BORRADORES DE ECONOMÍA

# Tweeting Inflation: Real-Time measures of Inflation Perception in Colombia.

By:
Jonathan Alexander Muñoz-Martínez
David Orozco
Mario A. Ramos-Veloza

No. 1256
2023


BANCO DE LA REPUBLICA
COLOMBIA

# Tweeting Inflation: Real-Time measures of Inflation Perception in Colombia.

Jonathan Alexander Muñoz-Martínez
*jmunozma@banrep.gov.co*

David Orozco
*dmorozco@icesi.edu.co*

Mario A. Ramos-Veloza*
*mramosve@banrep.gov.co*



**Abstract**

This study follows a novel approach proposed by Angelico et al. (2022) using Twitter to measure inflation perception in Colombia in real time. By applying machine learning techniques, we implement two real-time indicators of inflation perception and show that both exhibit a dynamic similar to inflation and inflation expectations for the sample period January 2015 to March 2023. Our interpretation of these results suggests that our indicators are closely linked to the underlying factors that drive inflation perception. Overall, this approach provides a valuable instrument for gauging public sentiment towards inflation and complements the traditional inflation expectations measures used in the inflation–targeting framework.

**JEL classification**: E31, E37, E52

***Keywords***— Inflation perceptions, Twitter, Real-time data, Central banks.

*The authors are members of the Office for Monetary Policy and Economic Information at Banco de la República de Colombia, Central Bank of Colombia. The authors gratefully acknowledge the valuable comments of Ana Maria Iregui, Franz Hamann, and the participants at the XX research workshop of Banco de la República to previous versions of the document. Any remaining errors are the exclusive responsibility of the authors.

# Tuiteando la Inflación: Medidas en Tiempo Real de la Percepción de la Inflación en Colombia



**Resumen**

Este estudio sigue un enfoque novedoso propuesto por Angelico et al. (2022) para la medición en tiempo real de la percepción de la inflación en Colombia utilizando Twitter. Mediante la aplicación de técnicas de aprendizaje automático, calculamos dos indicadores en tiempo real de la percepción de la inflación y mostramos que exhiben una dinámica comparable a la inflación y las expectativas de inflación, lo que sugiere que nuestros indicadores están estrechamente relacionados con los factores subyacentes que impulsan la percepción de la inflación entre enero de 2015 y marzo de 2023. En general, este enfoque proporciona un medio valioso para evaluar el sentimiento público hacia la inflación y ofrece una perspectiva complementaria a las medidas de expectativas de inflación tradicionales utilizadas en el marco de la política de inflación objetivo.

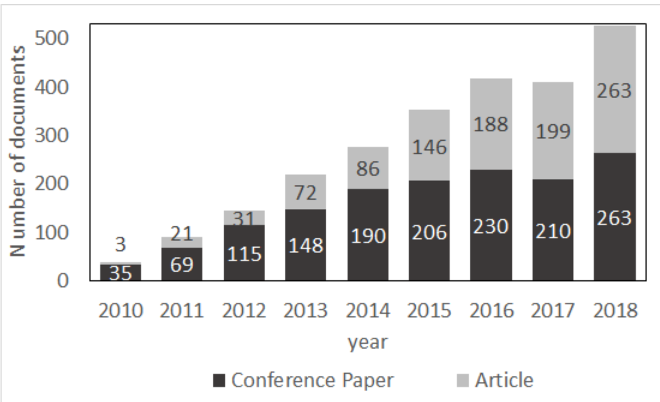
**Clasificación JEL**: E31, E37, E52

***Palabras clave***: Percepción de inflación, Twitter, medición en tiempo real, Bancos centrales.

# 1 Introduction

Inflation expectations are a critical component of the inflation targeting framework adopted by many central banks. Policymakers regularly monitor multiple measures of inflation expectations derived from economic surveys and financial market instruments to understand inflationary pressures in the short-run. As proposed by Angelico et al. (2022) and Bricongne et al., 2022, information on social media networks can be obtained in real time and can complement measures of inflation expectations by gauging collective perceptions of inflation. Carefully selected information in social media may complement inflation expectations measures available to policymakers at central banks by providing a measure of the public's inflation perceptions.

Twitter users primarily share their daily activities or seek information, making this social network a natural source of information on beliefs about politics, economics, and current events. Active users enjoy being a part of the community, tweeting regularly, and connecting with others. In particular, in the policymaking process, Twitter may provide valuable real-time and detailed information, as discussed by Vydra and Kantorowicz (2021). Evidence of the increasing usefulness of Twitter data in generating economic indicators in different fields is provided by Salvatore et al. (2021), as depicted in Figure 1. Their research show that Twitter can provide valuable information to policymakers and researchers. Furthermore, text analytics and big data have revolutionized the construction of indicators. For instance, Antenucci et al. (2014) use Twitter to construct labor market indices for job loss, job search, and job posting. Bailliu et al. (2019) analyze labor market conditions using text analytics applied to Mainland Chinese-language newspaper. Glaeser et al. (2019) uses Yelp to nowcast the local economy. Taking advantage of these characteristics, Becerra and Sagner (2020) constructed an EPU index and Indaco (2020) show that the volume of tweets is a valid proxy for estimating GDP.




Source: Salvatore et al. (2021).

Figure 1: Articles produced between 2010 - 2018 related with Twitter

In summary, social media platforms, such as Twitter offer policymakers an opportunity to access real-time and detailed information. It also shows how individuals perceive the inflation rate and how it affects their daily lives. This perception may be influenced by several factors, including personal experience, media coverage, and economic education. As Bricongne et al., 2022 shows, using Twitter to construct an inflation indicator offers several advantages. First, it provides timely information in real time and delivers insights more quickly than the other indicators. Second, it is cost-effective because there is no need to design, develop, or conduct an entire survey. Additionally, Twitter includes a wide range of opinions from users with diverse backgrounds, offering a more comprehensive view of inflation perceptions than traditional surveys do. As a non-traditional indicator, Twitter data complements the information collected from traditional survey-based indicators and financial markets. Moreover, Twitter is more reactive to short-term events and may be useful for providing early signals of changes in inflation. By leveraging these new data sources, policymakers can better understand inflation perceptions and expectations, and make informed decisions.

Using Twitter as an indicator of inflation expectations has some drawbacks. First, Twitter audiences are neither conventional nor permanent in nature. Therefore, the obtained measures may be subject to significant volatility. Second, it is difficult to ascertain whether Twitter users refer to future or present expectations or whether they are discussing price levels or inflation itself. Nonetheless, as Angelico et al. (2022) suggest, current judgments of economic conditions and inflation expectations are closely related, and similarities may be due to various factors including historical inflation experience, inflation perception, and anchoring to actual inflation rates. Various indicators provide signals, based on specific methodologies and data. Constructing and monitoring them may provide an unbiased measure of inflation that is unaffected by these differences and uncertainties. Therefore, constructing new indicators is useful for complementing actual measures and providing policymakers with a robust analysis of the current dynamics of inflation, perceptions, and expectations.

In this study, we construct two inflation perception indicators for Colombia following previous works by Angelico et al. (2022) and Bricongne et al. (2022) and a new Word2Vec approach to summarize tweet signals. As we focused on Colombia and Spanish as the official languages in over 20 countries, we collected tweets from January 2015 to March 2023 and used an algorithm that combines user profiles and tweet information to determine whether the tweet is Colombian. This is the first attempt to follow consumers' real-time inflation perceptions in Colombia to monitor household decisions. Additionally, this is the first perception indicator to complement existing indicators of inflation and expectations. We find that our perception indicators have a dynamic similar to expectations and observed inflation, which confirms the connection between expectations and observed inflation, and that our indicators provide a reliable measure of inflation perceptions. We also find that they provide additional information than the traditional AR(1) model forecasting inflation of alternative measures of inflation up to six months ahead; therefore, they convey additional information than just using the past values of the inflation rate in forecasting.

This document comprises six sections, beginning with the introduction. The second section details our approach to extracting tweets and constructing our data. Given that Spanish is the official language


Figure 2: Words searched.

of more than 20 countries, we explain how to determine whether a tweet is Colombian in origin. In the third section, we outline our methodology for constructing indicators. First, we combine a dictionary-based and Latent Dirichlet Allocation techniques to eliminate tweets referring to promotions, popular sayings, and crypto-currencies. Later, we construct two inflation perception indicators using the dictionary-based and Word2Vec approaches. In the fourth section, we compare the dynamics of our indicators with observed inflation and expectations and discuss the fact that all of these measures share some underlying structural factors that may reflect similar dynamics. In the fifth section, we perform an empirical exercise to show that our indicators provide additional information than past inflation in forecasting six measures of annual inflation. In the concluding section, we summarize our findings and highlight the potential implications of our research for future studies in the field.

# 2 Data Collection and Pre-processing

Twitter is a popular social media platform in Colombia. According to Datareportal, 2022, as of January 2022, there were approximately four million Twitter users in Colombia, representing approximately 10% of the population. Additionally, internet users in Colombia spend an average of 2 hours and 41 minutes per day on social media platforms, with Twitter being one of the most popular. Therefore, as Twitter is an important social media platform in Colombia, with a significant number of users, businesses, and organizations, it can leverage Twitter's reach and popularity to connect with potential customers and engage with their audience. To access Twitter's vast collection of tweets and to search for specific keywords within a given timeframe, we used the Twitter API v2 standard search endpoint. The query included tweets in Spanish that included at least one of the words related to prices, inflation, and cost of living, as presented in Figure 2. We did not constrain the tweet location and searched for tweets between January 2015 and March 2023.

