

Just Words? A Quantitative Analysis of the Communication of the Central Bank of Brazil*

Carlos Carvalho[†], Fernando Cordeiro[‡] and Juliana Vargas[§]

Contents: 1. Introduction; 2. Measuring the informational content of COPOM statements; 3. Data; 4. Regression specification; 5. Results; 6. Conclusion.

Keywords: Brazil, COPOM, central bank, communication, semantic, term-structure of interest rates, monetary policy.

JEL Code: E43, E52, E58.

We quantify the informational content of statements issued by the interest-rate setting committee of the Central Bank of Brazil (COPOM), building on the methodology developed by Lucca and Trebbi (2011). Using Google search queries, we measure the extent to which each COPOM statement is perceived to be associated with more “hawkish” or “dovish” language. This allows us to construct a time-series of the informational content of COPOM statements, which we then use in regressions to explain changes in the term-structure of interest rates around COPOM meetings – together with a market-based measure of interest-rate surprises. We find that, during Governor Tombini’s tenure, interest-rate surprises started to be “passed through” one-to-one (or more) even at long maturities, as markets seem to have bought into the idea that the interest-rate cuts that began in mid-2011 would lead to lower yields in Brazil into the foreseeable future. Most importantly, changes in the informational content of COPOM statements seem to have meaningful effects on yields at short-to-medium maturities. However, this result only holds for the period prior to Tombini’s tenure.

Neste artigo nós quantificamos o conteúdo informacional dos comunicados do Comitê de Política Monetária do Banco Central do Brasil (COPOM), com base na metodologia desenvolvida por Lucca and Trebbi (2011). Por

*Prepared for the workshop “Política Monetária: Estudos para o Brasil”, held at IPEA in Rio de Janeiro, June 2013. We thank participants and an anonymous referee for their comments and suggestions. Remaining errors are our own. This paper incorporates results from Vargas (2012) and Cordeiro (2013). E-mails: cvianac@econ.puc-rio.br, fernandolpcordeiro@gmail.com, juliana.tvargas@hotmail.com.

[†]PUC-Rio and Kyros Investimentos

[‡]PUC-Rio

[§]BR-Investimentos



meio de buscas no Google, medimos em que grau cada comunicado do COPOM está associado a uma linguagem mais indicativa de política monetária apertada ou frouxa. Isto nos permite construir uma série de tempo do conteúdo informacional dos comunicados do COPOM, a qual usamos em regressões que visam explicar mudanças na estrutura a termo das taxas de juros após reuniões do COPOM – juntamente com uma medida de surpresas de taxa de juros extraída dos preços de ativos financeiros. Nossos resultados indicam que, durante a gestão do Presidente Tombini, surpresas de juros passaram a ser integralmente incorporadas até mesmo nas taxas de juros de longo prazo. Isto sugere que os participantes dos mercados financeiros podem ter se convencido, ainda que temporariamente, que os cortes de juros iniciados em meados de 2011 produziram taxas de juros mais baixas por um longo período de tempo. Além disso, nosso resultado central sugere que o conteúdo informacional dos comunicados do COPOM tem efeitos significativos sobre taxas de juros de curto e médio prazos. Entretanto, este resultado só se aplica ao período anterior à gestão do Presidente Tombini.

1. INTRODUCTION

“There is no doubt that communication is an integral and important part of the monetary policy process. The messages conveyed by the Central Bank have, per se, induced relevant changes in financial conditions, in general.” (Tombini, 2013).¹

Central bank communication has evolved over time from being infrequent and often obscure to being a tool that is used regularly and in an increasingly transparent fashion by many central banks around the world. This is quite clear in the case of countries that have adopted an inflation targeting regime (e.g., Brazil), but it also applies to central banks with different – sometimes multidimensional – mandates, such as the Fed.

On top of the tendency towards more transparent communication, which started mainly during the 1990s, the global financial crisis of 2008/09 and associated recessions brought communication to the forefront of the monetary policymaking process in countries whose central bank lowered policy rates by so much that they became effectively constrained by the so-called zero lower bound. In those cases, communication – along with balance sheet operations – became a key policy instrument.

