Citi Research, Factset, MSCI

Overall these measures provide encouraging, if not compelling results. However, it is stark that many of these measures are derived from the same underlying ratios, making us question if we can construct a more domain specific set of metrics which captures complexity while considering the nature of earnings calls being well-informed domain specific experts conversing. We suggest measuring complexity in such a setting requires a more nuanced approach.

Complexity

So as to garner additional insight into the proposed complexity measures in the literature this section decomposes the readability measures into their individual components. The prior readability metrics are generally composed of:

- Word count to sentence count
- Character count to word count
- Syllable count to word count

with larger values being indicative of more complex text. Along with these measures we also examine;

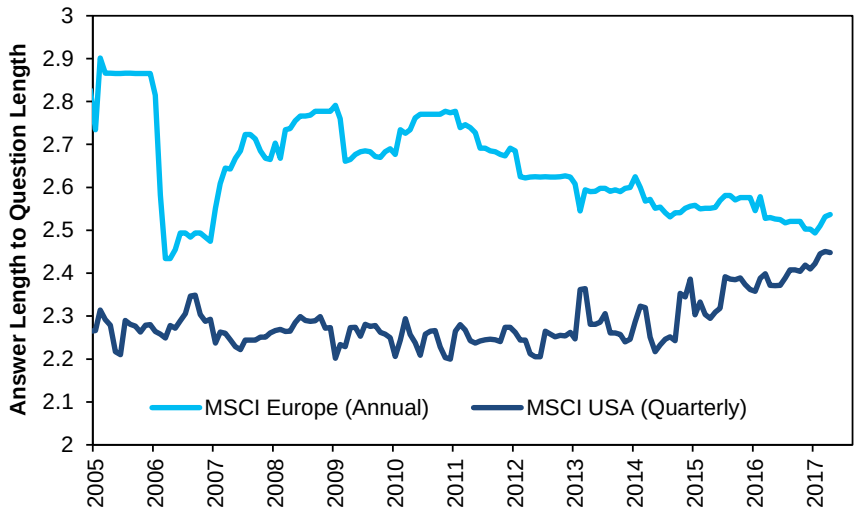
- Character count to sentence count
- Syllable count to sentence count

Which are simple intuitive extensions of measures used in the readability studies rather than representing any form of evolution.

Leveraging the structure of earnings calls, also of interest is the relative length (both words and number of characters) in the answer and question section. This is a very direct measure of complexity which uses the length of the question as a control for its difficulty with the length of the answer representing the way in which it was answered. Short questions being met with long answers is a sign of needless complexity whereas short answers relative to question length, we suggest, is indicative of concise, clear answers. It is hoped that this signal controls for some of the natural complexity found in discussing specific topics among experts.

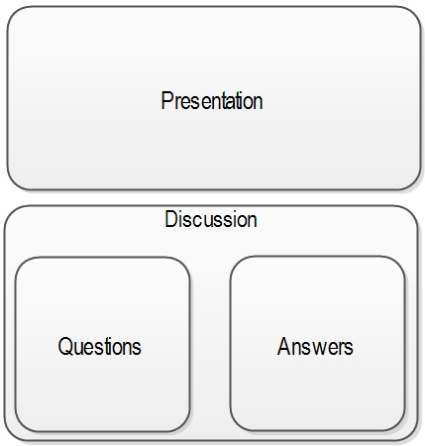
Examining answer to question length, Figure 51 shows that European corporate management appear to be more verbose when answering questions than their US peers. US management answer questions using approximately twice the number of words as the question, whereas European management are closer to three times. Interestingly, this trend seems to have disappeared as of late, with US management offering more wordy answers to questions. One possible explanation for this difference is that US firms report more information, more frequently, than their European peers requiring less explanation on each call.

Figure 51. Answer Length Relative to Question



Source: Citi Research, Factset, MSCI

Figure 52. Call Transcript Section Decomposition



Source: Citi Research

Intuitively, some might perceive a longer response by company management as an attempt to provide greater detail through transparency. While a valid argument, prior empirical work such as Li [2008] find more complex management communication is representative of 'cheap talk' or obfuscation. As such, we have a prior that simplicity is a stronger predictor of future performance than complexity.

Constructing the aforementioned signals and testing using our standard backtesting procedure, Figure 53 and Figure 54 present backtest summary statistics.

When decomposing into individual measures there is less of a distinction between the presentation section and the answer section with good performance being seen in both sections of the earnings transcript. Interestingly, individual component of the readability measures seem to perform better than the readability metrics themselves. This perhaps suggests that the complex weighting schemes imposed by the readability measures – which enable interpretations such as 'years of education' – are actually hurting the accuracy of the complexity estimate.

Figure 53. EU Summary Statistics

	Presentation					Answers					Discussion	
	Sll_Word	Word_Sent	Char_Sent	Sll_Sent	Char_Word	Sll_Word	Word_Sent	Char_Sent	Sll_Sent	Char_Word	AsQsWord	AsQsChar
Return	2.5%	2.2%	4.1%	4.5%	2.5%	5.0%	4.8%	1.0%	3.3%	3.5%	3.9%	1.1%
Risk	5.0%	4.8%	4.5%	4.4%	5.2%	6.4%	6.1%	5.9%	4.1%	4.5%	4.8%	5.6%
IR	**0.51**	**0.46**	**0.92**	**1.01**	**0.48**	**0.78**	**0.78**	**0.17**	**0.80**	**0.78**	**0.82**	**0.19**
Max	6.3%	3.6%	3.7%	3.6%	9.2%	12.2%	11.1%	6.3%	4.2%	4.9%	4.8%	7.3%
Min	-3.8%	-3.3%	-2.8%	-2.8%	-4.6%	-4.6%	-4.4%	-3.8%	-2.3%	-2.6%	-2.7%	-3.1%
Hit Rate	58%	57%	66%	65%	55%	55%	58%	49%	58%	55%	58%	47%

Source: Citi Research, Factset, MSCI

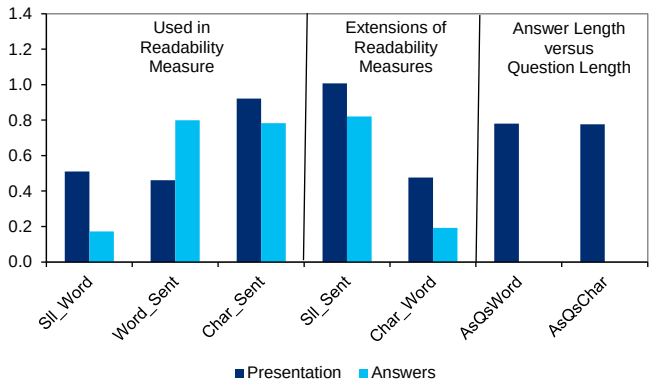
Figure 54. US Summary Statistics

	Presentation					Answers					Discussion	
	Sll_Word	Word_Sent	Char_Sent	Sll_Sent	Char_Word	Sll_Word	Word_Sent	Char_Sent	Sll_Sent	Char_Word	AsQsWord	AsQsChar
Return	0.9%	1.9%	1.2%	1.9%	0.4%	0.1%	1.9%	1.8%	2.0%	0.8%	1.7%	2.1%
Risk	5.6%	4.1%	4.0%	4.1%	5.6%	6.3%	4.3%	4.6%	4.6%	6.1%	4.4%	4.4%
IR	**0.17**	**0.46**	**0.30**	**0.46**	**0.07**	**0.02**	**0.43**	**0.40**	**0.44**	**0.14**	**0.38**	**0.47**
Max	5.0%	3.4%	3.2%	3.4%	7.9%	4.7%	3.4%	3.8%	4.0%	5.8%	4.3%	3.4%
Min	-8.8%	-4.8%	-4.5%	-4.8%	-6.3%	-8.6%	-4.7%	-6.1%	-5.6%	-9.7%	-4.0%	-4.0%
Hit Rate	52%	57%	54%	57%	54%	49%	50%	55%	53%	51%	57%	59%

Source: Citi Research, Factset, MSCI

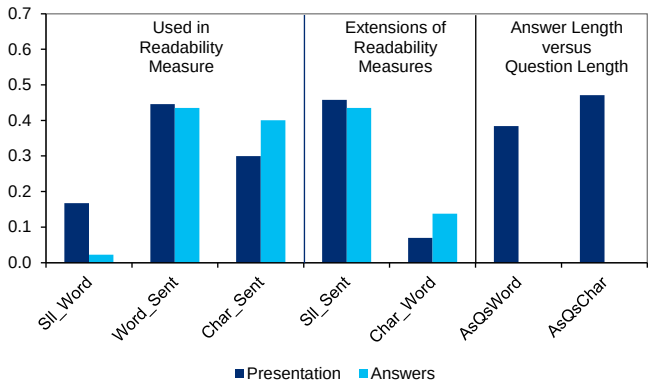
On an IR basis it is interesting that our simple extensions of the components of readability measures (characters and syllables per sentence) performed just as well, if not better, than the components which are widely used. This suggests broadening the readability measures results in better complexity scores. More broadly, standardizing number of words, characters or syllables by number of sentences rather than number of words seems to result in better risk adjusted performance. We hypothesize number of sentences is a more informative denominator as it offers some insight into phrase formation whereas counting words offers no insight into the skill with which they are put together.

Figure 55. EU IR



Source: Citi Research, Factset, MSCI

Figure 56. US IR



Source: Citi Research, Factset, MSCI

The newly formulated answer to question length ratio performs well in both regions. Results are also consistent using either relative words or characters which should be expected due to them having an intuitively high correlation as can be seen in Figure 57. Interestingly, answer to question length has significantly higher correlation to sentence-denominated measures rather than measures commonly used in the prior section's readability measures.

Figure 57. Return Correlation (EU Lower diagonal, US upper diagonal)

	Char_Word	Char_Sent	Sll_Word	Sll_Sent	Word_Sent	AsQsWord	AsQsChar
Char_Word		0.42	**0.91**	0.47	0.27	0.12	0.16
Char_Sent	0.41		0.46	**0.97**	**0.95**	0.44	0.45
Sll_Word	**0.86**	0.41		0.52	0.33	0.08	0.13
Sll_Sent	0.47	**0.96**	0.50		**0.91**	0.40	0.41
Word_Sent	0.17	**0.87**	0.15	**0.80**		0.44	0.44
As_Qs_WC	0.42	0.08	0.24	0.12	-0.01		**0.97**
As_Qs_CHAR	0.46	0.13	0.26	0.17	0.02	**0.98**	

Source: Citi Research, Factset, MSCI

Waffle

Being simultaneously inspired by both the Fog score's use of complex words to total words and having spent hours of our lives sat in unproductive meetings this next section attempts to define a measure of concision. We consider repetition an unnecessary evil indicative of corporate participants either not having much to say, or saying what they do have to say in an elongated, excessively verbose, fashion implying a low level of informational content in the earnings call.

To measure repetition, or waffle as we term it, we divide number of syllables, word and sentences by the number of unique words (or syllables in one case) in the document section. Lots of unique words relative to total words imply a greater degree of informational content, offering more clarity to investors and offering less management obfuscation. Scaling by uniqueness offers some insight into the natural level of complexity required when conversing with a room full of domain experts while hopefully capturing the degree of useful informational content in the call.

