Leveraging Digital Data Streams: The Development and Validation of a Business Confidence Index

Gabriele Piccoli
Louisiana State University
University of Pavia
gpiccoli@lsu.edu

Joaquin Rodriguez
Engineering Ingegneria Informatica Spa
University of Pavia
joaquinalfredo.rodriguez@eng.it

Richard T. Watson
Terry College of Business
University of Georgia
rwatson@terry.uga.edu

Abstract

Quarterly Earnings Calls by public companies represent an example of a rich digital data stream. In this study, we intercept this data stream and use it to create a new business confidence index, the P&W Index™. This index correlates with existing business confidence indices, both consumer and business, and correlates with US GDP percentage change. The main advantage is that it can be updated as Earnings Calls are released, and there are typically multiple such releases in a week. In contrast, traditional measures are computed on a monthly or quarterly basis. Further, the index can be computed by industry or region. Initial evidence supports further research to validate and refine the proposed business confidence index.

1. Introduction

Organizations, governments, and business leaders use predictions of the future to make decisions today. They rely on indicators, such as a business confidence index, to help determine investments, hiring numbers, advertising, and organizational priorities in general. Now, the proliferation of digital data streams enables access to real-time or near real-time data that were not available before. We provide a definition, a framework for analyzing digital data streams, and demonstrate an application of data stream analysis in the context of business confidence index development. You can think of a business confidence index, or the better-known consumer confidence index, as a wisdom-of-the-crowd barometer. It is designed to abstract, and make usable, the current level of confidence of the population of interest.

However, current confidence indices have some limitations. First, because they rely on surveying very busy informants (e.g., CEOs), they can only be computed at fixed intervals (e.g., monthly, quarterly), following the constraints of human opinion data collection. Second, because they require ad-hoc polling, they rely on data from a limited subset of the population of interest. Improvements on these two dimensions, frequency and breadth, would likely improve information quality and predictive power of the index.

In this research, we present an early attempt to address the shortcomings of traditional confidence indices. We extract information from a high-quality financial digital data stream to automate the computation of a business confidence index. The proposed method and index: a) do not require the direct participation of business leaders, relying instead on a publicly available digital data stream that captures their thinking; b) has greater comprehensiveness because it enables automatic polling of the thousands of organizations that report their earnings each quarter; c) can be revised every day based on the earnings calls released that day; d) is based on expressed opinions and specific topics of relevance to the organizations, rather than responses to a multi-choice questionnaire with fixed items. For example, the data stream we use includes 3914 firms for 2013. For this study, we use a subset of the data stream providing information on 322 firms for an average of 700 documents per year with each document containing an average of 54,000 words.

The paper is organized as follows: in the next section we introduce the concept of a digital data stream (DDS) and discuss earning calls as an example of a DDS. We then introduce economic index theory and explain how it informs the creation of our proposed business confidence index. The following section explains how we intercept the DDS, its preprocessing, and how we compute the index. We then report initial validation of the proposed index against more established economic activity indicators, and we conclude our work with future research implications.

2. Digital Data Streams

"An event is anything that happens or is thought of as happening"[15]. An event is therefore an occurrence at a specific point in time and space. An event can be thought of as discrete, like the collision of two cars, or continuous, like the journey of a car from its departure to its destination point. Critical to this definition is the observer's unit of analysis, as well as the frame of reference adopted. This approach leads to definition of levels or layers of events [15]. For example, the marketing department of a supermarket chain performs sales analyses at the basket-of-goods level. However, this level is inadequate for capturing the order in which items are scanned, which might be of interest for implementing an optimal packing sequence to speed throughput and reduce required packing supplies (e.g., scan and pack the large items first and let the small items fill the spaces). In this case, an item scan is the unit of analysis and thus the event definition should focus at that level. Higher-level events can be derived from lower-level ones a posteriori (e.g., the individual scans for a basket can be aggregated to create a single sale event). While it is not possible to decompose events after encoding, it is usually feasible to redefine the unit of analysis to capture greater detail for future events.

We use the prior definition to define a digital data stream as the continuous digital encoding and transmission of data describing a related class of events. The DDS concept builds on the idea of digital data genesis, the digital representation of a discrete event, or of a continuous event at a particular point in time [17]. The transmission or flow of these digital representations of events is a digital data stream.

DDSs are an important driver behind the current big data trends, contributing to increasing both volume and velocity. The cost of collecting, streaming, and storing data is inversely related to the level at which events are tracked. However, this relationship is less penalizing for digital data streams than for static data that require collection, storage, and aggregation to deliver value. Given the layered nature of events, accumulating data at a low level quickly leads to the need for massive storage and analytical power. Conversely DDS streaming enables real time processing and analysis. The challenge is to match event level analysis and business significance from among all the events being generated and streamed.

