iSent: Itaú's Central Bank sentiment classifier

- This report presents the iSent, the Itaú's Central Bank sentiment index for Brazil and Chile, a sentiment classifier based on GPT-4, developed by our data science team using sentences published in central bank's official documents labeled by our economists.
- iSent-BCB has a good adherence to current and future interest rate shifts in Brazil (correlation around 0.8). The index identifies a hawkish shift in BCB's communication at the margin, which would be consistent with policy firming, contrary to our current expectation.
- For Chile, the iSent-BCCh also adheres to current and future policy moves (correlation of 0.7). The index now shows a shift to a neutral tone, suggesting that the BCCh is likely to pause in the near term, broadly consistent with our call.
- We also performed a simple back test using our iSent-BCB as a trading signal for the Brazilian DI market and found that the index has a positive information ratio on the sample as a whole.

Context

Central bank (CB) communications are crucial for guiding expectations (especially in inflation targeting frameworks), managing the economy and implementing monetary policy. A vast literature has explored the use of natural language processing (NLP) to analyze central bank communications. In fact, good NLP models have the potential to solve several problems faced by central bank watchers – they could mitigate human biases, automate classification of documents written in various languages and time zones, and create a fast and historically comparable assessment of intricate messages.

The first versions of capturing sentiment in central bank documents were based on algorithms that extract features and sentiments from texts with 'bag of words' (BoW) representation (for example Correa at al., 2017). BoW algorithms count how many times a word (or expression) appears in a document. For sentiment analysis tasks, the algorithm typically uses a dictionary that associates each word/expression with a class of sentiment, counts the number of instances of each class and estimates the sentiment in the document. They are simple and computationally inexpensive, but they have some caveats: 1) they ignore context, i.e., fail to take into consideration the order of sequence or dependency between words; and 2) they can't deal with new words/expressions.

Recent studies use deep-learning-based NLP, specifically large language models (LLMs), which consider contextual nuances, bringing a great performance improvement. In fact, in 2022 Itaú launched its algorithm for reading Brazil's central bank meeting minutes using the RoBERTA model (an optimized version of Google's BERT model) with good adherence to the interest rate scenario (see full report here). Optimized versions of these models use LLMs associated with a task-specific dictionary (for example, Xia, 2021), or manually labeled sentences (Pfeifer and Marohl, 2023) to improve accuracy.

For us, language and EM-specific context were barriers. We faced some difficulties to incorporate the advancements of the NLP literature in the last couple of years in the task of reading sentiment from official documents of Latin American countries and other EMs. First, the top-notch models for financial texts (like FinBERT) and dictionaries focused on central bank communications were trained in English – and we wanted to build a framework that was flexible on language. And even more importantly, the background of these pre-trained models and dictionaries in English is the reality of developed economies, where (at least until the pandemic era) inflation is lower and less volatile than in emerging economies, and fiscal policy, exchange rates and the external scenario tend to be a less important part in monetary policy discussions.

With the emergence of GPT and its ability to chat in different languages, it became easier to read the sentiment of LatAm central banks, without having to retrain models in the local language. To improve accuracy, we built our own labeled dataset with our economists with expertise in tracking Brazil and Chile's monetary policy classifying a sample of sentences of statements and minutes of monetary policy meetings of the respective Central Banks. The economists were not informed of the date the documents referred to. Our labeled dataset is composed by around 1k sentences of official documents of the Brazilian Central Bank, and another ~1k sentences of official documents of the Central Bank of Chile. The teams classified each sentence from the selected documents as dovish, neutral, hawkish, or out-of-context, helping our model to classify the sentences.

Methodology

Our classification unit is a sentence. A sentiment class (ranging from dovish, neutral, hawkish, or out-of-context) is attributed to each sentence, and the index is constructed on the relative presence of each class.

The core of our strategy consisted in building a GPT prompt that could do a good classification, while trying to minimize data contamination. We thus built a general persona described as a central bank watcher and basic scenarios that should be viewed as hawkish (high inflation, strong activity, etc.) and dovish (low inflation, weak activity, etc.).

To classify sentences using GPT-4, we employed an approach inspired by Retrieval-Augmented Generation (RAG) concepts (Lewis, Patrick, *et al.* 2020). We constructed a specialized dataset using the FAISS (Johnson, Douze and Jégou, 2017) indexing system, containing our training dataset. In this dataset the sentences are represented in vectors, and similar vectors are defined as those that are nearby in the Euclidean space. For classifying new sentences, we utilized the similarity search capabilities of FAISS to extract the most relevant pre-labeled sentences from our dataset. These retrieved examples are systematically integrated into the input prompt for GPT-4, thereby enhancing the contextual grounding of the model prior to classification. This method enhances the power of contextual understanding, allowing the model to more accurately infer the classification based on a targeted set of example sentences that are related to the sentence in question. [1,2]

Figure 1 shows the structure of the prompt.