The increasing importance attributed to communication by policymakers has been accompanied by research that tries to attest its relevance.² In particular, one strand of this literature tries to measure the effects of central bank communication on asset prices by analyzing and attaching numerical values to central bank statements.³ For example, Rosa and Verga (2008) do so by resorting to a subjective quantification of the content of ECB statements. Costa Filho and Rocha (2009, 2010) apply this method

¹Originally in Portuguese: “Não resta dúvida de que a comunicação é parte integrante e importante do processo de condução da política monetária. As mensagens passadas pelo Banco Central, per se, determinaram mudanças relevantes nas condições financeiras, de modo geral.”

²For a thorough survey of the literature on central bank communication see Blinder et al. (2008).

³There is also work that tries to quantify the effects of language on asset prices in other contexts. For example, Tetlock et al. (2008) document that individual stock prices react to the linguistic content of news stories about the company. Incidentally, the title of our paper is inspired by theirs.

to analyze the minutes of the meetings of the interest-rate setting committee of the Central Bank of Brazil (COPOM).⁴

In this paper we analyze the effects of COPOM statements on the term-structure of interest rates in Brazil. To that end, we quantify the informational content of those statements by adapting the methodology developed by Lucca and Trebbi (2011) to specificities of the “monetary policy jargon” in Portuguese.⁵ Using Google search queries, we measure the extent to which each COPOM statement is perceived to be associated with more “hawkish” or “dovish” language.⁶ This allows us to construct a time-series of the informational content of COPOM statements, which we then use in regressions – together with a market-based measure of interest-rate surprises – to explain changes in the term-structure of interest rates around COPOM meetings. We also investigate whether changes in language precede changes in the policy rate – the so-called SELIC rate.

The main advantage of the methodology proposed by Lucca and Trebbi (2011) over other existing approaches is that it minimizes the subjectivity inherent in classifying statements as hawkish or dovish. Their idea is to measure how economic agents – not the researchers – perceive the informational content of any given statement. This is achieved by measuring the relative frequency with which segments of the given statement are associated with the terms hawkish and dovish in a dataset that contains many articles, blogs etc. that (potentially) comment on or analyze that statement. Lucca and Trebbi (2011) go the extra mile and partition the statements using grammatical rules that further mitigate the role of judgement.

We adapt their methodology to the context of how monetary policy decisions are alluded to in Brazil. In the absence of words in Portuguese that play the roles of “hawkish” and “dovish”, we replace the latter with expressions that indicate tighter or looser monetary policy, such as “hike in interest rates” or “interest rate cut.” We also deviate from Lucca and Trebbi (2011) in that we partition the COPOM statements ourselves.⁷

Our results are as follows. As expected, interest-rate surprises affect the yield curve. Prior to Governor Tombini’s tenure these effects decreased with maturity, with long rates eventually pricing in a reversal of interest-rate surprises. This pattern changed dramatically during Tombini’s tenure, as even long rates started to respond one-to-one – sometimes more – to interest-rate surprises. That is, markets seem to have bought into the idea that the interest-rate cuts that began in mid-2011 would lead to a lower level for nominal rates into the very distant future. Finally, changes in COPOM language seem to have an effect on yields at short-to-medium maturities. But this result only holds prior to Tombini’s tenure. During the latter period, yields appear not to respond to changes in COPOM language.

Quantitatively, our estimates for the pre-Tombini period imply that a one standard deviation shift in COPOM language in the hawkish direction increases 4-month to 2-year yields by between 2 and 5 basis points. As a reference point, during the same period a positive one standard deviation interest-rate surprise increases 4-month yields by about 12 basis points, and 2-year yields by less than 2 basis points. Because we measure changes in language and not pure communication surprises, we argue below that our estimates likely understate the true causal effect of communication surprises on the yield curve. It thus appears that the effects of COPOM communication on medium-term maturities were large, at

⁴There is also a strand of the literature that tries to document the effects of communication without attempting to measure its content directly. For example, Janot and Mota (2012) rely on dummy variables for COPOM communication dates to document that the volatility of the term structure of interest rates is different after COPOM dates and also after the release of the quarterly Inflation Report. Montes (2012) also studies COPOM communication through the use of COPOM communication dummies.

⁵We build on Vargas (2012) and Cordeiro (2013), who, to the best of our knowledge, were the first to analyze COPOM statements using a methodology adapted from Lucca and Trebbi (2011).

⁶The words “hawkish” and “dovish” are most commonly used to classify Fed actions or language that are perceived as being associated with, respectively, tighter and looser monetary policy stances.