Streaming an event's data is, of course, not a new concept. The transmission of encoded information has occurred since humans developed recording capabilities, such as writing. Consider the journal kept by Captain James Cook to chronicle the voyages of *HMS Endeavour* [22] or the famous 1937 radio broadcast chronicling the *Hindenburg* disaster [21].

While data streaming has occurred for some millennia, the recent evolution of digital computers and networks has laid the foundation for *digital* data streaming, thus dramatically increasing the real-time and near real-time streamability of event data.

We define streamability as the degree to which an event is amenable to being encoded and transmitted. It is a function of an event's characteristics and available technology capabilities. Events differ in some intrinsic properties, namely a) **detectability** – the threshold necessary to sense the occurrence of the event. If a person is not present at a small meeting, the individual's absence, an event, is highly detectable by those present. If a person is stealing a negligible amount of money from her company (e.g., penny shaving or salami slicing), the event may be very difficult to detect. Discrete undetectable events cannot be streamed. For continuous events, fluctuating detectability results in missing data and limited streamabilty; b) **measurability** – the degree of difficulty in accurately establishing the magnitude of relevant attributes of the event. For example, a firm's quarterly profits have high measurability, while pain has low measurability and is typically estimated in medical settings by a patient's self-referential rating on a subjective ten-point scale. Measurability is important because low measurability limits the accuracy of the representation of the event (i.e., its encoding), reducing its suitability for streaming.

Technological advances continually increase streamabilty by lowering both the threshold for event detection of the particular properties of an event. They also increase the accuracy with which a property can be measured. Thus, while pain is currently a subjective measure it is conceivable that a technology will be developed that will measure pain on a continuous scale just as readily and reliably as it is done today with blood pressure. In effect, we can expect that new technology will increase the opportunity to stream more events, to stream more data about events, and to increase the accuracy of a data stream's properties.

2.1. Elements of a DDS

A digital data stream consists of digitally encoded elements that describe a discrete event (e.g., a retail sale, the berthing of a ship at a port of call), or the current state of an entity (e.g., the current location of a car, the current mood of a person). We identify seven elements that a DDS can capture (Table 1). Each element contributes to a full description of an event (e.g., when, where). Some elements are atomic, while others can hold multiple data item for the same event.

For example, a retail sale might have two 'what' elements, a credit card number and a 10% coupon. Some elements might be null. For instance, elements in a car's GPS DDS may not contain any value for the 'why' element, not because it doesn't exist but because there is no technology for currently assessing motivation without asking the driver.

Table 1. DDS elements

Element	Description	Example
When	The time when the data segment was created	A timestamp with date, time, and time zone
Where	The location of the entity when the segment was created	Latitude, longitude, elevation
Who	The unique identifier of the entity that caused the data segment to be created	Person's customer number, RFID of a pallet, URL of a web site
What	The activity that caused the segment to be created	The identifier of an item in a sales transaction, the arrival of a ship in a port
How	The means by which the event was initiated, authorized, or completed	Credit card number for payment, status of arriving flight (e.g., safe landing)
Why	Motivation for the action related to data segment creation	Birthday gift, planned destination
Outcome	The result of an action related to data segment creation	Revenue, cost, resources used, customer satisfaction

2.2. Genesis of a DDS

In order to become available for use, a digital data stream goes through three stages: birthing, streaming, and harvesting [16]. Birthing occurs either through digital data genesis, when the digital representation of the event is created concurrently with its birth. Examples include a digital picture or a Google search. Birthing can also result from the digital encoding of past events. Examples include the scanning of a physical image or the recording of a past event in a database.

Streaming occurs when the digital representations of the elements of an event are transmitted in a shared intelligible format on a channel that authorized people or machines can access. Thus, a public data stream will be typically transmitted on the Internet with JSON or XML format. A private stream might also be streamed on the Internet, but access will require a password and account and the stream's data might be encrypted.

Harvesting occurs when an organization taps into a DDS to extract value, as when TripIt or Traxo, takes an airline's email regarding a reservation, extracts the relevant facts, and creates a calendar entry. During

harvesting, the firm may also need to contextualize the data stream. Contextualizing is the process of merging static data (e.g., a standard airport code, the location of a hotel) or data from multiple DDSs to create an integrated set of facts.

After tapping into (or harvesting) a DDS, a firm can either process the single DDS or link and integrate it with existing data streams or stored data. An insurance company that monitors a weather forecast data stream and sends a text message to its customers in an area where hail is expected in the next 30 minutes illustrates the immediacy of a process-and-actuate approach [16]. Other companies like Inrix and TomTom merge and analyze data from multiple streams to provide traffic volume predictions and compute dynamic routes to avoid current congestion points, thus engaging in an analytic approach [16].