The decision to simultaneously examine words, syllables and sentences in the numerator stems from completeness rather than a preference for one particular measure – we expect them to be highly correlated.

The following tables outline summary statistics for our Waffle measures. As postulated, results are relatively consistent across measures with, arguably, the most intuitive measures of words to unique words and syllables to unique syllables offering strong and robust performance across regions and sections of the document.

Figure 58. EU Summary Statistics

	Presentation				Answers			
	Word_unWord	Sll_unSll	Sent_unWord	Sll_unWord	Word_unWord	Sll_unSll	Sent_unWord	Sll_unWord
Return	3.0%	2.2%	1.5%	3.9%	3.8%	3.7%	0.3%	4.0%
Risk	5.2%	5.1%	4.8%	5.4%	5.0%	5.1%	4.5%	5.8%
IR	**0.58**	**0.43**	**0.30**	**0.73**	**0.75**	**0.73**	**0.06**	**0.68**
Max	4.4%	4.3%	3.7%	4.0%	5.3%	7.1%	4.5%	6.6%
Min	-3.1%	-3.2%	-3.5%	-3.0%	-2.8%	-2.8%	-4.7%	-3.2%
Hit Rate	59%	55%	53%	61%	57%	59%	51%	57%

Source: Citi Research, Factset, MSCI

Figure 59. US Summary Statistics

	Presentation				Answers			
	Word_unWord	Sll_unSll	Sent_unWord	Sll_unWord	Word_unWord	Sll_unSll	Sent_unWord	Sll_unWord
Return	1.7%	1.3%	-0.1%	1.6%	1.3%	1.1%	-0.2%	1.3%
Risk	4.4%	4.5%	4.6%	4.4%	4.9%	4.5%	3.6%	5.6%
IR	**0.38**	**0.30**	**-0.02**	**0.35**	**0.27**	**0.24**	**-0.07**	**0.23**
Max	3.5%	3.7%	4.3%	4.0%	4.2%	3.4%	2.7%	4.3%
Min	-3.2%	-3.2%	-3.4%	-3.0%	-5.3%	-5.1%	-2.4%	-9.2%
Hit Rate	55%	53%	47%	58%	52%	53%	48%	57%

Source: Citi Research, Factset, MSCI

Also, as hypothesized, signal performance is highly correlated, suggesting examining all of these signals is a little excessive – Words to unique words is probably sufficient by itself. Sentences to unique words does seem to be the outlier, but given its poor performance the measure is not really of interest.

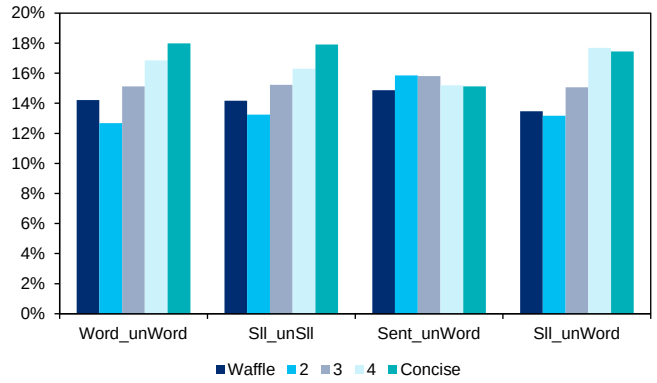
Figure 60. Return Correlation (EU Lower diagonal, US upper diagonal)

	Word_unWord	Sll_unSll	Sll_unWord	Sent_unWord
Word_unWord		0.92	0.87	-0.03
Sll_unSll	0.96		0.78	0.01
Sll_unWord	0.91	0.89		0.42
Sent_unWord	0.53	0.56	-0.17	

Source: Citi Research, Factset, MSCI

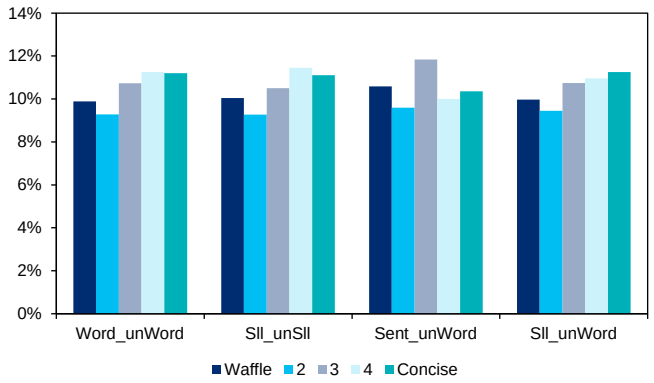
Satisfyingly, performance across baskets is reassuringly (approximately) monotonic offering credence to our backtest summary statistics. Sentence to unique words is again the exception reaffirming our assertion that this metric is the least attractive of those tested.

Figure 61. EU Return by Basket – Answers Section



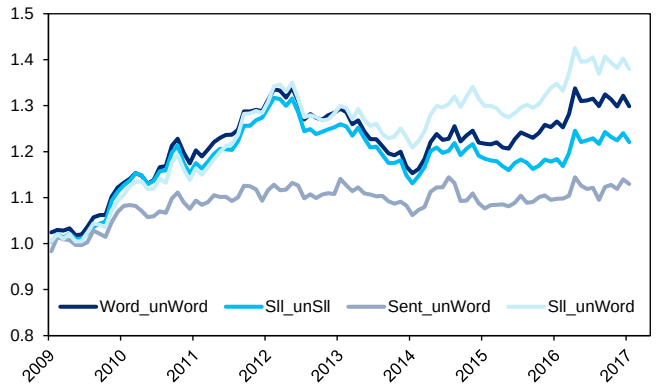
Source: Citi Research, Factset, MSCI

Figure 62. US Return by Basket – Answers Section



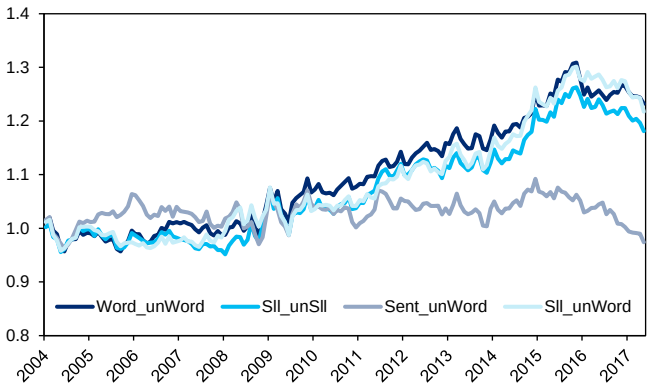
Source: Citi Research, Factset, MSCI

Figure 63. EU Cumulative Return – Presentation Section



Source: Citi Research, Factset, MSCI

Figure 64. US Cumulative Return – Presentation Section



Source: Citi Research, Factset, MSCI

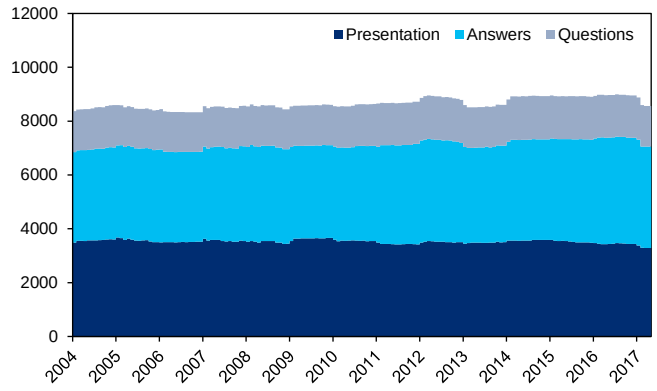
Length

The final characteristic we attempt to exploit is the length of the document as examined by Matsumoto et al [2006], Li [2008] and Loughran McDonald [2014]. Unlike those prior studies which are based on written text, earnings calls are spoken and scheduled to persist for a set period of time. Length, as measured by word or character counts, is naturally based on the length of time scheduled for the call. As such, we consider these basic length metrics to be a good measure of complexity, having the interpretation: "How many words can you say in one hour."

Under this paradigm more length equates to more complexity. Again, one could argue that more length equates to more information however, as discussed in the previous section, we are interested in unique information. Furthermore, prior studies find short text to be less complex and more informative.

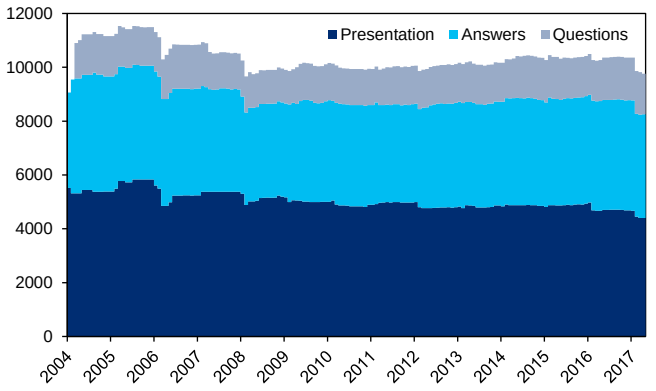
Splitting the earnings call by section, Figure 65 and Figure 66 show the average difference in call constructs across regions with Europe tending to spend longer on the presentation and relatively shorter period of time on the question and answer section. Word counts tend to be higher in Europe than the US possibly due to the focus on annual earnings calls in the region. Notably, earnings call lengths appear fairly stable through time, adding credence to the notion that lengths are based by time scheduling.

Figure 65. MSCI USA Median Earnings Call Word Count by Segment



Source: Citi Research, Factset, MSCI

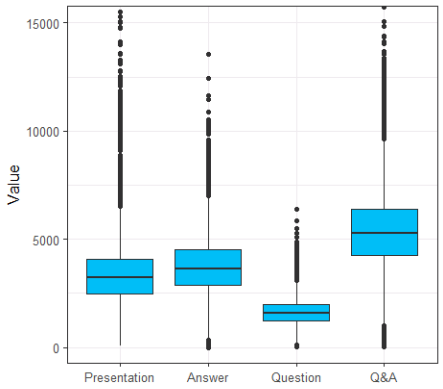
Figure 66. MSCI Europe Median Earnings Call Word Count by Segment



Source: Citi Research, Factset, MSCI

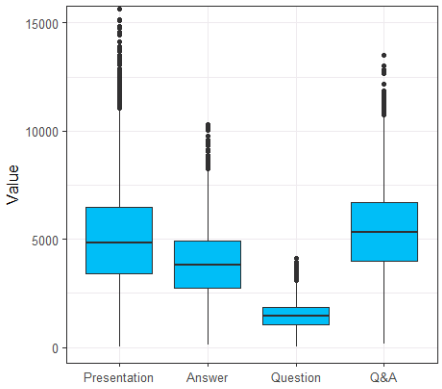
Figure 67 and Figure 68 provide a summary of word distributions across segments of a conference call. The interquartile ranges are greater in the Q&A section relative to the presentation segments for US conference calls, whereas the opposite holds true in Europe. There is also greater variance in the answer section in Europe relative to the US. Our initial exploratory analysis concurs with those shown by Matsumoto et al [2006].

