ISSN 2451-5728

Forecasting inflation with Twitter

J. Daniel Aromí
Martín Llada

Nº76 SERIE DOCUMENTOS DE TRABAJO DEL IIEP | SEPTIEMBRE DE 2022

CONICET UBA

I I E P

INSTITUTO INTERDISCIPLINARIO
DE ECONOMÍA POLÍTICA

ISSN 2451-5728

Autores

J. Daniel Aromí
aromi.daniel@gmail.com

Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina.
CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política (IIEP). Buenos Aires, Argentina.
Universidad Católica Argentina. Buenos Aires, Argentina.

Martín Llada
lladamartin@gmail.com

Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina.
CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política (IIEP). Buenos Aires, Argentina.

Como citar:

Aromí, J. D. y Llada, M (2022). Forecasting inflation with Twitter. Serie Documentos de Trabajo del IIEP, 76, 1-20. https://ojs.econ.uba.ar/index.php/DT-IIEP/article/view/2620

Los Documentos de Trabajo del IIEP reflejan avances de investigaciones realizadas en el Instituto y se publican con acuerdo de la Comisión de Publicaciones.
Los autores son responsables de las opiniones expresadas en los documentos.

Coordinación editorial **Ed. Hebe Dato**
Corrección de estilo **Ariana Lay y Ed. Hebe Dato**
Diseño **DG. Vanesa Sangoi**

Forecasting inflation with Twitter

Macroeconomic forecasting
Inflation rate
Social networks
Twitter

We use Twitter content to generate an indicator of the level of attention allocated to inflation in public discussions. The analysis corresponds to Argentina for the period 2012-2019. Estimated forecasting models show that the indicator provides valuable information regarding future levels of inflation. Out-of-sample exercises confirm that social media content allows for gains in forecast accuracy. Beyond point forecasts, the index provides valuable information regarding inflation uncertainty, that is, the size of forecast errors confidence intervals. The proposed indicator compares favorably with other indicators such as media content, media tweets, google search intensity and consumer surveys.

Pronósticos de inflación con Twitter

Pronósticos macroeconómicos
Inflación
Redes sociales
Twitter

Este trabajo utiliza el contenido de Twitter para generar un indicador del nivel de atención asignado a la inflación en las discusiones públicas. El análisis corresponde a Argentina para el período 2012-2019. Los modelos de pronóstico estimados muestran que el indicador proporciona información valiosa sobre los niveles futuros de inflación. Los ejercicios fuera de la muestra confirman que el contenido de las redes sociales permite obtener ganancias en término de precisión de los pronósticos. Más allá de los pronósticos puntuales, el índice proporciona información valiosa sobre la incertidumbre de la inflación, es decir, el tamaño de los intervalos de confianza de los errores de pronóstico. El indicador propuesto se compara favorablemente con otros indicadores como el contenido de los medios, los tweets de los medios, la intensidad de búsqueda en Google y las encuestas de consumidores.

JEL CODE E37, E71

Acknowledgements
We would like to thank Sergio De Raco and Fabio Gasparri for their advice on how to process large json files and Luciano Campos for useful comments. We would also like to thank María Laura Antunez, Juan Cabral y Santiago López for very valuable research assistance. We benefited from useful comments received at the UNLP Economic Department Seminar and the 3rd Workshop on Complex System hosted by IIEP.

Índice

1 Introduction

Inflation has become a central topic in macroeconomic analysis. In this context, there is high value in expanding our ability to monitor and, more importantly, anticipate inflation dynamics. Social media content emerges as a potentially valuable tool to advance this agenda. That is, the large volume of messages exchanged in public discussions can be used to extract information regarding the likely path of inflation dynamics.

In this study, we analyze public discussions on the micro-blogging site Twitter. Our study focuses on the case of Argentina. This is a particularly interesting case of study since this is an economy where inflation has been a recurrent and highly disruptive phenomenon.

The empirical evidence indicates Twitter content anticipates inflation. More specifically, a simple indicator of the level of attention allocated to inflation provides valuable information regarding inflation levels and inflation uncertainty. Estimated forecasting models indicate that an increment in the attention index are followed to statistically and economically significant increments in expected inflation. Out-of-sample forecasts confirm that the index allows for gains in forecast accuracy. Complementarily, higher values of the attention index anticipate increments in inflation uncertainty as approximated by the interquartile range of next month inflation forecasts.

The information gains are different from and compare favorably with the information provided by lagged inflation and lagged devaluation rate. Also, analyses show that these information gains are substantive compared to those that result from using traditional macroeconomic indicators such as the level of economic activity, monetary aggregates and interest rates. Furthermore, the information content of four alternative indicators of expectations are also evaluated: Google search Volume, newspaper content, mass media tweets and a consumer survey. These analyses confirm that social media data constitutes a particularly valuable source of information regarding future inflation.

From a broad perspective, the current study is motivated by the idea that inflation is the result of an emergent evolving process (Heymann & Leijonhufvud 1995, Arifovic 1995, De Grauwe & Ji 2019). The emergence of adaptive forward looking behavior implies that traditional variables, such as interest rates, monetary aggregates and levels of activity, might not summarize all available information regarding the evolution of the process. Hence, indicators of subjective states that co-evolve with the price level are likely to contain valuable information regarding subsequent dynamics. Furthermore, in the case of volatile emerging economies, such as the case we analyze in this work, changes in policy regimes and limited policy maker credibility augment the relevance of subjective indicators.