Prompt Structure

1	Persona and Task Description
2	Examples
3	Output Format

The index is calculated following the formula below, which is always between -1 and 1, and is larger when the perceived tone is more hawkish.

[1] Fine-Tuning (FT) is also a common method where a pre-trained model's weights are updated by training on a task-specific supervised dataset, typically involving a large set (thousands to hundreds of thousands) of labeled examples. The main advantage is strong performance on many benchmarks, but disadvantages include the need for large datasets and the potential for poor generalization out of distribution. See Brown et al. 2020 for a discussion.

[2] If the sentence to be classified appears in the labeled set, it is not used as a labeled example – we pick other different (yet similar) sentences. With that, we avoid that the prompt corpus might contain ground truth information of the evaluation data. Still, we are aware that pre-trained LLM models have data contamination issues, and our model is not different. For an interesting discussion, see Jiang at al. (2024).

$$iSent_{hawk/dove} = \frac{\#\, sentences\ hawk - \#\, sentences\ dove}{\#sentences\ hawk + \#sentences\ dove + \#sentences\ neutral}$$

Model Performance vs. Policy Decision

Both iSent-BCB and iSent-BCCh have good adherence to monetary policy shifts. In fact, we tested the ability of our sentiment classifiers in the task of anticipating monetary policy shifts, finding a correlation of ~0.8 for Brazil and ~0.7 for Chile, as shown in the tables below.

Table 1. Correlation Matrix – Brazil

	Δ Selic rate, t	Δ Selic rate, t+1
iSent BCB	0.79	0.77
Δ Selic rate, t	1	0.92
Δ Selic rate, t+1	0.92	1

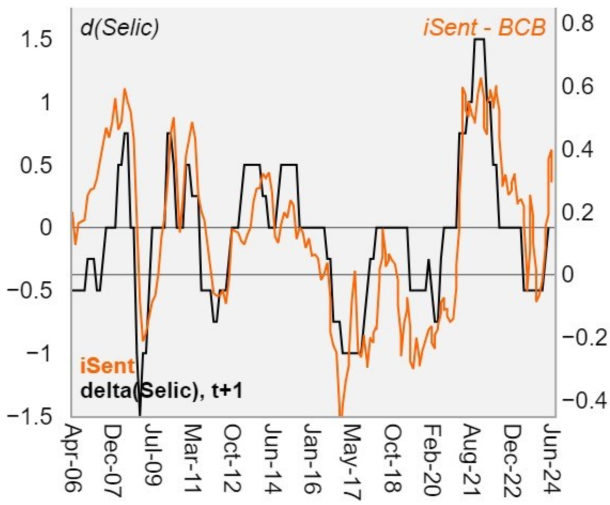
Table 2. Correlation Matrix – Chile

	Δ TPM, t	Δ TPM, t+1
iSent BBCh	0.71	0.68
Δ TPM, t	1	0.83
Δ TPM, t+1	0.83	1

For Brazil, the index suggests that the recent communication is consistent with imminent policy firming. A visual analysis confirms a good adherence of the index and the Selic rate shift one meeting ahead. In fact, the index did well in capturing most of the shifts in the past 18 years, most notably the hiking cycles in the late 2000's and early 2020's. Most recently, the sentiment captured in the latest meeting documents is consistent with tightening in the next Copom meeting, even though this is not our call.

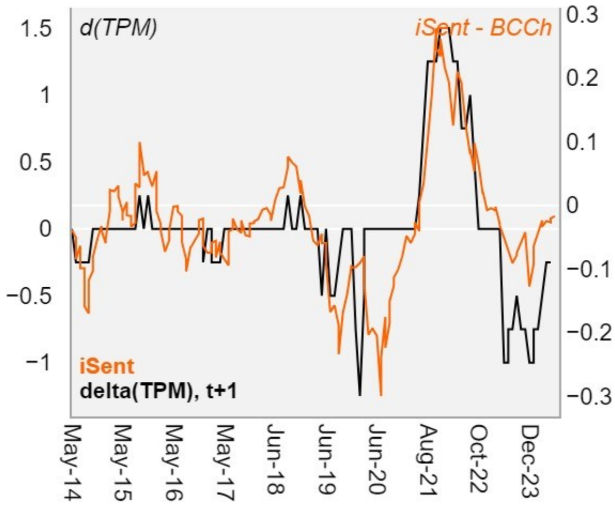
For Chile, the index suggests that recent communication signals a neutral shift. Upon visual inspection, the iSent index broadly anticipates the direction of the BCCh's policy decisions, albeit in a period in which the budget of policy cycles was relatively contained. Naturally, the pandemic period poses challenges for the index as the policy rate was at the effective lower bound. The index effectively matches the BCCh's 2021-2022 hiking cycle and the post-July 2023 easing cycle. The index now signals a neutral tone, suggesting that the BCCh is likely to pause in the near term, broadly consistent with our call.

