Measuring ECB's Communication: A "Media-Based" Automated Approach*

February 23, 2017

Abstract

I present a new measure of the monetary policy communication of the European Central Bank (ECB) with reference to future policy rate decisions of the Governing Council (GC) of the ECB. I construct an index based on the content of media articles published in days around monthly press conferences held by the ECB Presidents after meetings of the GC in the period between 1999 and 2013. I provide evidence that the index measures information related to the words pronounced by the ECB Presidents at press conferences rather than to policy rate levels set at GC meetings immediately before press conferences. I study the relationship between the ECB's monetary policy communication, as measured by the index, and future monetary policy decisions of the GC and I find it to be statistically significant. Finally, I find that "shocks" in the ECB's monetary policy communication affect market expectations of future levels of the policy rates.

JEL Classification Codes: E43, E52, E58

Keywords: central bank communication, monetary policy communication, ECB, ECB communication, monetary policy forecast, media-based approach, automated approach

1 Introduction

In this paper I study the monetary policy communication of the European Central Bank (ECB). I attempt to answer to the question "can (information contained in) monetary policy communication be measured, possibly in a not-subjective way?" by presenting an index aimed at proxying the informational

*I would like to thank Stefano Nardelli, who encouraged me to study the monetary policy communication of the European Central Bank. I am grateful for comments to Roberto Casarin, participants to the XXV International Rome Conference in Money, Banking, and Finance (in particular Stefano Marzioni, my discussant), and two anonymous referees.

content, with respect to future monetary policy decisions of the Governing Council (GC) of the central bank, of press conferences held by the ECB Presidents after monetary policy meetings of the GC.

I focus on decisions about levels of the ECB policy rates (key rates) set by the GC[1] at its regular monetary policy meetings in the period between January 1999 and December 2013, which corresponds to the first 15 years of operations of the ECB. The period includes 175 regular monetary policy meetings of the GC, each followed by a press conference, held with a monthly frequency (almost every month, see section 3).[2]

In section 2 I present the approach to the study of the ECB's monetary policy communication, which I develop further in the following sections: I discuss the novelty of the approach and I briefly review the literature to which the paper is related. In section 3 I explain how the index is constructed. Then I examine whether the index is an appropriate measure of monetary policy communication with reference to future monetary policy decisions by following two strategies. First (in section 4), I check whether the index captures information conveyed by "present" monetary policy decisions of the GC (which are announced shortly before press conferences and which might themselves be informative about future decisions) by studying correlations between the index and measures of information related to "present" decisions of the GC. Second (in section 5), I test if the index is predictive, at least in part, of future monetary policy decisions. Finally, in section 6, I test if monetary policy communication, as measured by the index, affects market expectations of future monetary policy decisions.

I find the correlation between the index and measures of information conveyed by "present" decisions of the GC to be quite low. Such a result indicates that the index does not capture information related to "current" decisions of the GC. Moreover I find the relationship between the index and future monetary policy decisions of the ECB to be positive and statistically significant. Finally I show that the ECB's monetary policy communication, as proxied by the index, affects market expectations of future monetary policy decisions.

2 A novel approach to the study of the ECB's monetary policy communication

2.1 The novelty of the approach

This paper contributes to the literature on central bank communication by presenting an "automated" approach to the analysis of the words pronounced by

[1] The GC sets levels of the three key ECB interest rates, which are: 1) the rate "on the main refinancing operations (MRO), which provide the bulk of liquidity to the banking system", 2) the rate "on the deposit facility, which banks may use to make overnight deposits with the Eurosystem", and 3) the rate "on the marginal lending facility, which offers overnight credit to banks from the Eurosystem" (see www.ecb.europa.eu/stats/monetary/rates/html/index.en.html).

[2] Since January 2015 monetary policy meetings and following press conferences are held at six-week intervals.

the ECB Presidents at monthly press conferences. Most of the studies on central bank communication are aimed at measuring the informational content of pieces of communication (either released by central banks or by other sources, such as media) related to central banks' present or future decisions. The measurement is simply the association of pieces of communication with values of indexes or variables which represent information on central banks' decisions. I distinguish between two approaches to the measurement of the content of central bank communication (and two corresponding types of indicators), according to the characteristics of the procedure through which pieces of communication are associated with values of indexes or variables representing information (measurement procedure). I define as subjective approaches (and subjective indicators) those based on measurement procedures which leave room for subjective preferences (judgement) of analysts who perform the measurement. Such procedures are replicable neither by different analysts nor by the same one(s) thereafter, since subjective preferences may vary across analysts and preferences of the same analyst(s) may vary across time. I define as automated approaches (and automated indicators) those based on measurement procedures which follow only objective criteria and leave no room for subjective judgement of analysts (although analysts choose such procedures and criteria according to their own subjective judgement). Since such procedures are based on objective criteria, they are replicable across time either by different analysts or by the same ones and can also be used to analyse (with minor modifications, if needed) "new" pieces of communication, provided that they are "similar" to those for which such procedures are initially designed.

The index is aimed at capturing information about the ECB's attitude toward its own key rates (see introduction), conveyed by the ECB Presidents at press conferences which follow monetary policy meetings of the GC. I define a central bank's (policy, key or interest) rate attitude as the willingness of the central bank to change or not to change its own policy rates in the near future until they reach desired levels (according to the preferences of the central bank toward levels of rates in the near future). More specifically, I define a hawkish attitude, or hawkishness, as the willingness of the central bank to raise rates in the near future and a dovish attitude, or dovishness, as the willingness to decrease them. I denote as neutral attitude, or neutrality, the absence of any readiness to change rates or, if it makes any difference, the tendency to keep them unchanged. Moreover I use the adjectives "hawkish", "dovish", and "neutral" (and the nouns "hawkishness", "dovishness", and "neutrality") also with reference to (the tone of) words, expressions, and (pieces of) communication which signal a hawkish, dovish, and neutral attitude, respectively.

The index is computed by counting occurences of predefined hawkish, dovish, and neutral expressions in media articles published in time intervals centred around days in which monetary policy meetings of GC and post-meeting press conferences take place (see section 3). If, in any given time interval, hawkish expressions outnumber dovish expressions, the index signals a hawkish attitude. In the opposite case the index signals a dovish attitude. If the number of hawkish and dovish expressions is the same or if articles contain only neutral expressions,

the index denotes a neutral attitude.

The approach presented in the paper is inspired by the methodology used by Lucca and Trebbi (2011), who count words in media articles to construct an indicator of Federal Reserve monetary policy communication.[3] As I noted earlier in this section, the novelty of the paper with respect to previous literature on central bank communication mainly consists in applying an automated approach based on the content of media articles (see also section 3) to the analysis of the ECB's communication on policy rate attitude. Previous studies on the ECB's communication about rate attitude rely on subjective approaches. To the best of my knowledge, the only exception is represented by Heinemann and Ullrich (2007) (see section 2.2), who construct an automated indicator aimed at capturing information on the ECB's policy rate attitude and based on words contained in the ECB Presidents' introductory statements to press conferences following GC meetings.[4] I believe that the approach which I present in the paper has some advantages over previous ones.

First, it is replicable, since: 1) the expressions whose occurrences are counted in media articles are predefined, i.e. chosen before reading the articles, and 2) such expressions are generically related to a central bank's interest rate attitude (e.g., "will raise interest rates") and are not referred to any specific feature of either the ECB Presidents or press conferences, or to any specific word or expression pronounced by the ECB Presidents at press conferences.[5] Therefore such an approach could be used to analyse pieces of central bank communication which are not considered in the paper, including press conferences, pieces of communication other than press conferences, and pieces of communication by other central banks. Moreover the transparency of the approach allows modifications which might be deemed necessary in such other applications.

Second, the approach should provide a better measure of the ECB's rate attitude, since it captures (averages) the perception of the ECB's rate attitude

[3] Lucca and Trebbi construct an index based on a word count in media articles published in three-days intervals centred around policy meetings of the Federal Reserve Open Market Committee (FOMC). After meetings, the FOMC releases statements which might contain information on the Federal Reserve policy rate attitude. For each three-day interval centred around a meeting they calculate two values of the index: an "ex-ante" value, based on articles published before the meeting, and an "ex-post" value, based on articles published after the meeting. The difference between the ex-post and the ex-ante value of the index should represent the unexpected component of the information on the Federal Reserve policy rate attitude, conveyed by the FOMC statement.

[4] Note that Jansen and De Haan (2007) also analyse the ECB's communication by employing an automated approach, albeit one aimed at measuring information on the ECB's inflation expectations in the Eurozone and not on the ECB's rate attitude (although inflation expectations could be seen as a signal for rate attitude, since a central bank might change reates as a reaction to changes in inflation expectations; namely, it might raise rates if expected inflation increases and cut them if expected inflation decreases). Jansen and De Haan count occurences of the word "vigilance" (or "vigilant") in pieces of central bank communication and news reports of different types and study if such occurences affect daily changes in market expectations of Eurozone inflation. However, they find that occurrences of "vigilance" (or "vigilant") tend to decrease and become less informative over time.