As an application of Natural Language Processing (NLP), our study employs a corpus[1] of 83.4 mil-

[1] A corpus is a collection of documents.

lion documents, each of which is a tweet. The first challenge we faced was identifying tweets from Colombian users. Given that only about 1% of tweets provide information on geographic location, and Spanish is the official language of 21 countries, we used a combination of tweet and user information to classify tweets as Colombian.[2] We established four criteria for identifying Colombian tweets: (i) tweets geolocated in Colombia, (ii) tweets containing words that refer to Colombia, (iii) users geolocated in Colombia, and (iv) users with descriptions that indicate Colombian origin.

With this dataset, we employed a list of main cities and demonyms to determine which tweets and user descriptions pertain to Colombia.[3] In total, we identify 3.5 million tweets as Colombian. The specific outcomes of our classification approach are presented in Table 1, where each cell shows the number of tweets that meet the conditions of the corresponding row and column. Diagonal elements denote the number of tweets identified solely by one criterion. The total count for each row signifies the number of tweets identified by the criteria, and may not necessarily sum up the numbers within the row. Among 176,619 tweets, the criterion of tweet location identified the smallest number of Colombian tweets. Hence, employing additional criteria for user information, encompassing geolocation and description, yields a greater number of identified Colombian tweets and, consequently, a more extensive database for our analysis. Notably, 29,336 tweets fulfilled both the tweet descriptions and user location criteria, confirming their origin in Colombia. It is evident that user location emerges as the most pivotal criterion for our classification, with over 2.7 million entries in our database corresponding to users located in Colombian cities, towns, or within the country.

| | Tweet Location | Tweet Description | User Location | User Description | Total |
|---|---|---|---|---|---|
| Tweet location | 34,862 | 27,984 | 129,851 | 29,336 | 176,619 |
| Tweet description | | 286,406 | 329,948 | 127,435 | 647,192 |
| User location | | | 1912,230 | 530,627 | 2758,944 |
| User description | | | | 250,176 | 809,275 |

Table 1: Classification of Colombian Tweets

To prepare our Colombian data for machine learning analysis, we cleaned it by removing any irrelevant or duplicated information, lowercasing all text, and removing stop words, hashtags, mentions, and punctuation marks. This step helped to standardize the data and reduce the number of features.

# 3 Constructing the inflation perception indicators

In this section, we discuss the methodology used to construct two alternative perception inflation indices. First, we removed tweets related to topics that did not contain relevant price signals. Next, we

[2]Spanish is the official language in the following countries: Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, Equatorial Guinea, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Spain, Uruguay, and Venezuela.

[3]This list is provided in Table A.1 in Appendix A.

develop our indicators, an N-gram indicator based on a dictionary that identifies whether each tweet indicates a price increase, decrease, or stability. Additionally, we constructed a Word2Vec (W2V) indicator based on the proximity of our tweets to keywords signaling price movements.

## 3.1 Removal of non price related tweets

To clean our data and obtain a signal related to inflation, we extract tweets that contain price content relevant to the Colombian CPI's basket of goods and services. Thus, we eliminated three categories of tweets that do not contain any useful price information: tweets containing popular sayings in Colombia, tweets related to cryptocurrency price dynamics, and tweets related to sales and promotions. We employed two classification approaches to remove non-inflation-related categories from our database. We search for n-grams that contain words related to sayings in Spanish, and we also use Latent Dirichlet Allocation (LDA) to identify tweets related to sales or promotions and eliminate them from our database.[4] First, we eliminate tweets containing a list of n-grams including the terms "precio", "caro ", or related words that belongs to phrases or popular expressions without price signal, Table 2.

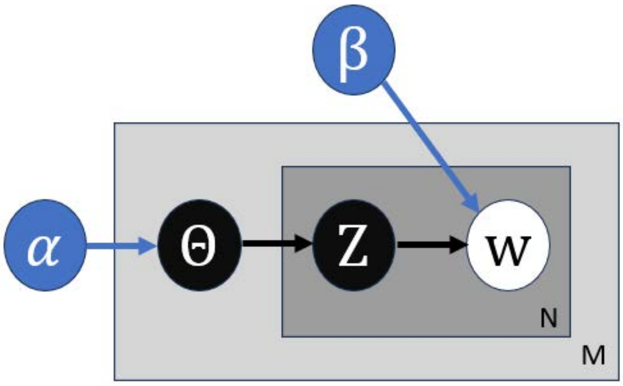
| | |
|---|---|
| lo barato sale caro | no tiene precio |
| a cualquier precio | al precio que sea |
| precio de la fama | a precio de gallina robada |
| precio de la libertad | precio de la rectitud |
| no importa el precio | por ningun precio |

Table 2: Popular sayings in Colombia.

Following the removal of popular sayings, we remove cryptocurrency and promotions using a probabilistic topic modeling approach (LDA) proposed by Blei et al., 2003, which is also one of the most popular topic models in the natural language processing (NLP) field. This is reinforced by its ease of use and the fact that it has been demonstrated to categorize text in the same way as people do (Chang et al., 2009). Other studies that employ this strategy to summarize the corpus in subjects include Larsen et al., 2021, who created 80 time-series measures of the news topics the media reported, that is, the various types of news reporting. Parra-Polanía, 2020 examine the underlying subjects in the Central Bank of Colombia's minutes and monetary policy reports and Gabrielyan et al., 2020 propose an index of inflation news that takes into account the intensity of a given topic using UK news.

LDA is used to identify the main topics discussed in a collection of documents ($N$) comprising a corpus ($M$). The algorithm assumes that each document is a mixture of a small number of topics ($Z$) and that each word ($W$) in the document is drawn from one of these topics. The LDA algorithm

[4] Other possible methods of topic classification include dictionary-based approaches using pre-trained ML algorithms like Robertuito Pérez et al. (2021), Xml-RoBerta Conneau et al. (2020), or developing and training our own model.




Source: Li et al., 2017.

Figure 3: Graphical Model of LDA

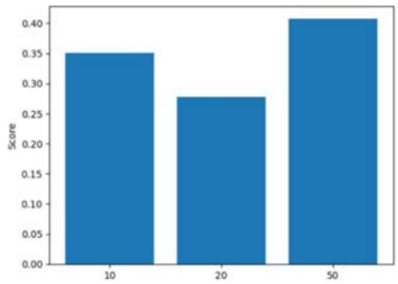



Figure 4: Coherence Score for alternative number of topics.

follows several steps to identify topics: first, it creates a dictionary of words; second, it randomly assigns topics in each document ($\alpha$) and words in each topic ($\beta$);[5] third, it iteratively updates the topic assignments based on the probabilities of the words belonging to each topic ($\Theta$) and identifies the topics and the distribution of words in each topic. A simple scheme of the algorithm is shown in Figure 3. The output of the LDA algorithm is a set of topics, each represented by a distribution of words and a probability distribution of topics for each document in the corpus. Overall, the LDA algorithm provides a useful tool for discovering underlying topics in a large corpus of text data. Finally, because the LDA estimation procedure does not provide the names of the topics, the assignment of labels to each topic is based on the most important terms within the topic.

To determine the optimal number of topics in our data, we applied the *Coherence Score* proposed by Röder et al., 2015. This metric involves grouping words into subsets, estimating the relative distance between words within each topic, and producing an aggregate coherence score. Thus, we can compare the consistency index calculated for different subject scores and choose the one that yields the highest value. In Figure 4 we compare the coherence score for classifying 10, 20, or 50 topics in our database, and we can see that the highest score is found with 50 topics.

Table 3 presents the terms with a high probability associated with each of the topics we removed from our database to provide information about promotions or cryptocurrencies. Each line in the

[5] $\alpha$ and $\beta$ are parameters of a Dirichlet distribution.

table corresponds to one topic, so we remove ten topics related to promotions and eight related to cryptocurrencies.

| |
|---|
| **Promotions** |
| venta, vende, info, informacion, excelente |
| bogota, ciudad, medellin, vendo |
| calidad, justo, pedidos, excelente |
| envio, pago, disponible, tienda, calidad, whatsapp, info |
| super, descuento, pagina, informacion, whatsapp, disponible, bogota |
| servicio, mes, calidad, whatsapp, excelente |
| productos, mejores, calidad, tienda, info, producto, servicio, dia |
| domicilio, pide, contactanos, metro, whatsapp, servicio, envio, calidad, ciudad |
| compra, cafe, final, via, venta, especial, colombiano, descuento |
| envio, gratis, cali, whatsapp, bogota, medellin, info |
| |
| **Cryptocurrencies** |
| usd, eth, ethereum, mar, final |
| usd, ltc, litecoin, mar, via, cara |
| usd, xmr, monero, mar, via |
| bitcoin, btc, tasa, fuente, usd |
| millones, dolares, colombianos, vale, pasa, empresa, mes |
| pesos, mil, colombianos, millones, vale |
| ana, final, bogota, despues, economia, mercado, gasolina, alza, informacion |
| xrp, usd, via, mas, bogota, mar, dolar, pais, alto, servicio, pagar, productos |

Table 3: Words from topics related to promotions and cryptocurrencies.