2.3. Earnings Calls as a DDS

An Earnings Call (EC) is an electronic meeting organized by a company through an external teleconferencing service to report its periodic performance [4,18]. ECs are examples of performance reporting events, in many cases voluntary, including CSR reports, letters to the shareholders, and new product launches. ECs are typically held in conjunction with required quarterly earning reports.

Analog broadcasting, once the standard for communication, has today migrated to the Internet – the preferred medium of communication of organizations with their customers and investors [12]. This is particularly true for financial communications. Traditional paper reports and annual meetings are increasingly giving way to the required transmission of XML encoded reports (e.g., 10-K) and webcasts. This shift has enabled organizations to improve their communication with their stakeholders, and the market in general [1].

A firm's top management generally conducts an EC with a few analysts of major global investment banks participating directly to ask questions, after the leadership has commented on the quarterly results. A digital transcript of the EC is produced and made available on the same day as the call. Earning calls are characterized by formal language [7] and are conducted using a standard format and a recognizable microstructure (Table 2). High content elements of the EC are sections three to six. While ECs are primarily a commentary on the firm's past three months of activity and the relative results, they can also include forward-looking statements and information.

Table 2. Earnings Call Structure

	Move/Steps	Speaker(s)
1	Opening Welcoming participants Giving technical instructions Passing the floor to the first corporate speaker	Teleconference operator
2	Introduction Greeting participants Introducing the team of speakers Outlining the presentation Stating the Safe Harbor disclaimer Passing the floor to the next corporate speaker	Investor Relations Officer (or: CEO, CFO, Vice President)
3	Financial overview Greeting participants Commenting on overall financial results Instilling confidence Passing the floor to the next corporate speaker	Chief Executive Officer (or: Co-Chairman/Co-CEO, CFO)
4	Financial details Elaborating on financial statements Forecasting future trends Instilling confidence Passing the floor to the next speaker	Chief Financial Officer (or: CAO, Treasurer)
5	Wrap-up Discussing remaining topics Instilling confidence Opening the Q&A sessions	Chief Executive Officer (or: CFO, COO)
6	Q&A sequences (repeated until all participating analysts have had the opportunity to ask a question)	Teleconference operator + Analyst + Member of Executive Team
7	Closing Announcing the end of the call Inviting participants to disconnect	Teleconference operator

As a DDS, ECs have the following relevant elements: when (e.g., Apr. 23, 2014 8:33 PM ET), who (e.g., AAPL (APPLE)), what (e.g., Second Quarter Fiscal Year 2014; prepared remarks, question and answer session). A medium level of streamability has historically characterized ECs due to their limited detectability and low measurability. Specifically, while ECs have always been publicly communicated to stockholders, this information was difficult to obtain because until several years ago, the quarterly earnings press release and accompanying conference call were not digitized or standardized. From March 2003 onwards, ECs became regulated under the Securities Exchange Act [3].

Measurement of an EC's content requires an individual to manually transcribe verbatim what the participants said. While the streamability of the EC event has traditionally been relatively low, technology has contributed to mitigate limitations of detectability and measurability. Today, new ECs are easily detected because organizations are required to publicize the EC date and make the call accessible to the public. Traded firms are required to file SEC Form 8-K "to announce major events that shareholders should know about" [23]. Earnings publications and associated commentary, such as ECs, are an example of an event that triggers the filing of Form 8-K. Often organizations issue a press release announcing the conference call, including date, time, and instructions on how to access it. Today, the detectability of EC events is very high thanks to financial community sites such as Yahoo Finance who provide automatic calendar alerts for upcoming calls. Technology has also contributed to mitigating the low measurability of EC content with the use of webcasts as the predominant medium and speech to text algorithms for transcription.

3. Economic index theory

The measurement of economic conditions has been a concern for humankind since the beginning of organized activity. For example, hyperinflation was one of the catalysts for the demise of the Roman Empire [19]. National Accounts were introduced to provide detailed economic data by measuring the economic activity of a country [5]. However, National Accounts are not enough to provide a complete view of the economic condition of a country, and they are not predictive in nature. These limitations spurred interest in the development of economic indices with low measurement error and high predictive accuracy. Specifically, an economic index is "a measure, a function, which maps, on the one hand, a set D of economically interesting objects into the set R of real numbers" [10].

$$F : D \rightarrow R$$

The set D contains the economic indicators that comprise the index. Economic indices should satisfy the following criteria [20]:

1. How well understood and how important in the business cycles are the variables represented by the data (*economic significance*)?
2. How well does the given series measure the economic variable or process in question (*statistical adequacy*)?
3. How consistently has the series led (or coincided or lagged) a business cycle's peaks and troughs (*timing of recessions and revivals*)?
4. How regularly have the movements in the specific indicator reflected expansions and contractions in the economy at large (*conformity to historical business cycles*)?
5. How promptly can a cyclical turn in the series be distinguished from directional change associated with shorter, irregular movements (*smoothness*)?