[5] Differently from Heinemann and Ullrich (2007), who examine expressions contained in introductory statments to press conferences following GC meetings.

from a plurality of sources: articles, and, in a broader sense, authors of the articles and media which published them.

The paper studies if the index predicts, at least in part, changes in policy rate levels and how monetary policy communication, as measured by the index, affects market expectations of such changes. It aims neither to present a measure of monetary policy shocks, meaning exogenous changes to monetary policy which cannot be predicted by a monetary policy rule, nor to study the effects of monetary policy shocks on the economy.[6] It would be interesting to investigate how the index might be used to identify monetary policy shocks and also if communication, as proxied by the index, is an effective monetary policy tool in a broader sense, i.e., if it affects macroeconomic variables such as inflation and (un)employement.[7] However, this line of investigation goes beyond the scope of the paper and is left to future research efforts.

2.2 Related literature

I distinguish between two streams of literature to which the paper is related.

The first one is represented by studies which focus on central bank communication by measuring its informative content (usually with reference to a central bank's rate attitude) and investigating both its role as a monetary policy tool and its relationship with financial variables, such as policy rates set by central banks, interest rates on (sovereign) bonds, prices of futures contracts, etc.[8] Within this stream of literature, the paper is more closely related to studies which examine the ECB's communication about its own policy rate attitude.[9] Most of these studies rely on subjective measures of the content of the ECB's communication: e.g., see Musard-Gies (2006), Rosa and Verga (2007 and 2008),[10] and Berger et al. (2011), who analyse the ECB Presidents' introductory statements to press conferences following GC meetings, and Gerlach (2007), who focuses on editorials of the ECB monthly bullettins. The KOF institute of the ETH Zürich constructs an indicator, called the monetary policy communicator (MPC), which is difficult to define as either subjective or automated on

[6] On the topic of the identification of monetary policy shocks and their effects on the economy, see Romer and Romer (1989 and 2004), Leeper (1997), Bernanke and Mihov (1998), and references cited in such articles.

[7] Neuenkirch (2013) investigates the role of ECB's communication in monetary policy transmission through a VAR analysis, and Lucca and Trebbi (2011) use the same type of technique to study if the Federal Reserve's communication affects the US economy. See their papers for more detailed discussions of the analyses and for results.

[8] Blinder et al. (2008) provide a detailed survey of the literature until the late 2000s.

[9] See De Haan (2008) for a survey of studies on this topic published until the late 2000s.

[10] Rosa and Verga compile a list of expressions contained in the ECB Presidents' introductory statements to press conferences. They assign to each expression a value (number) denoting its relative level of hawkishness/dovishness. Then they do the same with each introductory statement. In associating each statement with a value of hawkishness/dovishness, they take into consideration the mean value of the expressions in the list which are contained in the statement. However, the mapping between the mean value of the expressions in the list contained in the statement and the value assigned to the statement is not applied in a "mechanical way" and is based, at least in part, on the subjective judgment of the authors: see Rosa and Verga (2007).

the base of the information publicly released by the institute: see KOF (2007). The MPC should capture information contained in the ECB Presidents' introductory statements to press conferences.[11] Conrad and Lamla (2010), Sturm and De Haan (2011), and Neuenkirch (2013) use the MPC in their analyses of the ECB's communication. As I noted in section 2.1, Heinemann and Ullrich (2007) present an automated measure of communication about the ECB's rate attitude, which is used, with minor differences, by Ullrich (2008).[12]

The second stream of literature focuses on relationships between news reported by media and financial variables: e.g., see Beetsma et al. (2013) on news and European sovereign bond spreads and Casarin and Squazzoni (2013) on bad news and stock markets during the financial crisis of 2008/2009. Lucca and Trebbi (2011) (see footnote 3) and Jansen and De Haan (2007) (see footnote 4) could be considered at the intersection of the two lines of studies: the one on central bank communication and the one on relationships between news reports and financial variables.

3 The index

The index is computed by counting occurences of hawkish, dovish, and neutral expressions in media articles published around days in which regular monetary policy meetings of GC and following press conferences take place.

Let t, with $t = 1, \ldots, 175$, denote the event represented by the t-th monetary policy meeting of the GC and the following press conference held by any of the three ECB Presidents (Duisenberg, Trichet, and Draghi) in the period between January 1999 and December 2013, which correponds to the first 15 years of operations of the ECB. Throughout this period, monetary policy meetings of the

[11] The MPC is computed by the KOF institute after every press conference following a monetary policy meeting of the GC. It is aimed at capturing information which is contained in the ECB Presidents' introductory statements and refers to the direction (upward or downward) and the intensity of the risks to price stability in the Eurozone (KOF, 2007). Such information could be seen as a signal of the ECB's policy rate attitude, since the central bank is expected to increase its rates as a reaction to upward risks to price stability and to decrease them as a reaction to downward risks (KOF, 2007). According to the information publicly released by the KOF institute, analysts divide each introductory statement into parts and then compute the MPC with reference to the statement by weighting the proportions of parts pointing respectively to upside and to downside risks to price stability (KOF, 2007). No details are provided about the process through which statements are divided into parts and each part is classified according to the direction of risk to which it points.

[12] Heinemann and Ullrich compile a list of expressions from a sample of the ECB Presidents' introductory statements to press conferences following GC meetings and investigate the potential of each expression to be informative about the ECB's rate attitude by testing if its average frequency differs among introductory statements preceding periods in which policy rates increase, decrease or do not change. If differences exist and are statistically significant, they consider the expression to be informative. Expressions whose frequency is greater before periods in which rates increase are deemed to be hawkish; those whose frequency is greater before periods in which rates decrease are deemed to be dovish. Then Heinemann and Ullrich compute for each introductory statement a value denoting its level of hawkishness/dovishness based on the frequency of each informative expression in the statement.

GC and following press conferences were usually held once a month,[13] usually on the first Thursday of the month. For each t, I calculate three values of the index: an "ex-ante" value I_{t^-}, based on articles published in a short time window (one and a half days) just before the event; an "ex-post" value I_{t^+}, based on articles published in a time window of the same length just after the event; and the difference ΔI_t between the ex-post and the ex-ante value, which should proxy new (unexpected) information on the ECB's rate attitude conveyed by the press conference (see sections 3.3 and 4). I refer to the three values of the index and to the corresponding time series as to the "specifications" of the index.

I present the derivation of the index by describing: first, the sets of hawkish, dovish, and neutral expressions whose occurrences are counted in media articles; second, the sample(s) of media articles; and third, the procedure to count occurences of expressions in sample(s) of articles and the formula to calculate (the three values, or specifications of) the index.

3.1 Hawkish, dovish, and neutral expressions

I construct three sets **H**, **D**, and **N** of hawkish, dovish, and neutral expressions, respectively.[14]

The complete list of the expressions of each set is quite long and would occupy several pages. In order to present it in "tractable" way, I report it in a short-form in appendix A1.[15] Here I focus on some characteristics of the elements of the three sets:

1. elements of the sets are expressions, i.e., combinations/strings of words, and not single words;[16]
2. "negative" hawkish or dovish expressions are counted as neutral, i.e., are elements of **N** and of neither **H** or **D**;
3. there are no "negative" neutral expressions among the elements of the sets;
4. expressions of the sets are usually not too short;
5. expressions of the sets usually have a forward-looking meaning.

Such characteristics should reduce the probability of making "wrong" inferences when counting occurrences of elements of the sets in media articles. In order

[13] The exceptions to the monthly frequency of meetings and press conferences are the following: 1) no meeting/conference took place in August of the years 1999, 2000, 2002, 2003, 2004, and 2005, in January 2001, and in September of 2001 and 2006; 2) two meetings/conferences took place in March 2000, October 2000, June 2001, and August 2006.

[14] The choice of the expressions depends on subjective judgment, which itself can be influenced by many factors. Among the factors which influenced my judgement and therefore my choice, I am able to recognize conversations with economists during a "traineeship" period spent at the Monetary Policy Strategy Division of the ECB between 2013 and 2014.

[15] In appendix A1 I show how expressions can be constructed as combinations of words from sets smaller than **H**, **D**, and **N**.

[16] Throughout the paper I use "expression" precisely with the meaning of "combination/string of words."

to make the point more clearly, I present three examples with hypothetical sentences.

