

IFC-Bank of Italy Workshop on “Data Science in Central Banking: Applications and tools”

14-17 February 2022

## Machine learning for measuring central bank policy credibility and communication from news<sup>1</sup>

Muhammad Abdul Jabbar, Okiriza Wibisono,  
Anggraini Widjanarti and Alvin Andhika Zulen,  
Bank Indonesia

---

<sup>1</sup> This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

# Machine Learning for Measuring Central Bank Policy Credibility and Communication from News

Okiriza Wibisono<sup>1</sup>, Muhammad Abdul Jabbar<sup>2</sup>, Alvin Andhika Zulen<sup>3</sup>, Anggraini Widjanarti<sup>4</sup>

## Abstract

Central bank's monetary policy will be effective only if the central bank is perceived to be credible. The relationship of monetary policy credibility and effectiveness has long been established by economists and central bankers, including credibility's significance in determining central bank's ability to manage the expectation of economic actors. Empirically, credibility is a qualitative concept, and thus is not straightforward to measure. In our previous research, we have constructed a machine learning-based index of central bank credibility from news data. The index is based on 4 component indexes: credibility of policy formulation, policy effectiveness, coordination with the government, and policy communication. The communication index measures how well public's expectation is aligned with the central bank's forward guidance, which in our case is traditionally related mainly to the policy rate. With the COVID-19 pandemic, quantitative easing (QE), along with the policy rate, plays a greater role in achieving central bank's monetary policy objectives. In this paper, we describe our machine learning-based credibility measure and propose an improvement to the communication index to accommodate the higher significance of QE. We test the indexes' effect on economists' expectation of short-term inflation.

Keywords: central bank credibility, quantitative easing, central bank communication, inflation expectation

JEL classification: E52, E31, D84

<sup>1</sup> Department of Statistics, Bank Indonesia. email: okiriza\_w@bi.go.id

<sup>2</sup> Department of Statistics, Bank Indonesia, email: muhammad\_abdul@bi.go.id

<sup>3</sup> Department of Statistics, Bank Indonesia, email: alvin\_az@bi.go.id

<sup>4</sup> Department of Statistics, Bank Indonesia, email: anggraini\_widjanarti@bi.go.id

The author would like to thank a colleague for insightful discussion on the econometrics methodology.

## Contents

|                                                                    |    |
|--------------------------------------------------------------------|----|
| 1. Background.....                                                 | 3  |
| 2. Literature Review .....                                         | 4  |
| 2.1 Bank Indonesia's Policy Credibility Survey .....               | 4  |
| 2.2 Text mining of economic news.....                              | 5  |
| 3. Methodology .....                                               | 6  |
| 3.1. Data .....                                                    | 6  |
| 3.1.1. News articles.....                                          | 6  |
| 3.1.2. Inflation estimates.....                                    | 6  |
| 3.2. Policy Credibility Index .....                                | 8  |
| 3.2.1. Annotation .....                                            | 8  |
| 3.2.2. Data preprocessing.....                                     | 9  |
| 3.2.3. Model training .....                                        | 9  |
| 3.2.4. Text classification and index calculation .....             | 10 |
| 3.3. Regression Setting.....                                       | 11 |
| 3.3.1. Model and estimation .....                                  | 11 |
| 3.3.2. Control variables.....                                      | 12 |
| 4. Result and Discussion .....                                     | 13 |
| 4.1 Index Results.....                                             | 13 |
| 4.1.1 Policy credibility index and the 4 component indexes .....   | 13 |
| 4.1.2 Comparison of communication indexes.....                     | 15 |
| 4.2 Regression Results.....                                        | 16 |
| 5. Conclusion & Future Works .....                                 | 18 |
| 5.1. Conclusion .....                                              | 18 |
| 5.2. Future Works .....                                            | 18 |
| References.....                                                    | 19 |
| Appendix A: Example annotated sentences.....                       | 20 |
| Appendix B: Regression results for heterogeneity of estimates..... | 21 |

## 1. Background

Central bank's monetary policy will be effective only if the central bank is perceived to be credible. The role of credibility in supporting policy effectiveness has long been established by economists and central bank practitioners. This is in-line with the rational expectation hypothesis which is believed to guide the behavior of economic agents. The credibility of monetary policy will determine the ability of central bank to manage the expectation of economic actors (Łyziak & Paloviita, 2016) and will be reflected in the expectation of long-term interest rates and other asset prices (Blinder, 2000).

Based on the theory, credibility is about consistency in providing accurate and valuable information or always being responsible (Sobel, 1985). In the terminology of policy makers, credibility is the expectation that policies that are announced will be implemented, not only reflecting the policy makers' wishes, but also capturing the situation and conditions behind the policy so that if there is a shock beyond expectation, it will affect the credibility (Drazen & Masson, 1994). Central bank credibility is formed by 7 (seven) important factors i.e. honesty, independence, consistency (history of fighting inflation), transparency, maintaining fiscal deficit, the use of policy rules, and incentives (personal loss) (Blinder, 2000).

Central bank credibility also plays a pivotal role in building public's trust on the institution. Trust is important for reasons of political accountability, ensuring operationally independent central banks are meeting the terms of their social contract with wider society. Another reason to try to build trust is that trust helps manage expectations. But, evidence suggests that public may never engage with central bank communication because it is written in a way that they cannot understand, which contributes to a lack of trust in the central bank as an independent institution (Haldane et al., 2020). As argued in Haldane and McMahon (2018), one of the reasons that a central bank may want to communicate more directly with the general public is to try to build public understanding as a means of establishing trust and credibility about central banks and their policies.

Empirically, credibility is a qualitative concept, which is not easy to measure, and there are several approaches have been used in measuring credibility. In Cukierman & Meltzer (1986) and Faust & Svensson (2001), central bank independence is used as a proxy for credibility. In Blinder (2000), the credibility of a central bank is measured using survey where economists choose the main characteristics that contribute to the central bank credibility, which is then quantified to produce an index. While in Łyziak (2016) measurement is made using an index, which is built from several indicators including the achievement of the central bank's target, the deviation of inflation expectation, transparent targets and indicators, independence, and accountability. The credibility measure based on deviation of inflation expectation strikes close to the objective of inflation targeting framework. However, it can be difficult to gauge especially as concerns households, who are often not even informed of the central bank's inflation target itself.

In our case, Bank Indonesia used to regularly conduct survey to external stakeholders to measure policy credibility, i.e. Bank Indonesia's Policy Credibility Survey. The Policy Credibility Survey is based on 6 aspects of credibility: formulation, independence, communication, accountability, coordination, and effectiveness.

In practice, the survey method (in general) has several weaknesses:

1. Survey fatigue: respondents experiencing burnout if surveyed repeatedly, so that surveys cannot be carried out too often (especially since the pool of economists or other stakeholders as respondents can be quite limited);
2. Desirability bias: respondents giving a response that is favorable for the surveyor, which is the central bank, hence the results are less objective;
3. Recency bias: respondents generally providing responses based on recent policies and/or events, hence the survey results are very dependent on the execution time; and
4. Survey cost & time.

Based on these considerations, we develop an alternative policy credibility measurement (indexes) by utilizing Big Data Analytics. The indexes are constructed from text mining of public perceptions toward central bank policy credibility that are reported in news media. In this paper, we explain the methodology, some of which have been covered in (Zulen, 2020), as well as describe an improvement to one of the policy credibility indexes and evaluate the indexes in econometric models.

The paper is organized as follows. In section 2, we provide literature reviews on Bank Indonesia's Policy Credibility Survey, text mining of economic and financial news, and inflation estimates, which we use for evaluating the indexes. In section 3, we discuss the data and methodology. In section 4, we provide a summary of the results and evaluation of the model. In section 5, we conclude the paper and offer some thoughts for future works.

## 2. Literature Review

### 2.1 Bank Indonesia's Policy Credibility Survey

From 2013 to 2018, Bank Indonesia conducted Bank Indonesia's Policy Credibility Survey, a semi-annual survey to measure policy credibility for all 3 sectors of policy: monetary, macroprudential, and payment system. The survey was aimed to provide a measure for policy credibility that is objective, accurate, reflecting broad view of stakeholders (including general public), and available timely. The survey was used to determine the effectiveness of policy communication as well as feedback for formulating future policy communication strategies.

The target respondent of the survey was approximately 1,000 respondents in 20 major cities in Indonesia, consisting of government personnel, bankers, industry players, academics, and general public. The survey measured 6 aspects of policy credibility, from which our indexes are derived:

1. Formulation: whether our policies are formulated carefully according to their objectives
2. Independence: whether we formulate our policies independently, without intervention from any party
3. Communication: whether our policies are well-communicated to the public
4. Accountability: whether our policies are well accounted for
5. Coordination: whether we always coordinate well with the government
6. Effectiveness: whether our policies are effective in achieving their objectives

## 2.2 Text mining of economic news

Text data have been widely used for research in economics and finance. Nowadays, text mining algorithms are growing rapidly along with the adoption of big data and machine learning. These algorithms can automatically "read" and "extract" relevant information from texts, such as person's name, topics, and sentiment. Compared to manual approach, text mining allows us to make use of much larger text data faster, including news, social media, and press releases.