## 3.2 N-gram Indicator

To construct this indicator, we use a dictionary-based approach by classifying our tweet signal into inflation up, down, or constant according to the selected n-grams. Thus, a tweet is classified into these categories only when an n-gram appears in its contents. The list of n-grams is a modified version of the one proposed by Angelico et al., 2022 in which we include relevant expressions for the Colombian case. We employ 312 expressions to denote inflation up and 181 to denote inflation down, as presented in Tables B.1 and B.2 in Appendix B. To illustrate our findings, Figure 5 presents the most frequent n-grams in our data. Once we classify our tweets according to their signal, we add all the information to construct the daily indicators of increasing inflation $U_t$ and decreasing inflation $D_t$.

We compute the balances using the daily aggregates of each signal and the formula of **Indicator #4** proposed by Angelico et al. (2022), that is, the difference between the logarithms of the upward and downward tweets per day $B_t = ln(U_t + 1) - ln(D_t + 1)$.[6]

[6] In their study Angelico et al. (2022) present three other balances which we also implemented and have a close dynamic and results in the empirical analysis to our indicator. The balance of **Indicator #1** is calculated

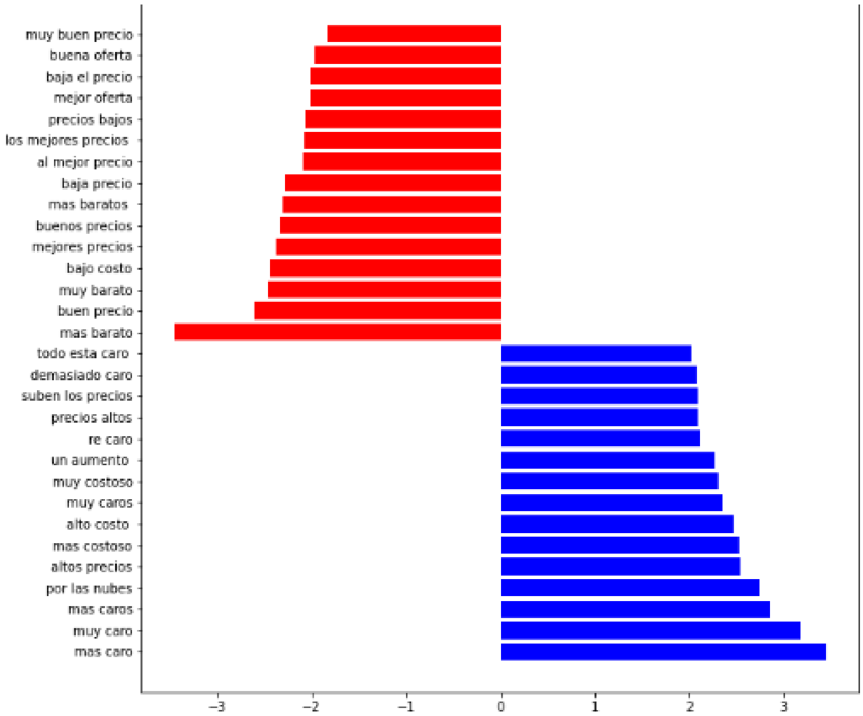



Figure 5: Most frequent n-grams

In order to remove the volatility we find in the balance $B_t$, we employ two techniques. First, we apply winsorization, which involves capping any value that exceeds three standard deviations from the mean. This helps mitigate the influence of extreme values on the indicator. Second, we implement a smoothing procedure by employing a backward-looking moving average over a period of 30 days. This moving average helps reduce short-term fluctuations and provides a more stable representation of the indicator's trend. Finally, our indicator is constructed as balanced divided by its standard deviation; that is, $\pi_t^T = \frac{B_t}{\sigma(B_t)}$. This strategy facilitates comparisons over time and across the variables. Moreover, this scaling approach not only maintains the sign of the initial balance between the upward and downward signals but also provides a scale based on the historical behavior of the indicator.
In Figure 6 we present our Twitter index in red and compare it with the index obtained without removing crypto-currencies, popular sayings, and promotions, which are depicted in black. The dynamics for both indices is similar. From 2015 to 2018, the values of both indices were relatively low, followed by an upsurge in 2019 and 2020, coinciding with the outbreak of the COVID-19 pandemic. Subsequently, there was a substantial increase in expectations, largely attributable to supply shocks resulting from cost hikes in transportation and the Russian invasion of Ukraine. However, there were some differences between the second and third quarters of 2016, and the third quarter of 2017.

as the difference between the number of upward and downward tweets per day, expressed as $B_t^1 = U_t - D_t$. **Indicator #2** adjusts the previous indicator by removing the impact of CPI releases and central bank meetings. To achieve this, we regress the balance of indicator #1 with respect to a dummy variable $R_t$ which captures CPI releases and central bank meeting dates. Therefore, the residual becomes the new expectations indicator $B_t^1 = R_t + B_t^2$, where $B_t^2$ captures the balance of inflation expectations due to other factors besides the central bank or the national institute of statistics (DANE). **Indicator #3** is constructed using exponential smoothing on the balance of indicator #1.

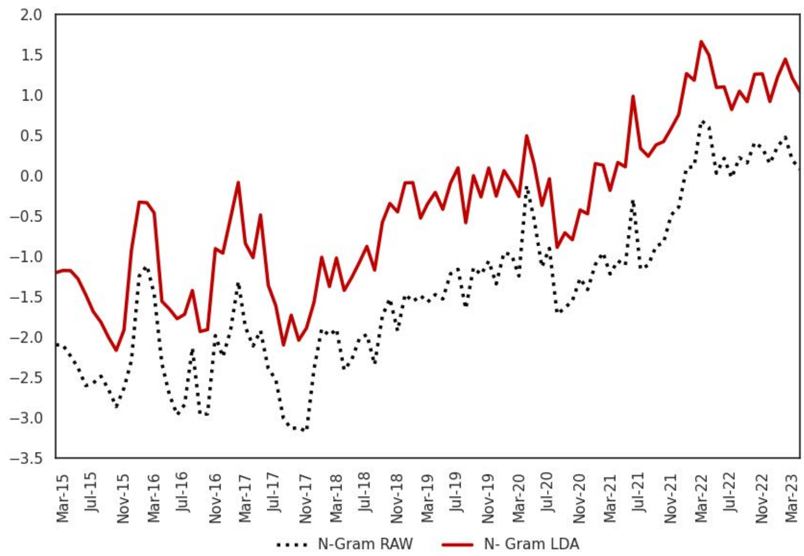



Figure 6: N-gram indicator Raw and after removing promotions, cryptocurrencies, and popular sayings.

As highlighted in the Introduction, one of the challenges encountered in classifying unstructured data is determining whether Twitter users express expectations regarding inflation or price levels. To address this issue, we exploit the known structure of the dictionary approach, classify our original n-grams, analyze whether each n-gram refers to prices or inflation, and compute the percentage of tweets that refer to price levels and inflation. Our analysis reveals that approximately 55% of tweets in Colombia, which indicates an increase in our inflation indicator, are related to inflation itself, while the remaining 45% pertain directly to price levels. Conversely, when the indicator signals a decrease, approximately 42% of the tweets specifically refer to inflation, whereas 58% refer to general price levels or specific products. This suggests that Twitter users do not distinguish between these two concepts.

## 3.3 Word2Vec Indicator

Another method for extracting a price signal from our Twitter corpus is to use the Word2Vec algorithm, which is a powerful tool for capturing semantic and syntactic relationships between words in a text corpus Mikolov et al. (2013). This technology extracts useful information from large amounts of unstructured text data and can be helpful in identifying patterns and trends that are difficult to notice using conventional approaches. This method is based on the idea that the meaning of a word can be inferred from its context. This method creates a vocabulary from text data and then learns the distributed representations or word embeddings of each word in the vocabulary. These embeddings encode the meaning of a word based on its co-occurrence patterns with other words in the text.

W2V addresses two main tasks: Continuous Bag-of-Words (CBOW) and Skip-Gram models. The first predicts the target words from the surrounding context words, and the second predicts the surrounding context words from the target words. We focus on Skip-Gram method under which to extract a rising price signal, we determine a set of terms linked to price rises, such as “expensive” or “price hike” and then compute the cosine similarity of the words in each tweet with the terms. This is a measure of vector similarity that ranges from 0 to 1, with 1 denoting a complete similarity. Tweets containing an increasing price signal are more likely to have a high similarity to price increase terms. As an example, consider two vectors, A and B, and the cosine similarity is defined as:

$$similarity_{A,B} = \cos\theta = \frac{A \cdot B}{\|A\| \, \|B\|}$$

where $\|A\|$ and $\|B\|$ denote the norm of each vector. $similarity_{A,B}$ is graphically represented in Figure 7.

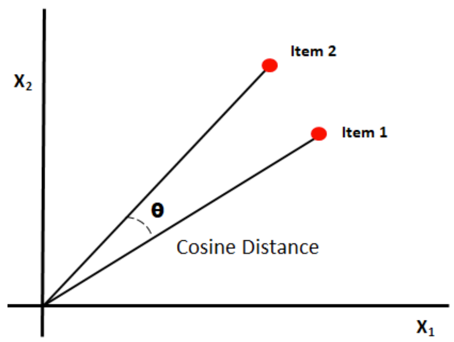



Figure 7: Cosine distance/Similarity.