First, consider the sentence "the ECB President ruled out an interest rate increase". If either the word "increase" or the expressions "rate increase" or "interest rate increase" were elements of **H**, their occurence would lead to infer a hawkish attitude of which the sentence is not suggestive. In order to avoid this type of wrong inference I do not include "increase", "rate increase" or "interest rate increase" in **H** and I include the negative hawkish expression "ruled out an interest rate increase" in **N**. Therefore I would count its appearance in the sentence as a neutral occurrence. If the sentence were "the ECB President did not rule out an interest rate increase", I would count a hawkish occurrence, since the expression "not rule out an interest rate increase" is an element of **H**.

Second, consider the sentence "the ECB will not keep rates steady". The negative neutral expression "will not keep rates steady" is not an element of any of the sets and, therefore, its occurence is not counted. If the sentence were "the ECB will keep rates steady", I would count it as a neutral occurrence, since "will keep rates steady" is an element of **N**.

Third, consider the sentence "the Governing Council decreased the interest rate". In the sentence I would not count any dovish (or hawkish or neutral) occurrence, since expressions of the three sets do not refer to monetary policy actions taken in the past (either close or distant). If the sentence were "the ECB will decrease the interest rate", I would count the occurrence of the dovish expression "will decrease the interest rate", which is an element of **D**.

Another characteristic of the expressions of the three sets is that they are not explicitly referred to the ECB. In order to avoid counting occurrences which are not related to the ECB and to count as many as possible of the occurrences related to it, I limit the count to sentences which contain at least a word or an expression related to the ECB, as I explain in section 3.2.

3.2 Media articles

For every t, I denote as d_t the day in which t took place, as $d_t - 1$ the day before d_t, and as $d_t + 1$ the day after d_t. Then I consider the set $\mathbf{S}_t$ of the articles, originally published on newspapers/newswires, with all the following characteristics in common:

1. they are retreived from three (mutually exclusive) digital sources: 1) the archive of Reuters, the news agency, 2) the section "Major Business Publications US" of Factiva, the database of newswire/newspaper articles, 3) the section "Major Business Publications Europe" of the same database;

2. they are written in English;

3. they were published in either one of two time intervals:

   (a) t^-, which consists of 1) $d_t - 1$ and 2) the hours of d_t before the announcement of the decision of the GC at around 1:45 p.m. Central

European Time (CET),[17]

(b) t^+, which consists of 1) the hours of d_t after the beginning of the press conference at 2:30 p.m. and 2) $d_t + 1$;[18]

4. they contain in the title: 1) both the word "ECB" (or the words "European Central Bank") and the word "rate" (or "rates" or both), if they are extracted from Reuters, 2) the word "ECB" (or the words "European Central Bank"), if they are extracted from Factiva.

Including in the sample only those articles whose title contains the words "ECB" or "European Central Bank" has the purpose of limiting the number of articles through which the search is perfomed and, at the same time, of including in the sample articles referred to the ECB and excluding articles referred (only) to other central banks. The choice of limiting the sample size has purely practical reasons, since each article retrieved both from Reuters and from Factiva had to be saved as a single file, then such files were merged with each other, obtaining 350 files (each file collects the articles published in a time interval, either t^-or t^+). Moreover the search and the count of expressions are quite time consuming, despite being performed by a computer program (see footnote 20 in section 3.3). The adoption of a more restrictive criterion for the selection of the Reuters articles than for the choice of the ones retrieved from Factiva (Reuters articles have to contain the word "rate", or "rates", in addition to "ECB", or "European Central Bank", in the title, Factiva ones do not have to) is also aimed at limiting the size of the sample of articles. Adopting a less restricitve criterion for the choice of the Reuters articles would have yielded a much larger sample as a result of the selection process,[19] but the sample would have been less tractable. It could be interesting to study if the results of paper would change by changing the words which articles have to contain in their title in order to be selected in the sample. However, I would expect that, apart from the sample size, only the probability of including in the sample articles about the ECB's monetary policy (and excluding articles about other central banks) could change (and that this would happen only by selecting articles which do not contain "ECB" or "European Central Bank" in the title). I would not expect any change in the probability of counting, within articles, expressions which are referred to the ECB and not to some other central bank, since I limit the search of expressions, within articles, to sentences which contain ECB-related words or expressions, as I explain in the next paragraph.

[17] Through the paper I use CET, which is the time of Frankfurt am Main, Germany, where the ECB is located and where most of the GC meetings and the following press conferences are held.

[18] For each article stored in either the Reuters or the Factiva database, both the date and the time of publication are available. Therefore it is possible to identify the articles which are published on different days and at different times of the same day.

[19] This might have to do with the fact that Reuters is a big news agency, which publishes articles much more frequently (many articles per day) than newspapers and smaller news agencies whose articles are collected by Factiva.

More formally, let an ECB-related word or expression be any element of the set $\mathbf{E} = \{$European Central Bank, ECB, Duisenberg, Trichet, Draghi, Governing Council$\}$. I define as sentence (of an article) any part of an article between two periods (I include in the definition of sentence the part of an article before the first period) and as ECB-related any sentence which contains at least one element of $\mathbf{E}$. For every t, let $\mathbf{S}_{t^-}$, with $\mathbf{S}_{t^-} \subseteq \mathbf{S}_t$, be the set of the articles published in t^- and $\mathbf{S}_{t^+}$, with $\mathbf{S}_{t^+} \subseteq \mathbf{S}_t$, the set of the articles published in t^+. Moreover, let $\mathbf{S}_{\mathbf{E},t^-}$ and $\mathbf{S}_{\mathbf{E},t^+}$ be the sets of the ECB-related sentences in the articles in $\mathbf{S}_{t^-}$ and $\mathbf{S}_{t^+}$, respectively. I limit the search and the count of hawkish, dovish, and neutral expressions to $\mathbf{S}_{\mathbf{E},t^-}$ and $\mathbf{S}_{\mathbf{E},t^+}$.

3.3 How to compute the values of the index

For every t, I compute first I_{t^-} (the ex-ante value of the index) and I_{t^+} (the ex-post value). The computation of both values is based on a count of occurrences of hawkish, dovish, and neutral expressions, which are elements of $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{N}$, respectively. I_{t^-} and I_{t^+} are derived as the ratio of a weighted sum on the unweighted sum of occurrences of such expressions in $\mathbf{S}_{\mathbf{E},t^-}$ and $\mathbf{S}_{\mathbf{E},t^+}$, respectively.

I denote the unweighted sums of the occurrences in the sentences in $\mathbf{S}_{\mathbf{E},t^-}$ of the expressions in $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{N}$ as H_{t^-}, D_{t^-}, and N_{t^-}, respectively, and the unweighted sums of the occurrences in the sentences in $\mathbf{S}_{\mathbf{E},t^+}$ of the same expressions as H_{t^+}, D_{t^+}, and N_{t^+}.[20] The weights used in the computation of the two weighted sums are 1, -1, and 0 for the occurrences of the expressions in $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{N}$, respectively. More formally, for every t, I compute

$$I_{t^-} = \frac{H_{t^-} - D_{t^-}}{H_{t^-} + D_{t^-} + N_{t^-}}$$

and

$$I_{t^+} = \frac{H_{t^+} - D_{t^+}}{H_{t^+} + D_{t^+} + N_{t^+}}.$$

Both I_{t^-} and I_{t^+} can assume any (real) value between -1 and 1, i.e., $I_{t^-}, I_{t^+} \in [-1, 1]$. A value of 0 indicates a "perfectly" neutral attitude. Values higher than 0 signal a hawkish attitude and the higher the value, the more hawkish the attitude signalled by that value. Conversely, values lower than 0 signal a dovish attitude and the lower the value, the more dovish the attitude signalled by that value.

[20] I wrote a short computer program which divides the articles (saved as text files) into sentences, identifies ECB-related sentences, and counts occurrences of hawkish, dovish, and neutral expressions in ECB-related sentences. The program is written in Python (a programming language) and employs a "tokenizer" program to partition the articles into sentences. The Python interface is freely downloadable at www.python.org. The tokenizer program is freely downloadable from NTLK (a platform of Python modules, i.e. programs which help to write other programs in Python) at www.nltk.org.

Then I calculate the difference between I_{t^+} and I_{t^-} which should proxy the new (unexpected) information on the ECB's rate attitude conveyed by the press conference in t (communication "shock"). More formally, I compute

$$\Delta I_t = I_{t^+} - I_{t^-}.$$

4 The index and "present" monetary policy decisions

In this section I examine the appropriateness of the index as a measure of monetary policy communication with reference to the ECB's policy rate attitude. I investigate whether the index measures information which is related not to the content of the ECB Presidents' press conferences, but to "present" monetary policy decisions, i.e. decisions taken by the GC on the same days in which press conferences take place. To this purpose, I study correlations between the index and measures of information related to "present" monetary policy decisions.