⁷While this is less desirable than their approach, we did not find a practical way to apply grammatical rules to perform that task.



least prior to Tombini's tenure. In the conclusion we offer possible explanations for why we fail to find a significant relationship between changes in COPOM language and changes in yields during Tombini's tenure.

2. MEASURING THE INFORMATIONAL CONTENT OF COPOM STATEMENTS

We start with a short introduction to the theoretical foundations of the methodology proposed by Lucca and Trebbi (2011) and a description of how we adapt it to analyze the monetary policy process in Brazil. The basic assumption is that it is possible to rank the intensity and direction of meaning of a given word or sentence according to an established metric. One way to formalize this idea is to use the concept of a semantic axis. Such an axis is defined by two words of opposite meaning. The position of a word on this axis indicates the direction of meaning relative to the defining antonyms, as well as its intensity. We shall refer to the position of a given word or sentence on this abstract axis as the semantic orientation of the word (or sentence).

For our purposes, we would like to build a numeric score that is capable of estimating the semantic orientation of a sentence over the semantic axis defined by the polar concepts of "restrictive" (or "tight") and "expansionary" (or "loose") monetary policy. We would also like our measure to increase with the degree of "monetary tightness" of the sentence.

The first step is to establish a measure of association between strings of text. Using this measure it is possible to compare a given string association with the two extremes that define our semantic axis, and thus infer some information about its semantic direction and intensity. We follow Lucca and Trebbi (2011) and employ a measure of association between words proposed by Church and Hanks (1990) for the study of natural languages. The measure is the pointwise mutual information (*PMI*) between the strings of text x and y , and is defined as:

$$PMI(x,y) = \log \left(\frac{\Pr(x,y)}{\Pr(x)\Pr(y)} \right). \quad (1)$$

Intuitively, *PMI* measures how much one word tells us about the other. If there is some statistical dependence between x and y , the probability of observing them together in a large corpus of text, $\Pr(x,y)$, will be greater than the product of the probabilities of observing them individually, $\Pr(x)\Pr(y)$. In this case, $PMI > 0$ would signal an association between x and y .

The semantic orientation of a string of text x over the semantic axis with the two polar concepts a_1 and a_2 is then defined as:

$$SO(x) = PMI(x,a_1) - PMI(x,a_2) = \log \left(\frac{\Pr(x,a_1)\Pr(a_2)}{\Pr(x,a_2)\Pr(a_1)} \right).$$

In our application, we consider the universe of all webpages indexed by Google as the corpus of text and estimate the relevant probabilities as the associated empirical frequencies using Google hit counts. More specifically, we compute the Google-based orientation score as follows:

$$GSO(x) = \log \left(\frac{hits(x,T_r) hits(T_e)}{hits(x,T_e) hits(T_r)} \right) = \log \left(\frac{hits(x,T_r)}{hits(x,T_e)} \right) + \gamma,$$

where T_r is a term indicating restrictive monetary policy, T_e is a term indicating expansionary policy, $hits(x,T)$ is a function returning the number of hits in a Google search of the string x combined with term T , and $\gamma = \log(hits(T_e)/hits(T_r))$ is a (unimportant) constant.

2.1. Choosing semantic direction terms

Our main objective is to estimate the position of selected strings of text from COPOM statements over an axis that runs from "expansionary" to "restrictive monetary policy." However, these polar con-

cepts are fairly complex. In particular, choosing a pair of antonyms in Portuguese that accurately describe them is a very difficult (if not impossible) task. Thus, we use an extended set of antonyms to construct a score measure that retains the intuition of the one proposed by Lucca and Trebbi (2011), but is better suited to the Portuguese language. If a given string of text is relatively more associated with words that describe monetary restriction rather than with the words that describe monetary expansion, our measure also assigns a relatively higher score to that string.

In order to define the semantic orientation terms that we use in our analysis, we select 15 sets of terms that could relate to movements in interest rates, trying to strike a balance between terms with a restrictive and those with an expansionist connotation. We then search for the groups of terms in Google, using quotation marks, and compute the number of hits in each Google search. We calculate the average number of hits for each set of terms and rank them accordingly. Finally, we select two sets of terms for our analysis (Tables 1 and 2):

Table 1: Word set in Portuguese

Expansionary	Restrictive
Corte de Juros	Alta de Juros
Redução de Juros	Elevação de Juros
Queda de Juros	Aumento de Juros

The set of corresponding terms in English is:

Table 2: Word set in English

Expansionary	Restrictive
Rate Cut	Rate Hike
Rate Reduction	Rate Elevation
Rate Decrease	Rate Increase

Let the set of restrictive monetary policy terms be $R = \{ \text{“alta de juros”, “elevação de juros”, “aumento de juros”} \}$, and the set of expansionary policy terms be $E = \{ \text{“corte de juros”, “queda de juros”, “redução de juros”} \}$. Then, building on Lucca and Trebbi (2011), we calculate the score of a given statement segment as:

$$GSO(x) = \log \left(\frac{T_{r \in R} \text{hits}(x, T_r)}{T_{e \in E} \text{hits}(x, T_e)} \right). \quad (2)$$

Finally, the semantic orientation of COPOM statement k is the average of the search units that compose the statement.

$$GSO_k = \sum_{x \in k} \frac{GSO(x)}{\#(x \in K)}. \quad (3)$$

In our regression analyses we make use of the (standardized) first difference of the score. So the level of the semantic orientation score does not matter. What matters is the fact that our measure is such that higher scores are associated with the perception of a tighter monetary policy stance.



2.2. Search units

We divide the sentences in each COPOM statement into shorter units, which we call “search units.” The main purpose of this step is to decompose the statement into the basic ideas that make the whole text meaningful. In principle, the punctuation used in the statement could be used as a delimiter for the search units. However, such rule would yield excessively long units with complex syntax and semantic structures. In order to form a set of words that might adequately represent the semantic content of the statements, we decompose sentences into any sequence that, in our view, has relevant meaning.

For concreteness, we present the following example based on the COPOM statement of November 28, 2012.

“O Copom decidiu, por unanimidade, manter a taxa Selic em 7,25% a.a., sem viés. Considerando o balanço de riscos para a inflação, a recuperação da atividade doméstica e a complexidade que envolve o ambiente internacional, o Comitê entende que a estabilidade das condições monetárias por um período de tempo suficientemente prolongado é a estratégia mais adequada para garantir a convergência da inflação para a meta, ainda que de forma não linear.”

Table 3 shows how the search units were defined, their Google search results, and the statements’ final score. Every search was implemented with the search units and semantic direction term using quotes, separated by a blank space. Quotes are used to make sure that Google searches for the exact sentence – otherwise it could return synonyms and only tangentially related text. For example, the query for “convergência da inflação para a meta” and the term “corte de juros” is [“convergência da inflação para a meta” “corte de juros”].

Table 3: Search units for the November 28, 2012 COPOM statement

Search unit “x”	
(1)	Considerando o balanço de riscos para a inflação
(2)	recuperação da atividade doméstica
(3)	complexidade que envolve o ambiente internacional
(4)	o Comitê entende que a estabilidade das condições monetárias por um período de tempo suficientemente prolongado é a estratégia mais adequada para garantir a convergência da inflação para a meta, ainda que de forma não linear
(5)	estabilidade das condições monetárias por um período de tempo suficientemente prolongado
(6)	estabilidade das condições monetárias
(7)	estratégia mais adequada para garantir a convergência da inflação para a meta
(8)	convergência da inflação para a meta, ainda que de forma não linear
(9)	convergência da inflação para a meta

"x"	corte *	queda *	redução *	alta *	elevação *	aumento *	GSO("x")
1	35	4	7	8	1	9	-0.407
2	64	5	8	1	0	1	-1.585
3	51	4	9	5	1	4	-0.806
4	178	178	174	174	178	175	-0.002
5	53	4	7	0	0	1	-1.806
6	57	4	8	1	0	1	-1.537
7	62	5	9	8	2	9	-0.602
8	59	2	8	0	0	1	-1.838
9	160	82	144	61	10	94	-0.369
GSO_k							-0.995

* = "de juros"

2.3. Methodological adjustments

To try to ensure that our score is aligned with market participants' information sets when they reacted to the statement, we impose a temporal constraint on all searches. Using Google's search tools, we restrict the search to include results only up to the date of the statement's release. As statements are released after trading hours, this restriction ensures that the score is not capturing later analyses that associate the statement with market movements on the following day – only the meaning agents might have attributed to the statement without observing the market's reaction.