To capture the tweets most closely related to each signal, we selected vectors with $\theta$ close to 0. In our application, we define three terms or centers to capture increasing price signals. “alto”, “alta”, and “alza”. The algorithm computes as nearest neighbors: ‘alto’, ‘alta’, ‘alza’, ‘aumento’, ‘subida’, ‘incremento’, ‘nubes’, ‘crecimiento’, ‘subiendo’, ‘caro’, ‘alimento’, ‘subieron’, ‘subido’, ‘encima’, ‘crisis’, ‘aumenta’, and ‘devaluacion’. With respect to the decreasing signals we select as centers: “bajo”, “bajo”, “caida”, which the algorithm associate as nearest neighbors: ‘bajo’, ‘baja’, ‘caida’, ‘debajo’, ‘bajan’, and ‘menor’. Next, we used words inferred from the corpus to classify the signal prices in our tweets. Finally, to construct our W2V inflation perception index, we summed the number of upward price signal tweets $U_t$ and downward tweets $D_t$, as well as their logarithmic balance, using the same procedure used in the N-gram indicator to remove volatility and provide a scale.

Figure 8 compares the W2V raw index computed using all available data with the index obtained after excluding promotions, crypto-currencies, and popular sayings from our database. The cleaning process resulted in lower volatility over time but in periods different from those found in the N-gram indicator.

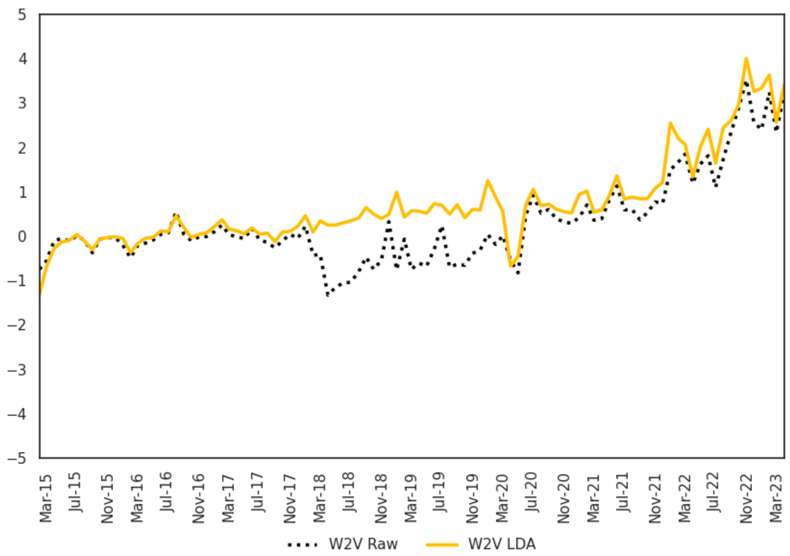


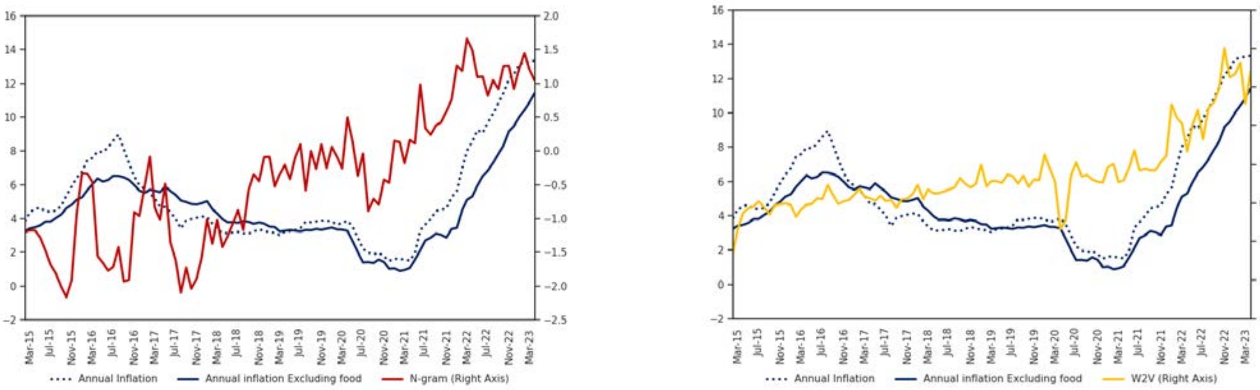

Figure 8: Word2Vec indicator Raw and after removing cryptocurrencies, promotions, and popular sayings.

# 4 A Comparative Analysis of our inflation perception indices, surveys, and market-based expectations.

In this section, we compare our inflation perception indicators with inflation and inflation expectation measures to determine whether common factors underlie their dynamics. First, we compare with annual headline inflation and inflation excluding food, this last measure eliminates volatility and is one of the main core inflation indicators.[7] Additionally, we compare our indices with two traditional measures of annual inflation expectations: The average response from a monthly survey that involves specialized forecasters from banks and other financial institutions and one-year break-even inflation expectation (BEI), which measures the one-year average expected inflation implicit in sovereign bond market data.[8] Specifically, BEI is computed as the difference between the yield of domestic currency bonds (TES in pesos) and that of inflation-linked bonds (TES in UVR). However, these estimates may be affected not only by inflation expectations but also by uncertainty about expectations as well as by liquidity frictions, and there are alternative strategies to obtain the component related to expected inflation, see for instance Espinosa-Torres et al., 2017. In our exercises, we preferred to use unadjusted estimates. In this way, our results are not dependent on the choice of a specific methodology to isolate the inflation signal and are based on information available to any market participant or professional forecaster.

[7] The Consumer Price Index is published monthly by DANE and can be accessed at https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc.

[8] The methodology and data from the survey for financial forecasters are available at: https://www.banrep.gov.co/es/estadisticas/encuesta-mensual-expectativas-analistas-economicos.




(a) N-gram indicator. (b) W2V indicator.

Figure 9: Comparison of our indices Headline inflation, and inflation excluding food.

## 4.1 Comparison with observed Inflation

Figure 9 compares our N-gram and W2V indicators with the inflation measures between 2015 and 2020. The left panel shows that our N-gram indicator exhibits a trend similar to that of the actual inflation rates between 2015 and 2020. A significant upsurge was observed during Q1-2016. The N-gram indicator effectively captures this spike, indicating a subsequent reduction in inflation levels during 2017 - 2018. However, as illustrated by the right panel graph, our W2V indicator shows only a small upward signal around June 2016, and remains stable between 2015 and the first quarter of 2018. In 2019, the N-gram and W2V indicators demonstrated distinct behaviors. Specifically, the former indicates a rising inflationary trend, whereas the latter suggests a constant level of inflation that aligns more closely with the observed measures. However, as we move into Q1 2020, both metrics display a marked increase in inflation rates; this uptick is not reflected by actual inflation measures, which indicates that any significant growth will only occur from 2021 onwards. Looking further ahead to late-2022, we observe interesting changes within these indicators: whereas n-grams signal stability for future periods, W2V may experience marginal reductions instead.

The similarity between our indices and observed inflation raises the question of whether we are measuring current perceptions or expectations or why there is a strong connection among them. Inflation expectations and actual inflation share many similarities due to various factors, including historical inflation experience, inflation perception, and anchoring to actual inflation rates. First, the historical inflation experience has a significant effect on expectations of future inflation rates, and individuals who have experienced higher inflation rates in the past are likely to have higher inflation expectations in the future Xu et al., 2016. Second, empirical evidence indicates that both inflation perceptions and past inflation significantly affect the formation of inflation expectations. This finding suggests that individuals' perceptions of inflation and their experiences with past inflation rates can influence their expectations of future inflation rates Xu et al., 2016. Finally, Grant and Thomas, 1999 found cointegration between survey measures of inflation expectations and actual inflation, indicating a strong relationship between the two variables, and inflation expectations have been anchored to both inflation targets and actual inflation, further emphasizing the similarity between these two variables Cicek and Akar, 2014. Furthermore, studies find that inflation expectations tend to increase as actual inflation accelerates, highlighting the importance of actual inflation rates in shaping individuals' expectations Feldkircher and Siklos, 2019. These findings suggest that individuals' expectations for future inflation rates are influenced by their past experiences, perceptions, and actual inflation rates.

## 4.2 Comparison with Inflation Expectations

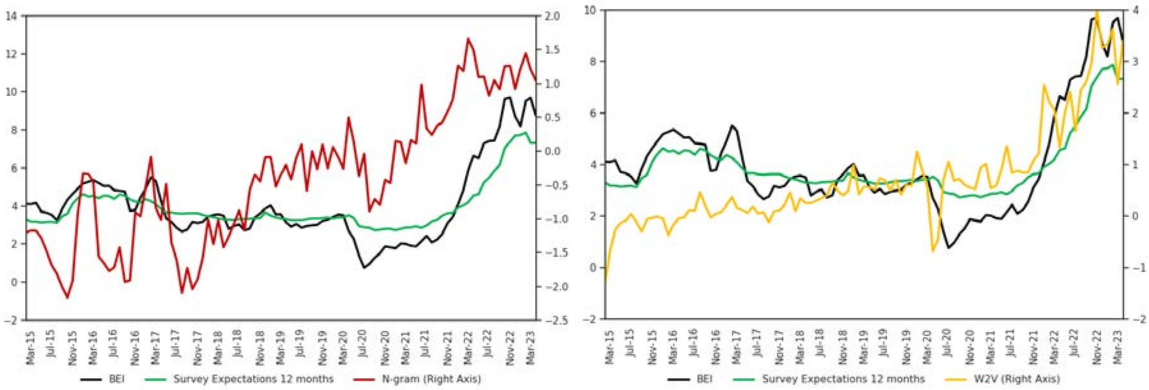
Figure 10 presents a comparison between our N-gram and W2V indicators and the traditional inflation expectation measures from January 2015 to March 2023. The upper panel displays our N-gram indicator, which demonstrates an increase in inflation expectations in 2016, followed by a subsequent decline in both 2017 and 2018, consistent with other inflation expectation measures such as break-even inflation or analysts' surveys. However, in 2019, it deviated from traditional indicators in terms of the observed rise in inflation perceptions. In contrast, our W2V indicator results showed stability similar to that of the two standard expected indication techniques mentioned earlier. Finally, as mentioned in the previous subsection, our indicators signal an increase in inflation perceptions by 2020. This finding is not consistent with the observed inflation or break-even inflation (BEI) expectations, but may be consistent with the survey-based indicator, which shows that inflation expectations did not decrease during 2020.