The role for the level of social media attention as an indicator of future inflation and inflation uncertainty is also linked to the insights formalized by the rational inattention literature. According to this perspective, the allocation of attention is adjusted with the value of incoming information (Sims 2003, Mackowiak & Wiederholt 2015). The allocation of more attention to inflation might point to optimal responses to the arrival of news regarding the future evolution of inflation.

Complementarily, this study is also connected to macroeconomic models of sunspots or multiple equilibria (Benhabib & Farmer 1999, Ascari et al. 2019). In these models, the macroeconomic trajectory is, in part, determined by the coordination of behavior and expectations. More public discussion of the inflation rate might be associated to innovations in the coordination of price setting behavior. This innovations are not necessarily captured in a rapid or accurate manner by traditional macroeconomic indicators. In this way, social media content emerges as a valuable indicator of price level uncertainty and expected trajectory.

This paper contributes to a substantive literature on inflation forecasts (see, for example, the survey by Faust & Wright 2013). This literature analyzes the information content of macroeconomic variables, financial indicators and surveys (Stock & Watson 1999, Ang et al. 2007, Sharpe et al. 2020). In addition, this literature considers alternative modeling techniques (Ang et al. 2007, Schorfheide & Song 2015) and changes in the forecasting ability of models (Rossi & Sekhposyan 2010). The evidence reported in this work evaluates the information content of a novel measure based on social media messages. The results suggest that this indicator can be a valuable tool both for the generation of point forecasts and for the assessment of inflation uncertainty.

Our paper is also related with studies that show how media content and policy making documents provide information regarding future macroeconomic dynamics (see Baker et al. 2016, Thorsrud 2018, Sharpe et al. 2020 and Aromi 2020). Relatedly, other analyses demonstrate the value of social media content in analyses of high frequency developments in financial markets (see Bollen et al. 2011, Azar & Lo 2016, Oliveira et al. 2017).

In the next section, the data and the methodology are described. Section 3 analyzes inflation forecast and the attention index. Inflation uncertainty is analyzed in the following section. Alternative indicators of subjective states are evaluated in section 5. Conclusions are presented in section 6.

2 Data and methodology

As previously mentioned, we want to evaluate if Twitter messages contain valuable information regarding the evolution of inflation and whether these data can be used to forecast inflation in Argentina. We use two different types of information: traditional data (such as past inflation and exchange rates) and subjective indicators which result from summarizing a large collection of messages. The sample period is January 2012 through September 2019.

The first set of data is given by the consumer price index and the exchange rate. The consumer price index data is from the National Institute of Statistics and Census (INDEC).[1] The Argentine peso-US dollar exchange rate time series is from the Central Bank.[2] Given the value of each time series on month t (y_t), we compute the monthly variation which is given by the log-difference between month t and month $t-1$ values: Δy_t=log(y_t)-log(y_{t-1}).

Table 1 shows descriptive statistics for the inflation rate (Δcpi_t) and the evolution of the exchange rate (Δer_t).[3] The table shows that the average monthly inflation rate for the period was 2%. The period was characterized by high volatility as indicated by the standard deviation of, approximately, 1%. The inflation rate reaches its maximum sample value of 6% on April 2016, a few months after an important devaluation and following nationwide increments in utility rates. The inflation rate achieved its minimum value (0.2%) on August 2016 as a consequence of a reversal of the previously mentioned increments in utility rates. The average monthly devaluation rate for the sample period is approximately 3%. This period was also volatile in terms of the foreign

[1]The period covered in this analysis includes a sub-period in which official government statistics were not reliable (January 2007-December 2015) or unavailable due to a transition period toward normalization of government statistics (December 2015-June 2016). As a consequence, official data is complemented with data from alternative sources: from January 2012 through August 2012 we use a general price level indicator from a consulting firm (Buenos Aires City), from September 2012 through May 2016 we use the time series from the Statistical Departament of the City of Buenos Aires (https://www.estadisticaciudad.gob.ar/eyc/?cat=66). Since June 2016, the data are from INDEC (https://www.indec.gob.ar/indec/web/Nivel3-Tema-3-5).

[2]We use the monthly average Wholesale Foreign Exchange Rate (ARS/USD) Com. A 3500 exchange rate. http://www.bcra.gov.ar/

[3]The exchange rate is expressed following the convention selected by the data provider: number of local currency units (Argentine pesos) per unit of foreign currency (U.S. dollar).

exchange rate with a standard deviation of 5%, a maximum value of 25% and a minimum of −4%.

The second type of data is social media content that is used to construct an indicator of attention allocated to inflation by users in Argentina. Twitter messages were collected for the period 2012-2019. The data source corresponds to a random sample of 1% of the messages that is provided free of charge by Twitter through its API. These tweets were collected and are distributed by the Internet Archive (`https://archive.org/details/twitterstream`). A corpus of “Argentine" social media content is created selecting those tweets for which the user-reported location includes the word “Argentina". This results in a subset of approximately 70 million tweets.[4]

Given the corpus of Argentine tweets, an indicator of attention is built computing the frequency of the noun “inflation" and the adjective “inflationary".[5] This is a simple and transparent strategy through which the large collection of unstructured data is summarized. More specifically, let i_t represent the number of times a keyword is detected in messages corresponding to month t and n_t represent the total number of tweets corresponding to month t. Then, the corresponding value of the inflation attention index is given by $I_t = i_t/n_t$. A higher number is interpreted as more attention being allocated to inflation.