When calculating the semantic scores, there are cases where the numerator or the denominator of the fraction in equation (2) is equal to zero. We impose the rule that when the sum of hits of all terms in the R set is equal to zero, the value of the score is set equal to the minimum value obtained in the whole series. Inversely, when the sum of the results of all terms in the E set is equal to zero, the value of the score is set to the maximum value of the whole score series. When both terms are equal to zero we drop the sentence from the statement's set of search units, since we believe that such occurrences indicate absence of reasonable meaning of the sentence in reference to our metric and the corpus of text.

The final semantic orientation score is characterized by the sets R and E , and the two methodological adjustments described above. For comparison purposes we also compute two variations of the semantic score. In the first one we weight the score of each search unit by its relative number of hits.⁸

In the second one we aggregate hawkish and dovish hit counts over all search units prior to taking logs.⁹

Although the results vary from one specification to the other, the main lessons from our analysis survive the use of these alternative measures of the informational content of COPOM communication.

3. DATA

Central Bank communication may occur over many different channels such as speeches, statements, reports, and minutes. As statements are a brief description of the Committee's views on the overall state of the economy and are systematically released after market sessions, they are well suited for

⁸Formally, $GSO^k = \sum_{x \in k} \alpha_x \frac{GSO(x)}{\#(x \in K)}$, where $\alpha_x = \frac{hits(x, T)}{[\sum_{x \in k} T \in \{R \cup E\} hits(x, T)]}$.

⁹Formally, $GSO^k = \log(\sum_{x \in k} T_r \in R hits(x, T_r)) - \log(\sum_{x \in k} T_e \in E hits(x, T_e))$.



an analysis of COPOM communication. The baseline score was constructed for a sample of 46 COPOM statements, from October 2007 to May 2013. These statements are generally composed of a justification for the SELIC target rate decision and an indication of the Committee's view of the economy. Market participants regard them as a valuable source of information to understand the COPOM's current thinking, as well as to construct possible scenarios for monetary policy. All statements are from the website of the Central Bank of Brazil.

We construct the yield curve using "Swap $DI \times pré$ " fixed-rate swap contracts. We include the following maturities: 1, 2, 3, 4, 6 months, and 1, 2, 3, 5, and 10 years. Data for the maturities up to 1 year are from the Central Bank of Brazil website, and the remaining maturities are from Bloomberg. The dependent variable in each of our regressions is the difference between the yield of a given maturity at the close of trading on the day following the COPOM meeting and the yield at the close of trading on the COPOM date. Table 4 reports summary statistics for the dependent variables.

Table 4: Summary statistics for the dependent variables

	Δ 1-month	Δ 2-months	Δ 3-months	Δ 4-months	Δ 6-months
Mean	-0.027	-0.028	-0.039	-0.024	-0.032
Median	-0.020	-0.020	-0.020	-0.010	-0.010
St. Dev.	0.101	0.110	0.124	0.169	0.150
Min	-0.390	-0.390	-0.490	-0.560	-0.600
Max	0.160	0.270	0.220	0.650	0.350
	Δ 1-year	Δ 2-years	Δ 3-years	Δ 5-years	Δ 10-years
Mean	-0.036	0.010	0.029	0.043	0.035
Median	0.000	0.014	0.032	0.040	0.042
St. Dev.	0.169	0.157	0.156	0.160	0.161
Min	-0.580	-0.491	-0.370	-0.332	-0.309
Max	0.410	0.408	0.420	0.539	0.520

4. REGRESSION SPECIFICATION

In order to estimate how changes in communication affect interest rates, we run the following regressions:

$$\Delta y_k^n = \beta_0 + \beta_1 MP_k \cdot (1 - Tomb) + \beta_2 \Delta GSO_k \cdot (1 - Tomb) + \beta_3 MP_k \cdot Tomb + \beta_4 \Delta GSO_k \cdot Tomb + \varepsilon_k, \quad (4)$$

where Δy_k^n is the change in the yield of maturity n between the date of statement k 's release and the following trading day; MP_k is a measure of interest-rate surprises to be discussed in more detail in the next section; $\Delta GSO_k \equiv GSO_k - GSO_{k-1}$ is the difference between the current statement's semantic score and the last statement's score; $Tomb$ is a dummy variable that equals 1 for all COPOM meetings under Tombini's presidency at the Central Bank of Brazil (i.e., from January 2011 onwards), and zero otherwise.¹⁰

4.1. Interest-rate surprises

Traditionally, COPOM decisions are announced at night, when Brazilian financial markets are closed. Assuming rational expectations, in the next business day yields should adjust in response to

¹⁰Replacing the constant in (4) with the $Tomb$ and $(1 - Tomb)$ dummies does not change our conclusions in any meaningful way.