The similarity of our indices with inflation expectations from various sources may be because the underlying economic factors that influence them are the same. Economic growth, monetary and fiscal policies, and global economic conditions all affect inflation. Therefore, they reflect the same underlying factors regardless of the source of inflation expectations. Furthermore, economic agents such as investors and consumers use news releases, economic indicators, and financial market development to form their expectations. For example, expectations of inflation derived from consumer surveys may be influenced by information about rising fuel prices, whereas financial market participants may form expectations based on changes in interest rates.

An empirical exercise to show that our indicators $\pi^T$ share common information with the traditional inflation expectations $\pi^E$ from break-even inflation and the monthly survey expectation is to estimate the following regression: we regress the expectation measures with a constant and our two indicators. Thus, if there is a relationship among these variables, the associated parameter $\gamma_1$ will be significant.

$$\pi_t^E = \gamma_0 + \gamma_1 \pi_t^T + \eta_t \tag{1}$$

Table 4 presents the estimates for Equation 1. Columns one and two show the results for the N-gram indicator while columns three and four for the W2V indicator. As we can see, all estimates of $\gamma_1$ are significantly different from zero; thus, our measures relate to the traditional inflation expectations.




(a) N-gram indicator. (b) W2V indicator.

Figure 10: Comparison of our indices with expectations measures.

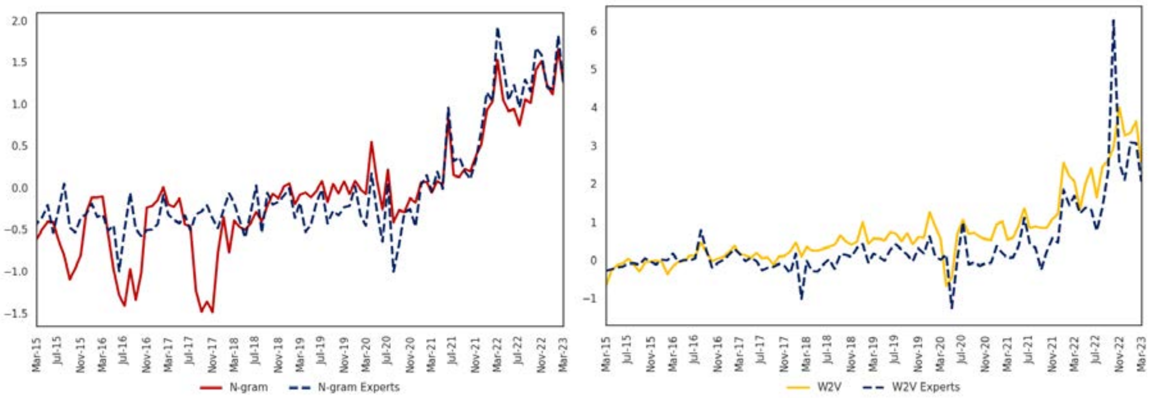
| | N-gram | | Word2Vec | |
|---|---|---|---|---|
| | BEI | EMEA | BEI | EMEA |
| $\gamma_1$ | 2.749*** | 0.016*** | 1.128*** | 0.007*** |
| | (0.61) | (0.002) | (0.144) | (0.001) |
| $\gamma_0$ | 4.727*** | 0.036*** | 3.805*** | 0.038*** |
| | (0.257) | (0.001) | (0.228) | (0.001) |
| Obs | 96 | 96 | 96 | 96 |
| $R^2$ | 0.165 | 0.181 | 0.394 | 0.479 |
| $R^2_{adj}$ | 0.188 | 0.267 | 0.388 | 0.483 |

Table 4: Relationship between our indicator and traditional inflation expectations.

In conclusion, our N-gram and W2V indicators show a close relationship between their dynamics and that of current inflation and inflation expectations. This finding is not surprising, given the strong connections we have discussed between actual events and how agents revise their inflation expectations based on the temporal nature of the shock. Specifically, whether the shock is expected to be permanent or transitory plays a significant role in shaping the agents' inflation expectations.

## 4.3 Robustness check - Perception Indicators for experts

One robustness check of our indicator is to construct it using only users with a better knowledge of economic conditions. Therefore, we classified the users as experts if they satisfied them. Using this strategy, we obtained a new database of 298,620 tweets, which was approximately 8.5% of the initial database. We then follow our strategy and construct two indicators using only the experts' tweets, in which promotions and popular sayings are not necessary, given that these tweets are clearly associated with price information. We checked and removed only five tweets using these criteria. However, some tweets are related with crypto-currencies, so we remove them using the same LDA model used for the




(a) N-gram indicator.

(b) W2V indicator.

Figure 11: Comparison of our indices and an expert subsample.

whole database. [9]

Figure 11 presents a comparison between our inflation perception indicators constructed from expert tweets and the results in the previous section. The left panel of the figure shows that until 2018, the expert indicator had a similar dynamic to the N-gram indicator but with a higher level. From 2019, the level and dynamics of the expert indicator became closer to those of the original indicator, and from 2021, the expert indicator once again had a higher value. Regarding the W2V indicator, the dynamics of the total database and those of experts are similar, except for the period between 2018 and 2020, where the expert indicator signals a higher inflation perception. However, our baseline indicators follow the behavior of inflation and traditional expectation measures more closely. It is worth noting that the two expert indices have variabilities similar to the original indices in terms of volatility.

# 5 Predictive accuracy of our indices.

In order to evaluate whether our Twitter indicators provide additional information than the inflation rate itself, and if it is useful in forecasting inflation rates, we conducted the following empirical exercise to determine whether including our indicators improves the forecasting accuracy of a reference model, which is a traditional AR(1). Following Fitchett and Robinson (2021), we evaluate the predictive power of our two Twitter indices on headline inflation and excluding food inflation by comparing direct forecasts from the two nested models. To do so, we compared the forecasts of our reference model.

[9]When we provide the whole database to the LDA algorithm we show enough information so it can learn terms related to crypto-currencies than if we only use the experts database to estimate the LDA.

$$\pi_{t+h} = \alpha_1 \pi_t + \epsilon_{1,t} \tag{2}$$

with those coming from a model that includes one of our Twitter indicators:

$$\pi_{t+h} = \alpha_1 \pi_t + \alpha_2 \pi_t^T + \epsilon_t \tag{3}$$

In equations 2 and 3, $\pi_{t+h}$ represents the annual inflation at horizon $h = 1, ..., 6$, and $\pi_t^T$ is the Twitter indicator. We use the direct forecast approach because it is less susceptible to bias resulting from incorrect model specifications, as suggested by Marcellino et al. (2006). Additionally, it does not require the forecasting of explanatory variables, thus avoiding uncertainty in their estimations. We conducted a rolling forecast evaluation using the last 39 months of our sample, that is, from January 2020 to March 2023. This forecasting period is marked by significant fluctuations in the inflation rates. Initially, there was a decrease due to the COVID pandemic in 2020, followed by a notable rise due to supply shocks, such as increased transport fees and Russia's invasion of Ukraine, which substantially increased costs globally. Given these conditions, the root-mean-square error (RMSE) may be affected in both models because it is a statistic susceptible to extreme values. Then, with forecasts from both models, we computed the RMSE and the relative RMSE as the quotient of the RMSE of the augmented model equation 3 and the RMSE of the reference model equation 2. A value of one indicates that both models have the same accuracy, a value lower than one suggests that adding the Twitter indicator improves forecast accuracy, and a value greater than one suggests that it worsens accuracy. Moreover, for models in which the relative accuracy is lower than one, we check whether the improvement in the RMSE is statistically significant using the Giacomini and White (2006) test.

We chose six distinct inflation indicators to evaluate the forecast accuracy. These indicators encompass various dimensions of inflation: headline inflation, core inflation, and four baskets that encapsulate diverse sources of inflationary pressures.[10] The structure of the Colombian Consumer Price Index (CPI) is illustrated in Figure 12, depicting the division into four distinct categories, along with the respective count of items and their corresponding weights. In the administrative category, we include products that are subject to government price control. The services category comprises goods whose pricing is less influenced by exchange rates, and is thus more contingent on internal demand dynamics. The goods basket includes items whose prices are shaped by international competition, thus reflecting both domestic and exchange rate-related factors. The food category encompasses products vulnerable to weather and transportation cost fluctuations that predominantly react to demand-side shocks. In our analysis, core inflation excludes food classification, which corresponds to the items in the first three baskets. Headline inflation is considered to be the sixth indicator.

