

# News Sentiment and Cross-Country Fluctuations\*

Samuel P. Fraiberger

Northeastern Network Science Institute  
177 Huntington Avenue, Boston, MA 02114  
Harvard Institute for Quantitative Social Sciences  
1737 Cambridge St, Cambridge, MA 02138  
s.fraiberger@neu.edu

## Abstract

What is the information content of news-based measures of sentiment? How are they related to aggregate economic fluctuations? I construct a sentiment index by measuring the net amount of positive expressions in the corpus of Economic news articles produced by Reuters over the period 1987 - 2013 and across 12 countries. The index successfully tracks fluctuations in Gross Domestic Product (GDP) at the country level, is a leading indicator of GDP growth and contains information to help forecast GDP growth which is not captured by professional forecasts. This suggests that forecasters do not appropriately incorporate available information in predicting future states of the economy.

## 1 Introduction

To date, there is limited cross-country evidence for the role that sentiment might play in explaining aggregate economic fluctuations. In this paper, I show how measures of aggregate beliefs extracted from news articles are related to GDP growth. I build upon a recent and fast-growing literature which relates information extracted from text to economic and financial variables ((Tetlock, 2007), (Baker et al., 2016), (Garcia, 2013)). The approach commonly used in this literature measures sentiment using pre-existing dictionaries.<sup>1</sup>

\*The author gratefully acknowledges financial support from the IMF Research Department.

<sup>1</sup>An alternative approach employed in (Choi and Varian, 2012) uses Google search results to forecast near-term values of economic indicators.

I build my sentiment index by measuring the net amount of positive expressions in the collection of Economic news articles from Reuters covering 12 countries over the period 1987 - 2013. The index successfully tracks GDP growth over time and across countries. Is sentiment a leading indicator of GDP growth? I estimate an autoregressive model GDP growth to which I add news-based sentiment measures. Coefficients on news-based sentiment measures are jointly significant at the country level for 10 out of 12 countries in my sample. Sentiment variables reduce in-sample forecast errors of GDP growth by 9.1% on average across countries compared to an autoregressive process. This indicates that news sentiment is a leading indicator of GDP growth.

Do news-based sentiment measures simply aggregate other well-established leading indicators? I test whether news-based sentiment measures contain information which is not reflected in professional forecasters' expectations. I run predictive regressions of annual GDP growth on consensus forecasts data at different forecasting horizons. I then add to the regressions my news sentiment index measured prior to the release of the consensus forecasts. Including sentiment reduces in-sample forecast errors by 19% on average across countries. News-based sentiment measures contain information which is not included in forecasters' expectations. Reductions in forecast errors are larger for longer forecasting horizons, which reflect the fact the long-term forecast are inherently hard. Reductions in forecast errors are also larger during bad times, which indicates that forecasters might be underreacting to bad news.

	# Articles	Avg. # Words	Start Date	End Date
Argentina	19007	338	6/1987	9/2013
Australia	31792	456	12/1998	7/2013
Brazil	32952	315	6/1987	9/2013
India	64306	384	6/1987	9/2013
Indonesia	22791	356	6/1987	9/2013
Japan	69607	408	12/1998	3/2012
New Zealand	20238	498	12/1998	8/2013
South Africa	30319	342	6/1987	9/2013
South Korea	32203	348	5/1987	9/2013
Sweden	10106	318	12/1998	11/2012
Switzerland	21499	351	12/1998	5/2013
United Kingdom	86182	422	12/2000	1/2013

Table 1: News Corpus Summary Statistics.

## 2 News-based Sentiment

### 2.1 Data Description

My dataset contains news articles extracted from Factiva.com, an online database which provides access to news archives from around the world. One can retrieve articles by querying a set of tags such as the source, the main topics and the locations associated with an article. A proprietary algorithm attributes topics and location tags to articles and is constant across the database.