Every event t conveys two public signals (pieces of public information) about the ECB's monetary policy:

- the levels of the policy rates decided at the GC meeting, which I denote as signal $i_{t,1}$,
- the tone (hawkish, dovish, or neutral) of the words of the ECB President at the press conference, which I denote as signal $i_{t,2}$.

Signal $i_{t,1}$ is an actual monetary policy decision, while signal $i_{t,2}$ is a piece of monetary policy communication. Although it seems plausible that, among the two signals, only $i_{t,2}$ is informative of the ECB's policy rate attitude, it would be arbitrary to reject in principle the hypothesis that also $i_{t,1}$ conveys information about future changes to levels of the key ECB interest rates. Moreover it would also be arbitrary to rule out that the index captures information related to the decision of the GC in t, rather than to the ECB's policy rate attitude. Therefore, at least in principle, both $i_{t,1}$ and $i_{t,2}$ might convey information which is captured by the index, with I_{t^-}, I_{t^+}, and ΔI_t measuring expected, actual (perceived), and unexpected information, respectively. However, it would be appropriate to consider the three specifications of the index as measures of monetary policy communication only if they captured information conveyed by $i_{t,2}$ without capturing information conveyed by $i_{t,1}$. In order to check whether the index captures or not information conveyed by $i_{t,1}$, I study correlations between the three specifications of the index and measures of information conveyed by $i_{t,1}$. Very low correlations should imply that the three specifications of the index do not capture information conveyed by $i_{t,1}$. Therefore, I would conclude that the three specifications are proxies for information conveyed by $i_{t,2}$ and, hence, appropriate measures of monetary policy communication (the tone of words of the ECB President at the press conference) with reference to the ECB's policy rate attitude.

As proxies of information related to $i_{t,1}$, apart from the actual values of the policy rates (see later in this section), I use measures based on two types of market rates: Euribor and Eonia rates. Euribor rates reflect rates in the euro-area interbank lending market for maturities from one week to twelve months and, as discussed by Rosa and Verga (2007), can be considered as good proxies of market expectations of levels of the key ECB rates in the short term.[21] Euribor rates are quoted every business day at 11 a.m.. For any business day d, let r_d be the value of the Euribor (spot) rate quoted on d for a maturity of one month. r_d can be considered as a good measure of market expectations in the morning on d (before 11 a.m.) of levels of the key ECB rates in place in the one-month period starting on d.[22] Euribor rates are equal to the ECB policy rates plus a risk premium. Risk premia of Euribor above ECB rates have not been constant over time, but have increased especially since the financial crisis of 2008/2009. An alternative measure of information related to $i_{t,1}$, based on rates with lower and relatively constant risk premia over the key ECB rates, is the Eonia Swap Index. The Eonia Swap Index with a given maturity measures market expectations of the average level of the Eonia, an index which reflects lending rates in the overnight interbank lending market in the euro-area, over the period up to the maturity date.[23] Let r'_d be the value of the Eonia Swap Index quoted on d for a maturity of one month. r'_d measures market expectations of the average level the Eonia in the one-month period starting on d, and, therefore, it can be considered as a measure, alternative to the one-month Euribor r_d, of market expectations of the key ECB rates in place in the one-month period starting on d.

First, I consider correlations between I_{t^-} and measures of expectations, updated just before t, of the levels of the ECB rates set at t in order to check whether I_{t^-} also captures such expectations. For every t, since d_t is a business day, then r_{d_t}, i.e. the one-month rate quoted on d_t, is available and captures expectations updated just a few hours before the ECB announces the values of its key interest rates set at the GC meeting and in place in the one-month period starting on d_t (Euribor rates are quoted at 11 a.m., see the above paragraph, and the decision of the GC is announced at 1:45 p.m., see section 3.2). Therefore I compute $corr\left(I_{t^-}, r_{d_t}\right)$. The same reasoning can be applied with reference to the Eonia Swap Index, which is also quoted every business day at 11 a.m. Since the Eonia Swap Index time series start from June 20, 2005, the first values quoted on a day of a GC meeting and the following press conference are relative to the event $t = 75$ which took place in July 2005. Let $r'_{d_{t'}}$, with $t' = t = 75, \ldots, 175$, be the one-month Eonia Swap Index quoted on $d_{t'}$. $r'_{d_{t'}}$

[21] For discussions and examples of measures of the "surprise" associated to monetary policy decisions of the FOMC of the Federal Reserve, see Kuttner (2001), Gürkaynak et al. (2005), and Lucca and Trebbi (2011).

[22] See Musard-Gies (2006) for a slightly different approach based on (implicit) Euribor forward rates.

[23] For more information on Euribor and Eonia rates and on the Eonia Swap Index, see www.emmi-benchmarks.eu, the website of the European Money Market Institute (EMMI), the organization which provides Euribor and Eonia rates and the Eonia Swap Index and whose members are national banking associations of European Union countries.

should capture expectations of the levels of the ECB rates set at t, updated just a few hours before the announcement of the decision of the GC. Therefore I compute $corr\left(I_{t'-}, r'_{d_{t'}}\right)$. Moreover, in order to allow a more meaningful comparison of $corr\left(I_{t'-}, r'_{d_{t'}}\right)$ with the correlation between the index and an Euribor-based measure of expectations, updated just before t, of the ECB rates at t, I also compute $corr\left(I_{t'-}, r_{d_{t'}}\right)$.

Second, I consider how I_{t^+} correlates with the key ECB rates set at t to check whether I_{t^+} captures actual (perceived) information conveyed by the policy rate decision taken at t-th meeting of the GC. I focus on the interest rate on the main refinancing operations (MROs) (which I also use as a proxy for the other two key ECB interest rates: the one on the deposit facility and the one on the marginal lending facility, see footnote 1 in the introduction), by computing $corr\left(I_{t^+}, MRO_t\right)$, where MRO_t is the level of the MRO rate set at the t-th GC meeting. Then I present correlations which can be more easily compared with the correlations, which I computed before, between the index and measures of expectations, updated just before t, of the ECB rates at t. Namely, I compute $corr\left(I_{t^+}, r_{d_t+1}\right)$, $corr\left(I_{t'^+}, r'_{d_{t'}+1}\right)$, $corr\left(I_{t'^+}, r_{d_{t'}+1}\right)$,[24] and $corr\left(I_{t'^+}, MRO_{t'}\right)$.

Third, I consider correlations between ΔI_t and measures of the unexpected component of the levels of the key interest rates set at t to check whether ΔI_t also captures the surprise related to the policy rate decision of the GC. I use the following measures of the unexpected component of the policy rate decision of the GC:

- the difference between the one-month Euribor rate quoted on the day after t and just before t: $\Delta r_t = r_{d_t+1} - r_{d_t}$;
- the difference between the level of the MRO rate set at t and the one-month Euribor rate quoted just before t: $\Delta R_t = MRO_t - r_{d_t}$.

Moreover, I compute correlations between $\Delta I_{t'}$ and the following measures of the unexpected components of the levels of the key interest rates set at t':

- the difference between the one-month Euribor rate quoted on the day after t' and just before t': $\Delta r_{t'} = r_{d_{t'+1}} - r_{d_{t'}}$;
- the difference between the one-month Eonia Swap Index quoted on the day after t' and just before t': $\Delta r'_{t'} = r'_{d_{t'+1}} - r'_{d_{t'}}$;
- the difference between the level of the MRO rate set at t' and the one-month Euribor rate quoted just before t': $\Delta R_{t'} = MRO_{t'} - r_{d_{t'}}$; and the one between the MRO rate set at t' and the one-month Eonia Swap Index just before t': $\Delta R'_{t'} = MRO_{t'} - r'_{d_{t'}}$.

[24] For every t, values of Euribor rates and the Eonia Swap Index quoted on $d_t + 1$ are available, since $d_t + 1$ is a business day.

All results are reported in table 1.

Table 1: Correlations between the specifications of the index and measures of information related to policy rate decisions of the GC in t and t'.

	r_{d_t}	$r_{d_{t'}}$	$r'_{d_{t'}}$			
I_{t-}	0.194					
$I_{t'-}$		0.375	0.454			
	r_{d_t+1}	$r_{d_{t'}+1}$	$r'_{d_{t'}+1}$	MRO_t	$MRO_{t'}$	
I_{t+}	0.217			0.211		
$I_{t'+}$		0.423	0.496		0.447	
	Δr_t	$\Delta r_{t'}$	$\Delta r'_{t'}$	ΔR_t	$\Delta R_{t'}$	$\Delta R'_{t'}$
ΔI_t	-0.124			-0.108		
$\Delta I_{t'}$		0.043	0.017		$-0,095$	0.163

Correlations over the whole period between January 1999 and December 2013 are quite low: they vary between -0.124 and 0.217. Over the shorter period between July 2005 and December 2013, correlations of $\Delta I_{t'}$ with measures of the unexpected components of "present" decisions of the GC are quite low and vary between -0.095 and 0.163, while correlations of the ex-ante and the ex-post specification of the index with measures of information relative to, respectively, expected and actual "present" decisions of the GC are higher than over the whole 1999-2013 period and vary between 0.375 and 0.496.