Table 5 presents the outcomes of our evaluation of the predictive accuracy for both indicators, covering a forecasting horizon of up to six months. Each column corresponds to a distinct alternative measure,

[10] For a comprehensive grasp of this categorization, refer to González-Molano et al. (2020).

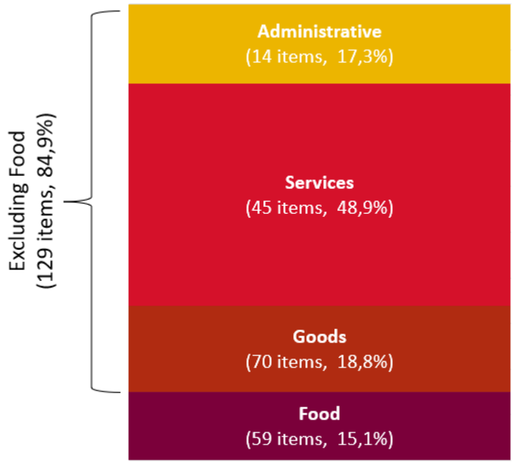



Figure 12: CPI decomposition for monitoring inflationary pressures.

showing the results of the indicators. The right panel of the table corresponds to the N-gram indicator. In terms of food inflation prediction, our findings indicate that the incorporation of this indicator improves forecasting accuracy by up to 14%, with statistical significance for forecasting four, five, and six months ahead. Concerning the other inflation indicators, we observe a reduction in the relative RMSE across all horizons, excluding the regulated basket where prices are under government control. Substantial improvements in forecasting accuracy are evident in forecasting food and goods inflation, while a reduction of approximately 10% is found in forecasting overall inflation for horizons between one and five months. Turning our attention to the W2V indicator, which is represented in the lower panel of the table, we observe a consistent reduction in the relative RMSE upon integrating this indicator. This reduction is particularly pronounced in the case of goods, services, excluding food, and total inflation measures, whereas no notable improvement is evident in forecasting the regulated inflation. Importantly, the gains achieved in terms of forecasting accuracy for the W2V indicator tend to surpass those observed for the N-gram indicator.

# 6 Conclusions

Inflation expectation and perception measures are crucial in the inflation-targeting regime. Having complementary indicators is useful given the limitations and advantages of all of them. Therefore, following Angelico et al. (2022), we construct two inflation perception indices using over three million tweets collected from 2015 to 2023 using the Twitter API. Social media platforms, such as Twitter, offer policymakers an opportunity to access real-time and detailed information on inflation expectations. This approach complements traditional measures of inflation expectations in several ways. First, social media platforms provide real-time information and deliver insights more quickly than other indicators. Second, it is cost-effective because there is no need to design, develop, or conduct an entire survey. In addition, social media platforms include a wide range of opinions from users with diverse

| | Food | Goods | Regulated | Services | Excluding Food | Total |
|---|---|---|---|---|---|---|
| | | | N-Gram | | | |
| 1 | 0.98 | 0.90** | 0.98 | 0.87** | 0.87** | 0.91** |
| 2 | 0.97 | 0.82** | 1.00 | 0.89* | 0.88* | 0.91* |
| 3 | 0.95 | 0.80** | 1.01 | 0.90 | 0.89 | 0.91* |
| 4 | 0.90** | 0.77** | 1.03 | 0.91 | 0.92 | 0.89* |
| 5 | 0.86** | 0.79** | 1.05 | 0.93 | 0.95 | 0.89* |
| 6 | 0.86** | 0.82** | 1.07 | 0.95 | 0.97 | 0.91 |
| | | | W2V | | | |
| 1 | 099 | 0.89** | 0.99 | 0.74*** | 0.79*** | 0.88** |
| 2 | 0.98 | 0.79** | 1.00 | 0.74*** | 0.77*** | 0.87*** |
| 3 | 0.97* | 0.80** | 1.00 | 0.77** | 0.80** | 0.87*** |
| 4 | 0.97 | 0.79** | 1.00 | 0.80** | 0.82** | 0.87** |
| 5 | 0.98 | 0.82** | 1.02 | 0.82** | 0.85** | 0.88** |
| 6 | 0.98 | 0.84** | 1.04 | 0.84** | 0.86** | 0.88** |

Table 5: Relative forecast accuracy of our perception indicators

backgrounds. Moreover, social media platforms are more reactive to short-term events and may be useful for providing early signals of changes in inflation.

Thus, in this study, we created two perception indicators using Twitter and machine learning techniques to extract a signal about price evolution in Colombia between January 2015 and March 2023. The first indicator, N-gram, is based on a supervised approach using a dictionary, whereas the second uses the Word2Vec approach to determine whether a tweet conveys rising or decreasing information on prices. Our indicators show a close dynamic with current inflation and traditional inflation expectations measures, which confirms the connection between current inflation, perceptions, and expectations because they have common factors. Finally, we provide evidence that our indicators provide additional information for forecasting inflation in the short run. On the one hand, the N-gram indicator notably improves food, goods and total inflation predictions by up to 18% over four to six months, with reduced relative RMSE observed for most inflation indicators. Additionally, the W2V indicator consistently leads to a substantial reduction in the relative RMSE, particularly for goods, services, and overall inflation measures, surpassing the gains in forecasting accuracy achieved by the N-gram indicator.

Our results should be interpreted carefully, as the use of social media platforms such as Twitter to construct indicators of inflation expectations has certain drawbacks. First, social media audiences are neither conventional nor permanent, and the measures obtained may be subject to significant volatility. Second, it is difficult to ascertain whether social media users refer to future or present expectations or whether they are discussing price levels or inflation itself. Nonetheless, as some studies suggest, current judgments of economic conditions and inflation expectations are closely related, and similarities may be due to various factors including historical inflation experience, inflation perception, and anchoring to actual inflation rates. Given that different indicators provide signals based on particular views, information, and methodologies, combining them may provide an unbiased and accurate estimate of inflation expectations. As a result, additional study is required to supplement our findings and the growing body of research in the field of measuring the public's inflation perceptions.

# References


Angelico, Cristina et al. (2022). "Can we measure inflation expectations using Twitter?" In: *Journal of Econometrics* 228.2, pp. 259–277. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2021.12.008 (cit. on pp. 1–4, 9, 19).

Antenucci, Dolan et al. (Mar. 2014). *Using Social Media to Measure Labor Market Flows.* NBER Working Papers 20010. National Bureau of Economic Research, Inc (cit. on p. 3).

Bailliu, Jeannine et al. (2019). "Can media and text analytics provide insights into labour market conditions in China?" In: *International Journal of Forecasting* 35.3, pp. 1118–1130. DOI: 10.1016/j.ijforecast.2019 (cit. on p. 3).

Becerra, Juan Sebastián and Andrés Sagner (July 2020). *Twitter-Based Economic Policy Uncertainty Index for Chile.* Working Papers Central Bank of Chile 883. Central Bank of Chile (cit. on p. 3).

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan, pp. 993–1022 (cit. on p. 7).

Bricongne, Jean-Charles et al. (Dec. 2022). *New indicators of perceived inflation in France based on media data.* https://blocnotesdeleco.banque-france.fr/en/blog-entry/new-indicators-perceived-inflation-france-based-media-data (cit. on pp. 3, 4).

Chang, Jonathan et al. (2009). "Reading tea leaves: How humans interpret topic models". In: *Advances in neural information processing systems* 22 (cit. on p. 7).

Cicek, Serkan and Cuneyt Akar (2014). "Do Inflation Expectations Converge Toward Inflation Target or Actual Inflation? Evidence from Expectation Gap Persistence". In: *Central Bank Review* 14.1, pp. 15–21. URL: https://ideas.repec.org/a/tcb/cebare/v14y2014i1p15-21.html (cit. on p. 15).

Conneau, Alexis et al. (July 2020). "Unsupervised Cross-lingual Representation Learning at Scale". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* Online: Association for Computational Linguistics, pp. 8440–8451. DOI: 10.18653/v1/2020.acl-main.747. URL: https://aclanthology.org/2020.acl-main.747 (cit. on p. 7).

Datareportal (2022). *DIGITAL 2022: COLOMBIA.* URL: https://datareportal.com/reports/digital-2022-colombia (visited on 03/31/2022) (cit. on p. 5).

Espinosa-Torres, Juan Andrés, Luis Fernando Melo-Velandia, and José Fernando Moreno-Gutiérrez (Mar. 2017). “Expectativas de inflación, prima de riesgo inflacionario y prima de liquidez: una descomposición del break-even inflation para los bonos del Gobierno colombiano”. In: *Revista Desarrollo y Sociedad* 1.78, pp. 315–365. DOI: 10.13043/dys.78.8. URL: https://revistas.uniandes.edu.co/index.php/dys/article/view/6702 (cit. on p. 13).

Feldkircher, Martin and Pierre L. Siklos (2019). “Global inflation dynamics and inflation expectations”. In: *International Review of Economics & Finance* 64.C, pp. 217–241. DOI: 10.1016/j.iref.2019.06.00. URL: https://ideas.repec.org/a/eee/reveco/v64y2019icp217-241.html (cit. on p. 15).