6. How timely are the statistics and how frequently are they reported *(currency or timeliness)*?
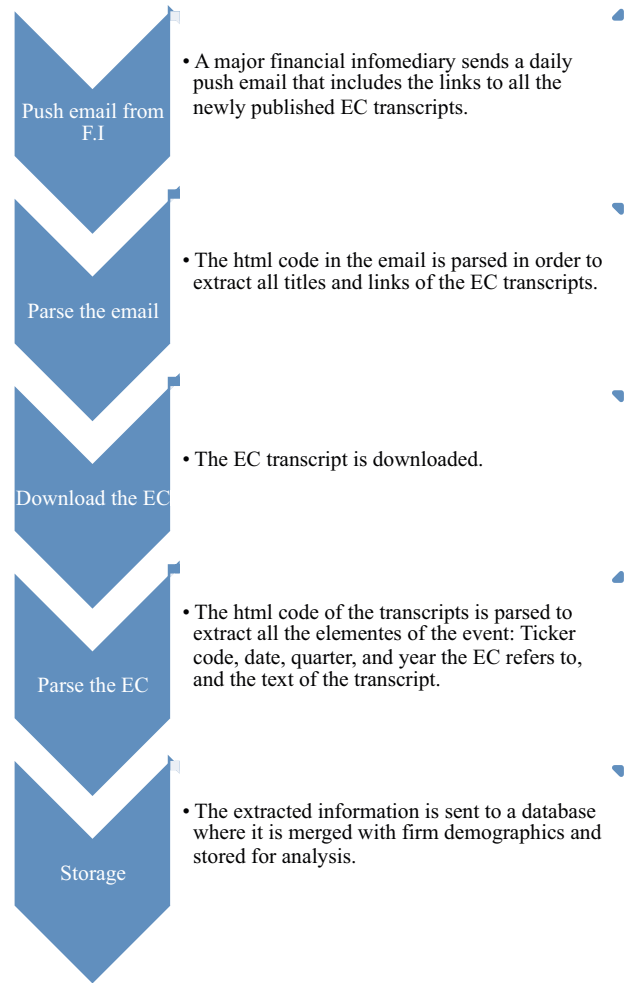
Given the embryonic nature of our research, we are unable at this stage to assess our proposed index against the six criteria (an agenda for such research is proposed later). Rather, in order to establish the potential value of the EC for economic prediction, we benchmark the historical performance of our proposed index against established indices – those that have exhibited the six characteristics and are well accepted by policy-makers, firms, and the academic community. Establishing the potential value of ECs as a source of data is important because an index that leverages them has some advantages over more established ones. For instance, our index would allow tracking of business sentiment weekly, or even daily. In addition, our index is more comprehensive than the established confidence indices based on sample surveys. Finally, ECs have the potential to produce multiple indices (e.g., a geographic, commodity, or industry index).

4. EC confidence index construction

4.1. Harvesting the earning calls stream

We harvest ECs daily by automatically downloading from a major financial infomediary[1] the transcripts released. An EC is available shortly after it is broadcast. We base the early validation of the index presented in this paper on ECs produced by the largest for-profit entities in the world. The list is the 2012 Financial Times 500 [9]. We chose the FT 500 because membership is global. Since our goal is to tap into the thinking of the organizations with the greatest weight on global economic activity, a transnational list is needed. To complete the research dataset, we retroactively downloaded all available ECs for the firms in the FT 500 list spanning the 2006-2014 time frame.

Once transcripts are downloaded in the form of an HTML page, they are automatically parsed using R 3.0 to extract the elements of each streamed event (i.e., the earnings call). Afterwards, the extracted data are written to a database that also includes basic demographic information (e.g., address, standard industry codes) about each firm's holding earning calls (Figure 1).

---

[1] We have permission for downloading.

Push email from F.I — • A major financial infomediary sends a daily push email that includes the links to all the newly published EC transcripts.

Parse the email — • The html code in the email is parsed in order to extract all titles and links of the EC transcripts.

Download the EC — • The EC transcript is downloaded.

Parse the EC — • The html code of the transcripts is parsed to extract all the elementes of the event: Ticker code, date, quarter, and year the EC refers to, and the text of the transcript.

Storage — • The extracted information is sent to a database where it is merged with firm demographics and stored for analysis.

Figure 1. Earnings call DDS workflow

Since the filing of ECs is not a binding legal requirement, only 322 of the 500 organizations in the list have issued earning calls in the last eight years. Our dataset for this study includes a total of 5,851 events over a ten-year period (Table 3).