As an example related to central banks, Sahminan (2008) identified keywords that reflect a tight, neutral, or loose monetary policy inclination in the press release statement of Bank Indonesia over the period from January 2004 to December 2007. Econometric analysis shows that monetary policy statements that contain loose or neutral policy inclination tend to lower interbank interest rates, while monetary policy statements with tight policy inclination tend to have no impact on interbank interest rates (asymmetric effect).

A closer research to ours is Tobback et al. (2017), who developed the Hawkish-Dovish (HD) index that measures media's perception of ECB communications. The HD index is computed by using two methods: semantic orientation (SO) and support vector machine (SVM). The HD index based on SO method is computed by counting the co-occurrences of strings with a pre-determined words/expressions that are normally associated with "hawkish" and "dovish" concepts to determine the tone of the document. For the SVM method, instead of using predefined set of keywords, the algorithm automatically looks for patterns in text documents to select the words with the highest discriminative power and determines the tone of a document based on them. Similar Hawkish-Dovish research has also been done earlier by Lucca & Trebbi (2009) for the FOMC statements, although it uses number of search hits, not directly mining (news) text data.

With regard to how we evaluate the policy credibility indexes, there are several research that test the impact of economic and central bank-related news on economic indicators. This is in line with the notion that sentiment affects economic conditions (self-fulfilling), in addition to the other way around that economic conditions affect sentiment (Algaba et al., 2020). For example, research by ter Ellen (2019) analyzes, among others, the impact of narrative monetary policy surprise in news media on macroeconomic indicators, including interest rate, CPI, and consumer confidence. The surprise is defined as the change in news topical coverage around the day of central bank policy announcements. A survey study by Hayo & Neuenkirch (2015) also shows that most financial market participants routinely monitor the media for central bank news, and thus the media may affect their decision-making.

Our evaluation methodology mostly refers the research by Lena (2011) which investigates the role of media on households' inflation perceptions and expectations, and research by Lamla & Maag (2012) which investigates the role of media on disagreement of inflation expectations across households and professional forecasters. Some of the main differences with our research are:

- We are interested in the effect of news on monetary policy credibility (more detail in the next section), while the previous research focus on news on inflation.
- We add two other explained variables: the accuracy of inflation expectations, and the anchoring of inflation expectations to inflation target of the government and Bank Indonesia. (We do not attempt to predict the inflation expectation itself).

- The coding of the whole set of news texts into credibility sentiment is done by machine learning models, after manually annotating several thousand example news sentences. In previous research, from what we understand the coding (of inflation news) was done manually or by a rules-based approach.

### 3. Methodology

#### 3.1. Data

##### 3.1.1. News articles

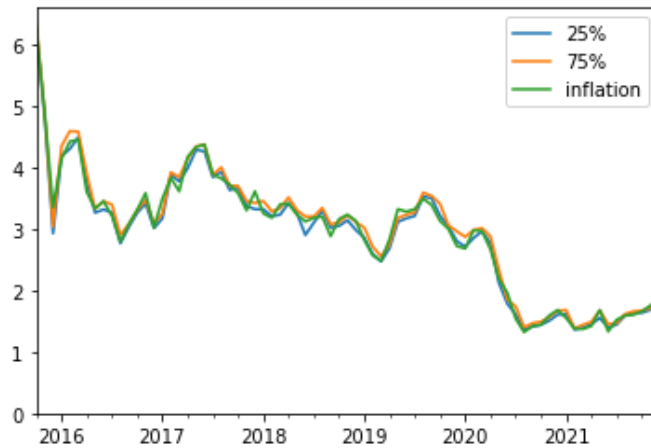
**Source:** News data serves as the main input for constructing the policy credibility indexes. We use news data from Bank Indonesia's Cyber Library, which is a curated internal repository of news articles related to economic and financial topics. There are more than 30 domestic news media, with an average of about 850 articles daily, although the number of news media and news articles can vary from month-to-month. The news data are available on a daily basis since 1999, but the news data that we use in this paper span from October 2015 to December 2021 (we will explain later about our choice of starting point). The news are in Bahasa Indonesia (Indonesian language).

**Filtering:** We filter out news that are not relevant for constructing the index. Specifically, we only keep news *sentences* that contain any of the keywords related to monetary policy: "inflation", "monetary", "exchange rate", "current account", "policy rate", "BI Rate", "BI 7-Day Reverse Repo Rate", and their variations in writing. Furthermore, the sentence or its previous/next sentence must mention "BI" or "Bank Indonesia".

**Additional keywords for communication index:** Starting from January 2020, we add more keywords in order to accommodate non-interest rate-related monetary policy, such as QE, which becomes more relevant since the onset of COVID-19 pandemic. Example keywords are "reserve requirement", "quantitative easing/QE", "accommodative", and "liquidity". We also include keywords related to Bank Indonesia's Joint Decrees with the Ministry of Finance on COVID-19 burden sharing mechanism: "burden sharing", "purchase of SBN (government securities)". These additional keywords are applied only to the communication index, since this index is the one most dependent on the specific monetary policy instruments being taken at the time.

##### 3.1.2. Inflation estimates

As one means to evaluate the policy credibility indexes, we test the indexes in several econometric models of inflation estimates. The estimates are obtained from Bloomberg Economist Estimates, which, every month, asks a number of domestic and foreign economists about their prediction of Indonesia's CPI inflation for the current month. The CPI data itself is usually released by Indonesia National Statistics Office (BPS) on the first week of the next month.



As can be observed from Figure 1, in terms of level, the estimates track actual monthly inflation pretty well, with the monthly mean estimates' having mean absolute deviation of only 9 basis points. ( $|\text{mean}(\text{estimate of inflation}_t) - \text{inflation}_t|$ , averaged across all months in the sample). However, out of the 75 months sample period, there are 56 times or three-fourths where actual inflation is out of the 25-75th percentile of the estimates.

Since we are interested in whether news articles that make up our policy credibility indexes affect inflation estimates and the news are in Indonesian, we limit the sample to only include domestic economists.

Distribution of inflation estimates

Table 2

|                                                                                              | Sample statistic |
|----------------------------------------------------------------------------------------------|------------------|
| <b>Mean number of economists</b>                                                             | 20.6             |
| <b>Mean number of economists, domestic sample</b>                                            | 10.2             |
| <b>Mean of monthly mean<sup>1</sup></b>                                                      | 2.951%           |
| <b>Mean of monthly standard deviation<sup>1</sup></b>                                        | 0.097%           |
| <b>Mean of monthly absolute deviation from actual inflation<sup>1</sup></b>                  | 9 bp             |
| <b>Number of times actual inflation lower than percentile 25 estimate<sup>1</sup></b>        | 33<br>(44%)      |
| <b>Number of times actual inflation higher than percentile 75 estimate<sup>1</sup></b>       | 23<br>(30.3%)    |
| <b>Number of times mean estimate lower than lower bound of inflation target<sup>1</sup></b>  | 21<br>(27.6%)    |
| <b>Number of times mean estimate higher than upper bound of inflation target<sup>1</sup></b> | 1<br>(1.3%)      |

<sup>1</sup> All sample statistics are calculated using the estimates of domestic economists, not the whole sample.

Source: Bloomberg, author's calculation

## 3.2. Policy Credibility Index

### 3.2.1. Annotation

A random sample of the filtered news sentences are manually annotated to construct training data for “teaching” machine learning classification models. Each sentence is labelled with 4 (four) information representing public’s perception on the credibility aspects. The possible labels are positive, negative, or irrelevant, except for communication index, for which the possible labels are accommodative/dovish, neutral, tight/hawkish, or irrelevant.

There are some adjustments with respect to the credibility aspects in Policy Credibility Survey as explained in section 2.1:

1. Independence aspect is merged into coordination aspect, as they both capture Bank Indonesia’s coordination with the government.
2. Accountability aspect is excluded, as it is rarely discussed in news media.
3. Communication aspect is defined as public’s perception/expectation of monetary policy stance. In the survey, it measured whether our policies had been well-communicated. Arguably this original definition is more difficult to learn from news media.

Annotation is done by the authors and subject matter experts on monetary policy communication and central bank credibility within Bank Indonesia. Prior to annotation, we write out the guidelines on how to annotate the news sentences including specific examples, so that the result is more consistent across annotators. Each sentence is annotated by 2-3 annotators to minimize bias.

A total of 12,560 sentences from January 2010 to August 2019 are annotated (first four rows of Table 1). For the additional keywords for communication index about 7,367 sentences are annotated, for the period January 2020 to May 2021. Example annotated sentences are provided in Appendix A.