Figure 67. Distribution of word counts by call segment – US



Source: Citi Research, Factset, MSCI

Figure 68. Distribution of word counts by call segment - EU



Source: Citi Research, Factset, MSCI

Translating these signals into actionable strategies, the following tables present summary statistics of long relative to short earnings calls. Performance in the presentation section is disappointing perhaps being indicative of the signals reliance on time being ineffective. Performance in the answer section is however, much stronger but possibly correlated to the complexity answers to question length measures.

Figure 69. EU Summary Statistics

	Presentation				Answers			
	CharCount	SllCount	WordCount	SentCount	CharCount	SllCount	WordCount	SentCount
Return	0.8%	1.1%	1.3%	0.5%	3.3%	3.3%	2.7%	2.0%
Risk	4.9%	4.8%	4.6%	4.5%	4.8%	4.8%	4.8%	5.1%
IR	**0.17**	**0.22**	**0.28**	**0.12**	**0.69**	**0.70**	**0.57**	**0.40**
Max	5.4%	5.4%	5.1%	3.0%	4.1%	4.1%	3.4%	3.1%
Min	-5.6%	-4.6%	-4.8%	-5.4%	-3.2%	-3.1%	-3.4%	-4.5%
Hit Rate	55%	55%	58%	59%	55%	54%	56%	55%

Source: Citi Research, Factset, MSCI

Figure 70. US Summary Statistics

	Presentation				Answers			
	CharCount	SllCount	WordCount	SentCount	CharCount	SllCount	WordCount	SentCount
Return	1.1%	1.4%	1.2%	0.9%	2.7%	2.8%	2.6%	2.0%
Risk	5.3%	5.4%	5.4%	5.2%	5.1%	5.0%	4.9%	4.2%
IR	**0.20**	**0.25**	**0.23**	**0.18**	**0.54**	**0.57**	**0.53**	**0.48**
Max	6.4%	6.4%	6.4%	5.6%	4.5%	4.4%	5.3%	3.5%
Min	-5.3%	-5.4%	-5.1%	-7.0%	-4.0%	-3.7%	-4.2%	-4.7%
Hit Rate	53%	54%	52%	52%	55%	58%	55%	50%

Source: Citi Research, Factset, MSCI

The specific length measure used seems relatively unimportant with words, syllables, sentences and characters all exhibiting similar performance and a high degree of correlation as exhibited in Figure 71. As with the waffle section, it is likely unnecessary to have so many measures capturing the same component in the earnings call transcript.

Figure 71. Return Correlation (EU Lower diagonal, US upper diagonal)

	CharCount	SllCount	WordCount	SentCount
CharCount		0.98	0.97	0.82
SllCount	0.98		0.96	0.82
WordCount	0.98	0.96		0.83
SentCount	0.86	0.83	0.87	

Source: Citi Research, Factset, MSCI

Composites: Bringing the Metrics Together

The above analysis shows a vast array of relatively simple signals to gain insight from an informed source of text, namely earnings call transcripts. We bring together readability and complexity measures from linguistic literature examining the base metrics that drive readability and complexity. We also attempted to quantify the amount of information in the call through examining repetition (Waffle) alongside simple length measures capturing concision.

While this is all interesting in isolation, in this section of the report we combine the different elements resulting in a final complexity score for each earnings call. Being cognizant of data-mining, as a first pass each signal is z-scored and equal weighted within each of the four subsections (Readability, Complexity, Waffle and Length). This signal is not a result of individual human judgment and so hopefully mitigates data-mining concerns.

Using these signals, Figure 72 and Figure 73 shows time series return correlation offering insight into the extent to which each of these four subsections are capturing different features in the document. Correlations are relatively high for Readability and Complexity which is intuitive as the tested complexity measures come from the Readability measures. In reality only one of these two sets of signals required.

Interestingly, Readability measures are also negatively correlated with Waffle in the presentation section for US companies. This is intuitive as the idea behind our Waffle measure stemmed from the fog index and its use of complex words.

Figure 72. EU (upper diagonal Presentation, lower Answers)

		Presentation		
	Waffle	Readability	Length	Complexity
Waffle		-0.12	0.06	0.15
Readability	0.30		0.02	**0.73**
Length	0.08	-0.18		0.12
Complexity	0.57	**0.75**	0.00	

Source: Citi Research, Factset, MSCI

Figure 73. US (upper diagonal Presentation, lower Answers)

		Presentation		
	Waffle	Readability	Length	Complexity
Waffle		-0.31	0.33	-0.13
Readability	0.33		0.07	**0.73**
Length	0.25	0.33		0.27
Complexity	0.52	**0.81**	0.46	

Source: Citi Research, Factset, MSCI

Examining performance, the following tables show positive information ratios for all composites except Readability, which exhibits a negative information ratio when utilized in the Answer segment for US calls. Given the relatively poor performance against other signals and the high correlation to Complexity we would drop Readability in favor of this.

There is surprisingly little difference between performance in the answer and presentation section, largely disproving our hypothesis that the un-prepared answer section would be more insightful. However, performance in Europe is better than the US, which is unlikely to surprise. The majority of academic work in this space has focused on the US. Europe is more challenging due to the different reporting rules in different countries however, simply using full-year earnings calls appears to have strong performance, which we postulate would only increase should interim reports be examined.

Figure 74. EU Summary Statistics

	Presentation				Answers			
	Waffle	Readability	Length	Complexity	Waffle	Readability	Length	Complexity
Return	3.2%	2.4%	1.4%	5.3%	3.9%	1.2%	2.6%	4.3%
Risk	5.2%	4.6%	5.0%	5.1%	5.6%	5.4%	4.8%	6.5%
IR	**0.62**	**0.52**	**0.28**	**1.04**	**0.71**	**0.23**	**0.54**	**0.66**
Max	4.8%	4.9%	5.7%	7.3%	7.5%	4.6%	3.2%	9.6%
Min	-2.9%	-2.5%	-5.6%	-2.5%	-3.5%	-4.6%	-4.1%	-4.4%
Hit Rate	55%	55%	54%	60%	57%	54%	59%	55%

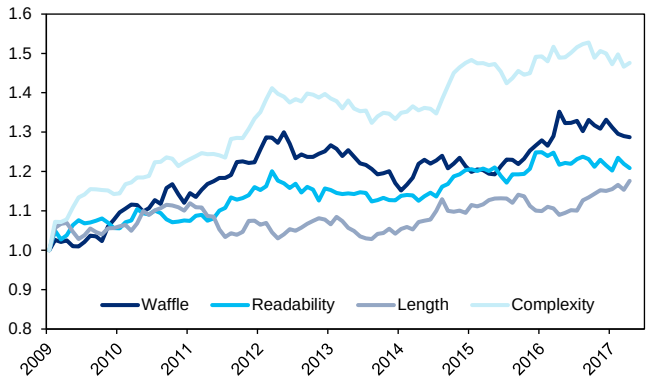
Source: Citi Research, Factset, MSCI

Figure 75. US Summary Statistics

	Presentation				Answers			
	Waffle	Readability	Length	Complexity	Waffle	Readability	Length	Complexity
Return	1.5%	1.3%	1.5%	2.0%	1.5%	-0.8%	2.2%	1.7%
Risk	4.5%	5.2%	5.2%	4.8%	4.8%	5.0%	4.8%	5.1%
IR	**0.33**	**0.25**	**0.28**	**0.42**	**0.31**	**-0.16**	**0.46**	**0.33**
Max	4.2%	5.7%	6.2%	3.9%	3.1%	4.4%	4.7%	4.4%
Min	-2.9%	-5.9%	-5.5%	-8.1%	-5.0%	-8.9%	-4.1%	-6.4%
Hit Rate	55%	53%	53%	52%	53%	48%	55%	55%

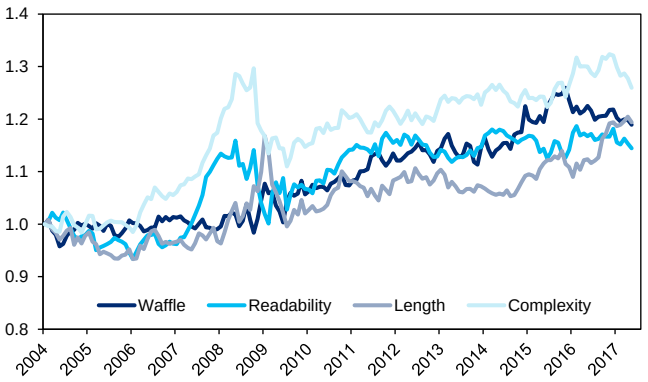
Source: Citi Research, Factset, MSCI

Figure 76. EU - Presentation



Source: Citi Research, Factset, MSCI

Figure 77. US - Presentation



Source: Citi Research, Factset, MSCI

Along with the performance of each subsection, Figure 78 equally weights each signal to arrive at a final earnings complexity scores for each call (All Signals).

While constructing composites through equally weighting across all signals mitigates data-mining concerns, it is not very practical as it requires the generation of lots of signals. Along with presenting a final composite equally weighting all signals Figure 78 presents the performance of a much simpler signal (Reduced Signals) equally weighting

- Character count
- Answer number of characters to question number of characters
- Word count divided by sentence count
- Word count divided by unique word count

These have been selected as they represent each of the examined subgroups using what we deem to be intuitive and simple to construct measures.

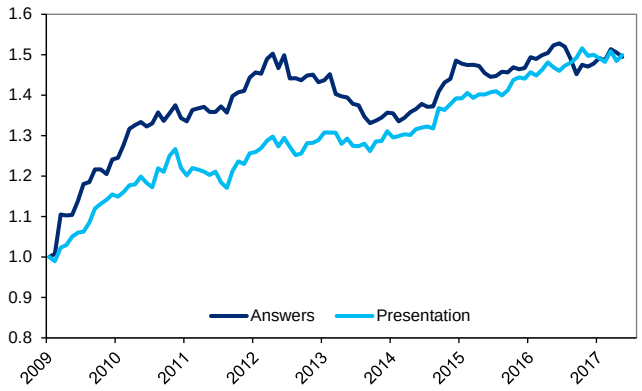
Performance is strong in both regions with positive and consistent returns being coupled with low risk, as one would expect from a country and sector neutral signal. Hit rates are attractive with minimum losses being inside twice volatility. The much reduced composite has very similar performance to the 'all signals' approach having generally lower levels of risk and marginally lower information ratios providing a feasible alternative approach.

Figure 78. Summary Statistics

	All Signals				Reduced Signals			
	US Answers	US Presentation	EU Answers	EU Presentation	US Answers	US Presentation	EU Answers	EU Presentation
Return	2.9%	2.3%	5.0%	5.0%	2.0%	1.4%	4.7%	3.5%
Risk	6.0%	5.1%	5.9%	4.8%	4.9%	4.3%	5.1%	4.8%
IR	**0.48**	**0.46**	**0.85**	**1.05**	**0.41**	**0.33**	**0.93**	**0.72**
Max	5.7%	5.2%	9.6%	4.0%	3.6%	4.9%	5.8%	4.0%
Min	-9.5%	-5.2%	-3.8%	-3.7%	-6.9%	-3.2%	-2.4%	-2.6%
Hit Rate	55%	58%	60%	62%	60%	55%	62%	59%

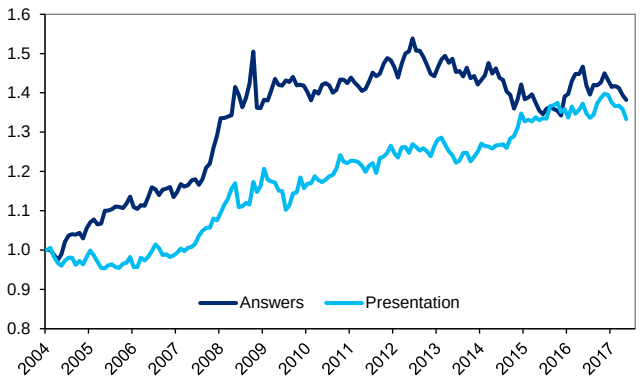
Source: Citi Research, Factset, MSCI

Figure 79. EU Composite Performance



Source: Citi Research, Factset, MSCI

Figure 80. US Composite Performance



Source: Citi Research, Factset, MSCI

Robustness Checks

The final section of this report examines the robustness of our 'All Signals' composite to our currently available competing suite of styles. While the above measures offer interesting performance in isolation, it is useless if our existing signals, from easier to collect data sources, subsume performance.