As shown in the last row of table 1, on average, there were three mentions to inflation per ten thousand tweets. An informal inspection of the index suggests that it is able to capture information related to inflation dynamics. The maximum value of the index, 9.75, corresponds to May 2018. This is a month of high volatility in the foreign exchange market and coincides with the start of a period marked by high inflation and a persistent economic and financial crisis. The minimum value of the index corresponds to July 2015. This is a period of relatively low inflation.

Table 1. Descriptive statistics

Sample period is 2012-2019. Data frequency is monthly. Δcpi_t: log difference of Consumer Price Index. Δer_t: log difference peso-dollar exchange rate. n_t : number of tweets (thousands). i_t : mentions of the term “inflation" or “inflationary". I_t : inflation attention index (multiplied by 10^4).

Variable	Mean	Median	St. Dev.	Q1	Q3	Minimum	Maximum
cpi_t	0.02	0.02	0.01	0.02	0.03	0.00	0.06
er_t	0.03	0.01	0.05	0.01	0.03	-0.04	0.25
n_t	711.95	713.53	260.82	450.77	929.01	192.90	1392.60
i_t	221.43	185.00	147.18	108.00	311.00	24.00	739.00
I_t	3.15	2.58	1.91	1.71	4.18	0.81	9.75

Figure 1 provides further evidence on the co-movement of inflation and the attention index. Three notable increments in the attention index coincide with three episodes of important increments in the inflation rate. Those three instances correspond to early 2014, the first half of 2016 and 2018. Additionally, the coefficient of correlation between inflation and the attention index is 0.47. Interestingly, suggesting that the index captures forward-looking information, the coefficient of correlation increases to 0.55 when the one-month-lagged attention index is considered.

[4] For a small subset of sample months (6 months), there are no Twitter data available from the Internet Archive collection. This is most likely due to random failures in the data collection process. To generate a complete time series, the information for these months was collected using Twitter´s advanced search tool.

[5] In spanish: “inflación" and “inflacionario/a/s".

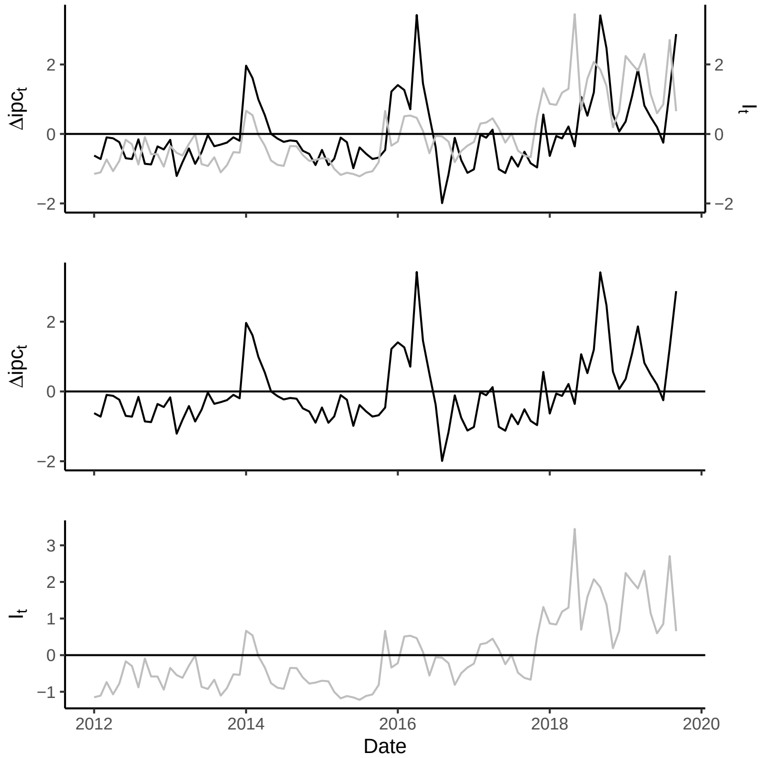

Figure 1. Inflation rate and Inflation Attention based on Twitter.
Note: To facilitate comparisons, both time series were standardized.

3 Inflation forecasts

In this section, the information content of the indicator of inflation attention is evaluated through a series of forecasting exercises. The forecasting models are given by an autoregressive specification that, depending on the specification, is complemented with an indicator of lagged Twitter content or the lagged change in the foreign exchange rate. The number of lags is selected minimizing the Bayesian Information Criterion.[6]

More formally, let cpi_t be the value of a Consumer Price Index in month t. The inflation rate computed on month t is given by $\Delta cpi_t = log(cpi_t) - log(cpi_{t-1})$. The baseline autoregressive model is given by:

$$\Delta cpi_{t+1} = \alpha + \sum_{s=0}^{p} \beta_s \Delta cpi_{t-s} + \mu_t \tag{1}$$

where μ_t is the error term.