Distribution of annotated sentences

Table 1

|                                   | Positive <sup>1</sup> | Negative <sup>2</sup> | Neutral        | Irrelevant                 |
|-----------------------------------|-----------------------|-----------------------|----------------|----------------------------|
| <b>Formulation</b>                | 1,595<br>(81.6%)      | 360<br>(18.4%)        | Not Applicable | 10,605<br>(84.5% of total) |
| <b>Effectiveness</b>              | 2,072<br>(81.1%)      | 482<br>(18.9%)        | Not Applicable | 10,006<br>(79.8% of total) |
| <b>Coordination</b>               | 303<br>(86.8%)        | 46<br>(13.2%)         | Not Applicable | 12,211<br>(97.2% of total) |
| <b>Communication</b>              | 493<br>(27.6%)        | 391<br>(33.0%)        | 706<br>(39.4%) | 10,770<br>(14.2% of total) |
| <b>Communication – additional</b> | 3<br>(0.1%)           | 3,358<br>(99.7%)      | 6<br>(0.2%)    | 4,000<br>(54.3% of total)  |

<sup>1</sup> Tight/hawkish for communication index. <sup>2</sup> Accommodative/dovish for communication index.

Sources: Authors’ calculation, from annotation

### 3.2.2. Data preprocessing

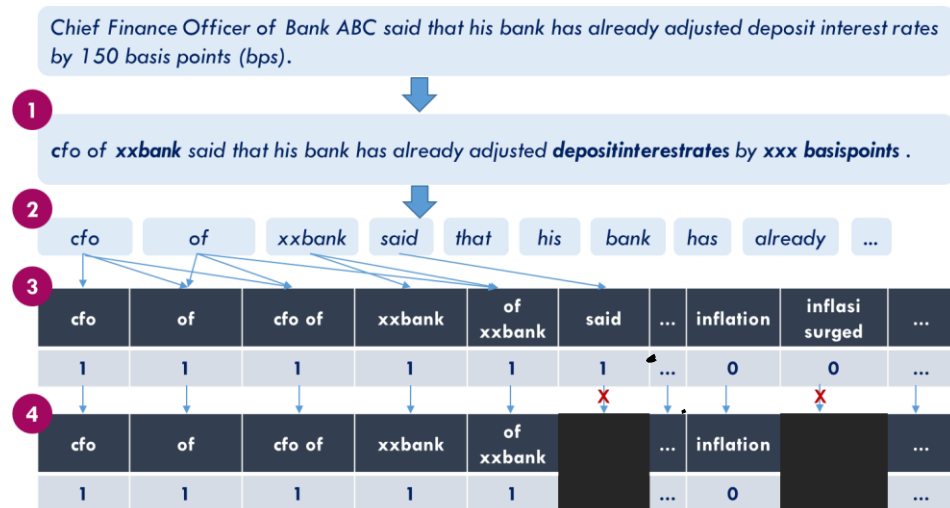
Each filtered sentence (not necessarily annotated) as described in section 3.1.1 is transformed from textual format into tabular-numeric so that it can be processed by machine learning algorithms, following the steps below:

1. Sentence cleansing: lowercasing, replacing synonyms, abbreviations, numbers, and common names in the sentence;
2. Tokenization: splitting sentence into words/tokens;
3. N-gram vectorization: creating n consecutive words (n-gram) as additional features; and
4. Sparse terms removal: removing rarely occurring terms from the feature list.

An example is shown in Figure 2 below.

Example data (sentence) preprocessing steps

Figure 2



### 3.2.3. Model training

From the preprocessed and annotated sentences, we train machine learning models to classify each sentence into one of possible labels (positive/negative/irrelevant, or accommodative/neutral/tight/irrelevant for communication index). We train one model for each aspect, so in total we have 4 sets of models representing the 4 aspects. We experimented with several learning algorithms: logistic regression, naïve bayes, decision tree, random forest, XGBoost, FastText, and deep learning – LSTM (long-short term memory network).

As can be seen from the annotation results (Table 1), the class distributions are quite imbalanced, with most sentences in the news media being irrelevant (even after filtering with keywords). For relevant sentences, the distributions are also imbalanced towards positive class. Considering these imbalances, we carry out the classification in 2 stages for each credibility aspect:

1. Classifying whether the filtered sentence contains perception about central bank credibility (credibility vs. non-credibility i.e. irrelevant); and

2. For relevant sentences, classifying the credibility sentiment in the sentence (positive vs. negative / hawkish vs. neutral vs. dovish).

More details on model training can be referred to in (Zulen, 2020).

**Improvement on model for communication aspect:** As an improvement to the previous procedure, we train a separate model for classifying sentences related to the additional keywords for communication index as described in section 3.1.1. Since most perceptions/expectations on monetary policy in news media since January 2020 is accommodative, which may reflect the monetary policy measures during COVID-19 pandemic, we simplify the classification model, i.e. the model only classifies whether each filtered sentence contains (accommodative) perception of monetary policy or whether it is irrelevant.

### 3.2.4. Text classification and index calculation

Having obtained the classification models for each credibility aspect, we apply the models to the whole (filtered) news data in Cyber Library. For each time period (monthly), we then tabulate the number of sentences classified as positive or negative for each aspect (formulation, effectiveness, and coordination). Besides on historical data, we also calculate the index calculation for ongoing periods, without the need for more manual annotation as the sentence classification models have been trained.

The index for formulation, effectiveness, and coordination aspects is calculated as the net balance of the number of positive and negative sentences in each time period.

$$index_{aspect,t} = \frac{\#positive_{aspect,t} - \#negative_{aspect,t}}{\#positive_{aspect,t} + \#negative_{aspect,t}}$$

For communication aspect, the calculation is different although still similar. First, we compute the stance perception index using net balance of the number of sentences classified into tight, accommodative, and neutral classes as below (note that the number of accommodative sentences include those related to the additional keywords as explained in the previous section). The time period is also different in that for the 3 other indexes, the monthly index are constructed from news at beginning of month (date 1) to end of month. For communication index, the beginning period is 1 day after monthly Board of Governors meeting to 1 day prior to the next monthly Board of Governors meeting.

$$index_{stance\ perception,t} = \frac{\#tight_t - \#accommodative_t}{\#tight_t + \#neutral_t + \#accommodative_t}$$

This index is then compared with the direction of forward guidance contained in Bank Indonesia's press release after each monthly Board of Governor meeting. In the formula below, forward guidance is 1 if the press release contains tight/hawkish stance, 0 if neutral, or -1 if accommodative/dovish. The codification of forward guidance is done manually.

$$index_{communication,t} = 1 - |forward\ guidance_t - index_{stance\ perception,t}|$$

The four indexes are averaged to obtain the (aggregate) policy credibility index, for each month. We refer to (Zulen, 2020) for more explanation on the indexes' characteristics.

### 3.3. Regression Setting

#### 3.3.1. Model and estimation

We test the indexes in a regression setting. The hypothesis is that perceptions towards central bank credibility, as captured in news media, are read by economists and may affect their internal models or beliefs about the economy. We are interested in economists' estimates of inflation, operationalized specifically as 3 variables below:

1. Inaccuracy of estimate: how much the mean estimate deviate from actual inflation, in absolute terms.
2. Unanchoring of estimate: how much the mean estimate deviate from the midpoint of Bank Indonesia and the government's yearly inflation target range, in absolute terms.
3. Heterogeneity of estimate: how much the estimates vary between economists, as coefficient of variation (CV). We use CV as normalization since the level of variation itself may depend on the prevailing level of inflation.

The regression model is as below.

$$y_{k,t} = \beta_0 + \sum_i \beta_i \text{indexes}_{i,t} + \sum_j \beta_j \text{controls}_{j,t} + u_t$$

where  $y_{k,t}$  are calculated as below:

$$y_{\text{inaccuracy},t} = |\hat{\pi}_{t,t-1} - \pi_t|$$

$$y_{\text{unanchoring},t} = |\hat{\pi}_{t,t-1} - \pi_t^*|$$

$$y_{\text{heterogeneity},t} = \frac{sd(\hat{\pi}_{t,t-1})}{|\hat{\pi}_{t,t-1}|}$$

$\hat{\pi}_{t,t-1}$  = an economist's estimate of inflation for month t made at month t-1,  $\pi_t^*$  is the midpoint of inflation target range, and  $sd(\dots)$  is the sample standard deviation function. We are interested in the significance of the  $\beta_i$ 's. Graphs of the explained variables are in Figure 3.

---

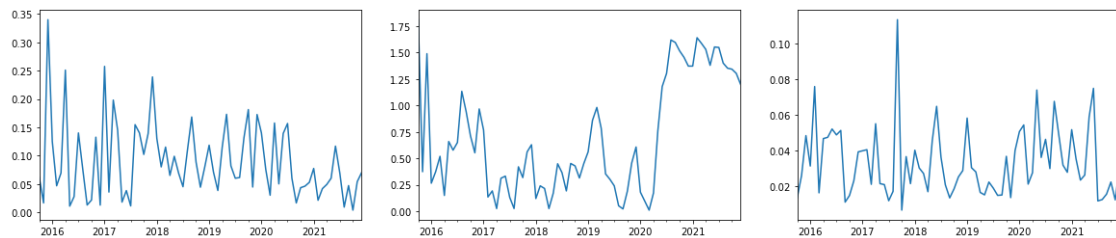
#### Dependent variables series

left: inaccuracy, middle: unanchoring, right: heterogeneity

Figure 3

The structural break in the middle graph coincides with one of the control variables (new COVID cases).