Figure 81 presents panel score rank correlations of our composite metrics to our standard Citi Style Composites. Results are remarkably robust across regions and sections with our earnings call information being most similar to small caps with high price and earnings momentum. This finding is mirrored in Figure 82 which presents score panel regressions of our earnings call signal against our standard style factors.

Figure 81. Panel Score Correlations

	Presentation		Answers	
	EU	US	EU	US
Estimates Momentum	0.069	0.029	0.029	0.022
Growth	0.063	-0.040	0.012	-0.083
Quality	0.033	0.008	0.011	-0.034
Risk	0.013	-0.040	0.031	0.007
Size	-0.081	-0.081	-0.178	-0.093
Value	-0.033	-0.004	-0.057	0.102
Price Momentum	0.079	0.027	0.062	0.029

Source: Citi Research, Factset, MSCI

Figure 82. Panel Regression t-Stats

	Presentation		Answers	
	EU	US	EU	US
(Intercept)	2.10	1.85	0.26	0.80
Estimates Momentum	1.49	4.67	-0.43	2.12
Growth	2.24	-4.62	-1.13	-8.59
Quality	0.52	5.40	-0.54	-1.58
Risk	0.09	-6.51	1.62	6.88
Size	-6.43	-17.55	-13.12	-15.44
Value	1.10	-4.70	-2.48	13.25
Price Momentum	3.99	3.98	3.62	9.56

Source: Citi Research, Factset, MSCI

We are comfortable with the small cap tilt given our base universe was large liquid stocks implying we will not have any problems with execution. Earnings revisions exposure is intuitive as analysts also respond to information in the earnings call. However, we anticipate a degree of nervousness with the extent to which this signal relates to price momentum, particularly in Europe.

Despite this high sensitivity to Price Momentum, the intercepts are positive and significant for the presentation section suggesting our signals have incremental information beyond that which is in our style composites. Results in the answer section are less robust consistent with smaller information ratios as observed above.

We are a little disappointed that the earnings call signal did not have a higher relationship to Quality. Processing text from management speaking is a possible way to incorporate management quality into a quantitative framework. Results are not completely disparaging with weakly positive results existing, giving us hope that a different set of measures may provide an insight into this missing dimension of stock returns. One potential explanation is that our current Quality composite does not include measures for management quality.

To examine the extent to which the earnings call composite adds value to already established, and related style composites on Price and Earnings Momentum, Figure 83 presents long short summary statistics country- and sector-neutral Price and Estimates Momentum including and excluding the presentation, and answer earnings signal in the style calculation.

It is clear that returns are increased and volatility reduced when earnings call signals are included in a Price Momentum composite. This is quite a strong result given only one additional signal has been added to the overall composite. Similar information ratio enhancements are seen when including the signals in Earnings Momentum, with benefits being attained from lower risk, implying the signal offers diversification benefits.

Figure 83. Summary Statistics of including composite into EU style factor

	Emom	Emom+As	Emom+Pres	Pmom	Pmom+As	Pmom+Pres
Return	4.1%	5.4%	4.0%	1.7%	3.0%	2.6%
Risk	9.9%	8.7%	8.7%	18.4%	16.9%	16.7%
IR	**0.42**	**0.62**	**0.46**	**0.09**	**0.18**	**0.16**
Max	6.8%	6.3%	5.4%	10.8%	10.9%	9.6%
Min	-13.2%	-9.9%	-10.4%	-27.5%	-20.5%	-22.3%
Hit Rate	58%	59%	57%	64%	61%	59%

Source: Citi Research, Factset, MSCI

Summary and Conclusions

It's clear from our recent research report - Searching for Alpha: Big Data: Navigating New Alternative Datasets – that there are a lot of interesting and potential alpha-adding alternative data sources available to investors. However, while Big Data brings huge potential for investors, in many cases it comes at a cost. This cost manifests itself through the overhead of data curation. For many investors that have typically dealt with standard structured data, ingesting this data through common data portals, much of this overhead is already dealt with, but with many Alternative Data sources, this is something that needs to be considered before analysis and implementation into strategies can start.

In the case of this research report, we have presented an easily accessible and relatively structured data source that has, in our empirical analysis, proven to add incremental value. Having said that, the data curation challenges were not trivial and much effort was put into building a clean data source. In this report we have tried to provide as much detail and context to the data curation such that investors have a sense of the effort required when working with unstructured or alternative data sources.

Having cleaned the data to a level that robust analysis can be conducted, for signal construction, our main hypothesis is based on the idea that simplicity of earnings calls is a positive predictor for future stock returns. Excessive obfuscation and lack of clarity from company insiders during earnings calls is indicative of them either hiding/withholding information, or alternatively, not being well informed enough about the businesses they run to offer well formulated insights.

To measure complexity we use readability measures commonly employed in the computational linguistics literature. Earnings calls represent a conversation between domain experts rather than for a general audience. As such, words which may be considered long or complex in a general setting, may be widely known and required in order for management to articulate their story concisely. We therefore look to modify the above Readability measures through decomposing them into their constituent parts (Complexity), deriving a measure of repetition (Waffle) and by examining the length of time they talk (Length / Rambling).

Placing the aforementioned measures into an overall earnings call composite number we find strong stock selection performance from buying simple and shorting complex stocks with information ratios of around 0.5 in the US and 0.8 in Europe. Interestingly we find little difference in performance from analysing the complexity of language in the scripted management presentation section versus the unscripted answering of analyst questions. This finding is in contrast to our initial prior that unscripted language would be more informative.

Overall we find that there is valuable information contained in earnings call transcripts and incremental performance to common style factors like price and earnings momentum. This analysis is based on relatively simple text mining methods so employing more sophisticated Natural Language Processing methods may add even more value. This will be area of focus in Part 2 of this series.

Appendix - Descriptive Statistics

The analysis was conducted using the MSCI World index as the base universe. The sample period started .01 Jan 2003 and ended 30 April 2017. Overall there were 3126 stocks in the universe but only 2113 that had call events.

Factset CallStreet Transcripts Summary

Figure 84. Events Downloaded and Selected for Processing

			Downloads	Parsed,In MSCI,Selected	Parsed and matched fundamentals	Parsed and matched fundamentals with Q&A
Call Street Earnings Events	Call Type	pdf	55,284	34,163	32,015	31,204
		html	27,889	20,522	19,589	18,746
	Total		83,173	54,685	51,604	49,950
Percentage Coverage	Call Type	pdf	66%	62%	62%	62%
		html	34%	38%	38%	38%
	Total		100%	100%	94%	91%

Source: Citi Research, Factset, MSCI,Thomson Reuters

Events Matched with Fundamentals and Estimates Databases

Figure 85. Conference Calls Mapped to WorldScope and IBES

	Mapped	Coverage (%)
WorldScope	51,703	94%
IBES	50,081	91%

Source: Citi Research, Factset, Thomson Reuters

Word descriptive statistics by call segment

Figure 86. Word Count Summary – MSCI USA

	Section	Obs	Min	1Q	Median	Mean	3Q	Max	StdDev	NA.Obs.	IQR
Annual	Presentation	6,777	67	2,721	3,524	3,713	4,449	24,266	1,597	11	1,728
	Questions	6,777	33	1,196	1,539	1,613	1,949	5,084	594	122	753
	Answers	6,777	6	2,846	3,608	3,739	4,534	10,204	1,316	116	1,688
	Discussion (Q&A)	6,777	39	4,191	5,213	5,352	6,376	15,069	1,737	122	2,185
Quarterly	Presentation	27,227	56	2,465	3,226	3,403	4,104	24,266	1,473	36	1,639
	Questions	27,227	31	1,223	1,565	1,630	1,965	6,381	594	552	742
	Answers	27,227	1	2,868	3,627	3,743	4,530	13,545	1,313	530	1,662
	Discussion (Q&A)	27,227	39	4,240	5,270	5,373	6,401	18,827	1,728	552	2,161

Source: Citi Research, Factset, MSCI

Figure 87. Word Count Summary – MSCI Europe

	Section	Obs	Min	1Q	Median	Mean	3Q	Max	StdDev	NA.Obs.	IQR
Annual	Presentation	3,971	32	3,424	4,865	5,127	6,476	28,309	2,573	10	3,052
	Questions	3,971	33	1,055	1,434	1,487	1,870	4,124	638	130	815
	Answers	3,971	95	2,754	3,819	3,918	4,929	10,301	1,644	106	2,175
	Discussion (Q&A)	3,971	142	4,003	5,311	5,410	6,710	13,502	2,104	130	2,707
Quarterly	Presentation	10,333	41	2,340	3,397	3,750	4,713	28,309	2,144	24	2,373
	Questions	10,333	6	1,098	1,516	1,555	1,962	4,543	657	245	864
	Answers	10,333	31	2,569	3,514	3,641	4,557	12,665	1,567	145	1,988
	Discussion (Q&A)	10,333	148	3,825	5,123	5,195	6,458	15,564	2,044	247	2,632

Source: Citi Research, Factset, MSCI

Appendix - Terms and Definitions

JSON – JavaScript Object Notation is a lightweight open standard content interchange format which consists of attribute-value information pairs.

jsonlite – R statistical package utilized for read/writing JSON data structures.

rvest – R statistical package utilized in parsing and manipulation of xml or xml like content.

xpdf – Xpdf is an open source PDF viewer, bundled within is the xpdf-utils package which includes a command line tool *pdftotext* which enables users to transform pdf documents into text.

tm – R statistical package to perform basic text mining tasks.

XML – Extensible Markup Language is a markup language that defines a set of rules for encoding document in a format that is both human and machine readable.

XPath – XML Path Language is a query language for selecting nodes within an XML document.

Appendix - References

Basu S.,1977, **Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis**, The Journal of Finance, Vol. 32, No. 3,pp. 663-682

Biddle G.C.,Hilary G.,Verdi S.R.,2009, **How does financial reporting quality relate to investment efficiency?** Journal of Accounting and Economics, Volume 48, Issues 2–3, December 2009, pp. 112-131

Bloomfield R.,2008, **Discussion of "Annual report readability, current earnings, and earnings persistence",**Journal of Accounting and Economics, Vol. 45, Issue. 2-3, pp. 248-252

Borochin P., Cicon J., Jared D.,Price S.M.,2008, **The Effects of Conference Call Tone on Market Perceptions of Value Uncertainty**, SSRN https://ssrn.com/abstract=2579907

Bowen R.M.,Davis A.K., Matsumoto D.A.,2002, **Do Conference Calls Affect Analysts Forecasts ?** The Accounting Review Vol. 77, No. 2, pp. 285-316

Bradshaw T.M., 2010, **Analysts' Forecasts:What Do We Know After Decades of Work?** SSRN https://ssrn.com/abstract=1880339

Davis A.K.,Piger J.M., Sedor L.M.,2011,**Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language**, SSRN https://ssrn.com/abstract=875399

Demers, E. and Vega, C., 2008, **Soft information in earnings announcements: News or noise?** Working paper, INSEAD.