4.2. Pre-processing

We performed preprocessing in order to facilitate the analysis of the EC content. We created a corpus using the tm package in R [11]. Prior to computing the index we performed the following steps:
- Transformation to lower case letters of all the text
- Removal of numbers
- Removal of English stop-words
- Removal of any special characters

Table 3. Number of earnings call by year

Year	Number of ECs
2005	16
2006	284
2007	514
2008	683
2009	671
2010	714
2011	823
2012	908
2013	1,058
2014	180
Total	**5,851**

4.3. Index construction

The P&W Index™, which we refer to as *the index* in this article, takes as input the frequencies of positive and negative words present in an EC transcript.[2] The positive and negative terms are taken from a modified version of a standard, validated dictionary of financial terms [13]. This dictionary is particularly suited for financial document analysis, since it takes into account the polarity of a term in the context of financial communications instead of its general English language meaning. For instance, a word like liability, which commonly has a negative meaning, is considered neutral by this dictionary.

The index for each document is computed as:

$$PW_d = \frac{\sum pos_d - \sum neg_d}{\sum pos_d + \sum neg_d}$$

Where:
PW_d is the sentiment of EC document d
pos_d is the count of positive terms in document d
neg_d is the count of negative terms in document d

The index is constructed by aggregating the counts of positive and negative terms encountered in EC for a specific time period, which is specifiable depending on

---

[2] The trademark, P&W Index™, refers to any of the family of indices derived by textual analysis techniques from quarterly earnings calls, annual letters to the shareholders, and other financial, or financially oriented, documents released by a company. An index can be calculated for a single document, a set of documents for a time period, an industry, country, or region. An index can be calculated using the full document or a portion thereof.

the purpose. Thus data can be aggregated by day, week, month, quarter, and so forth. The monthly and quarterly aggregation allows comparison to common monthly or quarterly indices and measures, such as Gross Domestic Product (GDP), the OECD Composite Leading Indicators, the Michigan Consumer Confidence Index (MCI), and other indices based on national accounts data. The index for a specific period would take the form:

$$PW_t = \frac{\sum_1^n pos - \sum_1^n neg}{\sum_1^n pos + \sum_1^n pos}$$

Where:
t is the time period of interest
n is the number of ECs for the period t
pos_t is the count of positive terms across all documents for period t
neg_t is the count of negative terms across all documents for period t

5. EC confidence index validation

The main assumption underpinning our work is that the information extracted from an earnings call contains valuable insight for understanding economic trends. The simplest form of corroboration of this proposition is to demonstrate that the index has significant correlation with other recognized indices. If this is the case, the ability to compute the index more frequently than conventional indices will provide a more timely and comparable instrument for analysis and prediction. Moreover, if correlation with other indices were less than 1, the index would contain non-overlapping insights that may be valuable for explanatory and/or predictive purposes.

5.1. Validation methodology

As a form of validation of the informational value of an EC-based index, we compare it to two classes of established indices: confidence indices and economic output measures. The first set is to show that our index is consistent with other confidence indices. The second is to illustrate that the index is correlated with the most established economic business cycles indicator. Finally, we compare our index to other established confidence indices with regard to this correlation with an economic output measure.

We use the Pearson product-moment correlation coefficient as the measure of consistency. We include the following indices and measures as benchmarks for our index:

- The Michigan Consumer Confidence Index (MCI)
- The European Industry/Business Climate Indicator (BCI)
- The OECD Composite Leading Indicators (CLI) for the US
- Quarter-over-quarter percentage change in GDP (US)

The Michigan Consumer Confidence Index (MCI) is one of the two most well known and widely used indices of consumer confidence. It is based on respondents answering approximately 50 core questions about personal finances, business conditions, and buying conditions [25].

The OECD Composite Leading Indicators (CLI) is composed from a wide range of key short-term economic indicators relative to a specific country [24].

The European BCI indicates the confidence level of all industry sectors in Europe. It is important to note that although it indicates the confidence of European firms, research shows that the European BCI incorporates US confidence [8].

With respect to business cycles and overall economic activity, we use quarter-over-quarter percentage change in Gross Domestic Product (GDP). GDP change is a good measure of the overall trend of the economy, as it measures all the added value produced in a specific country. We use US GDP because US companies dominate our data set.

5.2. Validation results

We analyze the result of the correlations between our index and the other indices chosen for comparison. We first report a monthly comparison for November 2005 through October 2013.