Higher correlation values over the sub-period between July 2005 and December 2013 might be explained as follows. Both the 1999-2013 period and the mid 2005-2013 sub-period comprise long time intervals in which either the MRO rate does not change or the direction of monthly changes in the MRO rate does not vary. During such time intervals, information about "present" monetary policy decisions is at least mildly correlated with information, captured by the index and contained in (the tone of) the words of the ECB Presidents at press conferences, about "future" decisions. However, the mid 2005-2013 sub-period comprises less of such intervals[25] (this is not surprising, given the fact that the sub-period is shorter than the whole period) and, therefore, the direction of monthly changes in the MRO rate varies less times in the mid 2005-2013 sub-period than in the whole 1999-2013 period. For this reason, information about "present" monetary policy decisions is more strongly correlated with information, captured by the index, about "future" decisions over the sub-period than over the whole period.

[25] The MRO rate increases in the period between December 2005 and June 2007 (during which months of 0.25% rises almost regularly alternate with months in which the rate does not change), never changes between June 2007 and October 2008 (with the exception of a 0.25% rise in July 2008), and decreases (every month) between November 2008 and May 2009. Afterward, i.e., in the four years and a half between June 2009 and December 2013, the monthly change in the MRO rate is equal to 0 with only six 0.25% exceptions (sparse in the period between April 2011 and November 2013; the first two upward and the last four downward).

However, with reference to the whole period between January 1999 and December 2013, I conclude that, given the low correlation values, the three specifications of the index do not measure information related to "present" monetary policy decisions (signal $i_{t,1}$) and, therefore, that they are appropriate measures of monetary policy communication, since they capture information about the tone (hawkish, dovish, or neutral) of the words pronounced by the ECB Presidents at monthly press conferences (signal $i_{t,2}$). Namely, I_{t^-} should measure expectations of the tone of the words of the ECB Presidents, I_{t^+} the actual (perceived) tone, and ΔI_t the unexpected information (communication "shock") conveyed by the words of the ECB Presidents. Values of I_{t^-} and I_{t^+} in the interval $(0, 1]$ indicate a hawkish press conference (expected in the case of I_{t^-} and perceived in the case of I_{t^+}). The higher (lower) the value, the more (less) hawkish is the tone of the words. Values of I_{t^-} and I_{t^+} in the interval $[-1, 0)$ indicate a dovish press conference (expected in the case of I_{t^-} and perceived in the case of I_{t^+}). The lower (higher) the value, the more (less) dovish is the tone of the words. Values equal to 0 denote a perfectly neutral tone.

5 The index and future monetary policy decisions

In this section, I study the consistency between the informational content (with reference to the ECB's rate attitude) of the words pronounced by the ECB Presidents at monthly press conferences and the ECB's actual monetary policy decisions in months which follow press conferences. I use I_{t^+}, the ex-post specification of the index, as a measure of the tone (hawkish, dovish, or neutral) of press conferences (see section 3) and I test if I_{t^+} has predictive power on changes in levels of the policy rates in the short term. In the analysis I focus on the rate on the main refinancing operations (MROs), which I also use as a proxy for the other two key interest rates whose levels are set by the GC of the ECB (as in section 3). For every event t, I observe the difference between the level of the MRO rate set at the monetary policy meeting of the GC at $t + m$, with $m = 1, \ldots, 6$, and the level set at the meeting at t. Then I regress the changes in the MRO rate between t and $t + m$ on I_{t^+}. For each m, the regression equation is

$$MRO_{t+m} - MRO_t = \alpha_m + \beta_m I_{t^+} + u_m.$$

Table 2 summarizes the results of the OLS estimation of the equations.

Table 2: Perception of the ECB's communication and MRO rate changes in the short term.

		$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
α_m		-0.228^{***}	-0.394^{***}	-0.546^{***}	-0.663^{***}	-0.744^{***}	-0.831^{***}
		(***)	(***)	(***)	(***)	(***)	(***)
I_{t^+}	β_m	0.197^{***}	0.338^{***}	0.465^{***}	0.561^{***}	0.625^{***}	0.694^{***}
		(***)	(***)	(***)	(***)	(***)	(***)
	SE	0.024	0.038	0.051	0.065	0.078	0.091
		(0.043)	(0.070)	(0.086)	(0.105)	(0.129)	(0.143)
R^2		0.28	0.31	0.32	0.30	0.27	0.25
Adj. R^2		0.28	0.31	0.32	0.30	0.26	0.25

Note: *, **, and *** denote significance at the 0.1, 0.05, and 0.01 level, respectively. Heteroschedasticity robust (HAC) standard errors and significance of coefficients estimated with HAC errors are reported within brackets. SE stands for "standard error".

The coefficient β_m is positive and significant at the 0.01 level for each m. Such result suggests that I_{t^+} has predictive power over future changes in levels of the policy rates in the short term. Moreover, the positive linear relationship, measured by β_m, between I_{t^+} and $MRO_{t+m} - MRO_t$ implies that, on average:

- if $I_{t^+} \in (0, 1]$, the higher the value of I_{t^+}, the more the MRO rate increases in the months after t and, conversely, the lower the value, the less the MRO rate increases;
- if $I_{t^+} \in [-1, 0)$, the lower the value of I_{t^+}, the more the MRO rate decreases in the months after t and, conversely, the higher the value, the less the MRO rate decreases.

This is what one would expect since, given the procedure through which the ex-post specification of the index is contructed (see sections 3 and 4), higher positive values of I_{t^+} should capture a more hawkish tone of the words of the ECB President (and, conversely, lower positive values should capture a less hawkish tone), and lower negative values should capture a more dovish tone (and, conversely, higher negative values should capture a less dovish done).

It is worth noting that the coefficient β_m increases with m, that is, the "effect" of (a unit variation of) I_{t^+} on the MRO rate total variation is larger for longer periods after t than for shorter ones (if one considers periods up to six months).[26] This (apparently counterintuitive) result might be explained as follows. As noted at the end of section 4, the period between January 1999 and December 2013 comprises long time intervals in which the direction of monthly changes in the MRO rate does not vary. During any of such time intervals, either upward or downward monthly variations of the MRO rate "accumulate" and the rate might either increase or decrease more between, e.g., t and $t+6$ than between t and $t+1$. In other words, the average across the 175 t-observations

[26] Rosa and Verga (2007, see section 2.2) obtain a similar result for the coefficient of their indicator of the tone of the ECB Presidents' introductory statements to press conferences.

of $|MRO_{t+m} - MRO_t|$ (the absolute value of the total change in the MRO rate between t and $t+m$) is larger as m is larger. For this reason, the "effect", measured by the coefficient β_m, of (a unit variation of) I_{t^+} on $MRO_{t+m} - MRO_t$ increases with m.

However, as one would expect, the "effect" of (a unit variation of) I_{t^+} on the MRO rate (average) monthly variation is larger for shorter periods after t than for longer ones, as shown in table 3, which reports the results of the OLS estimation, for each m, of the following regression equation:

$$\sum_{m'=1}^{m} \frac{MRO_{t+m'} - MRO_{t+m'-1}}{m} = \alpha'_m + \beta'_m I_{t^+} + u'_m.$$

Table 3: Perception of the ECB's communication and MRO rate average monthly changes in the short term.

		$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
α'_m		-0.228^{***}	-0.197^{***}	-0.182^{***}	-0.166^{***}	-0.148^{***}	-0.139^{***}
		(***)	(***)	(***)	(***)	(***)	(***)
I_{t+}	β'_m	0.197^{***}	0.169^{***}	0.155^{***}	0.140^{***}	0.125^{***}	0.116^{***}
		(***)	(***)	(***)	(***)	(***)	(***)
	SE	0.024	0.019	0.017	0.016	0.016	0.015
		(0.043)	(0.035)	(0.029)	(0.026)	(0.026)	(0.23)
R^2		0.28	0.31	0.32	0.30	0.27	0.25
Adj. R^2		0.28	0.31	0.32	0.30	0.26	0.25

Note: *, **, and *** denote significance at the 0.1, 0.05, and 0.01 level, respectively. Heteroschedasticity robust (HAC) standard errors and significance of coefficients estimated with HAC errors are reported within brackets. SE stands for "standard error".