---



We try several variations of the policy credibility indexes as explanatory variables:

- Indexes = 4 component indexes
- Indexes = communication index
- Indexes = average policy credibility index

All policy credibility index variables are contemporaneous, except communication index which is lagged by 1 month due to difference in time period as explained in section 3.2.4.

In total, there are 9 regression models that we test. The models are estimated with OLS, with heteroskedasticity and autocorrelation consistent (HAC) standard errors for hypothesis testing<sup>5</sup>.

### 3.3.2. Control variables

We use several control variables in line with the literature, including lagged dependent variable as in (Lamla & Maag, 2012) and (Lena, 2011). We add USD/IDR (percent change, mtm) as exchange rate is one of the important macro indicators for EM economies. We also add new COVID cases variable, as developments in the number of COVID cases impact mobility and economic activity considerably.

The list of explained and explanatory variables, as well their transformation, lag, and unit of measurement are summarized in Table 3 below. The lag structure assures that the explanatory variables reflect the information set that is available to economists at the time they make their monthly inflation estimates.

Regression variables

Table 3

| Variable                             | Data Source                   | Transformations                                                                     | Lag (months) | Unit    | Mean   | Std. Deviation |
|--------------------------------------|-------------------------------|-------------------------------------------------------------------------------------|--------------|---------|--------|----------------|
| <b>y, inaccuracy<sup>1</sup></b>     | Bloomberg Economist Estimates | absolute value of (monthly mean estimate – actual inflation)                        | -            | %       | 9 bp   | 6.6 bp         |
| <b>y, unanchoring<sup>1</sup></b>    | Bloomberg Economist Estimates | absolute value of (monthly mean estimate – midpoint of inflation target)            | -            | %       | 70 bp  | 54 bp          |
| <b>y, heterogeneity<sup>1</sup></b>  | Bloomberg Economist Estimates | monthly standard deviation of estimates / absolute value of (monthly mean estimate) | -            | %       | 3.4 bp | 1.9 bp         |
| <b>Formulation index<sup>2</sup></b> | Cyber Library + our ML model  | -                                                                                   | -            | decimal | 82.8%  | 12.9%          |
| <b>Effectiveness index</b>           | Cyber Library + our ML model  | -                                                                                   | -            | decimal | 63.8%  | 10.0%          |
| <b>Coordination index</b>            | Cyber Library + our ML model  | -                                                                                   | -            | decimal | 93.4%  | 11.6%          |

<sup>5</sup> R `lm` function for OLS, `coeftest` function in `lmtest` package for single variable hypothesis testing, `waldtest` function in `lmtest` package for joint hypothesis testing, and `vcovHAC` function in `sandwich` package for HAC standard errors.

|                                          |                                  |                                                 |                |                |       |       |
|------------------------------------------|----------------------------------|-------------------------------------------------|----------------|----------------|-------|-------|
| <b>Communication index</b>               | Cyber Library + our ML model     | -                                               | 1              | decimal        | 75.8% | 14.7% |
| <b>Avg credibility index<sup>2</sup></b> | Cyber Library + our ML model     | -                                               | - <sup>3</sup> | decimal        | 79.0% | 7.8%  |
| <b>Lag of y</b>                          | Bloomberg Economist Estimates    | -                                               | 1              | %              | -     | -     |
| <b>BI7DRR (policy rate)</b>              | Bank Indonesia                   | -                                               | -              | %              | 4.78% | 0.86% |
| <b>Inflation</b>                         | Bank Indonesia                   | -                                               | 1              | %              | 3.01% | 1.08% |
| <b>GDP</b>                               | BPS (National Statistics Office) | Growth (yoy) + interpolation to monthly         | 4              | decimal        | 3.72% | 2.92% |
| <b>USD/IDR</b>                           | Yahoo! Finance                   | Growth (mtm), last day close                    | -              | decimal        | 0.00% | 2.75% |
| <b>New COVID cases</b>                   | WHO website                      | Natural log (set to 0 before COVID-19 pandemic) | -              | -(natural log) | 3.30  | 5.22  |

Period: October 2015 to December 2021, monthly (75 observations).

<sup>1</sup> Only including domestic economists.

<sup>2</sup> Mean and std. deviation of formulation and avg. credibility indexes exclude outlying observation of formulation index in May 2017.

<sup>3</sup> Communication index is lagged one month. The other 3 component indexes are not lagged.

Sources: Various, authors' calculation

## 4. Result and Discussion

### 4.1 Index Results

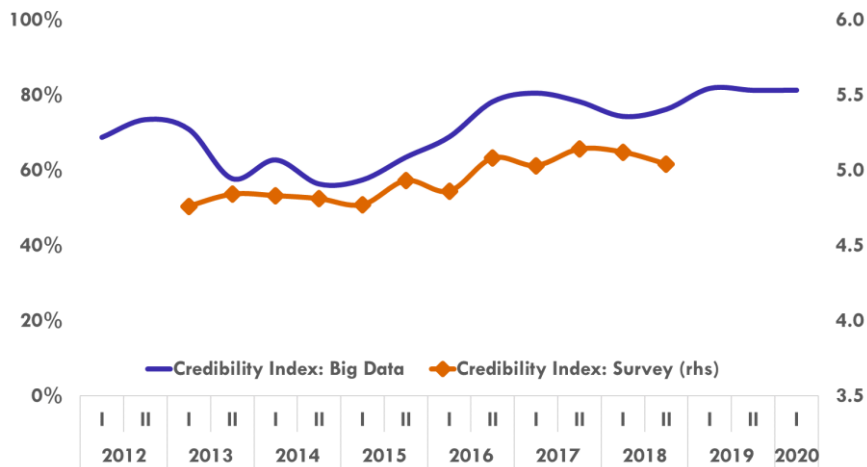
#### 4.1.1 Policy credibility index and the 4 component indexes

For result evaluation, we calculate the correlation between the aggregate policy credibility index generated from news using Big Data Analytics and from our survey (which was decommissioned since 2018). The overall out-of-sample classification accuracy is 63.4% F1 score (averaged across the 4 indexes).

Graph of both indices from is presented in Figure 4. Although both indexes show upward trend, we do not expect them to have such (upward) trend since the indexes should change from time to time based on public's perception towards monetary policy credibility. The indexes have a correlation of 79.7%. The high correlation value indicates that the policy credibility indexes from news have potential to be used as a measure of public's perception on monetary policy credibility.

Aggregate policy credibility index comparison with survey

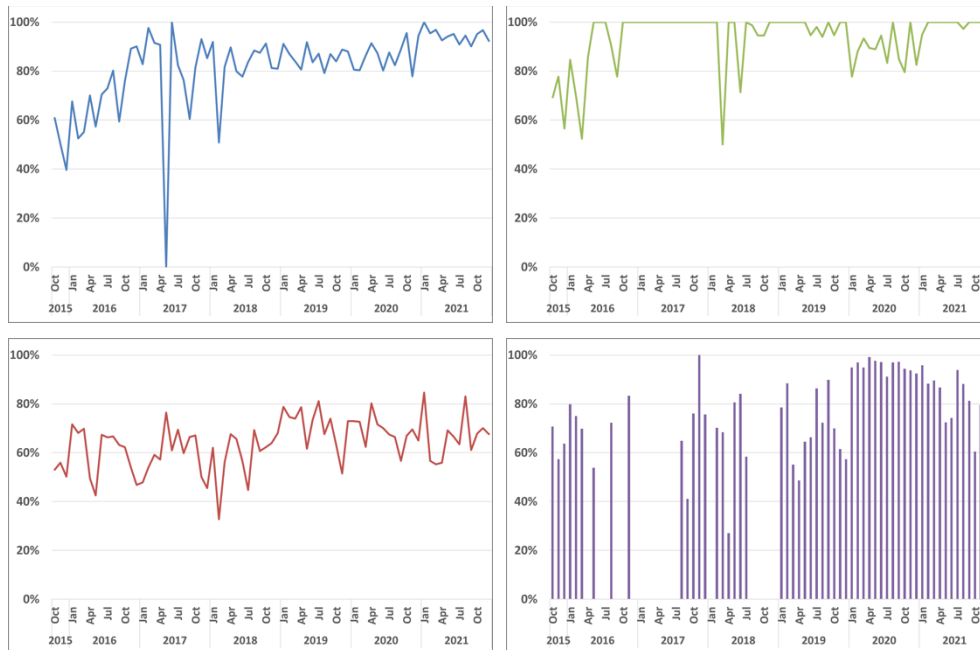
Figure 4



## Policy credibility indexes

top left (blue): formulation, top right (green): coordination,  
bottom left (red): effectiveness, bottom right (purple): communication

Figure 5



Referring to Figure 5 for the component indexes, we comment on some of the indexes' movements:

- All indexes are always positive. The coordination index is the highest, and it often touches the maximum possible value (100%). The effectiveness index is the lowest although still quite positive, with mean 64%.