Fitchett, Hamish and Finn Robinson (Jan. 2021). *Down to business: Which QSBO measures are the best at forecasting?* Reserve Bank of New Zealand Analytical Notes series AN2021/01. Reserve Bank of New Zealand. URL: https://ideas.repec.org/p/nzb/nzbans/2021-01.html (cit. on p. 17).

Gabrielyan, Diana, Jaan Masso, and Lenno Uusküla (2020). “Mining News Data for the Measurement and Prediction of Inflation Expectations”. In: *Theory and Applications of Time Series Analysis*. Ed. by Olga Valenzuela et al. Cham: Springer International Publishing, pp. 253–271. ISBN: 978-3-030-56219-9 (cit. on p. 7).

Giacomini, Raffaella and Halbert White (2006). “Tests of Conditional Predictive Ability”. In: *Econometrica* 74.6, pp. 1545–1578. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/4123083 (visited on 05/08/2023) (cit. on p. 18).

Glaeser, Edward L., Hyunjin Kim, and Michael Luca (June 2019). “Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity”. In: *Big Data for Twenty-First-Century Economic Statistics*. NBER Chapters. National Bureau of Economic Research, Inc (cit. on p. 3).

González-Molano, Eliana R. et al. (July 2020). *Nueva Clasificación del BANREP de la Canasta del IPC y revisión de las medidas de Inflación Básica en Colombia*. Borradores de Economia 1122. Banco de la Republica de Colombia. DOI: https://doi.org/10.32468/. URL: https://ideas.repec.org/p/bdr/borrec/1122.html (cit. on p. 18).

Grant, Alan P. and Lloyd B. Thomas (1999). “Inflationary expectations and rationality revisited”. In: *Economics Letters* 62.3, pp. 331–338. URL: https://EconPapers.repec.org/RePEc:eee:ecolet:v:62:y:1999:i:3:p:331-338 (cit. on p. 14).

Indaco, Agustín (2020). “From twitter to GDP: Estimating economic activity from social media”. In: *Regional Science and Urban Economics* 85.C. DOI: 10.1016/j.regsciurbeco.20 (cit. on p. 3).

Larsen, Vegard H., Leif Anders Thorsrud, and Julia Zhulanova (2021). "News-driven inflation expectations and information rigidities". In: *Journal of Monetary Economics* 117, pp. 507–520. ISSN: 0304-3932. DOI: https://doi.org/10.1016/j.jmoneco.2020.03.004. URL: https://www.sciencedirect.com/science/article/pii/S030439322030026X (cit. on p. 7).

Li, Guowen et al. (2017). "Using LDA Model to Quantify and Visualize Textual Financial Stability Report". In: *Procedia Computer Science* 122. 5th International Conference on Information Technology and Quantitative Management, ITQM 2017, pp. 370–376. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2017.11.382. URL: https://www.sciencedirect.com/science/article/pii/S187705091732625X (cit. on p. 8).

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson (2006). "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series". In: *Journal of Econometrics* 135.1, pp. 499–526. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2005.07.020. URL: https://www.sciencedirect.com/science/article/pii/S030440760500165X (cit. on p. 18).

Mikolov, Tomas et al. (2013). "Distributed Representations of Words and Phrases and their Compositionality". In: *Proceedings of NIPS* (cit. on p. 11).

Parra-Polanía, Julián Andrés (Mar. 2020). *Descripción de las Minutas e Informes de Política Monetaria a partir de herramientas de Lingüística Computacional*. Tech. rep. DOI: https://doi.org/10.32468/be.1108. URL: https://repositorio.banrep.gov.co/handle/20.500.12134/9826 (cit. on p. 7).

Pérez, Juan Manuel et al. (2021). "Robertuito: a pre-trained language model for social media text in spanish". In: *arXiv preprint arXiv:2111.09453* (cit. on p. 7).

Röder, Michael, Andreas Both, and Alexander Hinneburg (2015). "Exploring the Space of Topic Coherence Measures". In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. WSDM '15. Shanghai, China: Association for Computing Machinery, pp. 399–408. ISBN: 9781450333177. DOI: 10.1145/2684822.2685324. URL: https://doi.org/10.1145/2684822.2685324 (cit. on p. 8).

Salvatore, Camilla, Silvia Biffignandi, and Annamaria Bianchi (Aug. 2021). "Social Media and Twitter Data Quality for New Social Indicators". In: *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement* 156.2, pp. 601–630. DOI: 10.1007/s11205-020-02296- (cit. on p. 3).

Vydra, Simon and Jaroslaw Kantorowicz (June 2021). "Tracing Policy-relevant Information in Social Media: The Case of Twitter before and during the COVID-19 Crisis". In: *Statistics, Politics and Policy* 12.1, pp. 87–127. DOI: 10.1515/spp-2020-0013 (cit. on p. 3).

Xu, Yingying et al. (2016). "Modeling heterogeneous inflation expectations: empirical evidence from demographic data?" In: *Economic Modelling* 57, pp. 153–163. ISSN: 0264-9993. DOI: https://doi.org/10.1016/j.econmod.2016.04.017. URL: https://www.sciencedirect.com/science/article/pii/S0264999316301080 (cit. on p. 14).

# A Terms related with Colombian tweets

Table A.1: Denonyms that classify a user or tweet as Colombian

| | | | |
|---|---|---|---|
| alvaradense | amazonas | amazonense | amazonense |
| antioqueña | antioqueño | antioquia | arauca |
| araucana | araucano | armenia | armerita |
| asiseña | asiseño | atlanticense | atlantico |
| barrameja | barramejo | barranquilla | barranquillera |
| barranquillero | bellanita | belumbrense | bogota |
| bogotana | bogotano | bolivar | bolivarense |
| bonaverense | boyaca | boyacense | bucaramanga |
| bugueña | bugueño | bumangues | bumanguesa |
| cachaca | cachaco | cafetera | cafetero |
| caldas | caldense | caleña | caleño |
| cali | cañasgordense | caqueta | caqueteña |
| caqueteño | carmeluna | carmeluno | cartagena |
| cartagenera | cartagenero | cartagüeña | cartagüeño |
| casanare | casanareña | casanareño | cauca |
| caucana | caucano | cesar | cesarense |
| choco | chocoana | chocoano | colombia |
| colombiana | colombiano | cordoba | cordobes |
| costeña | costeño | cucuta | cucuteña |
| cucuteño | cundinamarca | cundinamarques | cuyabra |
| cuyabro | dabeibana | dabeibano | dosquebradense |
| envigadeña | envigadeño | envigado | escudores |
| escudores | espinaluna | espinaluno | facatativeña |
| facatativeño | falanense | falanense | florencia |
| florideña | florideño | fresnense | fusagasugueña |
| fusagasugueño | girardoteña | girardoteño | guainarense |
| guainia | guainiana | guainiano | guajira |
| guajiro | guamaluna | guamaluno | guaviare |
| guaviarense | herreruna | herreruno | hincha del verde |
| hincha de millonarios | hincha del rojo | huila | huilense |
| ibague | ibaguereña | ibaguereño | icononzuna |
| icononzuno | ipialeña | ipialeño | itagui |
| itagüiseña | itagüiseño | la guajira | laboyana |
| laboyano | magdalena | magdalenense | magdalenico |
| manizaleña | manizaleño | manizales | mariquiteña |

Table A.1: Continued

| | | | |
|---|---|---|---|
| mariquiteño | medellin | melgarense | meta |
| metense | mituana | mituano | mocoana |
| mocoano | monteria | monteriana | monteriano |
| nariñense | nariño | neiva | neivana |
| neivano | norte de santander | norte santandereana | norte santandereano |
| opita | paisa | paisana | paisano |
| palmira | palmirense | palocabildense | parce |
| parcera | parcero | pasqueña | pasqueño |
| pasto | pastuso | payanes | pereira |
| pereirana | pereirano | petrista | piedecuestana |
| piedecuestano | popayan | porteña | porteño |
| putumayense | putumayo | quibdo | quindiana |
| quindiano | quindio | remediana | remediano |
| riohacha | riohachera | riohachero | rionegrera |
| rionegrero | rionegro | risaralda | risaraldense |
| rola | rolo | rovirense | samaria |
| samario | san andres | san andresana | san andresano |
| santamarta | santander | santandereana | santandereano |
| santoto | santo tomas | sergio arboleda | sibateña |
| sibateño | sincelejana | sincelejano | sincelejo |
| soachuna | soachuno | soledeña | soledeño |
| sucre | sucreña | sucreño | tolima |
| tolimense | toludeña | toludeño | tulueña |
| tulueño | tumaqueña | tumaqueño | tunja |
| tunjana | tunjano | umbitana | umbitano |
| universidad de los andes | universidad del bosque | universidad del rosario | universidad del valle |
| universidad eafit | universidad eai | universidad ean | universidad icesi |
| universidad javeriana | universidad nacional | universidad tadeo | urabaense |
| uribista | valduparense | valle del cauca | vallecaucana |
| vallecaucano | valledupar | valluna | valluno |
| vaupense | vaupes | veleña | veleño |
| vichada | vichadense | villavicencio | villavicense |
| yaguareña | yaguareño | yopaleña | yopaleño |
| yumbeña | yumbeño | | |