6 ECB communication and policy rates expectations

In this section I investigate the relationship between the ECB's communication and market expectations of future levels of the policy rates in the short term. Namely, I test if market expectations are affected by unexpected information (about the ECB's policy rate attitude) conveyed by the words pronounced by the ECB Presidents at monthly press conferences. To this purpose, I regress differences between measures of market expectations after and before press conferences on ΔI_t, which captures the unexpected component of the tone of the words pronounced by the ECB Presidents at press conferences (see sections 3 and 4), and on control variables, which proxy the "surprise" components of monetary policy decisions of the GC (see section 4). I discuss various regression specifications in section 6.2 and in appendix A2.

I assume that market participants have rational expectations, i.e., that they form expectations using all available information and continuously update them on the base of new information. As discussed in section 4, every event t conveys two pieces public information (signals) about the ECB's monetary policy: the levels of the policy rates decided at the GC meeting ($i_{t,1}$) and the perceived tone (hawkish, dovish, or neutral) of the words of the ECB President at the press conference ($i_{t,2}$).

Both signals might affect market expectations of future levels of the policy rates, since $i_{t,1}$ represents the "starting point" of the future path of the policy rates from t on[27] and $i_{t,2}$ indicates (at least approximately) the direction of the policy rate path. Moreover no other type of news systematically reaches the market on press conference days. For such reasons, I use measures of the "surprise" component of $i_{t,1}$ as (the only) control variables in the regression specifications which I present in section 6.2 and in appendix A2. Measures of the unexpected components of policy rate decisions have been discussed in section 4. In section 6.1 I discuss measures of market expectations of future levels of the policy rates.

6.1 Measures of expectations of future policy rates

I present three alternative types of measures of expectations of future levels of the policy rates.

The first one is represented by Euribor (implicit) forward rates. I denote as $f_d^{n,1}$ the one-month Euribor n-month (implicit) forward rate relative to day d. $f_d^{n,1}$ can be considered a good proxy of market expectations on d of levels of the policy rates in the one-month period starting n months after d (see the discussion on Eurbor rates in section 4 and Rosa and Verga, 2007). Let r_d^n and r_d^{n+1} be the values of, respectively, the n-month and the $n+1$-month Euribor (spot) rate quoted on d, and D_d^n and D_d^{n+1} the maturities, expressed in number of calendar days, for which r_d^n and r_d^{n+1} are respectively quoted. $f_d^{n,1}$ is computed as follows:

$$f_d^{n,1} = \left(\frac{100 + r_d^{n+1} D_d^{n+1}/360}{100 + r_d^n D_d^n/360} \right) \frac{360}{D_d^{n+1} - D_d^n}.$$

I use $f_{d_t}^{n,1}$ as a measure of market expectations of future levels of the policy rates just before t and $f_{d_t+1}^{n,1}$ as a measure of expectations updated just after t.[28] Then I compute the difference

$$\Delta f_t^{n,1} = f_{d_t+1}^{n,1} - f_{d_t}^{n,1}$$

[27] If the levels of the policy rates set at t are higher (lower) than expected, market participants might revise their expectations of levels of the rates in the near future upward (downward), since the rate path starts from a level higher than expected.

[28] Since Euribor (spot) rates are quoted every (business) day at 11 a.m. (see section 4).

and I use it as a measure of the change in market expectations due to information (the montetary policy decision of the GC and the words of the ECB President at the press conference) conveyed by t.

The second type of measures of expectations is analogous to the first one, with the difference that (implicit) forward rates are calculated from values of the Eonia Swap Index. Let $f_d^{'n,1}$ be the n-month (implicit) forward value of the Eonia Swap Index for a maturity of one month relative to day d. I use

$$\Delta f_{t'}^{'n,1} = f_{d_{t'}+1}^{'n,1} - f_{d_{t'}}^{'n,1}$$

as a measure of the change in market expectations due to information conveyed by t.

The third type of measures of expectations is represented by Euribor futures rates. I consider Euribor futures contracts which are traded at LIFFE (London International Financial and Options Exchange).[29] Prices are quoted every business day and futures rates are equal to 100 less futures prices.[30] Contracts are cash-settled based on the value of the three-month Euribor rate quoted on the delivery date, which is the first business day after the last trading day.[31] Contracts are traded until two business days prior to the third Wednesday of the delivery month. There are four "main" delivery months, known as quarterly sequence (or quarterly expiries or expirations): March, June, September, and December. Contracts expiring in the quarterly sequence are characterized by higher liquidity than those expiring in months outside the sequence. For this reason, I focus on quarterly expiries. For every t, I focus on the contract which expires first in the quarterly sequence, unless t takes place in one of the four delivery months. In such a case I focus on the contract which expires second in the sequence, since the first contract would expire only few days after t, and the corresponding futures rate would probably be affected much more by the monetary policy decision of the GC in t than by the signal on the ECB's rate attitude which might be conveyed by the ECB President at the press conference following the GC meeting. Let F_{d_t} and F_{d_t-1} be the three-month Euribor futures rates calculated as 100 less the closing price on d_t and d_t+1, respectively. I use the difference

$$\Delta F_t = F_{d_t} - F_{d_t-1}$$

as a measure of the change in market expectations due to information conveyed by t.

6.2 Results

I test if market expectations are affected by unexpected information (with reference to the ECB's rate attitude) conveyed by the words pronounced by the

[29] LIFFE is now part of Intercontinental Exchange (ICE) after a series of takeovers.

[30] Prices are available at www.quandl.com.

[31] For information on Euribor futures contracts, see Bernoth and von Hagen (2004), and Rosa and Verga (2008), www.quandl.com, and www.theice.com (the website of ICE, see footnote 29).

ECB Presidents at monthly press conferences by regressing differences between measures of market expectations after and before press conferences on ΔI_t and on control variables which proxy the unexpected components of the levels of the policy rates set at GC meetings which take place immediately before press conferences. In this section I discuss two regression specifications. I present other two specifications in the appendix A2.

In the first specification the dependent variable is the difference $\Delta f_t^{n,1}$ in Euribor (implicit) forward rates and the unexpected component of the monetary policy decision of the GC in t is measured by Δr_t (difference in Euribor spot rates). The regression equation is

$$\Delta f_t^{n,1} = \alpha_n + \beta_n \Delta r_t + \gamma_n \Delta I_t + u_n,$$

for $n = 1, \ldots, 6$.[32] Table 4 summarizes the results of the OLS estimation of the equations.

Table 4: ECB communication's and policy rate expectations measured by Euribor (implicit) forward rates (controlling for unexpected components of monetary policy decisions, measured by differences in Euribor spot rates).

		$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
α_n		−0.000	−0.001	0.000	0.006*	0.005	0.005
		()	()	()	(**)	()	()
Δr_t	β_n	0.730***	0.577***	0.655***	0.616***	0.720***	0.740***
		(***)	(***)	(***)	(***)	(***)	(***)
	SE	0.038	0.053	0.060	0.066	0.082	0.083
		(0.074)	(0.139)	(0.117)	(0.110)	(0.112)	(0.138)
ΔI_t	γ_n	0.010*	0.017**	0.022***	0.032***	0.046***	0.038***
		(**)	(**)	(***)	(***)	(***)	(***)
	SE	0.005	0.007	0.008	0.003	0.011	0.011
		(0.005)	(0.008)	(0.007)	(0.009)	(0.013)	(0.012)
R^2		0.68	0.41	0.41	0.35	0.34	0.33
Adj. R^2		0.68	0.41	0.41	0.34	0.33	0.32

Note: *, **, and *** denote significance at the 0.1, 0.05, and 0.01 level, respectively. Heteroschedasticity robust (HAC) standard errors and significance of coefficients estimated with HAC errors are reported within brackets. SE stands for "standard error".

Coefficient γ_n is positive for each n and it is significant at the 0.01 level for most values of n and at either the 0.05 or the 0.1 level for the others. Such results suggest that unexpected information (communcation "shock") conveyed by the words pronounced by the ECB Presidents at monthly press conferences

[32] Since the correlation between the two independent variables Δr_t and ΔI_t is quite low (see section 4), I conclude that there is no problem of multicollinearity in this regression specification and in the following one (see later in this section). The two alternative regression specifications presented in appendix A2 should also present no multicollinearity, since the correlation between the two dependent variables $\Delta r'_{t'}$ and $\Delta I_{t'}$ is also quite low (see section 4).

affects market expectations of future (with respect to each press conference) levels of the policy rates in the short term. Namely, the positive linear relationship, measured by γ_n, between ΔI_t and $\Delta f_t^{n,1}$ implies that, on average, a press conference perceived as more hawkish (or less dovish) than anticipated leads to an upward revision of market expectations of future levels of the policy rates and, conversely, a press conference perceived as more dovish (or less hawkish) than anticipated leads to a downward revision of expectations. Moreover, the more ΔI_t differs from zero in absolute value (i.e., the greater the "shock" in the perceived tone of the words of the ECB President, either in a hawkish or in a dovish direction), the greater the revision of market expectations.