Table 4. Correlation between indices

	P&W Index	MCI	BCI	CLI
P&W Index	1.00	0.56	0.66	0.68
MCI	0.56	1.00	0.56	0.65
BCI	0.66	0.56	1.00	0.91
CLI	0.68	0.65	0.91	1.00

All the correlations in Table 4 are significant at α = 0.05 level. Using Cohen's criteria for effect size [6], we have a large effect size ($\rho \geq 0.5$) for the correlation of our index with the other measures. The P&W Index appears to correlate better with the other business climate indices than the Michigan Consumer Index

does. This result is true for the CLI (p = 0.0349; n = 96), but not for the BCI (p = 0.0899; n = 96). The EU Business Climate Indicator and the US Composite Leading Index display the strongest pairwise correlation of all indices. From the charts in the appendix we can visually compare the trend of the different indices. Such examination corroborates the notion that our index is congruent with the more established indices. The relationship between the four measures is shown visually in Figure 2 and Figure 3 (Appendix A).

We then compare the correlation between the various indices and the quarterly change in GDP. All correlations are significantly different from zero at the α = 0.05 level and are considered large in magnitude (Table 5).

Table 5. Correlation of indices with ΔGDP

	P&W Index	MCI	BCI	CLI
Δ GDP	0.717	0.491	0.642	0. 639

A significance test for the difference in magnitude of the correlation of the P&W Index versus the benchmarks indicates that it is significantly larger than MCI (p = 0.0179; n = 31). This result is encouraging. It suggests that household consumer confidence provides an inferior correlation to GDP change than business climate indices.

The P&W Index has a higher correlation coefficient than the other consumer confidence indices with GDP change, but this difference is not significant at the α = 0.05 level with either BCI (p = 0.0806; n = 31) or CLI (p = 0.0752; n = 31). This result indicates that our proposed index correlates better than the MCI and is on par with the other established indices for GDP change.

6. Building on early results

While certainly preliminary, our results make us optimistic about the possibilities that ECs offer as a foundation for a new index for economic predictions. In order to be accepted as a full-fledged economic index, however, the P&W Index should be tested against the six criteria described above. Future work should validate our underlying assumption that the overall sentiment expressed by organizations in ECs is a proxy for confidence. This can be accomplished by refining the text analysis algorithm. One line of work would continue using a dictionary and focusing on refining it around the confidence concept. Another approach consists of leveraging emerging text-analytic techniques (e.g., topic modeling [2]). Another important avenue for development relates to the

validation of the predictability of the index. Instead of measuring how well the P&W Index tracks other confidence indices, we can measure how well it correlates with economic variables of interest. In this early version of the paper we have focused on GDP change. However, our work requires broadening the scope of the analysis to other important indicators (e.g., industrial production, unemployment) as well as methodological refinement. A lot of work remains to be done to reliably extract valuable predictive information from earning calls. Future research should focus on identifying what other predictive data, besides overall business confidence, is contained in an EC. This extension could be vertical, designed to increase the quality of the P&W confidence index, or horizontal, designed to create other indices. For example, extracting specific topics like geographical information, brand, or product data, or general trends information (such as the discussion of sustainability initiatives, the firm's investment level, the employment level, etc.).

Finally, prediction is the most exciting new avenue for development of indices based on EC. The nature of an EC as a DDS makes it well suited for forecasting. The typical EC contains both prepared remarks as well as a question and answer session. Differentiating between the tone and directionality of prepared remarks versus the Q&A section is also a possible avenue for future research. Interesting questions in this area pertain to the prediction potential of different elements of ECs: prepared statements, question and answers, and forward-looking statements. While an EC is primarily a commentary on a firm's past three months of activity and the relative results, the most valuable element is the inclusion of explicit or implicit forward-looking information. It can be in the form of a discussion or commentary on the strategic moves of the company, or general statements about the expected future state of the economy. Identifying this type of forward-impacting information and using it for prediction of indicators, rather than just correlation with current indices, is a promising direction for future research. With this in mind, there is a need for a specialized dictionary that includes only words that are highly relevant to determining the state of the business cycle. An alternative consists is to use classifiers, or other machine learning algorithms, to discriminate forward-looking statements [14].

7. Conclusions

The emergence of DDSs related to economic agents (customers, companies, public administrations) has opened the way for a new set of indicators and indices. Among the different kinds of data, we have chosen to focus attention on a promising set of financial documents: Earnings Calls (EC). ECs have a number of valuable qualities: they are released on a regular basis, they are formal in nature, they are relatively structured, and they are thorough, providing significant content.

In this paper, we provide an early attempt in developing an automated business confidence index leveraging an EC's content. In doing so, we use relevant elements of the EC DDS: Who, When, What. We show that with these three elements we can validate our index, and we show that it is significantly correlated with other well-established consumer and business confidence indices. In addition, we evaluate the correlation of the index with an economic output measure (i.e., US GDP change). The validation confirms our expectations about the potential of the information included in an EC to provide a valuable business confidence barometer. It also provides some promise in relation to other indices by being as correlated or significantly more correlated to GDP change than the Michigan Consumer Confidence Index (MCI), the European Industry/Business Climate Indicator (BCI), or the OECD Composite Leading Indicators (CLI) for the US.