Before incorporating the indicator of attention, we propose a second model with the the monthly devaluation rate as predictor. In this way, estimated expected inflation is conditioned on a larger information set. Let Δer_t represent the monthly devaluation rate corresponding to month t. Then, the second baseline model satisfies:

$$\Delta cpi_{t+1} = \alpha + \sum_{s=0}^{p} \beta_s \Delta cpi_{t-s} + \beta_{er} \Delta er_t + \mu_t \tag{2}$$

[6] Several unit root test are carried out in order to evaluate whether the inflation rate variable is non-stationary and possesses a unit root. In all cases, we have rejected the null hypothesis.

The predictive ability of social media content is evaluated through extended models that incorporate, as predictor, one of two specifications of the indicator of attention. The first specification is the original indicator described above. The second indicator can be interpreted as an adjusted metric of attention. More specifically, the second indicator is given by: $\hat{I}_t = I_t - \frac{\sum_{k=1}^{12} I_{t-k}}{12}$. That is, the indicator measures differences between the latest value of the indicator of attention and its average level during the previous year. This alternative specification allows for robustness tests. At the same time, it permits a simple exploration of the appropriateness of alternative specifications. Formally, the forecasting models used to estimate the information content of social media indices are given by the following equation:

$$\Delta cpi_{t+1} = \alpha + \sum_{s=0}^{p} \beta_s \Delta cpi_{t-s} + \beta_{er} \Delta er_t + \beta_I \Delta A_t + \mu_t \tag{3}$$

where A_t is equal to I_t or $\hat{I}_t$ and in some model specifications β_{er} is set equal to 0. The parameter ofinterest is β_I. μ_t is the error term. Models are estimated for the period 2012-2019. The predictors are standardized to facilitate the comparison of the economic significance of different estimated parameters.

Table 2 shows the estimations of the different specifications of the forecasting models. The baseline models indicate that lagged monthly inflation and lagged devaluation are statistically and economically significant predictors ofinflation. Adjusted R^2's suggest that these variables contain substantive information regarding subsequent levels ofinflation. [7]

The estimated extended models indicate that social media content adds information regarding future inflation levels. A one standard deviation increment in the indicator of attention anticipates a mean increment of approximately 0.4% in monthly inflation. It is worth noting that these observations do not depend on the baseline model under consideration or the specification of the social media content indicator. Adjusted R^2's point to noticeable gains in anticipatory ability. For example, in the first baseline model, the adjusted R^2 increases from 0.427 to more than 0.50 as summaries of social media content are incorporated as predictors.

Having established that social media provides valuable information regarding future levels ofinflation, we present additional analyses that characterize this information in more detail. First, we estimate a nonlinear model to explore the value of alternative specifications. Then, we evaluate if the information provided by the attention index overlaps with the information provided by traditional macroeconomic indicators such as interest rates, monetary aggregates or economic activity.

Are increments in attention as informative as drops in attention? To provide an answer to this question we use the adjusted metric of attention, $\hat{I}_t$, to compute a second indicator of attention that is equal to the adjusted metric of attention if positive and zero otherwise: $\hat{I}_t^+ = \max\{0, \hat{I}_t\}$. Then we estimate forecast models in which this new indicator is used as a predictor. In table 3 the estimated models suggest that increments in attention are particularly informative. More specifically, when the indicator ofincrements in attention, $\hat{I}_t^+$, is used as the predictor, the estimated coefficients and the metrics of model fit increase. Additionally, when the adjusted metric, $\hat{I}_t$, is incorporated as a second indicator, its coefficient is not significantly different from zero. While our ability to identify the best forecast model is constrained by sample size, these evaluations suggest that there are gains associated to considering more flexible specifications.

Is the information provided by social media content different from the information provided by traditional indicators? We consider three traditional indicators to address this question: economic activity, interest rates, and monetary aggregates. The monthly indicator of economic activity, ea_t, is from the national statistics office.[8] The monetary aggregate, m_t, is the monetary base.[9] Interest rates, ir_t, corresponds to 30 through 59-day term deposits.[10]

[7] The results remains when we incorporate the inflation rate given by the monthly variation of the consumer prices index seasonally adjusted.

Table 2. Forecasting models

Sample period is 2012-2019. cpi_t: difference in logs of Consumer Price Index. er_t: difference in logs of the exchange rate. I_t : ratio between total mentions of the term "inflation" and total of Tweets. $\hat{I}_t$: relative level of inflation attention. Standardized coefficients.

-0.2cm

	(1)	(2)	(3)	(4)	(5)	(6)
cpi_t	0.008***	0.006***	0.006***	0.006***	0.005***	0.006***
	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.0004)
er_t		0.003***			0.002***	0.002***
		(0.001)			(0.001)	(0.001)
I_t			0.004***		0.004***	
			(0.001)		(0.001)	
$\hat{I}_t$				0.004***		0.003***
				(0.001)		(0.001)
Constant	0.025***	0.025***	0.025***	0.025***	0.025***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	92	92	92	80	92	80
R^2	0.433	0.495	0.531	0.516	0.560	0.539
Adjusted R^2	0.427	0.483	0.520	0.504	0.545	0.521
F Statistic	68.678***	43.580***	50.305***	41.095***	37.360***	29.614***

Note: standard errors in parentheses are estimated following Newey & West (1987, 1994).* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 3. Nonlinear models

Sample period is 2012-2019. cpi_t: difference in logs of Consumer Price Index. er_t: difference in logs of the exchange rate. $\hat{I}_t$: relative level ofinflation attention. $\hat{I}^+_t$: positive values of the adjusted metric of attention. Standardized coefficients.