Druz M.,Wagner A.F.,Zeckhauser R.J.,2016, **Reading Managerial Tone: How Analysts and the Market Respond to Conference Calls**, CEPR Discussion Paper

Feldman R., Govindaraj S., Livnat J., Segal B., 2008, **Management's Tone Change, Post Earnings Announcement Drift and Accruals,** SSRN https://ssrn.com/abstract=1287083

Franco G.D., Hope O., Dushyankumar V., Zhou Y.,2015, **Analyst Report Readability**, Contemporary Accounting Research Volume 32, Issue 1,2015 pp. 76–104

Frankel R.M.,Johnson F.M. 1999, **An Empirical Examination of Conference Calls as a Voluntary Disclosure Medium**, Journal of Accounting Research, Vol 37, No 1

Guay W.,Samuels D.,Taylor D.,2016,**Guiding Through the Fog: Financial Statement Complexity and Voluntary Disclosure**, SSRN https://ssrn.com/abstract=2564350

Huang, X., Teoh, S.H., and Zhang Y., 2011, **Tone Management**. The Accounting Review: May 2014, Vol. 89, No. 3, pp. 1083-1113.

Irani A.J., 2004,**The Effect of Regulation Fair Disclosure on the Relevance of Conference Calls to Financial Analysts**, Review of Quantitative Finance and Accounting: 2004, Vol. 22, Issue 1,pp 15-28

Kearney C., Liu S., 2014, **Textual sentiment in finance: A survey of methods and models**, International Review of Financial Analysis: May 2014, Vol 33, pp. 171-185.

Larcker D.F., Zakolyukina A.A., 2012, **Detecting Deceptive Discussions in Conference Calls**, Journal of Accounting Research, Vol. 50, Issue. 2, pp. 495-540

Li S., 2010, **Does Mandatory Adoption of International Financial Reporting Standards in the European Union Reduce the Cost of Equity Capital?** The Accounting Review: 2010, Vol. 85, No. 2, pp. 607-636.

Loughran, T., McDonald, B., 2011, **When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks**, The Journal of Finance: February, Vol. 66, Issue 1, pp. 35-65.

Loughran, T., McDonald, B., 2016, **Textual Analysis in Accounting and Finance: A Survey**, Journal of Accounting Research, Vol. 54, Issue. 4, pp. 1187-1230

Matsumoto D.A.,Roelofsen E.,Pronk M.,2006, **Managerial Disclosure vs. Analyst Inquiry: An Empirical Investigation of the Presentation and Discussion Portions of Earnings-Related Conference Calls**, SSRN https://ssrn.com/abstract=943928

Mayew W.M., Venkatachalam M., 2012, **The Power of Voice: Managerial Affective States and Future Firm Performance**, Vol. 67, Issue. 1, pp. 1-43

Michaely R.,Rubin A.,Vedrashko A.,2016, **Further evidence on the strategic timing of earnings news: Joint analysis of weekdays and times of day** Journal of Accounting and Economics, Vol. 62, Issue 1, pp. 24-45

Miller B.P.,2010, **The Effects of Reporting Complexity on Small and Large Investor Trading**, The Accounting Review: November 2010, Vol. 85, No. 6, pp. 2107-2143.

Price S.M.,Doran J.,Peterson D.R.,Bliss B.A.,2010, **Earnings Conference Calls and Stock Returns: The Incremental Informativeness of Textual Tone**, Journal of Banking and Finance, Vol. 36, No. 4, pp. 992-1011 2012

Schumacher E.,Eskenazi M., 2016, **A Readability Analysis of Campaign Speeches from the 2016 US Presidential Campaign** https://arxiv.org/ftp/arxiv/papers/1603/1603.05739.pdf

Sunder S.V.,2002, **Investor Access to Conference Call Disclosures: Impact of Regulation Fair Disclosure on Information Asymmetry**, SSRN https://ssrn.com/abstract=298653

Tetlock P.C. 2007, **Giving Content to Investor Sentiment: The Role of Media in the Stock Market,** The Journal of Finance: June, Vol. 62, Issue 3, pp. 1139-11686.

Tvesky A.,Kahnerman D.,1981, **The Framing of Decisions and the Psychology of Choice** American Association for the Advancement of Science: 1981, Vol. 211,Issue 4481,pp. 453-358.

Previously Published Research

Figure 88. Research Reports (since 2009)

Report	Date Published
Searching for Alpha: Big Data – Navigating New Alternative Datasets	10-Mar-2017
Searching for Alpha: Betting Against Accurate Beta	13-Feb-2017
Searching for Alpha: Competitive Advantage – Survival of the Fittest	19-Sep-2016
Searching for Alpha: Financial Strength Redux	9-Sep-2016
Searching for Alpha: Dynamic Style Weighting – Risk-Based Equity Style Allocation	14-Apr-2016
No Shorts Please: Long-Only Pure Style Portfolios	4-Mar-2016
Industry Alpha insights: Banks – One Size Does Not Fit All	18-Feb-2016
Under the Microscope: Stock Momentum Conflation	21 Sep-2015
Searching for Alpha: Macro Moves Markets	15-Sep-2015
Searching for Alpha: Style Performance, Trading Volumes and Investor Agreement	23-Mar-2015
World Radar Screen: Refining Our Global Search for Alpha	13-Mar-2015
Searching for Alpha: Networking with Analysts: Modelling Analyst Forecast Dependence	18-Feb-2015
The Rise of Low Risk Investing: Is it Getting Crowded Out There	01-Oct-2014
Under the Microscope: Five Innovations in Momentum Investing	27-Mar-2014
Searching for Alpha: Timing Price Momentum	07-Mar-2014
Equity Risk Premia Investing: A New Methodology For Monitoring Style Performance	27-Nov-2013
Stock Market Country Selection: Changes to a Well Established Model	23-Jul-2013
Searching for Alpha: Digging for Dividends – QUARI QUality with A Reliable Income	02-Jul-2013
Global Theme Machine: An Objective Way of Identifying Attractive Investment Themes	24-Jun-2013
Searching for Alpha: Purifying Analyst Recommendations – Removing Beta to get to the Alpha	25-Mar-2013
Searching for Alpha: Tangible Benefits of Intangibles – Brand, Respect & Intellectual Capital	06-Nov-2012
Global Quantitative Research Conference: 2012 Conference Proceedings	26-Sep-2012
Low-Risk Portfolio Strategies: Sharpe Ratio Maximisation and Multi-Asset Applications	09-Mar-2012
Macro Risk and Style Rotation: A Guide Rather than a Prescription	28-Feb-2012
Searching for Alpha: Accruals Volatility – A New Approach to Quality Investing	14-Sep-2011
Industry Alpha Insights: Four Approaches to Tactical Industry Selection	24-Aug-2011
Industry Alpha Insights: Quantifying Industry Specific Fundamentals	17-Mar-2011
Low-Risk Equity Portfolios: More than just Minimum Variance	18-Nov-2010
Under the Microscope: Measuring Systemic Risk – The Absorption Ratio	15-Nov-2010
Under the Microscope: Optionality in Valuation	14-Jun-2010
Searching for Alpha: Earnings Surprise – Still Profiting from Surprises	31-Mar-2010
Searching for Alpha: Optimising Style Rotation Strategies	15-Oct-2009

Source: Citi Research

Figure 89. What Works in Equity Markets

Report	
Top of the Market?	08-Aug-2017
Waiting for Earnings Delivery	27-Jan-2017
Style Purification – Does it Help?	24-Jan-2017
Single Factor or Multi-Factor Investing	22-Jul-2016
Are Spin-offs Money Spinners?	10-May-2016
Income and Low Risk Crowding	25-Apr-2016
Look for Persistent Earnings Leadership	24-Feb-2016
Macro Influence: Are Spreads Getting the Credit They Deserve?	09-Feb-2016
Protection from Macro Volatility: Stock Specific Risk	08-Jan-2016
Quant Road Ahead 2016	10-Dec-2015
Brand Value Powers On	08-Dec-2015
Can Small-Caps be Reignited?	28-Oct-2015
Quality Works, most of the time…	16-Sep-2015
Macro Still Matters, But Move Towards Micro	27-Jul-2015
Price Momentum: What's Trending?	03-Jul-2015

Source: Citi Research

Citi Quantitative Teams

Figure 90. Citi Quantitative Teams

For Informational Purposes Only

Global Quantitative Research		
Chris Montagu[1]	+44-20-7986-3958	chris.montagu@citi.com
Liz Dinh[3]	+61-2-8225-4896	liz.dinh@citi.com
Jason Kim[2]	+852-2501-2735	jason.jh.kim@citi.com
Helen Krause[1]	+44-20-7986-8653	helen.krause@citi.com
James Murray[1]	+44-20-7986-4579	james.murray@citi.com
Chris Ma[2]	+852-2501-2404	chris.ma@citi.com
David Chew[1]	+44-20-7986-7698	david.chew@citi.com
Zee Yusuf[1]	+44-20-7986-4228	zee.yusuf@citi.com
Bhavik Bochar[4]	+91-22-4277-5019	bhavik.bochar@citi.com
Ying Zhu[1]	+44-20-7986-8325	ying.zhu@citi.com
Academic Consultant		
Jose M Carabias[1]	+44-20-7986-3958	gqlondon@citi.com
Equity Quantitative Strategy (Sales and Trading Team, please see note below)		
Hong Li	+1-212-723-2881	hong.li@citi.com
Jason Li	+1-212-816-6033	jason.li@citi.com
Tiffany Zhou	+1-212-816-4659	tiffany.zhou@citi.com
Richard Schlatter	+1-212-816 6279	richard.w.schlatter@citi.com

1 Citigroup Global Markets Ltd; 2 Citigroup Global Markets Asia Ltd.; 3 Citigroup Pty Limited, 4 Citigroup Global Markets India Private Ltd.

NON-US RESEARCH ANALYST DISCLOSURES: The non-US research analysts listed above (i.e., the research analysts listed above other than those identified as employed by Citigroup Global Markets Inc.) are not registered/qualified as research analysts with FINRA. Such research analysts may not be associated persons of the member organization and therefore may not be subject to the NYSE Rule 472 and NASD Rule 2711 restrictions on communications with a subject company, public appearances and trading securities held by a research analyst account. Unless indicated in Appendix A-1 of this document or any of the referenced documents, the analysts listed above have not contributed to this document or any of the referenced documents.

SALES AND TRADING TEAMS: People designated as working in the sales or trading departments of Citi are not independent research analysts and have not contributed to this document or any of the referenced documents.