Table (1) presents summary statistics of the news articles in my corpus. I focus on articles produced by Reuters News, which is the most comprehensive stream of news over time and across countries. I have collected all the news articles containing the topics tag "Economic news" and covering a set of 12 countries: Argentina, Australia, Brazil, India, Indonesia, Japan, New Zealand, Sweden, Switzerland, South Africa, South Korea, and the United Kingdom. The time window is December 1998 to September 2013 for developed countries (Australia, Japan, New Zealand, Sweden, Switzerland, and the United Kingdom), and June 1987 to September 2013 for developing countries (Argentina, Brazil, India, Indonesia, South Africa and South Korea). For each article, I have automatically extracted the timestamp, the title, the main text, the topics and the locations tags.

Positive	IDF	Negative	IDF
strong	1.95	limited	1.10
support	2.08	debt	1.61
help	2.08	cut	1.79
recovery	2.30	crisis	1.95
good	2.48	late	2.30
gains	2.56	deficit	2.30
boost	2.56	risk	2.30
easing	2.64	concerns	2.56
agreement	2.64	recession	2.64
highest	2.64	unemployment	2.64
better	2.77	weak	2.64
positive	2.77	decline	2.64
confidence	2.83	slowdown	2.83
steady	2.83	problems	2.83
agreed	2.83	risks	2.89
balance	2.83	concern	2.94
helped	2.83	slow	3.00
open	2.89	losses	3.04
strength	3.09	fears	3.04
stronger	3.09	negative	3.04
provide	3.14	fail	3.04

Table 2: The 20 Most Frequent Negative and Positive Words.

### 2.2 Text Processing

I combine dictionaries of positive and negative words compiled by (Loughran and McDonald, 2011) for financial texts and by (Young and Soroka, 2012) for political and economic texts. I search for inflections of each word in these dictionaries which are present in my corpus. Given a root tonal word (e.g. "lose"), I retrieve all the inflected words in the news corpus ("losing", "looser", "lost", "loss", etc ...) and add them to the dictionaries. I check the relevance of the most frequent words and eliminate the ones which are irrelevant. My dictionary of positive words contains 3,527 items and the one with negative words contains 7,109 items.

Table 2 shows that the most frequent positive and negative words indeed reflect the sentiment typically associated with economic and financial outcomes.

Here is an example of an article in which the main location tag is Argentina (in bold) and one of the topic tags is "Economic news" (in bold):<sup>2</sup>

<sup>2</sup>The words included in my dictionaries are underlined.

Title: Argentina's Peronists defend Menem's labor reforms.

Timestamp: 1996-09-02

Text: BUENOS AIRES, Sept 2 (Reuters) - The Argentine government Monday tried to counter criticisms of President Carlos Menem's proposals for more flexible labor laws, arguing that not just workers would contribute to new unemployment insurance. Menem angered trade unions, already in disagreement over his fiscal austerity programmes, by announcing a labor reform package Friday including suspending collective wage deals and replacing redundancy payouts with unemployment insurance.

Topics: Labor/Personnel Issues, Corporate/Industrial News, Economic/Monetary Policy, **Economic News**, Political/General News, Labor Issues, Domestic Politics

Locations: **Argentina**, Latin America, South America

The dictionary-based approach is straightforward and transparent, yet some words are not properly classified. To improve accuracy, I normalize 373 negative forms such as "no longer", "neither", "not having", etc ... to "not" as proposed in (Young and Soroka, 2012). I then build a second pair of lists of positive and negative expressions which appear preceded by a "not". A positive (negative) word preceded by a "not" is classified as negative (positive). Finally, I normalize 783 ambiguous expressions to correctly account for their tone. For instance, the expression "lose support" would be classified as neutral, so I normalize it to be counted as negative.

### 2.3 Sentiment Index

Using this classification of tonal expressions, a simple measure sentiment is the difference between the fraction of positive expressions and the fraction of negative expressions in each article. This measure is unlikely to capture all the nuances of a text, but it is likely to give a good indication of how news tone varies across country and over time.