- i) surprises in the SELIC rate decision – i.e., the difference between market expectations for the decision and the actual decision;
- ii) new information provided by the statement;
- iii) any other relevant information. We are implicitly assuming that in the day following each COPOM meeting in our sample there was no disclosure of other relevant economic information which could affect the asset prices that we analyze.

Therefore, in our regressions we only control for interest-rate surprises. We estimate these surprises as the difference between the 30-day *Swap DI* \times *pré* on the first trading day after the COPOM decision and the 30-day rate on the decision day. As COPOM meetings are more than 30 days apart, if we assume negligible risk premia for very short-term contracts, 30-day contracts on the day of the announcement essentially reflect market expectations of the upcoming interest-rate decision.

4.2. Semantic orientation series

Under the assumption of rational expectations, yields along the term-structure of interest rates should reflect agent's outlooks for the future path of the policy interest rate, and possibly risk premia. Therefore, we should expect the *non-anticipated* component of the semantic content of statements to affect asset prices.

Because ΔGSO_k is not a communication surprise, in general the OLS estimates of the coefficients β_2 and β_4 in equation (4) are likely to be biased toward zero, relative to the true causal effects of communication surprises on yields. The reason is that the regressor ΔGSO_k includes both the component that should affect our dependent variables (i.e., the communication surprises), as well as a component that should not. That component – the expected change in the informational content of the statement relative to the previous COPOM meeting – works as measurement error in that regressor, and thus tends to bias the associated coefficients toward zero relative to the true causal effects of surprises.

Short of being able to quantify expectations of the informational content of the statements, we consider the difference between the current statement score and the last statement score as a noisy proxy for communication surprises.

5. RESULTS

Before presenting results from our regressions – which can speak to the question of whether yields react to changes in COPOM language – we provide a very simple answer to the more basic question of whether markets *should* react to COPOM statements. The answer would be negative if changes in language were not informative of future changes in the policy rate. We therefore run Granger causality tests between changes in our semantic score index and changes in the SELIC rate.

Figure 1 shows the time series of our semantic score index and of the SELIC rate. It suggests that changes in the former precede changes in the latter. This result is confirmed by formal Granger-causality tests based on a bivariate VAR with changes in these two variables. There is strong statistical evidence that changes in our semantic score index precede changes in the policy rate by one COPOM meeting, while we fail to reject the null of Granger causality in the other direction.¹¹ This result holds both prior to, and during Tombini's tenure.

Table 5 reports our main results. We emphasize three findings. First, as expected, yields respond to interest-rate surprises. This is true both during, and prior to Governor Tombini's tenure. Second, prior to Tombini's tenure the effects of interest-rate surprises decreased with maturity. Starting from

¹¹This results is robust to estimating the system with one or two lags, which are the prescriptions of various information criteria. Results are available upon request.