Appendix A-1

Analyst Certification

The research analysts primarily responsible for the preparation and content of this research report are either (i) designated by "AC" in the author block or (ii) listed in bold alongside content which is attributable to that analyst. If multiple AC analysts are designated in the author block, each analyst is certifying with respect to the entire research report other than (a) content attributable to another AC certifying analyst listed in bold alongside the content and (b) views expressed solely with respect to a specific issuer which are attributable to another AC certifying analyst identified in the price charts or rating history tables for that issuer shown below. Each of these analysts certify, with respect to the sections of the report for which they are responsible: (1) that the views expressed therein accurately reflect their personal views about each issuer and security referenced and were prepared in an independent manner, including with respect to Citigroup Global Markets Inc. and its affiliates; and (2) no part of the research analyst's compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views expressed by that research analyst in this report.

IMPORTANT DISCLOSURES

Disclosure for investors in the Republic of Turkey: Under Capital Markets Law of Turkey (Law No: 6362), the investment information, comments and advices given herein are not part of investment advisory activity. Investment advisory services are provided by authorized institutions to persons and entities privately by considering their risk and return preferences. Whereas the comments and advices included herein are of general nature. Therefore, they may not fit to your financial situation and risk and return preferences. For this reason, making an investment decision only by relying on the information given herein may not give rise to results that fit your expectations. Furthermore, Citi Research is a division of Citigroup Global Markets Inc. (the "Firm"), which does and seeks to do business with companies and/or trades on securities covered in this research reports. As a result, investors should be aware that the Firm may have a conflict of interest that could affect the objectivity of this report.

Analysts' compensation is determined by Citi Research management and Citigroup's senior management and is based upon activities and services intended to benefit the investor clients of Citigroup Global Markets Inc. and its affiliates (the "Firm"). Compensation is not linked to specific transactions or recommendations. Like all Firm employees, analysts receive compensation that is impacted by overall Firm profitability which includes investment banking, sales and trading, and principal trading revenues. One factor in equity research analyst compensation is arranging corporate access events between institutional clients and the management teams of covered companies. Typically, company management is more likely to participate when the analyst has a positive view of the company.

For securities recommended in the Product in which the Firm is not a market maker, the Firm is a liquidity provider in the issuers' financial instruments and may act as principal in connection with such transactions. The Firm is a regular issuer of traded financial instruments linked to securities that may have been recommended in the Product. The Firm regularly trades in the securities of the issuer(s) discussed in the Product. The Firm may engage in securities transactions in a manner inconsistent with the Product and, with respect to securities covered by the Product, will buy or sell from customers on a principal basis.

For important disclosures (including copies of historical disclosures) regarding the companies that are the subject of this Citi Research product ("the Product"), please contact Citi Research, 388 Greenwich Street, 28th Floor, New York, NY, 10013, Attention: Legal/Compliance [E6WYB6412478]. In addition, the same important disclosures, with the exception of the Valuation and Risk assessments and historical disclosures, are contained on the Firm's disclosure website at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures. Valuation and Risk assessments can be found in the text of the most recent research note/report regarding the subject company. Pursuant to the Market Abuse Regulation a history of all Citi Research recommendations published during the preceding 12-month period can be accessed via Citi Velocity (https://www.citivelocity.com/cv2) or your standard distribution portal. Historical disclosures (for up to the past three years) will be provided upon request.

Citi Research Equity Ratings Distribution

	12 Month Rating			*Catalyst Watch*		
Data current as of 30 Jun 2017	Buy	Hold	Sell	Buy	Hold	Sell
Citi Research Global Fundamental Coverage	47%	39%	13%	2%	97%	1%
% of companies in each rating category that are investment banking clients	64%	64%	58%	57%	64%	51%
Citi Research Quantitative World Radar Screen Model Coverage	30%	40%	30%			
% of companies in each rating category that are investment banking clients	41%	33%	33%			
Citi Research Asia Quantitative Radar Screen Model Coverage	20%	60%	20%			
% of companies in each rating category that are investment banking clients	39%	28%	24%			
Citi Research Australia Radar Model Coverage	52%	0%	48%			
% of companies in each rating category that are investment banking clients	51%	0%	34%			

Guide to Citi Research Fundamental Research Investment Ratings:
Citi Research stock recommendations include an investment rating and an optional risk rating to highlight high risk stocks.
Risk rating takes into account both price volatility and fundamental criteria. Stocks will either have no risk rating or a High risk rating assigned.
Investment Ratings: Citi Research investment ratings are Buy, Neutral and Sell. Our ratings are a function of analyst expectations of expected total return ("ETR") and risk. ETR is the sum of the forecast price appreciation (or depreciation) plus the dividend yield for a stock within the next 12 months. The Investment rating definitions are: Buy (1) ETR of 15% or more or 25% or more for High risk stocks; and Sell (3) for negative ETR. Any covered stock not assigned a Buy or a Sell is a Neutral (2). For stocks rated Neutral (2), if an analyst believes that there are insufficient valuation drivers and/or investment catalysts to derive a positive or negative investment view, they may elect with the approval of Citi Research management not to assign a target price and, thus, not derive an ETR. Analysts may place covered stocks "Under Review" in response to exceptional circumstances (e.g. lack of information critical to the analyst's thesis) affecting the company and / or trading in the company's securities (e.g. trading suspension). As soon as practically possible, the analyst will publish a note re-establishing a rating and investment thesis. To satisfy regulatory requirements, we correspond Under Review and Neutral to Hold in our

ratings distribution table for our 12-month fundamental rating system. However, we reiterate that we do not consider Under Review to be a recommendation. Investment ratings are determined by the ranges described above at the time of initiation of coverage, a change in investment and/or risk rating, or a change in target price (subject to limited management discretion). At other times, the expected total returns may fall outside of these ranges because of market price movements and/or other short-term volatility or trading patterns. Such interim deviations from specified ranges will be permitted but will become subject to review by Research Management. Your decision to buy or sell a security should be based upon your personal investment objectives and should be made only after evaluating the stock's expected performance and risk.

Prior to May 1, 2014 Citi Research may have also assigned a three-month relative call (or rating) to a stock to highlight expected out-performance (most preferred) or under-performance (least preferred) versus the geographic and industry sector over a 3 month period. The relative call may have highlighted a specific near-term catalyst or event impacting the company or the market that was anticipated to have a short-term price impact on the equity securities of the company. Absent any specific catalyst the analyst(s) may have indicated the most and least preferred stocks in the universe of stocks under consideration, explaining the basis for this short-term view. This three-month view may have been different from and did not affect a stock's fundamental equity rating, which reflected a longer-term total absolute return expectation.

Catalyst Watch Upside/Downside calls:
Citi Research may also include a Catalyst Watch Upside or Downside call to highlight specific near-term catalysts or events impacting the company or the market that are expected to influence the share price over a specified period of 30 or 90 days. A Catalyst Watch Upside (Downside) call indicates that the analyst expects the share price to rise (fall) in absolute terms over the specified period. A Catalyst Watch Upside/Downside call will automatically expire at the end of the specified 30/90 day period; the analyst may also close a Catalyst Watch call prior to the end of the specified period in a published research note. A Catalyst Watch Upside or Downside call may be different from and does not affect a stock's fundamental equity rating, which reflects a longer-term total absolute return expectation. For purposes of FINRA ratings-distribution-disclosure rules, a Catalyst Watch Upside call corresponds to a buy recommendation and a Catalyst Watch Downside call corresponds to a sell recommendation. Any stock not assigned to a Catalyst Watch Upside or Catalyst Watch Downside call is considered Catalyst Watch Non-Rated (CWNR). For purposes of FINRA ratings-distribution-disclosure rules, we correspond CWNR to Hold in our ratings distribution table for our Catalyst Watch Upside/Downside rating system. However, we reiterate that we do not consider CWNR to be a recommendation. For all Catalyst Watch Upside/Downside calls, risk exists that the catalyst(s) and associated share-price movement will not materialize as expected.

Guide to Citi Research Quantitative Research Investment Ratings:
Citi Research Quantitative Research World Radar Screen recommendations are based on a globally consistent framework to measure relative value and momentum for a large number of stocks across global developed and emerging markets. Relative value and momentum rankings are equally weighted to produce a global attractiveness score for each stock. The scores are then ranked and put into deciles. A stock with a decile rating of 1 denotes an attractiveness score in the top 10% of the universe (most attractive). A stock with a decile rating of 10 denotes an attractiveness score in the bottom 10% of the universe (least attractive).
Citi Research Asia Quantitative Radar Screen model recommendations are based on a regionally consistent framework to measure relative value and momentum for a large number of stocks across regional developed and emerging markets. Relative value and momentum rankings are equally weighted to produce a global attractiveness score for each stock. The scores are then ranked and put into quintiles. A stock with a quintile rating of 1 denotes an attractiveness score in the top 20% of the universe (most attractive). A stock with a quintile rating of 5 denotes an attractiveness score in the bottom 20% of the universe (least attractive).
Citi Research Australia Quantitative Radar Screen model recommendations are based on a robust framework to measure relative value and momentum for a large number of stocks across the Australian market. Stocks with a ranking of 1 denotes a stock that is above average in terms of both value and momentum relative to the stocks in the Australian market. A ranking of 10 denotes a stock that is below average in terms of both value and momentum relative to the stocks in the Australian market.

For purposes of FINRA ratings-distribution-disclosure rules, a Citi Research Quantitative World Radar Screen recommendation of (1), (2) or (3) most closely corresponds to a buy recommendation; a recommendation from this product group of (4), (5), (6) or (7) most closely corresponds to a hold recommendation; and a recommendation of (8), (9) or (10) most closely corresponds to a sell recommendation. An (NR) recommendation indicates that the stock is no longer in the screen.
For purposes of FINRA ratings distribution disclosure rules, a Citi Research Asia Quantitative Radar Screen recommendation of (1) most closely corresponds to a buy recommendation; a Citi Research Asia Quantitative Radar Screen recommendation of (2), (3), (4) most closely corresponds to a hold recommendation; and a recommendation of (5) most closely corresponds to a sell recommendation. An (NR) recommendation indicates that the stock is no longer in the screen.
For purposes of FINRA ratings-distribution-disclosure rules, a Citi Research Quantitative Research Australia Radar Screen recommendation of "attractive" (1) most closely corresponds to a buy recommendation. All other stocks in the sector are considered to be "unattractive" (10) which most closely corresponds to a sell recommendation. An (NR)/(0) recommendation indicates that the stock is no longer in the screen.
Recommendations are based on the relative attractiveness of a stock, thus can not be directly equated to buy, hold and sell categories. Accordingly, your decision to buy or sell a security should be based on your personal investment objectives and only after evaluating the stock's expected relative performance.

NON-US RESEARCH ANALYST DISCLOSURES
Non-US research analysts who have prepared this report (i.e., all research analysts listed below other than those identified as employed by Citigroup Global Markets Inc.) are not registered/qualified as research analysts with FINRA. Such research analysts may not be associated persons of the member organization and therefore may not be subject to the FINRA Rule 2241 restrictions on communications with a subject company, public appearances and trading securities held by a research analyst account. The legal entities employing the authors of this report are listed below:

Citigroup Global Markets Ltd	Chris Montagu; Helen H Krause, CFA; James J Murray, CFA; David T Chew; Zee Yusuf, CFA; Ying Zhu
Citigroup Global Markets Asia	Chris Ma; Jason Kim

Citigroup Global Markets Australia Pty Limited Liz Dinh, CFA

OTHER DISCLOSURES

Any price(s) of instruments mentioned in recommendations are as of the prior day's market close on the primary market for the instrument, unless otherwise stated.