The digitization of business has given rise to digital data streams that enable real-time analysis of current events. The transcription and electronic publication of earnings calls has created a digital data stream that can be mined to gain current insights into the opinions of the world's business leaders on the state and direction of the economy. By describing the P&W Index, this research illustrates the potential of the earnings call digital data stream to provide valuable economic indicators.

8. References

[1] Bhatia, V.K. Interdiscursivity in Professional Communication. *Discourse & Communication 4*, 1 (2010), 32–50.

[2] Blei, D.M. Probabilistic topic models. *Magazine Communications of the ACM 5*, 4 (2012), 77–84.

[3] Bochner, S.E. and Blake, R.C. The Earnings Release: Legal Requirements and Best Practices. *Insights: The Corporate and Securities Law Advisor 22*, 3 (2008).

[4] Bushee, B.J., Matsumoto, D., and Miller, G.S. Open versus Closed Conference Calls: The Determinants and Effects of Broadening Access to Disclosure. *Journal of Economics and Accounting 34*, 1–3 (2003), 149–180.

[5] Carson, C.S. The history of the united states national income and product accounts: The development of an analytical tool. *Review of Income and Wealth 21*, 2 (1975), 153–181.

[6] Cohen, J. A power primer. *Psychological Bulletin 112*, 1 (1992), 155–159.

[7] Crawford Camiciottoli, B. Earnings calls: Exploring an emerging financial reporting genre. *Discourse & Communication 4 (4)*, (2010), 343–359.

[8] Dées, S. and Brinca, P.S. Consumer confidence as a predictor of consumption spending. Evidence for the United States and the Euro area. *European Central Bank*, 1349 (2011).

[9] Dullforce, A.-B. FT 500 2012. *Financial Times*. http://www.ft.com/cms/s/0/e9d0412e-ca82-11e1-89f8-00144feabdc0.html#axzz33ePfDD73.

[10] Eichhorn, W. What is an Economic Index ? An Attempt of an Answer. In *Theory and Applications of Economic Indices*. Springer, 1978, 3–42.

[11] Feinerer, I., Hornik, K., and Meyer, D. Text Mining Infrastructure in R. *Journal of Statistical Software 25*, 5 (2008).

[12] Garcés-Conejos Blitvich, P. Conference Calls, the Business Meetings of the 21st Century: A Multi-Layered Approach to an Emerging Genre. *Pragmática, Discurso y Sociedad. Quaderns de Filologia Estudis Lingüístics XII*, (2007), 57–75.

[13] Loughran, T. and Mcdonald, B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance 66*, 1 (2011), 35–65.

[14] Lucia, N., Alessandro, Z., Ignazio, G., Gabriele, P., and Joaquin, R. Automatic Prediction of Future Business Conditions. *9th International Conference on Natural Language Processing*, (2014).

[15] Luckham, D.C. *Event Processing for Business: Organizing the Real-Time Enterprise*. John Wiley & Sons, 2011.

[16] Piccoli, G. and Pigni, F. Harvesting External Data: The Potential of Digital Data Streams. *MIS. Quarterly Executive 12*, 1 (2013), 143–154.

[17] Piccoli, G. and Watson, R.T. Profit from Customer Data by Identifying Strategic Opportunities and Adopting the "Born Digital" Approach. *MIS Quarterly Executive 7*, 3 (2008), 113–122.

[18] Saatchi, E. The Discourse of Voluntary Disclosures in Quarterly Conference Calls: Implications for Investor Relations. *Discourse and Identity in Specialized Communication*, Milan: Lubrina Editore (2007).

[19] Temin, P. The Economy of the Early Roman Empire. *Journal of Economic Perspectives 20*, 1 (2006), 133–151.

[20] Zarnowitz, V. Composite Indexes of Leading, Coincident, and Lagging Indicators. *National Bureau of Economic Research Business Cycles: Theory, History, Indicators, and Forecasting*, (1992), 316–356.

[21] Top 10 Historic Radio Broadcasts. *Listverse*, 2007. http://listverse.com/2007/07/19/top-10-historic-radio-broadcasts/.

[22] Captain Cook's Journal During His First Voyage Round the World by James Cook. *Project Gutenberg*, 2014. http://www.gutenberg.org/files/8106/8106-h/8106-h.htm.

[23] Form 8-K. *Sec*, 2014. http://www.sec.gov/answers/form8k.htm.

[24] Composite Leading Indicators. *OECD*, 2014. http://stats.oecd.org/Index.aspx?DataSetCode=MEI_CLI.