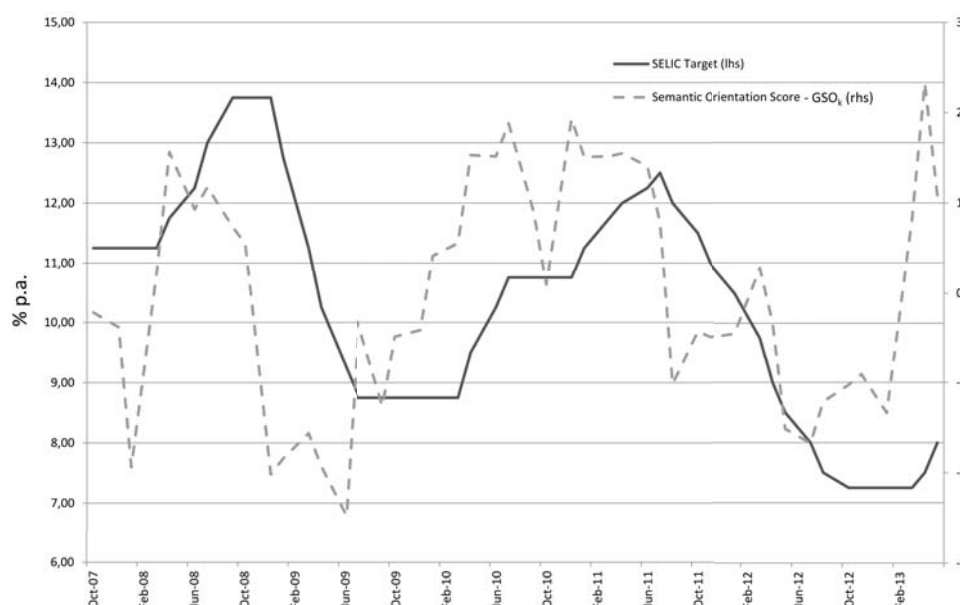


Figure 1: SELIC and Semantic Orientation Score

around a one-to-one response at very short maturities, coefficients decline in a roughly monotonic fashion, with long rates eventually pricing in a reversal of interest-rate surprises. This pattern changed dramatically under Governor Tombini, as even long rates started to respond one-to-one – sometimes more – to interest-rate surprises. That is, markets seem to have bought into the idea that the interest-rate cuts that began in mid-2011 would lead to a lower level for nominal rates into the very distant future.

Finally, changes in COPOM language seem to have an effect on yields at short-to-medium maturities – up to one, perhaps two years. But this result only holds in the period prior to Tombini's tenure. Quantitatively, if taken at face value, the estimated coefficients imply that a one standard deviation increase in the semantic orientation score of a COPOM statement – which corresponds to a unitary increase in the level of the score – increases 4-month to 2-year yields by between 2 and 5 basis points. As a reference point, a positive one standard deviation interest-rate surprise during the same period (of approximately 10 basis points) increases 4-month yields by about 12 basis points and 2-year yields by less than 2 basis points.

These results are, of course, subject to the caveat that changes in our semantic orientation score do not correspond to pure communication surprises. However, as argued previously, the OLS coefficients estimated with our semantic score are likely to understate the true causal effect of communication surprises. It thus appears that the effects of COPOM communication on medium-term maturities were large, at least in the period prior to Tombini.

5.1. An important remark

An important drawback of using Google queries to construct the semantic orientation score is its difficult replication. Because Google is continuously searching the web in order to keep its index of pages up to date and because webpages might be removed over time, it is quite possible that two identical queries made with a reasonable time interval between them will have different results. As statements get older it is more likely that webpages quoting them disappear from the results.

We tried to handle this obstacle by forcing all queries to use the same Google IP address. Google spreads its service among different data centers, and each one has its own web cache that is not necessarily identical to the others. Using the same IP address at every query is a way to try to assure that one is using the same corpus of text every time. In addition, after some point we started saving printouts of the webpages with the number of hits returned by each query. This is our de facto dataset. Unfortunately we did not implement this procedure since the beginning of this research project, and initially only recorded the number of hits in a spreadsheet. Our dataset of printouts is thus incomplete.

Regarding alternative search engines, there is no guarantee that using Yahoo (Bing), or any other search engine, would yield the same result. Because each one has its proprietary algorithm for web crawling and indexing pages, results are likely to differ. In fact, our experience with Yahoo (Bing) shows that hit counts are often quite different.

The ideal corpus of text to perform the kind of analysis we undertake in this paper is a dataset of articles, blog entries etc. that only expands over time, such as Factiva's. Unfortunately we did not have access to such a resource when conducting our research for this paper.

Table 5: Regression results

Dep. variable	$MP_k \cdot (1 - Tomb)$	$\Delta GSO_k \cdot (1 - Tomb)$	$MP_k \cdot Tomb$	$\Delta GSO_k \cdot Tomb$	R^2
Δ 2-mo. yield	1.022 [0]***	0.007 [.238]	1.047 [0]***	0.000 [.997]	0.910
Δ 3-mo. yield	0.735 [.003]***	-0.001 [.896]	1.267 [0]***	0.008 [.557]	0.702
Δ 4-mo. yield	1.256 [0]***	0.044 [.051]*	1.561 [0]***	0.010 [.577]	0.747
Δ 6-mo. yield	0.889 [0]***	0.021 [.034]**	1.671 [0]***	0.010 [.67]	0.800
Δ 1-yr. yield	0.577 [.008]***	0.052 [.062]*	1.673 [0]***	0.007 [.83]	0.597
Δ 2-yr. yield	0.160 [.614]	0.045 [.101]	1.546 [0]***	0.001 [.976]	0.537
Δ 3-yr. yield	-0.105 [.791]	0.036 [.249]	1.240 [0]***	-0.012 [.582]	0.355
Δ 5-yr. yield	-0.292 [.489]	0.039 [.258]	1.136 [0]***	-0.014 [.441]	0.315
Δ 10-yr. yield	-0.474 [.296]	0.017 [.711]	0.756 [0]***	-0.007 [.751]	0.170

Note: p-values in brackets are based on robust standard errors. Superscripts *, **, and *** denote statistical significance at, respectively, 10%, 5% and 1% levels.