iSent – Brazil



Source: BCB, Itaú

iSent – Chile



Source: BCCh, Itaú

ISent as a Trading Signal for Brazilian Rates

As we have seen, variations in our sentiment indicator have a strong historical correlation with future monetary policy decisions. Nonetheless, the next obvious question one should ask is if the sentiment indicator extracted by the model has predictive power for the future rates market. In the case of Brazil, COPOM statements and minutes are always released to the public outside market hours. Hence, market participants theoretically have the same amount of information and time of processing it before the next trading session starts. On the other hand, people have cognitive and behavior biases and market participants are no different. One key advantage of statistical and mathematical models is precisely to remove human biases and draw direct conclusions from data.

In order to share some light on this subject, we performed a simple back test using our iSent-BCB as a trading signal for Brazilian DI market. We first take the expanding average of the indicator (using at least 8 documents) as a way to remove a possible level bias. Then, starting in June 2006, we calculate the 2-document moving average of the indicator (including minutes and statements when available) to smoothen the trading signal, since the document-by-document classification can be noisy. Next, the trading signal is as simple as it could be: when this demeaned indicator is above (below) zero, we interpret it as a hawkish (dovish) signal. We initially tested two strategies, one trading the 1-year equivalent DI contract and another trading the slope between the 2-year and 5-year DI[3]. Hawkish (dovish) signals are translated in a paying (receiving) position on the 1-year and flattening (steepening) the curve. Each period the portfolio is rebalanced to target a R$1 million daily ex-ante VaR[4]. We assume, naturally, that the signal calculated after the release of the last document can only be available at the next trading session, and trades are always assumed to be executed at market close[5]. Standard brokerage costs are accounted for.

Both simple strategies have positive information ratio on the whole sample, although with significant drawdowns. The largest drawdown for the 1-year strategy happened during 2012-2015 years, when following the signal perceived by our indicator alone didn't result in profitable strategies. The last easing cycle is also a notable drawdown for this strategy, but the performance of the slope strategy was satisfactory. We note that even though both strategies use the same trading signal, the historical correlation between the two stands only at around 0.2. Hence, trading a combined portfolio of both strategies outperforms both.

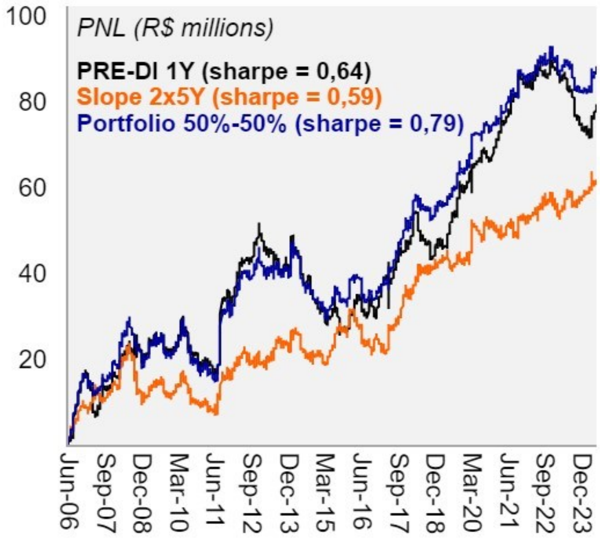
One should always remember that past performance is by no means a guarantee of future performance. Besides, some documents classified were used by our analysts to reinforce GPT model, that by itself is also a model trained on countless documents around the world, including possibly past central bank communication. Even so, we believe our back testing exercise can provide intuitive insights and strengthens the argument that favors increasing the usage of quantitative tools to enhance economics and market analysis.

[3] We build the equivalent 1-year and the slope by historically rolling DI contracts, linearly interpolating expiration dates to keep the duration approximately constant. Due to liquidity issues, we only trade the January and June expirations for the 1-year equivalent, and only January expirations for the longer durations.

[4] We use the realized ewma vol of the equivalent contracts.

[5] Trades are exactly at the market close of the next trading session after the release of the document. Hence, if it is a minute (available at early morning) the strategy trades at the same day of the release. If it is a statement (released after market close) the strategy rebalances at the next trading day. We assume exactly market close prices, hence there is no slippage.