The completion and first dissemination of any recommendations made within this research report are as of the Eastern date-time displayed at the top of the Product. If the Product references views of other analysts then please refer to the price chart or rating history table for the date/time of completion and first dissemination with respect to that view.

European regulations require that where a recommendation differs from any of the author's previous recommendations concerning the same financial instrument or issuer that has been published during the preceding 12-month period that the change(s) and the date of that previous recommendation are indicated. For fundamental coverage please refer to the price chart or rating change history within this disclosure appendix or the issuer disclosure summary at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures.

European regulations require that a firm must establish, implement and make available a policy for managing conflicts of interest arising as a result of publication or distribution of investment research. The policy applicable to Citi Research's Products can be found at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures.

The proportion of all Citi Research fundamental research recommendations that were the equivalent to "Buy","Hold","Sell" at the end of each quarter over the prior 12 months (with the % of these that had received investment firm services from Citi in the prior 12 months shown in brackets) is as follows: Q2 2017 Buy 32% (70%), Hold 45% (63%), Sell 24% (57%); Q1 2017 Buy 32% (70%), Hold 45% (63%), Sell 24% (56%); Q4 2016 Buy 31% (71%), Hold 45% (64%), Sell 24% (58%); Q3 2016 Buy 32% (68%), Hold 44% (64%), Sell 24% (61%).

Citigroup Global Markets India Private Limited and/or its affiliates may have, from time to time, actual or beneficial ownership of 1% or more in the debt securities of the subject issuer.

Citi Research generally disseminates its research to the Firm's global institutional and retail clients via both proprietary (e.g., Citi Velocity and Citi Personal Wealth Management) and non-proprietary electronic distribution platforms. Certain research may be disseminated only via the Firm's proprietary distribution platforms; however such research will not contain changes to earnings forecasts, target price, investment or risk rating or investment thesis or be otherwise inconsistent with the author's previously published research. Certain research is made available only to institutional investors to satisfy regulatory requirements. Individual Citi Research analysts may also opt to circulate published research to one or more clients by email; such email distribution is discretionary and is done only after the research has been disseminated. The level and types of services provided by Citi Research analysts to clients may vary depending on various factors such as the client's individual preferences as to the frequency and manner of receiving communications from analysts, the client's risk profile and investment focus and perspective (e.g. market-wide, sector specific, long term, short-term etc.), the size and scope of the overall client relationship with the Firm and legal and regulatory constraints.

Pursuant to Comissão de Valores Mobiliários Rule 483, Citi is required to disclose whether a Citi related company or business has a commercial relationship with the subject company. Considering that Citi operates multiple businesses in more than 100 countries around the world, it is likely that Citi has a commercial relationship with the subject company.

Securities recommended, offered, or sold by the Firm: (i) are not insured by the Federal Deposit Insurance Corporation; (ii) are not deposits or other obligations of any insured depository institution (including Citibank); and (iii) are subject to investment risks, including the possible loss of the principal amount invested. The Product is for informational purposes only and is not intended as an offer or solicitation for the purchase or sale of a security. Any decision to purchase securities mentioned in the Product must take into account existing public information on such security or any registered prospectus. Although information has been obtained from and is based upon sources that the Firm believes to be reliable, we do not guarantee its accuracy and it may be incomplete and condensed. Note, however, that the Firm has taken all reasonable steps to determine the accuracy and completeness of the disclosures made in the Important Disclosures section of the Product. The Firm's research department has received assistance from the subject company(ies) referred to in this Product including, but not limited to, discussions with management of the subject company(ies). Firm policy prohibits research analysts from sending draft research to subject companies. However, it should be presumed that the author of the Product has had discussions with the subject company to ensure factual accuracy prior to publication. All opinions, projections and estimates constitute the judgment of the author as of the date of the Product and these, plus any other information contained in the Product, are subject to change without notice. Prices and availability of financial instruments also are subject to change without notice. Notwithstanding other departments within the Firm advising the companies discussed in this Product, information obtained in such role is not used in the preparation of the Product. Although Citi Research does not set a predetermined frequency for publication, if the Product is a fundamental equity or credit research report, it is the intention of Citi Research to provide research coverage of the covered issuers, including in response to news affecting the issuer. For non-fundamental research reports, Citi Research may not provide regular updates to the views, recommendations and facts included in the reports. Notwithstanding that Citi Research maintains coverage on, makes recommendations concerning or discusses issuers, Citi Research may be periodically restricted from referencing certain issuers due to legal or policy reasons. Citi Research may provide different research products and services to different classes of customers (for example, based upon long-term or short-term investment horizons) that may lead to differing conclusions or recommendations that could impact the price of a security contrary to the recommendations in the alternative research product, provided that each is consistent with the rating system for each respective product.

Investing in non-U.S. securities, including ADRs, may entail certain risks. The securities of non-U.S. issuers may not be registered with, nor be subject to the reporting requirements of the U.S. Securities and Exchange Commission. There may be limited information available on foreign securities. Foreign companies are generally not subject to uniform audit and reporting standards, practices and requirements comparable to those in the U.S. Securities of some foreign companies may be less liquid and their prices more volatile than securities of comparable U.S. companies. In addition, exchange rate movements may have an adverse effect on the value of an investment in a foreign stock and its corresponding dividend payment for U.S. investors. Net dividends to ADR investors are estimated, using withholding tax rates conventions, deemed accurate, but investors are urged to consult their tax advisor for exact dividend computations. Investors who have received the Product from the Firm may be prohibited in certain states or other jurisdictions from

purchasing securities mentioned in the Product from the Firm. Please ask your Financial Consultant for additional details. Citigroup Global Markets Inc. takes responsibility for the Product in the United States. Any orders by US investors resulting from the information contained in the Product may be placed only through Citigroup Global Markets Inc.

Important Disclosures for Bell Potter Customers: Bell Potter is making this Product available to its clients pursuant to an agreement with Citigroup Global Markets Australia Pty Limited. Neither Citigroup Global Markets Australia Pty Limited nor any of its affiliates has made any determination as to the suitability of the information provided herein and clients should consult with their Bell Potter financial advisor before making any investment decision.

The Citigroup legal entity that takes responsibility for the production of the Product is the legal entity which the first named author is employed by. The Product is made available in Australia through Citigroup Global Markets Australia Pty Limited. (ABN 64 003 114 832 and AFSL No. 240992), participant of the ASX Group and regulated by the Australian Securities & Investments Commission. Citigroup Centre, 2 Park Street, Sydney, NSW 2000. Citigroup Global Markets Australia Pty Limited is not an Authorised Deposit-Taking Institution under the Banking Act 1959, nor is it regulated by the Australian Prudential Regulation Authority. The Product is made available in Australia to Private Banking wholesale clients through Citigroup Pty Limited (ABN 88 004 325 080 and AFSL 238098). Citigroup Pty Limited provides all financial product advice to Australian Private Banking wholesale clients through bankers and relationship managers. If there is any doubt about the suitability of investments held in Citigroup Private Bank accounts, investors should contact the Citigroup Private Bank in Australia. Citigroup companies may compensate affiliates and their representatives for providing products and services to clients. The Product is made available in Brazil by Citigroup Global Markets Brasil - CCTVM SA, which is regulated by CVM - Comissão de Valores Mobiliários ("CVM"), BACEN - Brazilian Central Bank, APIMEC - Associação dos Analistas e Profissionais de Investimento do Mercado de Capitais and ANBIMA – Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais. Av. Paulista, 1111 - 14º andar(parte) - CEP: 01311920 - São Paulo - SP. If the Product is being made available in certain provinces of Canada by Citigroup Global Markets (Canada) Inc. ("CGM Canada"), CGM Canada has approved the Product. Citigroup Place, 123 Front Street West, Suite 1100, Toronto, Ontario M5J 2M3. This product is available in Chile through Banchile Corredores de Bolsa S.A., an indirect subsidiary of Citigroup Inc., which is regulated by the Superintendencia de Valores y Seguros. Agustinas 975, piso 2, Santiago, Chile. The Product is distributed in Germany by Citigroup Global Markets Deutschland AG ("CGMD"), which is regulated by Bundesanstalt fuer Finanzdienstleistungsaufsicht (BaFin). CGMD, Reuterweg 16, 60323 Frankfurt am Main. Research which relates to "securities" (as defined in the Securities and Futures Ordinance (Cap. 571 of the Laws of Hong Kong)) is issued in Hong Kong by, or on behalf of, Citigroup Global Markets Asia Limited which takes full responsibility for its content. Citigroup Global Markets Asia Ltd. is regulated by Hong Kong Securities and Futures Commission. If the Research is made available through Citibank, N.A., Hong Kong Branch, for its clients in Citi Private Bank, it is made available by Citibank N.A., Citibank Tower, Citibank Plaza, 3 Garden Road, Hong Kong. Citibank N.A. is regulated by the Hong Kong Monetary Authority. Please contact your Private Banker in Citibank N.A., Hong Kong, Branch if you have any queries on or any matters arising from or in connection with this document. The Product is made available in India by Citigroup Global Markets India Private Limited (CGM), which is regulated by the Securities and Exchange Board of India (SEBI), as a Research Analyst (SEBI Registration No. INH000000438). CGM is also actively involved in the business of merchant banking and stock brokerage in India, and is registered with SEBI in this regard. CGM's registered office is at 1202, 12th Floor, FIFC, G Block, Bandra Kurla Complex, Bandra East, Mumbai – 400098. CGM's Corporate Identity Number is U99999MH2000PTC126657, and its contact details are: Tel:+9102261759999 Fax:+9102261759961. The Product is made available in Indonesia through PT Citigroup Securities Indonesia. 5/F, Citibank Tower, Bapindo Plaza, Jl. Jend. Sudirman Kav. 54-55, Jakarta 12190. Neither this Product nor any copy hereof may be distributed in Indonesia or to any Indonesian citizens wherever they are domiciled or to Indonesian residents except in compliance with applicable capital market laws and regulations. This Product is not an offer of securities in Indonesia. The securities referred to in this Product have not been registered with the Capital Market and Financial Institutions Supervisory Agency (BAPEPAM-LK) pursuant to relevant capital market laws and regulations, and may not be offered or sold within the territory of the Republic of Indonesia or to Indonesian citizens through a public offering or in circumstances which constitute an offer within the meaning of the Indonesian capital market laws and regulations. The Product is made available in Israel through Citibank NA, regulated by the Bank of Israel and the Israeli Securities Authority. Citibank, N.A, Platinum Building, 21 Ha'arba'ah St, Tel Aviv, Israel. The Product is made available in Italy by Citigroup Global Markets Limited, which is authorised by the PRA and regulated by the FCA and the PRA. Via dei Mercanti, 12, Milan, 20121, Italy. The Product is made available in Japan by Citigroup Global Markets Japan Inc. ("CGMJ"), which is regulated by Financial Services Agency, Securities and Exchange Surveillance Commission, Japan Securities Dealers Association, Tokyo Stock Exchange and Osaka Securities Exchange. Shin-Marunouchi Building, 1-5-1 Marunouchi, Chiyoda-ku, Tokyo 100-6520 Japan. If the Product was distributed by SMBC Nikko Securities Inc. it is being so distributed under license. In the event that an error is found in an CGMJ research report, a revised version will be posted on the Firm's Citi Velocity website. If you have questions regarding Citi Velocity, please call (81 3) 6270-3019 for help. The Product is made available in Korea by Citigroup Global Markets Korea Securities Ltd., which is regulated by the Financial Services Commission, the Financial Supervisory Service and the Korea Financial Investment Association (KOFIA). Citibank Building, 39 Da-dong, Jung-gu, Seoul 100-180, Korea. KOFIA makes available registration information of research analysts on its website. Please visit the following website if you wish to find KOFIA registration information on research analysts of Citigroup Global Markets Korea Securities Ltd. http://dis.kofia.or.kr/websquare/index.jsp?w2xPath=/wq/fundMgr/DISFundMgrAnalystList.xml&divisionId=MDIS03002002000000&serviceId=SDIS03002002000. The Product is made available in Korea by Citibank Korea Inc., which is regulated by the Financial Services Commission and the Financial Supervisory Service. Address is Citibank Building, 39 Da-dong, Jung-gu, Seoul 100-180, Korea. The Product is made available in Malaysia by Citigroup Global Markets Malaysia Sdn Bhd (Company No. 460819-D) ("CGMM") to its clients and CGMM takes responsibility for its contents. CGMM is regulated by the Securities Commission of Malaysia. Please contact CGMM at Level 43 Menara Citibank, 165 Jalan Ampang, 50450 Kuala Lumpur, Malaysia in respect of any matters arising from, or in connection with, the Product. The Product is made available in Mexico by Acciones y Valores Banamex, S.A. De C. V., Casa de Bolsa, Integrante del Grupo Financiero Banamex ("Accival") which is a wholly owned subsidiary of Citigroup Inc. and is regulated by Comision Nacional Bancaria y de Valores. Reforma 398, Col. Juarez, 06600 Mexico, D.F. In New Zealand the Product is made available to 'wholesale clients' only as defined by s5C(1) of the Financial Advisers Act 2008 ('FAA') through Citigroup Global Markets Australia Pty Ltd (ABN 64 003 114 832 and AFSL No. 240992), an overseas financial adviser as defined by the FAA, participant of the ASX Group and regulated by the Australian Securities & Investments Commission. Citigroup Centre, 2 Park Street, Sydney, NSW 2000. The Product is made available in Pakistan by Citibank N.A. Pakistan branch, which is regulated by the State Bank of Pakistan and Securities Exchange Commission, Pakistan. AWT Plaza, 1.1. Chundrigar Road, P.O. Box 4889, Karachi-74200. The Product is made available in the Philippines through Citicorp Financial Services and Insurance Brokerage Philippines, Inc., which is regulated by the Philippines Securities and Exchange Commission. 20th Floor Citibank Square Bldg. The Product is made available in the Philippines through Citibank NA Philippines branch, Citibank Tower, 8741 Paseo De Roxas, Makati City, Manila. Citibank NA Philippines NA is regulated by The Bangko Sentral ng