	(1)	(2)	(3)
cpi_t	0.006***	0.006***	0.006***
	(0.0004)	(0.001)	(0.001)
er_t	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)
$\hat{I}_t$	0.003***		−0.002
	(0.001)		(0.002)
$\hat{I}^+_t$		0.004***	0.005*
		(0.001)	(0.003)
Constant	0.025***	0.025***	0.025***
	(0.001)	(0.001)	(0.001)
Observations	80	80	80
R^2	0.539	0.556	0.557
Adjusted R^2	0.521	0.538	0.534
F Statistic	29.614***	31.670***	23.602***

Note: standard errors are estimated following Newey & West (1987, 1994).* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 4 shows the estimated models that incorporate these traditional variables as predictors after standardization. The estimated coefficients associated to these variables have the expected sign. Nevertheless, statistical significance is consistently observed only in the case ofinterest rates. Importantly, for all model specifications, the coefficient corresponding to the indicator of Twitter content is positive and significant. Additionally, the estimated coefficient remains mostly unaltered as different macroeconomic variables are incorporated to the model. These results suggest that the attention index provides valuable information that is different from the information provided by traditional economic indicators.

Table 4. Attention indices vs. traditional variables

Sample period is 2012-2019. cpi_t: difference in logs of Consumer Price Index. er_t: difference in logs of exchange rate. $\hat{I}_t$: relative level ofinflation attention. ea_t : difference in logs ofindicator of economic activity. m_t : difference in logs of monetary base. ir_t : interest rates for term deposits. Standardized coefficients.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
cpi_t	0.006***	0.006***	0.008***	0.007***	0.008***	0.006***	0.006***	0.004***
	(0.0004)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
er_t	0.003***	0.002***						
	(0.001)	(0.001)						
$\hat{I}_t$		0.003***		0.004***		0.004***		0.004***
		(0.001)		(0.001)		(0.001)		(0.001)
ea_t			0.001	0.002				
			(0.001)	(0.001)				
m_t					0.001	0.001		
					(0.001)	(0.001)		
ir_t							0.003**	0.004***
							(0.001)	(0.001)
Constant	0.025***	0.025***	0.025***	0.025***	0.025***	0.025***	0.025***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	92	80	91	80	91	80	92	80
R^2	0.495	0.539	0.440	0.540	0.433	0.519	0.475	0.580
Adjusted R^2	0.483	0.521	0.428	0.522	0.420	0.500	0.463	0.563
F Statistic	43.580***	29.614***	34.609***	29.765***	33.557***	27.340***	40.244***	34.948***

Note: standard errors are estimated following Newey & West (1987, 1994).* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Summarizing, the evidence reported above suggests that indices based on social media content have valuable information regarding future levels ofinflation. These findings are robust to changes in the set of predictors and the specification of the indicator. Also, the information provided by the attention index is different from that provided by traditional macroeconomic indicators.

[8] The seasonally adjusted monthly economic activity estimator (EMAE according to its Spanish acronym). https://www.indec.gob.ar/indec/web/Nivel4-Tema-3-9-48 .

[9] More specifically, the indicator is the seasonally adjusted monetary base. M0 (monetary base): Sum of monetary circulation (notes and coins issued by the Central Bank of Argentina), and deposits in pesos held by financial institutions in a current account with this Institution. End-of-Month Stock. http://www.bcra.gov.ar/. The results are basically the same when we use M1.

[10] Monthly average interest rate. We decided to use interest rates for term deposits because in the case of Argentina there is no interest rate that was used consistently as an instrument of monetary policy during the period under analysis.

3.1 Out-of-sample Forecast

To provide further insights on the information content of social media we implement out of sample forecasts exercises in which models are trained recursively with past information. The performance of forecast generated by the baseline autoregressive model are compared to forecasts produced by models that incorporate an additional predictor. Four predictors are considered: monthly devaluation rate (er_t), the inflation attention index (I_t), the adjusted level of attention ($\hat{I}_t$) and increments in attention ($\hat{I}_t^+$).

Each forecast model is evaluated computing the root-mean-square prediction error (RMSE). For extended models these measure of accuracy is also expressed as a fraction of the RMSE of the baseline model. The performance of the models is assessed using two alternative starting dates for pseudo out-of-sample forecast exercise. The starting dates are selected so that the smallest training subsample represents 60% and 80% of the full sample respectively. Following Faust et al.(2013), resampling techniques are implemented to compute the statistiscal significance of the differences in accuracy.[11]

Table 5 shows the results for out-of-sample forecast exercises. For all extended models, the estimated forecast accuracy is higher than that observed in the case of the baseline model. These differences are statistically significant in all but one case. For forecasts generated by a single predictor, the strongest performance is observed in the case of the original inflation attention index (I_t). The last row of the table shows that forecast combinations allow for further gains in accuracy. In summary, these out-of-sample forecast exercises provide further support to the idea that social media content provides valuable information regarding future levels of inflation.

Table 5. Out-of-sample forecasts

Sample period is 2012-2019. Data frequency is monthly. cpi_t: difference in logs of Consumer Price Index. $\hat{er}_t$: difference in logs of exchange rate considering its average accumulated value during the last semester. I_t : ratio between total mentions of the term "inflation" and total of Tweets. $\hat{I}_t$: relative level of inflation attention. $\hat{I}_t^+$: positive changes on the relative level of inflation attention. Forecast combinations are implemented through simple averages. p-values in brackets.