- There is a stark outlier on formulation index in May 2017 (value: 0%). We double-checked the resulting sentence classifications and find that the machine learning model do correctly classify all sentences during this period.
- Only the formulation index shows upward trend for the whole study period (disregarding the outlying observation). The other indexes do not have clear upward/downward trend.
- In some periods communication index is blank. This is the case if there is no forward guidance in the press release of the previous period's Board of Governors monthly meeting.

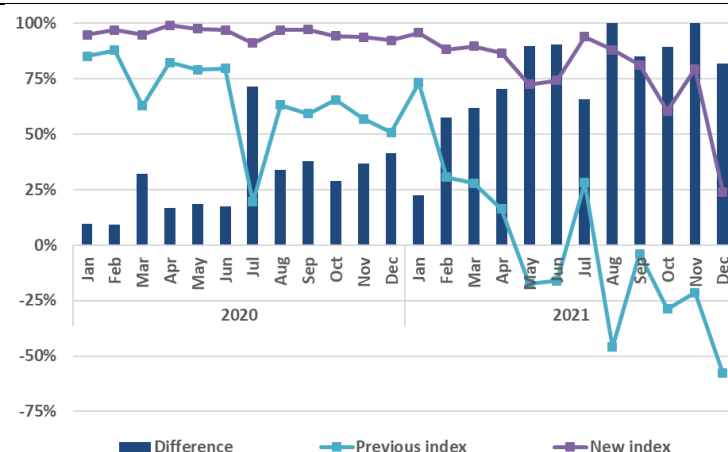
We refer to (Zulen et al., 2020) for more discussion on results of classification accuracy and event analyses of the indexes.

### 4.1.2 Comparison of communication indexes

In sections 3.1.1 and 3.2.4, we described an improvement to the communication index, namely by adding more keywords related to non-interest rate monetary policy and burden sharing arrangements since January 2020 (onset of COVID-19 pandemic).

Comparison of communication indexes

Figure 5



The new communication index is visibly higher than the previous index. The monthly difference is on average 54.3%, and the difference is as high as 134% (on August 2021). On the surface, this means that the new communication index is more aligned with forward guidance in the press releases of Board of Governors monthly meeting. The difference also has an increasing trend (mean difference for 2020 = 29.5%, for 2021 = 79.1%).

These observations show that the new indexes' potential ability to capture public's perceptions and expectations on the broader set of monetary policy, including not only policy rate but also reserve requirement and QE, as well as other coordinative policies with the government. As a note, Bank Indonesia's policy rate was last changed (decreased) in February 2021 to 3.50%. Public in the news media expect that the policy rate is going to be increased in 2022, thus the previous index has been negative since the latter half of 2021.

We have not analyzed these differences further e.g., in an econometric model.

## 4.2 Regression Results

Here we report the results of the regressions as described in section 3.3.

Regression results

Table 4

| Variable                         | $y_{inaccuracy}$  |                     |                     | $y_{unanchoring}$    |                     |                     |
|----------------------------------|-------------------|---------------------|---------------------|----------------------|---------------------|---------------------|
|                                  | 4 indexes         | Comm. index         | Credib. index       | 4 indexes            | Comm. index         | Credib. index       |
| Formulation index (%)            | 0.063<br>(0.053)  | -                   | -                   | -0.446***<br>(0.144) | -                   | -                   |
| Effectiveness index (%)          | -0.087<br>(0.083) | -                   | -                   | -0.864*<br>(0.473)   | -                   | -                   |
| Coordination index (%)           | -0.089<br>(0.078) | -                   | -                   | -0.159<br>(0.337)    | -                   | -                   |
| Communication index (%<br>lag 1) | -0.044<br>(0.077) | -0.026<br>(0.076)   | -                   | -0.190<br>(0.290)    | -0.120<br>(0.282)   | -                   |
| Avg credibility index (%)        | -                 | -                   | -0.068<br>(0.211)   | -                    | -                   | -1.457*<br>(0.780)  |
| Intercept (%)                    | 0.239<br>(0.199)  | 0.1455**<br>(0.068) | 0.188<br>(0.183)    | 1.346<br>(0.888)     | 0.220<br>(0.327)    | 1.453<br>(0.791)    |
| Lag of $y$ (%)                   | -0.113<br>(0.125) | -0.113<br>(0.115)   | -0.118<br>(0.114)   | 0.445***<br>(0.133)  | 0.482***<br>(0.128) | 0.458***<br>(0.121) |
| BI7DRR (policy rate) (%)         | -0.001<br>(0.014) | -0.003<br>(0.0148)  | -0.003<br>(0.015)   | 0.094<br>(0.063)     | 0.056<br>(0.057)    | 0.056<br>(0.064)    |
| Inflation (%<br>lag 1)           | 0.001<br>(0.018)  | 0.004<br>(0.011)    | 0.000<br>(0.017)    | -0.111**<br>(0.055)  | -0.033<br>(0.049)   | -0.097**<br>(0.048) |
| GDP (yoy, decimal, lag 4)        | -0.086<br>(0.266) | -0.176<br>(0.243)   | -0.129<br>(0.26405) | -1.000<br>(0.843)    | -1.941*<br>(1.076)  | -1.472*<br>(0.807)  |
| USD/IDR (mtm, decimal)           | -0.127<br>(0.251) | -0.074<br>(0.275)   | -0.092<br>(0.255)   | -1.335<br>(1.145)    | -0.553<br>(0.928)   | -0.871<br>(0.968)   |
| New COVID cases (ln)             | -0.004<br>(0.004) | -0.004<br>(0.004)   | -0.004<br>(0.003)   | 0.048<br>(0.021)     | 0.039*<br>(0.022)   | 0.043**<br>(0.018)  |
| $R^2$                            | 0.135             | 0.095               | 0.096               | 0.785                | 0.752               | 0.775               |
| Adjusted $R^2$                   | -0.000            | 0.000               | 0.002               | 0.751                | 0.726               | 0.752               |
| Wald test p-value                | 0.088*            | -                   | -                   | 0.013**              | -                   | -                   |

Significant at: \* 10% level, \*\* 5% level, \*\*\* 1% level

From the table above, several observations can be made:

- For  $y = (\text{in})\text{accuracy}$  of estimates:
  - The models and the indexes do not well explain the accuracy of economists' inflation estimates. The highest  $R^2$  is 0.135 when using all 4 indexes as explanatory variables.
  - In terms of significance, the 4 indexes are only jointly significant at 10% significance level.

- None of the control variables are also significant.
- It should be noted that the estimates are summarized as the sample mean across economists. It may be interesting to analyze these effects for individual economists, e.g. in a panel setting, to know whether the indexes affect estimate accuracy for some economists.
- For  $y = (\text{un})\text{anchoring of estimates}$ :
  - In general this variable is more readily explained by the indexes and the control variables, with  $R^2$ 's above 0.75.
  - For the model with 4 indexes, the formulation and effectiveness indexes are statistically significant, with the correct sign (negative). The model with the single aggregate policy credibility index is also significant, albeit only at 10% level. We may view this as indicating that the more positive perceptions on central bank credibility in news media, namely on monetary policy formulation and effectiveness, economists' inflation estimates in the very short run become more anchored to (midpoint of) inflation target.
  - In terms of magnitude, it could be interpreted that an increase of 20 percentage points in formulation index *ceteris paribus* would increase anchoring of inflation estimates by 9 bp on average, and similarly by 17 bp for effectiveness index, or by 29 bp if we take the single policy credibility index model.
  - For effectiveness index, the relevant sentences include news about inflation being within or outside inflation target, in expectation or realization (past inflation). In the case of the sentence discussing past inflation being within or outside inflation target, this information is similar to the lag of the explained variable, which is very statistically significant.
- For  $y = \text{heterogeneity of inflation estimates}$ , the results are reported in Appendix B. In summary, none of the index and control variables are significant for predicting heterogeneity of estimates, with  $R^2$  as low as 0.061.

It should be noted that the results above are sensitive to the outlying observation of formulation index in May 2017. We also run the regressions with formulation index in May 2017 interpolated from April and June 2017. In this setting, we find that we cannot reject the null hypotheses for any of the 3 explained variables (results not included in this paper).

In light of these results and discussions, we side with the conclusion that the 4 indexes and the aggregate policy credibility index have, at best, weak effect on the accuracy of economists' inflation estimates. For the anchoring of inflation estimates, the formulation and effectiveness indexes are statistically significant. Further analysis is warranted since all the results are sensitive to the one outlying observation of formulation index.

## 5. Conclusion & Future Works

### 5.1. Conclusion

We develop a methodology for measuring Bank Indonesia's monetary policy credibility by utilizing news articles data and machine learning-based technique. From the out-of-sample evaluation results, we achieve an average F1-score of 63.4%. The aggregate policy credibility index also moves in-line with the index generated from our survey, with a correlation of 79.7%. The high correlation value indicates that the policy credibility indexes from news have potential to be used as a measure of public's perception on monetary policy credibility.

From the previous research (Zulen, 2020), we improved on the communication index by adding more keywords to capture non-interest rate policy such as QE and reserve requirement, as well as coordinative policies with the government related to COVID-19 pandemic. The new communication index is largely more aligned with forward guidance in the press releases of Board of Governors' monthly meeting.