# B Terms related with inflation and deflation

Table B.1: n-grams for Downward Signal

| | | |
|---|---|---|
| al mejor precio | costos bajando | poco paga |
| baja costo | costos caen | pocos los precios |
| baja el costo | costos cayendo | precio a la mitad |
| baja el precio | costos disminuyen | precio absequible |
| baja en el costo | costos menores | precio baja |
| baja en el precio | deflacion del salario | precio bajando |
| baja en los costos | deflacion en el precio | precio bajara |
| baja inflación | deflacion en los precios | precio bajo |
| baja la inflacion | deflacion en los salarios | precio barato |
| baja precio | demasiado barato | precio cae |
| bajada de precio | demasiado baratos | precio caera |
| bajada de precios | descuento en el precio | precio cayendo |
| bajan costos | descuento en los precios | precio de ganga |
| bajan los costos | desploma el petroleo | precio de huevo |
| bajan los precios | desploma el precio | precio descontado |
| bajando el costo | desploman los precios | precio diminuto |
| bajando el precio | desplome en el costo | precio disminuyendo |
| bajando los costos | desplome en el precio | precio especial |
| bajando los precios | desplome en los costos | precio favorito |
| bajar el costo | desplome en los precios | precio inferior |
| bajar el precio | disminuye costo | precio inmejorable |
| bajar la inflacion | disminuye el costo | precio justo |
| bajar los precios | disminuye el petroleo | precio mas bajo |
| bajar precios | disminuye el salario | precio mas competitivo |
| bajara el precio | disminuyen los costos | precio minimo |
| bajaran los precios | el mejor precios | precio razonable |
| bajaran los salarios | increible precio | precio reducido |
| bajaron de costo | inferior precio | precio va a caer |
| bajaron de precio | inferiores precios | precios a la baja |
| bajaron los costos | la deflacion golpea | precios bajan |
| baje el precio | los mejores precios | precios bajando |
| bajen los precios | mas barato | precios bajaran |
| bajo costo | mas baratos | precios bajos |
| bajo costo | mejor oferta | precios baratos |
| bajo el precio | mejor precio | precios caen |
| bajo precio | mejores precios | precios caeran |

Table B.1: Continued

| | | |
|---|---|---|
| bajos costos | menor costo | precios cayendo |
| bajos los precios | menor inflacion | precios de ganga |
| bajos precios | menor precio | precios de oferta |
| barato el costo | menores costos | precios descontados |
| barato el precio | menores costos | precios diminutos |
| baratos los costos | menores precios | precios disminuyen |
| baratos los precios | menos caro | precios disminuyendo |
| buen precio | menos caros | precios inferiores |
| buena oferta | menos costo | precios inigualables |
| buenos precios | menos costoso | precios mas bajos |
| cae el costo | menos costosos | precios minimos |
| cae el precio | menos inflacion | precios razonables |
| caen los costos | minimo precio | rebaja de precio |
| caera el precio | minimos precios | rebajas de precio |
| caeran los precios | mitad de precio | reduccion en el precio |
| caidas en el precio | mitad de precios | reducen el precio |
| caidas en los precios | muy barato | reducen los costos |
| costo bajo | muy buen precio | reducen los precios |
| costo barato | muy buenos precios | reduciendo el precio |
| costo disminuye | no es caro | reduciendo los costos |
| costos a la baja | no es costoso | reduciendo los precios |
| costos bajan | oferta en los precios | reducir el precio |
| poco el precio | paga poco | salario a la baja |
| una ganga | poco costoso | salarios a la baja |
| se llama deflacion | | |

Table B.2: n-grams for Upward Signal

| | | |
|---|---|---|
| absurdo el precio | elevado pago | por las nubes |
| absurdo los precios | elevado precio | precio absurdo |
| abusan con el precio | elevados los precios | precio abusivo |
| abusan con los precios | elevados pagos | precio al triple |
| abusivo precio | elevados precios | precio altisimo |
| abusivos precios | elevan los precios | precio alto |
| abusos precios | elevan precios | precio aumentara |
| acelera la inflacion | elevando el precio | precio caro |
| altas inflaciones | elevando los precios | precio crece |
| altisimo precio | elevando precio | precio creciente |
| altisimos precios | elevando precios | precio de oro |
| alto costo | escalada de los precios | precio disparado |
| alto el precio | escalada de precio | precio elevado |
| alto pago | escalada de precios | precio es altisimo |
| alto precio | escalada del precio | precio exorbitante |
| altos costos | expectativa de inflacion | precio impagable |
| altos los precios | explosion de los precios | precio inflado |
| altos pagos | factura cara | precio mas alto |
| altos precios | factura costosa | precio miserable |
| alza de los precios | facturas caras | precio muy alto |
| alza de precios | facturas costosas | precio nunca visto |
| alza del precio | fuerte alza | precio record |
| alza el precio | fuerte alzas | precio sube |
| alza en el precio | fuerte creciemiento | precio subira |
| alza en los precios | fuerte inflacion | precio tan alto |
| alzan los precios | gasolina cara | precio va a subir |
| aumenta el costo | gasolina costosa | precio va aumentando |
| aumenta inflacion | gasolina impagable | precios a la alza |
| aumenta precio | guerra a la inflacion | precios absurdos |
| aumentado el precio | guerra contra la inflacion | precios abusivos |
| aumentado los precios | hiper inflacion | precios al doble |
| aumentan los costos | hiperinflacion | precios al triple |
| aumentan precios | increiblemente caro | precios altisimos |
| aumentando precio | increiblemente caros | precios altos |
| aumentara precio | incremento de los precios | precios arriba |
| aumentaron precios | incremento de precios | precios aumentaran |
| aumento de la inflacion | incremento del costo | precios caros |
| aumento de los precios | incremento en el precio | precios crecen |

Table B.2: Continued

| | | |
|---|---|---|
| aumento de precios | incrementos en los precios | precios crecientes |
| aumento del costo | inflación alta | precios disparados |
| aumento del precio | inflacion aumenta | precios elevados |
| aumento del salario | inflacion crece | precios están subiendo |
| aumento en el costo | inflacion creciente | precios estan super inflados |
| aumento en el precio | inflacion de dos digitos | precios exorbitantes |
| aumento en los precios | inflacion dura | precios impagables |
| aumento precio | inflacion elevada | precios mas altos |
| aumento precios | inflacion elevado | precios mas caros |
| aumentos de la gasolina | inflacion es un problema | precios muy altos |
| aumentos de los precios | inflacion fuerte | precios por las núbes |
| aumentos del petroleo | inflacion mas alta | precios record |
| aumentos del precio | inflacion muy alta | precios son altisimos |
| bajar el precio | inflacion no perdona | precios suben |
| cara factura | inflacion perdura | precios subiendo |
| cara la gasolina | inflacion persiste | precios subiran |
| caras facturas | inflacion se acelera | precios tan altos |
| caro el combustible | inflacion se dispara | precios tan altos |
| caro el precio | inflaciones altas | precios van a subir |
| caro pago | inflaciones muy altas | precios van arriba |
| caro precio | inflan el precio | precios van aumentando |
| caro recibo | inflan los precios | re caro |
| caros los precios | inflan precios | recibo caro |
| caros pagos | inflaran precio | recibos caros |
| caros precios | inflaran precios | repunte en costos |
| caros recibos | lo caro que esta | sale mas barato |
| combustible caro | lo caro que estan | sale mas caro |
| combustibles caros | los precios caros | sorprendentemente caro |
| como esta de caro | lucha contra inflacion | sorprendentemente caros |
| costo abismal | lucha contra la inflacion | suba de precios |
| costo impagable | mas caro | sube el costo |
| costo tan elevado | mas caros | sube el precio |
| costos crecientes | mas costoso | sube precio |
| costos impagables | mas costosos | suben los costos |
| costos por las nubes | mas han encarecido | suben los precios |
| costosa factura | mayor costo | suben precios |

Table B.2: Continued

| | | |
|---|---|---|
| costosas facturas | mayor inflacion | subida de los precios |
| crea inflacion | mayor precio | subida de precios |
| crea la inflacion | mayores costos | subida del precio |
| crece el costo | mayores precios | subido de precio |
| crece el precio | miserable precio | subido el precio |
| crece inflacion | miserables precios | subido los precios |
| crece la inflacion | mucho mas caro | subiendo el precio |
| crecen el precio | mucho mas caros | subiendo los precios |
| crecen los costos | mucho mas costoso | subieron de precio |
| crecen los precios | mucho mas costosos | subir el precio |
| creciente inflacion | muchos mas caros | subir los precios |
| crecimiento fuerte | muy cara | subira precio |
| cuesta mucho | muy caro | subiran precios |
| culpa de la inflacion | muy costosa | subirnos el precio |
| demasiado caro | muy costoso | subirnos los precios |
| demasiado costoso | nace inflacion | super inflacion |
| demasiados caros | nace la inflacion | tan caro |
| demasiados costosos | pagan caro | todo esta caro |
| dispara costo | pagando caro | triple del precio |
| dispara precio | pagara caro | triplica precio |
| disparan costos | pago alto | triplicado su precio |
| disparan precios | pago caro | triplicado sus precios |
| doble de precio | pago elevado | un aumento |
| duplicado su precio | pagos altos | un crecimiento |
| duplicado sus precios | pagos caros | un precio altisimo |
| dura la inflacion | pagos elevados | unos aumentos |
| el precio caro | peor inflacion | unos crecimientos |
| eleva el precio | peor precio | veces el precio |
| eleva precio | peores precios | venezuela tiene menos inflacion |
| elevado el precio | perdura la inflacion | persiste la inflacion |