Forecasts begin	01/2017 (60%)		05/2018 (80%)	
	RMSE	Ratio	RMSE	Ratio
Baseline	0.0100		0.0125	
er_t	0.0097	0.966 [0.05]	0.0122	0.978 [0.118]
I_t	0.0090	0.897 [0.00]	0.0102	0.812 [0.02]
$\hat{I}_t$	0.0096	0.954 [0.00]	0.0112	0.90 [0.04]
$\hat{I}_t^+$	0.0096	0.952 [0.00]	0.0110	0.877 [0.04]
Forecast combination	0.0088	0.874 [0.00]	0.0104	0.831 [0.00]

[11] We estimate two models: (a) a restricted model that involves estimating an AR(4) process for cpi_t and (b) an unrestricted model that consists of a regression of cpi_t on four lags of itself and two predictors: er_t and I_t. In each bootstrap replication (500 replications), we then resample the residuals of the unrestricted model using wild bootstrap and construct a bootstrap sample of cpi_t (cpi_t^{boot}) using these resampled residuals, together with the coefficients from the restricted model.

4 Inflation uncertainty

Beyond inflation forecasts, social media can be conjectured to provide information regarding inflation uncertainty. This is a question of interest since inflation uncertainty has important economic consequences (Huizinga 1993, Heymann & Leijonhufvud 1995 and Elder 2004). Our metric seems adequate for this task since higher inflation uncertainty is likely accompanied by a higher level of attention that can be inferred from social media messages. More specifically, under rational inattention, the allocation of attention is adjusted with the value of incoming information (Sims 2003, Mackowiak & Wiederholt 2015). Hence, the allocation of more attention to inflation is consistent with the arrival of valuable news regarding the future evolution of inflation. Complementarily, considering the literature on multiple equilibria or sunspots (Benhabib & Farmer 1999, Ascari et al. 2019), more public discussion of the inflation might linked to innovations in the coordination of price setting behavior.

Our first evaluation involves estimating quantile regressions that characterize the distribution of inflation shocks as a function of three variables of interest. Formally, shocks are approximated by residuals of an autoregressive model:

$$\hat{e}_{t+1} = \Delta cpi_{t+1} - [\hat{\alpha} + \hat{\beta}_{cpi}\Delta cpi_t + \hat{\beta}_I \hat{I}_t + \hat{\beta}_{er}\Delta er_t]$$

Then, quantile $\tau \in [0,1]$ of the shock $\hat{e}_{t+1}$ conditional on the value of indicator x_t is modeled as an affine function:

$$Q_{\hat{e}_{t+1}|x_t}(\tau) = \alpha_\tau + \beta_\tau x_t$$

where the x_t can be one of three indicators: inflation (Δcpi_t), devaluation rate (Δer_t), or the adjusted indicator of inflation attention $\hat{I}_t$.

Figure 2 and table 6 provide information on the estimated quantiles. The estimations suggest that the attention index is associated to important changes in uncertainty. In particular, the interquartile range of the shock is estimated to increase from 0.47 to 0.83 as the attention index increases from one standard deviation below the mean to one standard deviation above the mean. This strong association between the attention index and uncertainty, proxied by forecast errors, are not observed when quantiles are estimated as a function of lagged inflation or lagged devaluation.

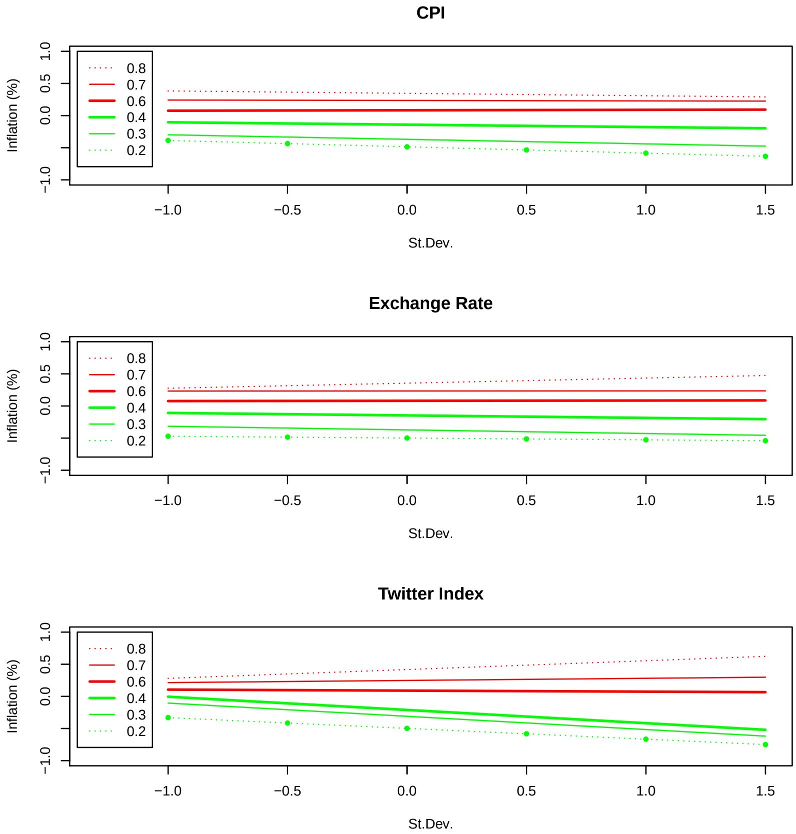

Figure 2. Quantile regressions

Note: Given $\tau \in \{0,1\}$ quantile $Q_{\hat{e}_{t+1}|x_t}(\tau)$ is modeled as $Q_{\hat{e}_{t+1}|x_t}(\tau) = \alpha_\tau + \beta_\tau x_t$. Estimated quantiles are conditioned on x_t with $x_t \in \{\Delta cpi_t, \Delta er_t, \hat{I}_t\}$. To facilitate comparisons, in each case, the indicator on which quantiles are conditioned, x_t, was standardized.