Back test results



Source: Itaú

Priscilla Burity
Dávila Patrícia Ferreira Cruz
Marcel Chamarelli Gutierrez
Andrés Perez Morales

References

Correa, Ricardo, et al. "Sentiment in Central Banks' Financial Stability Reports." Review of Finance, vol. 25, no. 1, 2021, pp. 85-120. Oxford University Press, https://academic.oup.com/rof/article/25/1/85/5824804.

Itaú Unibanco. "Introduzindo um algoritmo para leitura de atas de bancos centrais: aplicação para o Brasil indica fim do ciclo." 23 Aug. 2022, https://www.itau.com.br/content/dam/ibba/analises-economicas/pdfs/macro-visao/23082022_MACRO_VISAO_NPL-BCB.pdf.

Xia, Ruizhe. "An Automated Approach On Generating Macro-Economic Sentiment Index Through Central Bank Documents." Imperial College, 2023, https://www.imperial.ac.uk/media/imperial-college/faculty-of-natural-sciences/department-of-mathematics/math-finance/Xia_Ruizhe_01201752.pdf.

Pfeifer, Marohl. "CentralBankRoBERTa: A Fine-tuned Large Language Model for Central Bank Communications." ScienceDirect, 2023, https://www.sciencedirect.com/science/article/pii/S2405918823000302#dfig1.

Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." 2020, https://arxiv.org/abs/2005.11401.

Johnson, Jeff, Matthijs Douze, and Hervé Jégou. "Billion-scale Similarity Search with GPUs." 2017, https://arxiv.org/abs/1702.08734.

Brown, Tom, et al. "Language Models are Few-Shot Learners." 2020, https://arxiv.org/abs/2005.14165.

Lee, Kenton, Jason Lee, Jing Zhao, and Zhouhan Zhao. "Investigating Data Contamination for Pre-training Language Models." 2024, https://arxiv.org/pdf/2401.06059.

Macro Research – Itaú

Mario Mesquita – Chief Economist

To access our reports and forecast visit our website:
https://www.itau.com.br/itaubba-pt/macroeconomic-analysis

Relevant Information

1. This report has been prepared and released by the Macro Research Department of Itaú Unibanco S.A. ("Itaú Unibanco"). This report is not a product of the Equity Research Department of Itaú Unibanco or Itaú Corretora de Valores S.A. and shall not be construed as a research report ("relatório de análise") for the purposes of Article 1 of the CVM Instruction NR. 20, dated 2021.
2. The exclusive purpose of this report is to provide macroeconomics information and it does not constitute and shall not be construed as an offer to buy or sell or a solicitation of an offer to buy or sell any financial product, or to participate in any particular trading strategy in any jurisdiction. The information herein is believed to be reliable as of the date on which this report was released and it has been obtained from public sources believed to be reliable. However, Itaú Unibanco does not make any explicit or implied representation or warranty as to the completeness, reliability or accuracy of such information, nor does this report intend to be a complete statement or summary of the markets or developments referred to herein. Itaú Unibanco has no obligation whatsoever to update, modify or amend this report and inform the reader accordingly.
3. The opinions contained herein reflect exclusively the personal views of the analyst responsible for this report and were prepared independently and autonomously, including in relation to Itaú Unibanco, Itaú Corretora de Valores S.A. and any other companies within their economic group.
4. This report may not be reproduced or redistributed to any other person, in whole or in part, for any purpose, without the prior written consent of Itaú Unibanco. Additional information on the financial products mentioned in this report may be available upon request. Itaú Unibanco and/or any other company within its economic group is not and shall not be liable for any investment decisions (or otherwise) based on the information provided herein.

Additional Note: This material does not take into consideration the objectives, financial situation or specific needs of any particular client. Clients must obtain financial, tax, legal, accounting, economic, credit and market advice on an individual basis, based on their personal characteristics and objectives, prior to making any decision based on the information contained herein. By accessing the material, you represent and confirm that you understand the risks related to the financial instruments described in this material and the laws in your jurisdiction relating to the provision and sale of financial service products. You acknowledge that this material contains proprietary information and you agree to keep this information confidential for your exclusive use.

For inquiries, suggestions, complaints, criticisms and compliments, talk to Itaú's CSCC: 0800 728 0728. Or contact us through our portal https://www.itau.com.br/atenda-itau/para-voce/. If you are not satisfied with the proposed solution, please contact the Itaú Corporate Ombudsman: 0800 570 0011 (on weekdays from 9 AM to 6 PM) or our PO Box 67.600, São Paulo-SP, Zip Code 03162-971. Hearing impaired, every day, 24h, 0800 722 1722.