We test the policy credibility indexes in econometric model of economists' inflation estimates, namely whether the indexes help explain (1) accuracy, (2) anchoring, or (3) heterogeneity of estimates. Initial results show that the indexes are at best weakly significant in explaining accuracy of estimates, but formulation and effectiveness indexes are significant in explaining anchoring of estimates. However, the results are sensitive to the outlying observation of formulation index, and more analyses should be conducted before drawing further conclusions. Heterogeneity of estimates are not explained by the indexes.

### 5.2. Future Works

Some possible research directions include:

- Analysis on other economic estimates

This paper focuses on estimates of inflation in the very short run (next month by domestic economists). Other estimates that can be potentially explained by the policy credibility indexes are: estimates of inflation by the broader coverage of economists (including foreign economists), farther estimate horizon (e.g. end of year or 1 year), estimates of inflation by the general public (in line with the work by Munday & Brookes, 2021), or estimates of other macroeconomic indicators (e.g. GDP).

- More elaborate econometric models

We tested the policy credibility indexes using a simple linear regression model of the relationship between estimate variables and the indexes and control variables, with HAC standard errors. It could be interesting to test the indexes with more elaborate models, e.g. including more lags of the policy credibility index variables, panel data across economists.

- Data source/coverage addition

The current indexes cover news published by Indonesian media. A possible improvement is to add English news for constructing the indexes, which could then be tested against the estimates from foreign economists. In addition, currently the additional keywords are only applied to communication index. The formulation, effectiveness, and coordination indexes are also likely to change if we apply the additional keywords.

## References

- Algaba, A., et al. (2020). Econometrics Meets Sentiment: An Overview of Methodology and Applications. *Journal of Economic Surveys*, Vol. 34, Issue 3, pp. 512-547.
- Blinder, A. S. (2000). Central bank credibility: Why do we care? How do we build it? *The American Economic Review*, 90(5), 1421-1431.
- Cukierman, A., & Meltzer, A. H. (1986). The Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymmetric Information. *Econometrica*, 54(5), 1099-1128.
- Drazen, A., & Masson, P. R. (1994). Credibility of Policies Versus Credibility of Policymakers. *The Quarterly Journal of Economics*, 109(3), 735-754.
- Faust, J., & Svensson, L. E. (2001). Transparency and Credibility: Monetary Policy with Unobservable Goals. *International Economic Review*, 42(2), 369-397.
- Haldane, A., & McMahon, M. (2018). Central Bank Communications and the General Public. *AEA Papers and Proceedings*, 108, pp. 578-583.
- Haldane, A., Macaulay, A., & McMahon, M. (2020). The 3 E's of central bank communication with the public. *Bank of England Staff Working Paper* No. 847.
- Hayo, B., & Neuenkirch, M. (2015). Self-monitoring or reliance on media reporting: How do financial market participants process central bank news? *Journal of Banking & Finance*, vol. 59, pp. 27-37.
- Lucca, D. O., & Trebbi, F. (2009). Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements. *NBER Working Papers* 15367.
- Łyziak, T., & Paloviita, M. (2016). Anchoring of inflation expectations in the euro area: recent evidence based on survey data. *ECB Working Paper Series* No. 1945.
- Munda, T., & Brookes, J. (2021). Mark my words: the transmission of central bank communication to the general public via the print media. *Bank of England Staff Working Paper* No. 944.
- Sahminan. (2008). Effectiveness of Monetary Policy Communication in Indonesia and Thailand. *BIS Working Paper* No. 262.
- Sobel, J. (1985). A Theory of Credibility. *The Review of Economic Studies*, 52(4), 557-573.
- Tobback, E., Nardelli, S., & Martens, D. (2017). Between Hawks and Doves: Measuring Central Bank Communication. *ECB Working Paper Series* No. 2085.
- Zulen, A. A., & Wibisono, O. (2018). Measuring stakeholders' expectation on central bank's policy rate. *Proceedings of the Ninth IFC Conference*, pp. 1507-1534.
- Zulen, A. A., Wibisono, O., Widjanarti, A. (2020). Developing Machine Learning Technique for Measuring Central Bank Credibility. *2020 Data for Policy Conference*.

## Appendix A: Example annotated sentences

Example annotated sentences from news

Table B.1

| No. | Sentence*                                                                                                                                                                                                                                               | Credibility Aspect                              | Label                        |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|------------------------------|
| 1.  | <i>BI also decided to hold the policy rate as there are still external and domestic risks that need to be monitored.</i>                                                                                                                                | <b>Formulation</b>                              | <b>Positive</b>              |
| 2.  | <i>Due to BI's "wrong" exchange rate policy, rupiah (IDR) depreciated by more than 75% since September 2011 or about 15% per year.</i>                                                                                                                  | <b>Formulation</b>                              | <b>Negative</b>              |
| 3.  | <i>The increase in BI's policy rate managed to appreciate rupiah.</i>                                                                                                                                                                                   | <b>Effectiveness</b>                            | <b>Positive</b>              |
| 4.  | <i>Although BI Rate has been lowered, interest rate on bank loans in Indonesia is entirely not competitive.</i>                                                                                                                                         | <b>Effectiveness</b>                            | <b>Negative</b>              |
| 5.  | <i>This policy is in line with the government and Bank Indonesia's aim to manage current account deficit toward more healthy figures.</i>                                                                                                               | <b>Coordination</b>                             | <b>Positive</b>              |
| 6.  | <i>Some regard that there is political pressure on the monetary authority to lower BI Rate, related to the government's ambition to pursue 5.7% economic growth, even though this presumption has been denied by BI Governor in multiple occasions.</i> | <b>Coordination</b>                             | <b>Negative</b>              |
| 7.  | <i>Within 2013 there is still room for BI Rate hike, 50 bps at most.</i>                                                                                                                                                                                | <b>Expectation / Communication</b>              | <b>Tight/ Hawkish</b>        |
| 8.  | <i>BI Rate this year can be held steady at 5.75% as in last year, as long as there is no increase in inflation.</i>                                                                                                                                     | <b>Expectation / Communication</b>              | <b>Neutral</b>               |
| 9.  | <i>If inflation becomes more controlled, then BI Rate will likely be lowered again.</i>                                                                                                                                                                 | <b>Expectation / Communication</b>              | <b>Accommodative/ Dovish</b> |
| 10. | <i>Exchange rate stability policy and quantitative easing will be continued.</i>                                                                                                                                                                        | <b>Expectation / Communication – additional</b> | <b>Accommodative/ Dovish</b> |
| 11. | <i>The government and the central bank has agreed to share the burden through the National Economic Recovery Program (PEN), as a response to COVID-19 pandemic.</i>                                                                                     | <b>Expectation / Communication – additional</b> | <b>Accommodative/ Dovish</b> |

\*) Translated by the authors from Bahasa Indonesia to English

## Appendix B: Regression results for heterogeneity of estimates

| Regression results for $y$ = heterogeneity of estimates |                            |                   | Table B.1         |
|---------------------------------------------------------|----------------------------|-------------------|-------------------|
| Variable                                                | $y_{\text{heterogeneity}}$ |                   |                   |
|                                                         | 4 indexes                  | Comm. index       | Credib. index     |
| Formulation index (%)                                   | -0.017<br>(0.025)          | -                 | -                 |
| Effectiveness index (%)                                 | 0.009<br>(0.025)           | -                 | -                 |
| Coordination index (%)                                  | -0.020<br>(0.022)          | -                 | -                 |
| Communication index (%)                                 | 0.008<br>(0.018)           | 0.009<br>(0.020)  | -                 |
| Avg. credibility index (%)                              | -                          | -                 | -0.032<br>(0.035) |
| Intercept (%)                                           | 0.086<br>(0.051)           | 0.042<br>(0.034)  | 0.079<br>(0.048)  |
| Lag of $y$ (%)                                          | -0.063<br>(0.151)          | -0.010<br>(0.126) | -0.039<br>(0.137) |
| BI7DRR (policy rate) (%)                                | -0.003<br>(0.005)          | -0.001<br>(0.005) | -0.001<br>(0.005) |
| Inflation (%)                                           | -0.002<br>(0.004)          | 0.001<br>(0.003)  | -0.001<br>(0.004) |
| GDP (growth, decimal)                                   | -0.206<br>(0.097)          | -0.215<br>(0.094) | -0.220<br>(0.091) |
| USD/IDR (growth, decimal)                               | 0.025<br>(0.069)           | 0.017<br>(0.070)  | 0.017<br>(0.065)  |
| New COVID cases (ln)                                    | -0.001<br>(0.001)          | -0.001<br>(0.001) | -0.001<br>(0.001) |
| R2                                                      | 0.089                      | 0.061             | 0.066             |
| Adjusted R2                                             | -0.053                     | -0.038            | -0.031            |
| Wald test p-value                                       | 0.752                      | -                 | -                 |

# Machine Learning for Measuring Central Bank Credibility and Communication

Okiriza Wibisono, Muhammad Abdul Jabbar,  
Anggraini Widjanarti, Alvin Andhika Zulen

IFC-Bank of Italy Workshop: Data science in central banking  
14-17 February 2022

The views and results expressed here are those of the authors and do not necessarily represent Bank Indonesia.