6. CONCLUSION

We quantify the informational content of COPOM statements using a semantic orientation score and find that communication matters for monetary policymaking in Brazil. Our results suggest that at a given point in time communication is indeed a policy instrument that is (or should be) related to the policy rate, but has some life of its own. Of course the two cannot move independently over time, else economic agents will not perceive any relevant information in COPOM statements.

Our results indicate that the yield curve responds differently to interest-rate surprises and changes in communication under Governor Tombini than it did in the past. In particular, from our results it appears that markets had “bought” into the idea that the drop in interest rates that began in mid-2011 was going to be long-lived, as rates started to respond one-to-one (or more) to interest-rate surprises, even at the long end of the yield curve. It will be interesting to see what happens next, now that the COPOM has embarked on a tightening cycle.

Most importantly, our results suggest that the yield curve stopped responding to changes in COPOM language during Tombini’s tenure. It is important to keep in mind that our findings are only suggestive, because we do not measure pure communication surprises. However, as argued above it is likely that our estimates understate the causal effects of communication surprises on the yield curve. The reason is that expected changes in language should not induce changes in yields. Consequently, it might be the case that the reason we fail to find a significant relationship between changes in COPOM language and changes in yields during Tombini’s tenure is that all such changes in communication were perfectly anticipated. Unexpected shifts in COPOM language under Tombini would have continued to shape the yield curve as usual – had only such communication surprises occurred.

Needless to say, it may also be the case that communication surprises have occurred during Tombini’s tenure, but for some reason yields failed to respond in any systematic fashion. Poor communication and/or lack of credibility are possible explanations that come to mind.

In future research it would thus be valuable to push our analysis further, and try to construct measures of expectations of COPOM language that can be used to better estimate the effects of communication surprises. One would then be able to test whether the response of markets to COPOM communication has indeed changed over time, as our results indicate.

BIBLIOGRAPHY

- Blinder, A., Ehrmann, M., Fratzsche, M., Haan, J. D., & et al, D. J. (2008). Central Bank Communication and Monetary Policy a Survey of Theory and Evidence. ECB Working Paper 898.
- Church, K. W. & Hanks, P. (1990). Word association norms, mutual information and lexicography. *Computational Linguistics*, 16:22–19.
- Cordeiro, F. (2013). Quantificando os Efeitos da Comunicação do BCB. Monografia PET, Departamento de Economia da PUC-Rio.
- Costa Filho, A. & Rocha, F. (2009). Comunicação e Política Monetária no Brasil. *Revista Brasileira de Economia*, 63:405–422.
- Costa Filho, A. & Rocha, F. (2010). Como o Mercado de Juros Futuros Reage à Comunicação do Banco Central? *Economia Aplicada*, 14:265–292.
- Janot, M. & Mota, D. (2012). O Impacto da Comunicação do Banco Central do Brasil sobre o Mercado Financeiro. Trabalhos para Discussão 265, BACEN.
- Lucca, D. & Trebbi, F. (2011). Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements. Technical Report 15367, 2009 version, NBER WP.

- Montes, G. (2012). Financial market reaction to central bank monetary policy communications under an inflation-targeting regime: the case of Brazil. *CEPAL Review*, 107:165–181.
- Rosa & Verga (2008). The Impact of Central Bank Announcements on Asset Prices in Real Time: Testing the Efficiency of the Euribor Futures Market. *International Journal of Central Banking*, 4:175–217.
- Tetlock, P., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *Journal of Finance*, 63:1437–1467.
- Tombini, A. (2013). Speech by BCB Governor Alexandre Tombini before Congress, May 21, 2013. Available at http://www.bcb.gov.br/pec/appron/apres/Discurso_Alexandre_Tombini_CMO%_21-5-2013.pdf.
- Vargas, J. (2012). Mensurando a Comunicação do BACEN: Análise quantitativa dos comunicados do Copom. Monografia de graduação, Departamento de Economia da PUC-Rio.