In the second specification I use the difference $\Delta F_t = F_{d_t} - F_{d_t-1}$ in three-month Euribor futures rates as dependent variable. The unexpected component of the monetary policy decision of the GC in t is measured, as in the first specification, by Δr_t. The regression equation is

$$\Delta F_t = \alpha + \beta \Delta r_t + \gamma \Delta I_t + u.$$

Table 5 summarizes the results of the OLS estimation of the equation.

Table 5: ECB's communication and policy rate expectations measured by three-month Euribor futures rates (controlling for unexpected components of monetary policy decisions, measured by differences in Euribor spot rates).

α		-0.008^{*}
		(*)
Δr_t	β	0.534^{***}
		(***)
	SE	0.091
		(0.141)
ΔI_t	γ	0.035^{***}
		(**)
	SE	0.012
		(0.014)
R^2		0.19
Adj. R^2		0.18

Note: *, **, and *** denote significance at the 0.1, 0.05, and 0.01 level, respectively. Heteroschedasticity robust (HAC) standard errors and significance of coefficients estimated with HAC errors are reported within brackets. SE stands for "standard error".

Results are similar to those obtained by estimating the equations of the first specification, with few differences. Coefficient γ is greater than γ_n, while β is lower than β_n, and this is true for most values of n. R^2 and adjusted R^2 are lower than those obtained for all values of n; the opposite is true if we look at the standard error.

Regression specifications presented in the appendix yield results similar to those discussed in this section. Therefore I conclude that the results are quite

robust across different model specifications.[33]

7 Conclusion

In previous sections I presented a new measure of the ECB's communication, by constructing an index based on the content of media articles published in days around monthly press conferences held by the ECB Presidents after meetings of the GC in the period between 1999 and 2013 (section 3). Then I provided evidence that the index measures information related to the content of press conferences rather than to policy rate levels set at GC meetings which take place immediately before conferences (section 4). Moreover I used the index to study the relationship between the the ECB's monetary policy communication and future monetary policy decisions of the GC and I found evidence of the statistical significance of such relationship (section 5). Finally, I found that unexpected informational content of the ECB's monetary policy communication, as measured by the index, affects market expectations of future policy rates (section 6).

In the remaining of this section I briefly focus on lines of research along which the analysis presented in the paper could be expanded.

First, one could break the 1999-2013 period into shorter periods (e.g., post and after the "Lehman collapse") to repeat the analysis with reference to each sub-period and check if the results differ or not across sub-periods. This has already been done in an embryonic form by using Eonia-based measures of market expectations, which are available only for the period between 2005 and 2013, in sections 4 and 5 and in appendix A2, but a more thorough analysis might lead to interesting results.

Second, one could compare the "forecasting ability" of the index with that of different models, such as policy-rule models based on economic data and expectations about the economy (e.g., Taylor-rule types of models), or could add the index as an explanatory variable to such models and examine the statistical significance of differences in "predictive power" of models with the index and "benchmark" models without the index.

Third, one could use the index to investigate the role of ECB's communication in monetary policy transmission through VAR models (see end of section 2.1 and footnote 7).

Fourth, one could repeat the analysis on different samples of articles, such as samples of articles published in languages other than English or in media based in specific countries, to check if the perception of the ECB's monetary policy communication is sensitive to, so to say, "mother-tongue" and "geographical location" of media. This would require either to translate from English the list

[33] However, I investigated a regression specification in which I used the difference $\Delta R_t = MRO_t - r_{d_t}$ as the independent variable measuring the unexpected component of the policy rate decision of the GC at t. In such specification, the coefficient of neither independent variable is statistically significant.

of hawkish, dovish, and neutral expressions used in the present paper or to use a new list of expressions in languages other than English.

Fifth, one could apply automated measures of the ECB's communication to the analysis of events other than post-GC-meeting press conferences, such as speeches and interview of members of the Excutive Board and of the GC.

Appendix

A1

The sets **H**, **D**, and **N** have many elements, which are expressions, i.e. combinations/strings of words. Listing all of them one by one would take several pages (and reading them one by one would require a lot of time). In order to present the elements of **H**, **D**, and **N** in a more "tractable" way, I define the expressions contained in the three set as combinations of words and expressions which are elements of smaller sets. Note: elements of the sets are treated by the program which counts expressions in articles (section 3.3, footnote 20) as a strings of letters and blank spaces. For this reason, I consider as parts of elements of the sets some "incomplete words" such as in the case of "n't"+verb, which allows to count, for example, "don't+verb", "doesn't"+verb, "didn't"+verb, or in the case of "hik", which allows to count "hike", "hikes", and "hiking" (e.g., see smaller sets $\mathbf{Z}_3$, $\mathbf{H}_1$, and $\mathbf{D}_1$). I list here the smaller sets:

- $\mathbf{A}_1$ ={comment, remark, speech, statement, tone},
- $\mathbf{A}_2$ ={is, are, was, were, sound, sounds, sounding, sounded, more},
- $\mathbf{A}_3$ ={is considering, are considering, considered, discussed, hint at, hints at, hinted at, signal, signals, signaled, signalled, expect, expects, open to, open for, opens the door for, opens the door to, opened the door for, opened the door to, possibility of, toward, towards},
- $\mathbf{A}_4$ ={about to, poised to, prepared to, ready to, can, could, may, might, will, would},
- $\mathbf{A}_5$ ={not rule out, not ruled out, n't rule out, n't ruled out},
- $\mathbf{A}_6$ ={can be, could be, may be, might be},
- $\mathbf{Z}_2$ ={is not, are not, were not, not sounding, not sound, isn't, arent, werent, n't sounding, n't sound},
- $\mathbf{Z}_3$ ={is not considering, are not considering, not considered, not discuss, not hint at, not signal, do not expect, isnt considering, arent considering, n't considered, n't discuss, n't hint at, n't signal, n't expect},
- $\mathbf{Z}_4$ ={not about to, not poised to, not prepared to, not ready to, cannot, could not, may not, might not, will not, would not, n't about to, n't poised to, n't prepared to, n't ready to, cannot, couldn't, won't, wouldn't},

- $\mathbf{Z}_5$ ={rules out, ruled out},
- $\mathbf{B}_1$ ={a, a deposit rate, an interest rate, a policy rate, a rate, deposit rate, interest rate, policy rate, rate, any, another, another deposit rate, another interest rate, another policy rate, another rate, a further, a further deposit rate, a further interest rate, a further policy rate, a further rate, further, further deposit rate, further interest rate, further policy rate, further rate, a future, a future deposit rate, a future interest rate, a future policy rate, a future rate, future, future deposit rate, future interest rate, future policy rate, future rate, more, more deposit rate, more interest rate, more policy rate, more rate},
- $\mathbf{B}_2$ ={deposit rate, interest rate, policy rate, rate, further, further deposit rate, further interest rate, further policy rate, further rate, future, future deposit rate, future interest rate, future policy rate, future rate},
- $\mathbf{B}_3$ ={deposit rate, interest rate, policy rate, rate},
- $\mathbf{H}_1$ ={hik, increas, ris, tighten, rais},
- $\mathbf{H}_2$ ={hike, increase, rise, tightening, raise},
- $\mathbf{H}_3$ ={hikes, increases, rises},
- $\mathbf{D}_1$ ={cut, decreas, reduc, eas},
- $\mathbf{D}_2$ ={cut, decrease, reduction, ease, easing},
- $\mathbf{D}_3$ ={cuts, decreases, reductions},
- $\mathbf{N}_1$ ={remain on hold, remain steady, remain unchanged, stand pat},
- $\mathbf{N}_2$ ={no change, no hint, no hurry, no move, no signal, wait and see, wait-and-see}.

In what follows, let $\mathbf{X}$ be any of the sets in the list above and X be the total number of the elements of $\mathbf{X}$. Let also x, with $x = 1, \ldots, X$, be any element of $\mathbf{X}$. Given any two sets $\mathbf{X}$, $\mathbf{Y}$, with $\mathbf{X} \neq \mathbf{Y}$, I denote as x, y an expression in which any x is immediately followed a space, which is immediately followed by any y. Moreover, given any three sets $\mathbf{W}$, $\mathbf{X}$, and $\mathbf{Y}$, with $\mathbf{W} \neq \mathbf{X} \neq \mathbf{Y}$, I denote as $w, (x,)y$ both an expression in which any w is immediately followed a space, which is immediately followed by any x, which is immediately followed by a space, which is immediately followed by any y, and the expression in which that w is immediately followed a space, which is immediately followed by that y. In the notation, a sequence of characters representing a word or an expression (such as "hawkish" or "dovish") might subsitute either w, x or y.