Perception on **central bank (policy) credibility** plays a **pivotal role** in building public's trust on the institution and managing expectation.



Need an **objective** measure of central bank credibility. Existing approaches (direct survey, inflation expectation) have drawbacks.



Utilizing **Big Data Analytics** – text mining to gather public perception regarding central bank policy credibility.

This research:

- Constructing monetary policy credibility indexes from news
- Evaluating the indexes in a regression setting



## News articles

**Source:** Cyber Library (internal repository of curated economic and financial news)

~30 domestic news (in Bahasa Indonesia)  
~850 articles daily

Whole period: since Jan 1999  
Training data: Jan 2010 – May 2021  
For regression: Oct 2015 – Dec 2021

Keywords for filtering sentences:

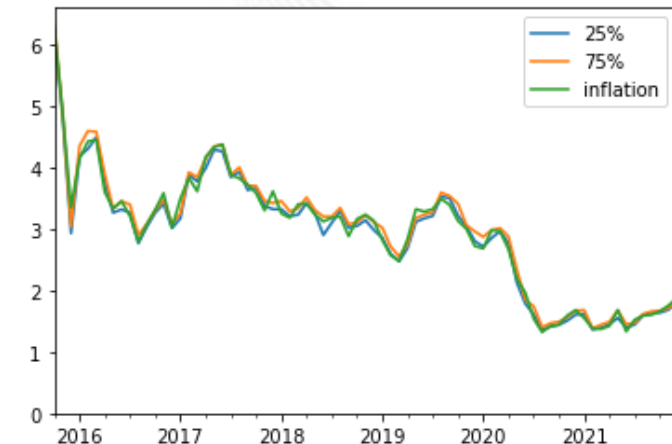
- inflation, monetary, exchange rate, current account, policy rate, BI Rate, BI 77-Day Reverse Repo Rate, (+ their variations)
- Addition for communication idx: QE, reserve requirement, accommodative, liquidity, burden sharing, government bond purchase
- AND the sentence must mention BI/Bank Indonesia (or its previous/next sentence)

## Inflation estimates

**Source:** Bloomberg (Economist Estimates)

Estimate of current month's inflation  
~20 economists per month, ~10 domestic

|                                                                                  |               |
|----------------------------------------------------------------------------------|---------------|
| <b>Mean of monthly mean</b>                                                      | 2.951%        |
| <b>Mean of monthly standard deviation</b>                                        | 0.097%        |
| <b>Mean of monthly absolute deviation from actual inflation</b>                  | 9 bp          |
| <b>Number of times actual inflation lower than percentile 25 estimate</b>        | 33<br>(44%)   |
| <b>Number of times actual inflation higher than percentile 75 estimate</b>       | 23<br>(30.3%) |
| <b>Number of times mean estimate lower than lower bound of inflation target</b>  | 21<br>(27.6%) |
| <b>Number of times mean estimate higher than upper bound of inflation target</b> | 1<br>(1.3%)   |



Sample statistics from domestic economists only

## 1. Annotation

A sample of filtered sentences are annotated as training data for ML classification models.

- Annotated by authors and experts within BI.
- Guidelines incl. examples
- 2-3 annotators per sentence

| Distribution               | Pos*             | Neg*             | Neutral        |
|----------------------------|------------------|------------------|----------------|
| Formulation                | 1,595<br>(81.6%) | 360<br>(18.4%)   | N/A            |
| Effectiveness              | 2,072<br>(81.1%) | 482<br>(18.9%)   | N/A            |
| Coordination               | 303<br>(86.8%)   | 46<br>(13.2%)    | N/A            |
| Communication              | 493<br>(27.6%)   | 391<br>(33.0%)   | 706<br>(39.4%) |
| Communication – additional | 3<br>(0.1%)      | 3,358<br>(99.7%) | 6<br>(0.2%)    |

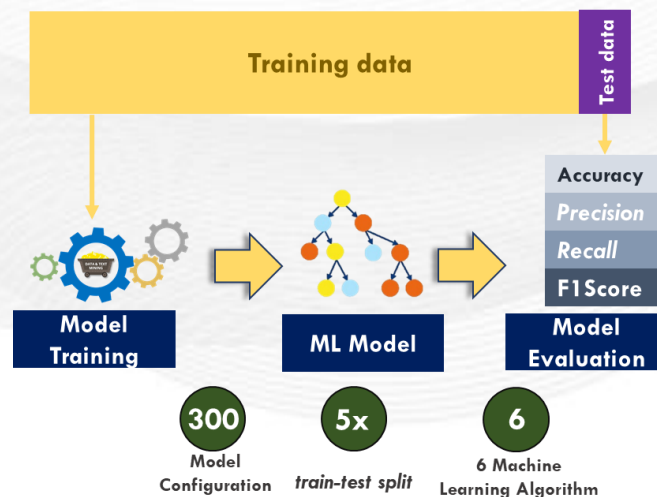
A total of ~20,000 sentences are annotated, majority are irrelevant (neither positive/ negative/ neutral)

\*) For communication index, pos = tight, neg = accommodative.

## 3. Model training

ML model is trained for classifying sentences into pos/neg labels, for each aspect.

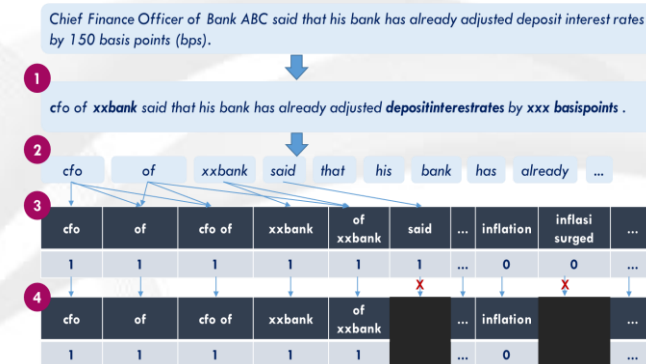
- 90-10 train test split, repeated 5 times
- >300 model configurations
- 6 ML algorithms
- Avg F1 of best model: 63.4%



## 2. Data preprocessing

Each sentence is transformed from text into tabular-numeric format for training ML models.

- Sentence cleansing
- Tokenization
- N-gram vectorization
- Remove sparse terms



## 4. Index calculation

The ML models are applied to all sentences, to construct monthly indexes.

$$Index_{aspect\ k,t} = \frac{\#positive_{k,t} - \#negative_{k,t}}{\#positive_{k,t} + \#negative_{k,t}}$$

$$Index_{stance\ perception,t} = \frac{\#tight_t - \#accommodative_t}{\#tight_t + \#neutral_t + \#accommodative_t}$$

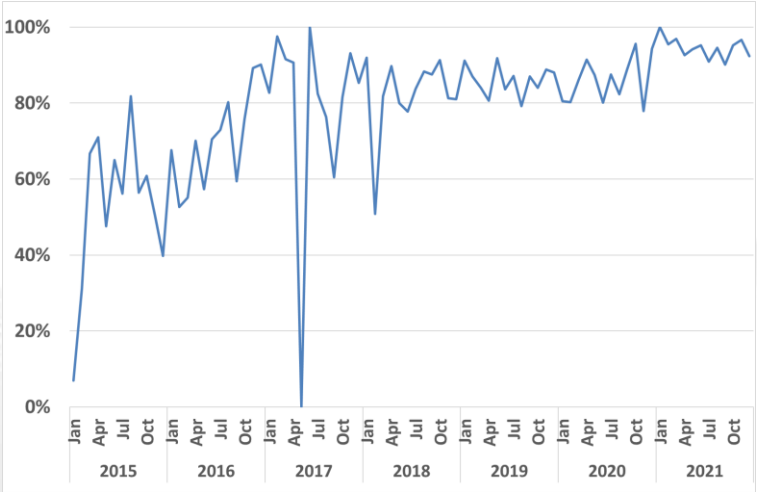
$$Index_{communication,t} = 1 - |Fwd\ Guidance_t - Indeks_{stance\ perception,t}|$$