To complement the previous results, we evaluate the association between inflation uncertainty and social media attention estimating simple forecasting models. Following Rossi & Sekhposyan (2015), inflation uncertainty is proxied using the mean absolute forecast error: $iu_t = |\hat{e}_{t+1}|$. Then, we propose simple empirical models in which uncertainty is function of indicators $x_t \in \{\Delta cpi_t, \Delta er_t, \hat{I}_t\}$. Formally, when all indicators are incorporated, the inflation uncertainty forecast model is given by:

$$log(iu_{t+1}) = \alpha + \beta_0 log(iu_t) + \beta_{cpi}\Delta cpi_t + \beta_I \hat{I}_t + \beta_{er}\Delta er_t + u_{t+1}$$

Table 6 confirms that the attention index provides valuable information regarding inflation uncertainty. A one standard deviation increment in the attention index is associated to a 30% increment in the expected mean absolute error. In contrast, no robust association is found when the other two indicators are considered. In the case of the devaluation rate, no statistically significant relationship is found. In the case of lagged inflation, a positive relationship is found in the case of the univariate model. This association loses statistical significance once the attention index is incorporated as a predictor in the model.

Table 6. Inflation uncertainty forecasts

Sample period is 2012-2019. cpi_t: difference in logs of Consumer Price Index. er_t: difference in logs of exchange rate. $\hat{I}_t$: relative level of inflation attention. Standardized coefficients.

	[1]	[2]	[3]	[4]
$log(iu_{t+1})_{t-1}$	−0.103	−0.075	−0.134	−0.146*
	(0.091)	(0.113)	(0.089)	(0.084)
Δcpi_t	0.195***			0.124
	(0.073)			(0.084)
Δer_t		0.093		−0.066
		(0.096)		(0.108)
$\hat{I}_t$			0.307***	0.295**
			(0.102)	(0.122)
Constant	−5.700***	−5.696***	−5.697***	−5.699***
	(0.112)	(0.111)	(0.112)	(0.113)
Observations	79	79	79	79
R^2	0.039	0.012	0.090	0.104
Adjusted R^2	0.013	−0.014	0.067	0.055
F Statistic	1.531	0.447	3.781**	2.138*

Note: standard errors in parentheses are estimated following Newey & West (1987, 1994).*p<0.1; **p<0.05; ***p<0.01

5 Evaluation of other subjective indicators

The previous analyses show that social media content provides valuable information regarding future levels of inflation and inflation uncertainty. One open question is whether other indicators of subjective states display similar capacity to provide information regarding future inflation. In this subsection, we estimate inflation forecast models that allow for a comparison of the information content of multiple alternative indicators of subjective states.

Three types of indicators are considered. First, we evaluate search intensity indices from Google Trends. We consider two keywords: “inflacion"(inflation) and “dolar" (dolar). The selection of “dolar" as a keyword responds to the strong association between the exchange rate and the rate of inflation. Also, we consider two indicators of media attention. One index is constructed processing the full text of the economic section of “La Nación" a prominent Argentine newspaper[12]. The second mass media index is constructed processing a collection of 1.4M tweets from 6 major news outlets: “ambitocom", “clarincom", “cronistacom", “infobae", “lanacion and “perfilcom". Finally, we also use the index of household inflation expectations from CFI-UTDT. More specifically, in the absence of an indicator of expected inflation for the next month, we use the median expected inflation over the next 12 months.

As in the previous exercises, we estimate simple autoregressive models that incorporate an additional predictor. Table 7 shows the estimated forecasting models for the alternative subjective indicators. To facilitate comparisons, each subjective indicator was standardized. Interestingly, mass media content and mass media tweets fail to provide information regarding future inflation. One plausible explanation for this result is that, in the case under analysis, mass media plays a minor role in the formation of macroeconomic expectations. Nevertheless, a more comprehensive analysis is needed to arrive to more informative evaluation of this expla-

[12]The results shown below remain mostly unaltered when we incorporate the full text of the economic section of “Pagina 12" and “Ambito Financiero".

nation. Negative results are also observed in the case of household surveys. While preliminary, this evidence points to limits in our capacity to extract subjective information through questionnaires.

In contrast, search intensity indicators are shown to contain valuable information. In particular, a strong performance is observed in the case of the indicator of search intensity for the keyword “dollar". In consistence with the previous sections, the evidence linked to search intensity suggest that indicators of online behavior of a large quantity of users constitutes a valuable tool to extract forward-looking macroeconomic information.