[25] Survey Information. *Surveys of Consumers. Thomson Reuters University of Michigan*. http://www.sca.isr.umich.edu/fetchdoc.php?docid=24774.
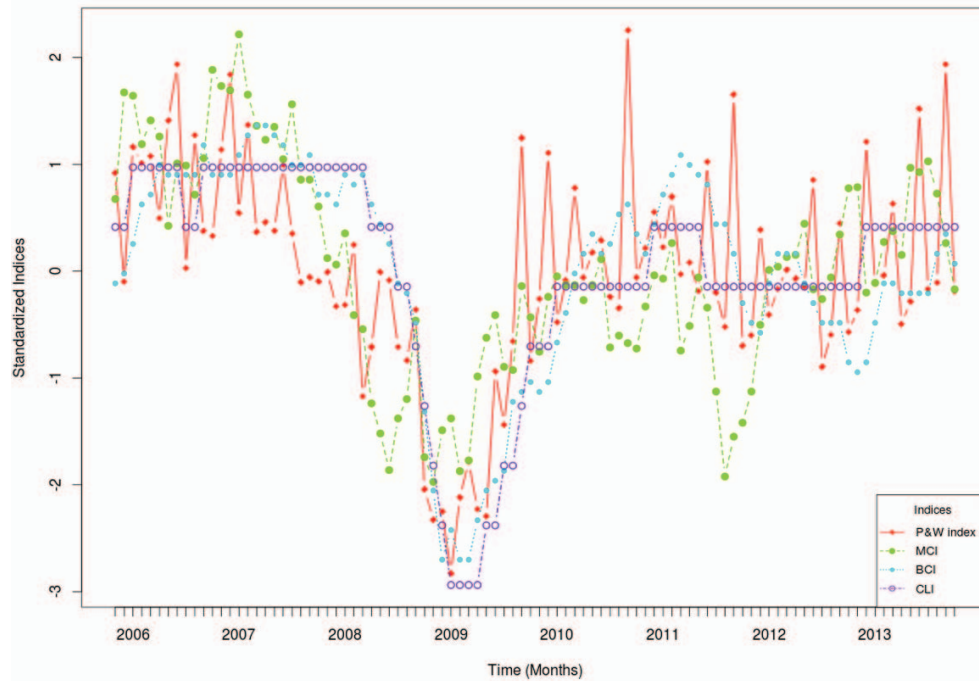
9. Appendix

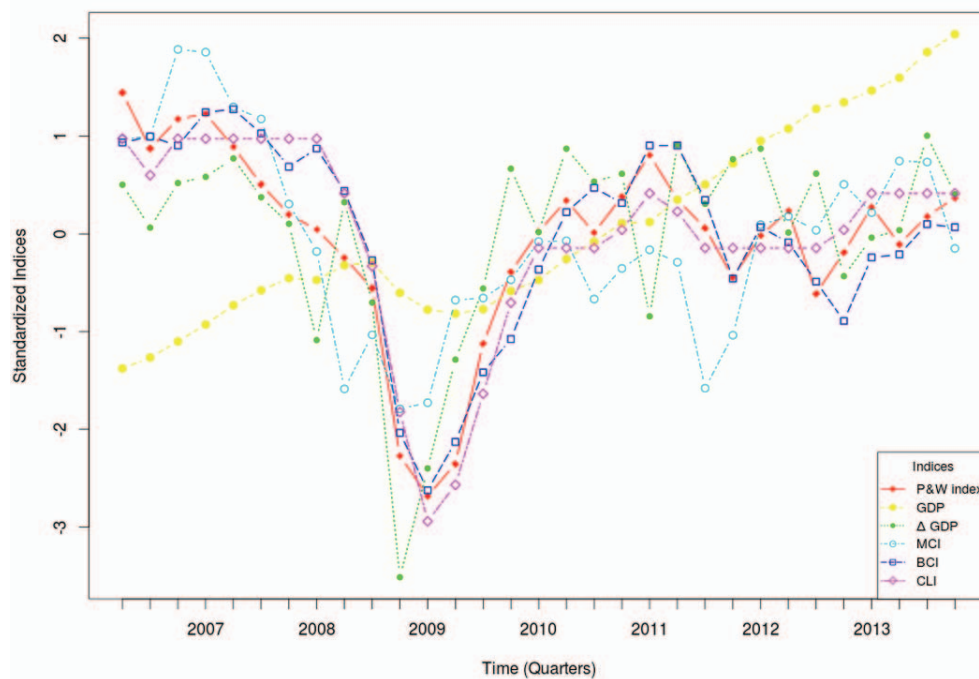**Figure 2. Monthly time series of the analyzed indices**

Figure 3. Quarterly time series of the analyzed indices and GDP