The elements of $\mathbf{H}$ are *all* the expressions:

- hawkish, a_1,
- a_2, hawkish,

- $a_2, (b_1,) h_1$,
- a_4, h_1,
- $a_5, (b_1,) h_1$,
- a_5, any, b_2, h_1,
- imminent, $(b_3,) h_1$,
- h_2, is on the cards,
- h_2, a_6, on the cards,
- h_3, a_6, on the cards,
- h_3, are on the cards,

excluding the expressions listed above if they are immediately preceded by (a space immediately preceded by) "not" or by a negative verb.

The elements of **D** are the expressions obtained by substituting in all the expressions listed above the word "dovish" to the word "hawkish", d_1 to h_1, d_2 to h_2, and d_3 to h_3.

The elements of **N** are *all* the expressions:

- z_2, hawkish,
- z_3, (any,) h_1,
- z_3, $(b_1,) h_1$,
- z_3, (any,) b_2, h_1,
- z_4, h_1,
- z_5, (any,) h_1,
- z_5, $(b_1,) h_1$,
- z_5, (any,) b_2, h_1,
- no, (more,) b_2, h_1,
- no imminent, $(b_3,) h_1$,
- h_2, z_5, on the cards,
- h_3, z_6, on the cards,
- obtained by substituting in all the expressions listed above the word "dovish" to the word "hawkish", d_1 to h_1, d_2 to h_2, and d_3 to h_3,
- neutral, a_1
- a_2, neutral,
- a_4, n_1,
- n_2.

A2

First alternative specification

$$\Delta f_{t'}^{'n,1} = \alpha_{1,n} + \beta_{1,n}\Delta r'_{t'} + \gamma_{1,n}\Delta I_{t'} + u_{1,n},$$

for $n = 1, \ldots, 6$. Table 6 summarizes the results of the OLS estimation of the equations.

Table 6: ECB's communication and policy rate expectations measured by (implicit) forward Eonia Swap Index (controlling for unexpected components of monetary policy decisions, measured by differences in values of the Eonia Swap Index).

		$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
$\alpha_{1,n}$		-0.003	-0.006	-0.004	-0.006	-0.008	-0.008
		()	()	()	()	()	()
$\Delta r'_{t'}$	$\beta_{1,n}$	0.860^{***}	1.007^{***}	1.105^{***}	1.106^{***}	1.150^{***}	1.188^{***}
		(***)	(***)	(***)	(***)	(***)	(***)
	SE	0.106	0.126	0.158	0.200	0.220	0.237
		(0.095)	(0.190)	(0.235)	(0.244)	(0.266)	(0.287)
$\Delta I_{t'}$	$\gamma_{1,n}$	0.031^{***}	0.036^{***}	0.052^{***}	0.061^{***}	0.073^{***}	0.076^{***}
		(**)	(**)	(**)	(**)	(**)	(**)
	SE	0.106	0.013	0.016	0.020	0.022	0.024
		(0.012)	(0.017)	(0.021)	(0.028)	(0.031)	(0.033)
R^2		0.46	0.45	0.41	0.32	0.31	0.29
Adj. R^2		0.45	0.44	0.40	0.31	0.30	0.28

Note: *, **, and *** denote significance at the 0.1, 0.05, and 0.01 level, respectively. Heteroschedasticity robust (HAC) standard errors and significance of coefficients estimated with HAC errors are reported within brackets. SE stands for "standard error".

Second alternative specification

$$\Delta F_{t'} = \alpha_1 + \beta_1\Delta r'_{t'} + \gamma_1\Delta I_{t'} + u_1.$$

Table 7 summarizes the results of the OLS estimation of the equations.

Table 7: ECB's communication and policy rate expectations measured by three-month Euribor futures rates (controlling for unexpected components of monetary policy decisions, measured by differences in values of the Eonia Swap Index).

α_1		-0.004
		(*)
$\Delta r'_{t'}$	β_1	0.321**
		(**)
	SE	0.124
		(0.126)
$\Delta I_{t'}$	γ_1	0.058***
		(***)
	SE	0.016
		(0.019)
R^2		0.17
Adj. R^2		0.16

Note: *, **, and *** denote significance at the 0.1, 0.05, and 0.01 level, respectively. Heteroschedasticity robust (HAC) standard errors and significance of coefficients estimated with HAC errors are reported within brackets. SE stands for "standard error".

References

[1] Beetsma, R., M. Giuliodori, F. de Jong, and D. Widijanto (2013), "Spread the News: The Impact of News on the European Sovereign Bond Markets during the Crisis", *Journal of International Money and Finance*, 34, pp. 83–101.

[2] Berger, H., J. de Haan, and J.-E. Sturm (2011), "Does Money Matter in the ECB Strategy? New Evidence Based on ECB Communication", *International Journal of Finance and Economics*, 16, pp. 16–31.

[3] Bernoth, K., and J. von Hagen (2004), "The Euribor Futures Market: Efficiency and the Impact of ECB Policy Announcements", *International Finance*, 7, pp. 1–24.

[4] Bernanke, B. S., and I. Mihov (1998), "Measuring Monetary Policy", *Quarterly Journal of Economics*, 113, pp. 896–902.

[5] Blinder, A. S., M. Ehrmann, M. Fratzscher, J. de Haan, and D.-J. Jansen (2008), "Central Bank Communication and Monetary Policy: A Survey of the Evidence", *Journal of Economic Literature*, 46, pp. 910–945.

[6] Casarin, R., and F. Squazzoni (2013), "Being on the Field When the Game Is still under Way. The Financial Press and Stock Markets in Times of Crisis", *PloS ONE*, 8: e67721. doi:10.1371/journal.pone.0067721.

[7] Conrad, C., and M. J. Lamla (2010), "The High-Frequency Response of the EUR-USD Exchange Rate to ECB Communication", *Journal of Money, Credit and Banking*, 42, pp. 1391–1417.

[8] de Haan, J. (2008), "The Effect of ECB Communication on Interest Rates: An Assessment", *Review of International Organizations*, 3, pp. 375–398.

[9] KOF (2007), "KOF Monetary Policy Communicator for the Euro Area – Supplementary Information", Press Release, Friday, November 2.

[10] Kuttner, K. N. (2001), "Monetary Policy Surprises and Interest Rates: Evidence From the Fed Funds Futures Market", *Journal of Monetary Economics*, 47, pp. 523–544.

[11] Gerlach, S. (2007), "Interest Rate Setting by the ECB, 1999–2006: Words and Deeds", *International Journal of Central Banking*, 3, pp. 1–46.

[12] Gürkaynak, R. S., B. Sack, and E. Swanson (2005), "Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements", *International Journal of Central Banking*, 1, pp. 55–93.

[13] Heinemann, F., and K. Ullrich (2007), "Does it Pay to Watch Central Bankers' Lips? The Information Content of ECB Wording", *Swiss Journal of Economics and Statistics*, 143, pp. 155–185.

[14] Jansen, D. J., and de Haan, J. (2007), "The Importance of Being Vigilant: Has ECB Communication Influenced Euro Area Inflation Expectations?", CESifo WP 2134.

[15] Leeper, E. M. (1997), "Narrative and VAR Approaches to Monetary Policy: Common Identification Problems", *Journal of Monetary Economics*, 40, pp. 641–658.

[16] Lucca, D. O., and F. Trebbi (2011), "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements", WP.

[17] Musard-Gies, M. (2006), "Do European Central Bank's Statements Steer Interest Rates in the Euro Zone?", *Manchester School*, 74, pp. 116–139.

[18] Neuenkirch, M. (2013), "Monetary Policy Transmission in Vector Autoregressions: A New Approach Using Central Bank Communication", *Journal of Banking and Finance*, 37, pp. 4278–4285.

[19] Romer, C. D., and D. H. Romer (1989), "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz", *NBER Macroeconomics Annual*, 4, pp. 121–170.

[20] ——— (2004), "A New Measure of Monetary Shocks: Derivation and Implications", *American Economic Review*, 94, pp. 1055–1084.

[21] Rosa, C., and G. Verga (2008), "The Impact of Central Bank Announcements on Asset Prices in Real Time", *International Journal of Central Banking*, 4, pp. 175–216.

[22] _____ (2007), "On the Consistency and Effectiveness of Central Bank Communication: Evidence from the ECB", *European Journal of Political Economy*, 23, pp. 146–175.

[23] Sturm, J-E., and de Haan, J. (2011), "Does Central Bank Communication Really Lead to Better Forecasts of Policy Decisions? New Evidence Based on a Taylor Rule Model for the ECB", *Review of world economics*, 147, pp. 41–58.

[24] Ullrich, K. (2008), "Inflation Expectations of Experts and ECB Communication", *North American Journal of Economics and Finance*, 19, pp. 93–108.