Table 7. Attention index vs. proxies of subjective states

This table shows the results associated with in-sample forecast exercise. Sample period is 2012-2019. Baseline : autoregressive benchmark model. I_t : ratio between total mentions of the term “inflation" and total of Tweets (a mention each 104 Tweets). $GT-inflation$: searches of term “inflation" in Google. $GT-dollar$: searches of term “dollar" in Google. $Newspaper$: frequency of term “inflation" in the newspaper La Nacion. $Mass\ media\ tweets$: ratio between total mentions of the term “inflation" and total of Tweets published by Mass media. $Cons.Surv.$: Inflation expectations reported by individuals in Di Tella Survey. Standardized coefficients.

	Baseline	I_t	GT-inflation	GT-dollar	Newspaper	Mass media tweets	Cons. Surv.
$\hat{\alpha}$	0.025***	0.025***	0.025***	0.024***	0.025***	0.024***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\beta}_0$	0.008***	0.005***	0.006***	0.005***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\beta}_{ind}$		0.004***	0.002**	0.004***	-0.001	0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Adj. R^2	0.427	0.511	0.451	0.494	0.428	0.410	0.414

Note: standard errors in parentheses are estimated following Newey & West (1987, 1994).*p<0.1; **p<0.05; ***p<0.01

6 Conclusions

This paper examines the information content of social media messages. More specifically, we analyze “Argentine" tweets for the period 2012-2019. The evidence indicates that Twitter messages provide valuable information regarding expected inflation and inflation uncertainty. The information content is economically significant. An increment of one standard deviation in the index is associated with by an increment of 0.4% in expected inflation in the following month. The information content of the index is different from that provided by traditional macroeconomic indicators. These findings are robust to changes in the specification of the forecast exercise.

There are several directions in which these exercises can be extended. First, in this work, attention to inflation was approximated using an extremely simple but transparent strategy to summarize unstructured information. The use of natural language processing models could allow for gains in the capacity to extract information from Twitter messages. In a similar direction, models of community detection could be used to discover clusters of users whose content is particularly informative. Finally, this study evaluated regularities using monthly time series. Analyses at higher frequencies can provide further insights regarding the relationship between social media content and inflation dynamics.

References

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better?. *Journal of monetary Economics*, 54(4), 1163-1212.

Arifovic, J. (1995). Genetic algorithms and inflationary economies. *Journal of Monetary Economics*, 36(1), 219-243.

Aromí, J. Daniel, (2020). Linking Words in Economic Discourse: Implications for Macroeconomic Forecasts, *International Journal of Forecasting*, 36(4), 1517-1530.

Ascari, G., Bonomolo, P., & Lopes, H. F. (2019). Walk on the wild side: Temporarily unstable paths and multiplicative sunspots. *American Economic Review*, 109(5), 1805-42.

Azar, P. D., & Lo, A. W. (2016). The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds. *The Journal of Portfolio Management*, 42(5), 123-134.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4), 1593-1636.

Benhabib, J., & Farmer, R. E. (1999). Indeterminacy and sunspots in macroeconomics. *Handbook of macroeconomics*, 1, 387-448.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8.

De Grauwe, P., & Ji, Y. (2019). Inflation targets and the zero lower bound in a behavioural macroeconomic model. *Economica*, 86(342), 262-299.

Faust, J., Gilchrist, S., Wright, J. H., & Zakrajšsek, E. (2013). Credit spreads as predictors of real-time economic activity: a Bayesian model-averaging approach. *Review of Economics and Statistics*, 95(5), 1501-1519.

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In *Handbook of economic forecasting* (Vol. 2, pp. 2-56). Elsevier.

Heymann, Daniel & Leijonhufvud, Axel (1995). High Inflation: The Arne Ryde Memorial Lectures, Oxford University Press.

Mackowiak, B., & Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99(3), 769-803.

Newey W. K. West K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703-708.

—-. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4), 631-653.

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, 73, 125-144.

Rossi, B., & Sekhposyan, T. (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when?. *International Journal of Forecasting*, 26(4), 808-835.

—-. (2015). Macroeconomic uncertainty indices based on nowcast and forecast error distributions. *American Economic Review*, 105(5), 650-55.

Schorfheide, F., & Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business Economic Statistics*, 33(3), 366-380.

Sharpe, S. A., Sinha, N. R., & Hollrah, C. A. (2020). The Power of Narratives in Economic Forecasts, *Finance and Economics Discussion Series*, Federal Reserve Board.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2), 293-335.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business Economic Statistics*, 1-17.

Sobre los Documentos de Trabajo

La serie de Documentos de Trabajo del IIEP refleja los avances de las investigaciones realizadas en el instituto. Los documentos pasan por un proceso de evaluación interna y son corregidos, editados y diseñados por personal profesional del IIEP. Además de presentarse y difundirse a través de la página web del instituto, los documentos también se encuentran disponibles en la biblioteca digital de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires, el repositorio digital institucional de la Universidad de Buenos Aires, el repositorio digital del CONICET y en la base IDEAS RePEc.

CONICET UBA

I I E P

INSTITUTO INTERDISCIPLINARIO DE ECONOMÍA POLÍTICA

Universidad de Buenos Aires | Facultad de Ciencias Económicas

Av. Córdoba 2122 - 2º piso (C1120 AAQ)
Ciudad Autónoma de Buenos Aires, Argentina
+54 11 5285-6578 | **www.iiep.economicas.uba.ar**

@iiep_oficial