Pilipinas. The Product is made available in Poland by Dom Maklerski Banku Handlowego SA an indirect subsidiary of Citigroup Inc., which is regulated by Komisja Nadzoru Finansowego. Dom Maklerski Banku Handlowego S.A. ul.Senatorska 16, 00-923 Warszawa. The Product is made available in the Russian Federation through AO Citibank, which is licensed to carry out banking activities in the Russian Federation in accordance with the general banking license issued by the Central Bank of the Russian Federation and brokerage activities in accordance with the license issued by the Federal Service for Financial Markets. Neither the Product nor any information contained in the Product shall be considered as advertising the securities mentioned in this report within the territory of the Russian Federation or outside the Russian Federation. The Product does not constitute an appraisal within the meaning of the Federal Law of the Russian Federation of 29 July 1998 No. 135-FZ (as amended) On Appraisal Activities in the Russian Federation. 8-10 Gasheka Street, 125047 Moscow. The Product is made available in Singapore through Citigroup Global Markets Singapore Pte. Ltd. ("CGMSPL"), a capital markets services license holder, and regulated by Monetary Authority of Singapore. Please contact CGMSPL at 8 Marina View, 21st Floor Asia Square Tower 1, Singapore 018960, in respect of any matters arising from, or in connection with, the analysis of this document. This report is intended for recipients who are accredited, expert and institutional investors as defined under the Securities and Futures Act (Cap. 289). The Product is made available by The Citigroup Private Bank in Singapore through Citibank, N.A., Singapore Branch, a licensed bank in Singapore that is regulated by Monetary Authority of Singapore. Please contact your Private Banker in Citibank N.A., Singapore Branch if you have any queries on or any matters arising from or in connection with this document. This report is intended for recipients who are accredited, expert and institutional investors as defined under the Securities and Futures Act (Cap. 289). This report is distributed in Singapore by Citibank Singapore Ltd ("CSL") to selected Citigold/Citigold Private Clients. CSL provides no independent research or analysis of the substance or in preparation of this report. Please contact your Citigold//Citigold Private Client Relationship Manager in CSL if you have any queries on or any matters arising from or in connection with this report. This report is intended for recipients who are accredited investors as defined under the Securities and Futures Act (Cap. 289). Citigroup Global Markets (Pty) Ltd. is incorporated in the Republic of South Africa (company registration number 2000/025866/07) and its registered office is at 145 West Street, Sandton, 2196, Saxonwold. Citigroup Global Markets (Pty) Ltd. is regulated by JSE Securities Exchange South Africa, South African Reserve Bank and the Financial Services Board. The investments and services contained herein are not available to private customers in South Africa. The Product is made available in the Republic of China through Citigroup Global Markets Taiwan Securities Company Ltd. ("CGMTS"), 14 and 15F, No. 1, Songzhi Road, Taipei 110, Taiwan, subject to the license scope and the applicable laws and regulations in the Republic of China. CGMTS is regulated by the Securities and Futures Bureau of the Financial Supervisory Commission of Taiwan, the Republic of China. No portion of the Product may be reproduced or quoted in the Republic of China by the press or any third parties [without the written authorization of CGMTS]. Pursuant to the applicable laws and regulations in the Republic of China, the recipient of the Product shall not take advantage of such Product to involve in any matters in which the recipient may have conflicts of interest. If the Product covers securities which are not allowed to be offered or traded in the Republic of China, neither the Product nor any information contained in the Product shall be considered as advertising the securities or making recommendation of the securities in the Republic of China. The Product is for informational purposes only and is not intended as an offer or solicitation for the purchase or sale of a security or financial products. Any decision to purchase securities or financial products mentioned in the Product must take into account existing public information on such security or the financial products or any registered prospectus. The Product is made available in Thailand through Citicorp Securities (Thailand) Ltd., which is regulated by the Securities and Exchange Commission of Thailand. 399 Interchange 21 Building, 18th Floor, Sukhumvit Road, Klongtoey Nua, Wattana ,Bangkok 10110, Thailand. The Product is made available in Turkey through Citibank AS which is regulated by Capital Markets Board. Tekfen Tower, Eski Buyukdere Caddesi # 209 Kat 2B, 23294 Levent, Istanbul, Turkey. In the U.A.E, these materials (the "Materials") are communicated by Citigroup Global Markets Limited, DIFC branch ("CGML"), an entity registered in the Dubai International Financial Center ("DIFC") and licensed and regulated by the Dubai Financial Services Authority ("DFSA") to Professional Clients and Market Counterparties only and should not be relied upon or distributed to Retail Clients. A distribution of the different Citi Research ratings distribution, in percentage terms for Investments in each sector covered is made available on request. Financial products and/or services to which the Materials relate will only be made available to Professional Clients and Market Counterparties. The Product is made available in United Kingdom by Citigroup Global Markets Limited, which is authorised by the Prudential Regulation Authority ("PRA") and regulated by the Financial Conduct Authority ("FCA") and the PRA. This material may relate to investments or services of a person outside of the UK or to other matters which are not authorised by the PRA nor regulated by the FCA and the PRA and further details as to where this may be the case are available upon request in respect of this material. Citigroup Centre, Canada Square, Canary Wharf, London, E14 5LB. The Product is made available in United States by Citigroup Global Markets Inc, which is a member of FINRA and registered with the US Securities and Exchange Commission. 388 Greenwich Street, New York, NY 10013. Unless specified to the contrary, within EU Member States, the Product is made available by Citigroup Global Markets Limited, which is authorised by the PRA and regulated by the FCA and the PRA. The Product is not to be construed as providing investment services in any jurisdiction where the provision of such services would not be permitted. Subject to the nature and contents of the Product, the investments described therein are subject to fluctuations in price and/or value and investors may get back less than originally invested. Certain high-volatility investments can be subject to sudden and large falls in value that could equal or exceed the amount invested. Certain investments contained in the Product may have tax implications for private customers whereby levels and basis of taxation may be subject to change. If in doubt, investors should seek advice from a tax adviser. The Product does not purport to identify the nature of the specific market or other risks associated with a particular transaction. Advice in the Product is general and should not be construed as personal advice given it has been prepared without taking account of the objectives, financial situation or needs of any particular investor. Accordingly, investors should, before acting on the advice, consider the appropriateness of the advice, having regard to their objectives, financial situation and needs. Prior to acquiring any financial product, it is the client's responsibility to obtain the relevant offer document for the product and consider it before making a decision as to whether to purchase the product. Citi Research product may source data from dataCentral. dataCentral is a Citi Research proprietary database, which includes the Firm's estimates, data from company reports and feeds from Thomson Reuters. The printed and printable version of the research report may not include all the information (e.g., certain financial summary information and comparable company data) that is linked to the online version available on the Firm's proprietary electronic distribution platforms.

by the recipient to any third party. Where included in this report, MSCI sourced information is the exclusive property of Morgan Stanley Capital International Inc. (MSCI). Without prior written permission of MSCI, this information and any other MSCI intellectual property may not be reproduced, redisseminated or used to create any financial products, including any indices. This information is provided on an "as is" basis. The user assumes the entire risk of any use made of this information. MSCI, its affiliates and any third party involved in, or related to, computing or compiling the information hereby expressly disclaim all warranties of originality, accuracy, completeness, merchantability or fitness for a particular purpose with respect to any of this information. Without limiting any of the foregoing, in no event shall MSCI, any of its affiliates or any third party involved in, or related to, computing or compiling the information have any liability for any damages of any kind. MSCI, Morgan Stanley Capital International and the MSCI indexes are services marks of MSCI and its affiliates. The Firm accepts no liability whatsoever for the actions of third parties. The Product may provide the addresses of, or contain hyperlinks to, websites. Except to the extent to which the Product refers to website material of the Firm, the Firm has not reviewed the linked site. Equally, except to the extent to which the Product refers to website material of the Firm, the Firm takes no responsibility for, and makes no representations or warranties whatsoever as to, the data and information contained therein. Such address or hyperlink (including addresses or hyperlinks to website material of the Firm) is provided solely for your convenience and information and the content of the linked site does not in any way form part of this document. Accessing such website or following such link through the Product or the website of the Firm shall be at your own risk and the Firm shall have no liability arising out of, or in connection with, any such referenced website.

ADDITIONAL INFORMATION IS AVAILABLE UPON REQUEST