| No | Sentence                                                                                                                                                                                                                                                | Credibility Aspect                       | Label                  |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|------------------------|
| 1  | <i>BI also decided to hold the policy rate as there are still external and domestic risks that need to be monitored.</i>                                                                                                                                | Formulation                              | Positive               |
| 2  | <i>Due to BI's "wrong" exchange rate policy, rupiah (IDR) depreciated by more than 75% since September 2011 or about 15% per year.</i>                                                                                                                  | Formulation                              | Negative               |
| 3  | <i>The increase in BI's policy rate managed to appreciate rupiah.</i>                                                                                                                                                                                   | Effectiveness                            | Positive               |
| 4  | <i>Although BI Rate has been lowered, interest rate on bank loans in Indonesia is entirely not competitive.</i>                                                                                                                                         | Effectiveness                            | Negative               |
| 5  | <i>This policy is in line with the government and Bank Indonesia's aim to manage current account deficit toward more healthy figures.</i>                                                                                                               | Coordination                             | Positive               |
| 6  | <i>Some regard that there is political pressure on the monetary authority to lower BI Rate, related to the government's ambition to pursue 5.7% economic growth, even though this presumption has been denied by BI Governor in multiple occasions.</i> | Coordination                             | Negative               |
| 7  | <i>Within 2013 there is still room for BI Rate hike, 50 bps at most.</i>                                                                                                                                                                                | Expectation / Communication              | Tight/Hawkish          |
| 8  | <i>BI Rate this year can be held steady at 5.75% as in last year, as long as there is no increase in inflation.</i>                                                                                                                                     | Expectation / Communication              | Neutral                |
| 9  | <i>If inflation becomes more controlled, then BI Rate will likely be lowered again.</i>                                                                                                                                                                 | Expectation / Communication              | Accommodative / Dovish |
| 10 | <i>Exchange rate stability policy and quantitative easing will be continued.</i>                                                                                                                                                                        | Expectation / Communication – additional | Accommodative / Dovish |
| 11 | <i>The government and the central bank has agreed to share the burden through the National Economic Recovery Program (PEN), as a response to COVID-19 pandemic.</i>                                                                                     | Expectation / Communication – additional | Accommodative / Dovish |

Formulation

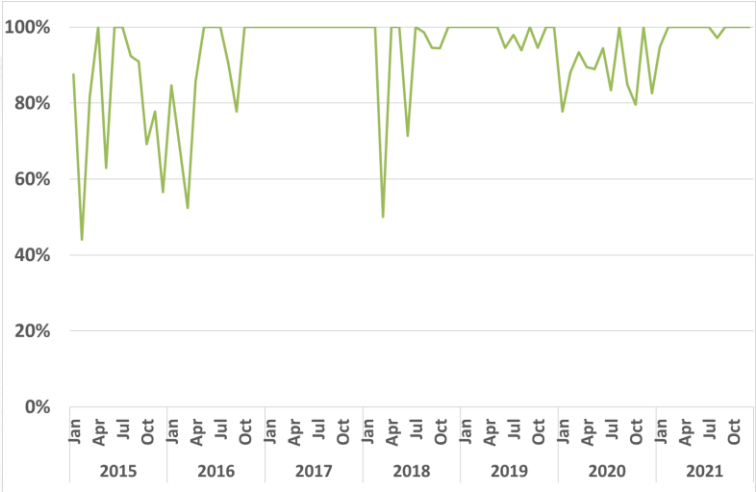
Correlation of avg. policy credibility index with survey results (2013-2018): 79.7%

Coordination



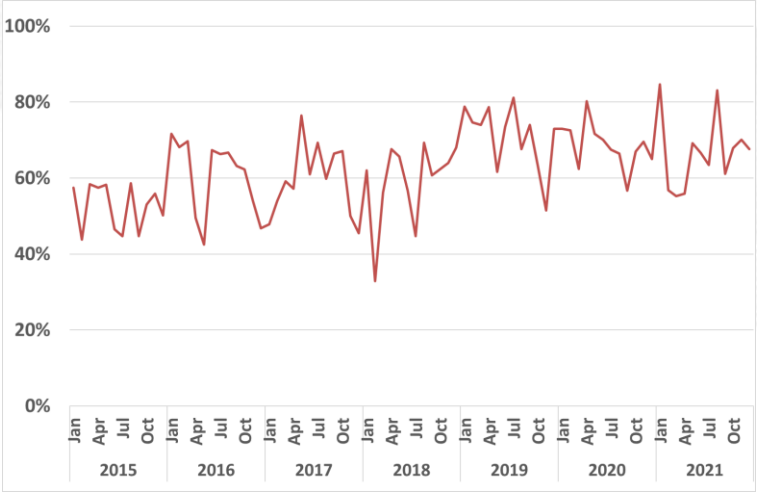
|       |        |
|-------|--------|
| Min   | 0.0%   |
| Mean* | 82.8%  |
| Max   | 100.0% |
| Std*  | 12.9%  |

\*) Excluding outlying observation in May 2017



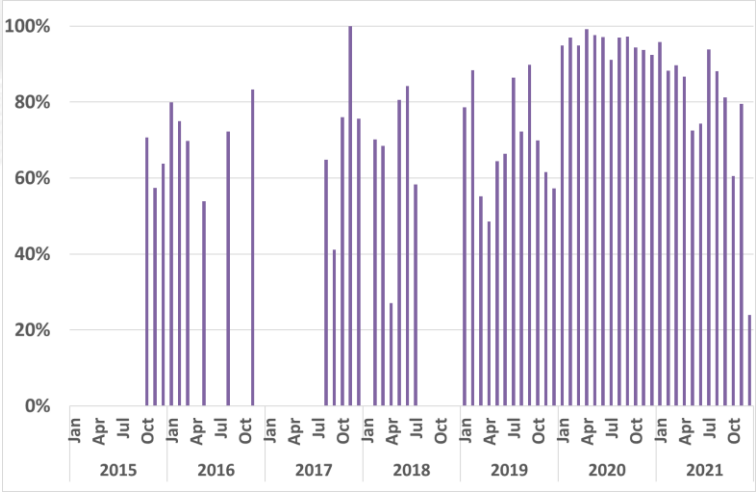
|      |        |
|------|--------|
| Min  | 50.0%  |
| Mean | 93.4%  |
| Max  | 100.0% |
| Std  | 11.6%  |

Effectiveness



|      |       |
|------|-------|
| Min  | 32.8% |
| Mean | 63.8% |
| Max  | 84.7% |
| Std  | 10.0% |

Communication



|      |        |
|------|--------|
| Min  | 27.0%  |
| Mean | 75.8%  |
| Max  | 100.0% |
| Std  | 14.7%  |

We test whether news content, as summarized by the four indexes and the average policy credibility index, are incorporated into economists' estimates of inflation in that they make the estimates more accurate and/or anchored\*.

$$y_{inaccuracy,t} = |\bar{\pi}_{t,t-1} - \pi_t|$$

$$y_{unanchoring,t} = |\bar{\pi}_{t,t-1} - \pi_t^*|$$

- Period: Oct 2015 to Dec 2021, monthly
- Variables of interest (3 sets of equations):
  - a. 4 policy credibility indexes
  - b. communication index (not shown\*)
  - c. average policy credibility index
- Controls: lag of y, BI7DRR, lag inflation, GDP growth yoy (interpolated), USD/IDR % change mtm, ln of new COVID cases
- Communication index lagged by 1 month to allow its distinctive time period (from one BoG monthly meeting to the next, instead of from date 1 to end of month)
- OLS + HAC standard errors
- Result sensitive to outlying observation of formulation index in May 2017

| Variable/Statistic                   | Y inaccuracy      |                   | Y unanchoring               |                            |
|--------------------------------------|-------------------|-------------------|-----------------------------|----------------------------|
| Formulation index                    | 0.063<br>(0.053)  | -                 | <b>-0.446***</b><br>(0.144) | -                          |
| Effectiveness index                  | -0.087<br>(0.083) | -                 | <b>-0.864*</b><br>(0.473)   | -                          |
| Coordination index                   | -0.089<br>(0.078) | -                 | -0.159<br>(0.337)           | -                          |
| lag(Communication index)             | -0.044<br>(0.077) | -                 | -0.190<br>(0.290)           | -                          |
| Avg. credibility index               | -                 | -0.068<br>(0.211) | -                           | <b>-1.457*</b><br>(0.780)  |
| lag(Y)                               | -0.113<br>(0.125) | -0.118<br>(0.114) | <b>0.445***</b><br>(0.133)  | <b>0.458***</b><br>(0.121) |
| lag(Inflation)                       | 0.001<br>(0.018)  | 0.000<br>(0.017)  | <b>-0.111**</b><br>(0.055)  | <b>-0.097**</b><br>(0.048) |
| R <sup>2</sup> , Adj. R <sup>2</sup> | 0.135, -0.000     | 0.096, 0.002      | 0.785, 0.751                | 0.775, 0.752               |
| F-statistic (Wald test)              | <b>0.088*</b>     | -                 | <b>0.013**</b>              | -                          |

\*) Results from only including communication index are all not statistically significant, as well as results on y = heterogeneity of estimates (as coefficient of variation).

## Conclusion

- 1 **Machine learning methodology for measuring perceptions of monetary policy credibility from news**
- 2 **Communication index includes recent policy developments, e.g. non-interest rate policy (QE, reserve requirement), coordinative policies with the government**
- 3 **Evaluation of the policy credibility indexes in econometric model of economists' inflation estimates. Results sensitive to outlier of formulation index**

## Future Works

- 1 **Analysis on effect on other economic estimates, e.g. further horizon, GDP, general public**
- 2 **More elaborate econometric models for evaluation, e.g. panel of economists**
- 3 **Data source/coverage addition, e.g. English news, additional keywords. Macroprudential and payment system policies?**