Let  $t_{ij}$  be the number of occurrences of word  $i$  in article  $j$ . Let  $n_{ij}$  ( $p_{ij}$ ) be the number of occurrences of negative (positive) word  $i$  in document  $j$ . Corre-

spondingly, let  $\bar{p}_{ij}$  ( $\bar{n}_{ij}$ ) the number of occurrences of negative (positive) word  $i$  in document  $j$  preceded by a "not".

The positivity of article  $j$  is given by:

$$\pi_j = \frac{\sum_i p_{ij} + \sum_i \bar{p}_{ij} - \sum_i \bar{n}_{ij}}{\sum_i t_{ij}}. \quad (1)$$

In the numerator, the first term corresponds to the weighted sum of all the positive words. The second term corresponds to the weighted sum of negative words preceded by a "not". The last term corresponds to the weighted sum of positive words preceded by a "not".

Similarly, the negativity of article  $j$  is given by:

$$\nu_j = \frac{\sum_i n_{ij} + \sum_i \bar{n}_{ij} - \sum_i \bar{p}_{ij}}{\sum_i t_{ij}}. \quad (2)$$

The net positivity of article  $j$  is given by:

$$s_j = \pi_j - \nu_j. \quad (3)$$

## 3 Sentiment and GDP Growth

Figure (1) shows that my sentiment index successfully tracks fluctuations in GDP growth at the country level. The first natural question is whether or not sentiment is a leading indicator of GDP growth.

### 3.1 Granger Causality Tests

To show this, I estimate the autoregressive distributed lag model described by equation (4):

$$y_{t,c} = \alpha + \sum_{i=1}^p \beta_{i,c} y_{t-i,c} + \sum_{i=1}^q \gamma_{i,c} \pi_{t-i,c} + \sum_{i=1}^q \zeta_{i,c} \nu_{t-i,c} + \epsilon_{t,c}, \quad (4)$$

where  $y_{t,c}$  is the log GDP growth between  $t$  and  $t+3$  months in country  $c$  and  $\epsilon_{t,c}$  is the error term. I first estimate an autoregressive process of GDP growth at a quarterly frequency and at the country level by choosing the number of lags  $p$  which minimizes the AIC criterion. I then add monthly lags of positive

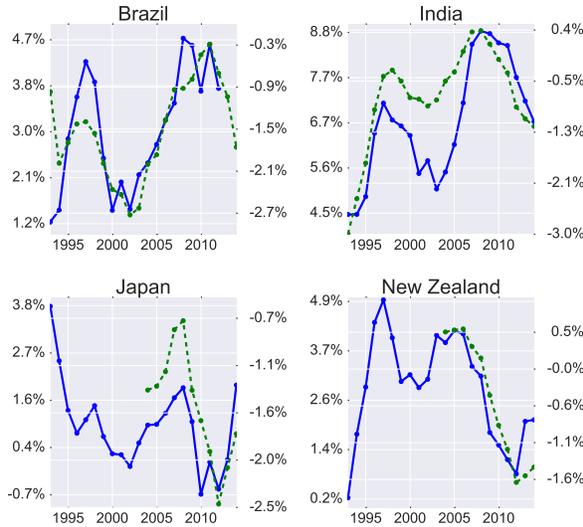


Figure 1: News Sentiment and GDP growth per Country. Time series of my benchmark sentiment index (dotted line) and GDP growth (full line) at a yearly frequency. Sentiment measures are computed at the article level and averaged at a yearly frequency. I then apply a 3-year moving average to smooth the series. News articles come from Factiva.com, GDP growth comes from the International Financial Statistics Database (IFS).

and negative sentiment (averaged at a monthly frequency), again choosing the number of lag values  $q$  using the AIC criterion.

Table (3) shows that lags of negative sentiment are a leading indicator of GDP growth at the country level for 9 out of the 12 countries in my sample. Lags of positive sentiment are a leading indicator of GDP growth for half of the country in my sample. This evidence is consistent with previous literature using news-based measures of sentiment which finds that most of the textual information is contained in negative words ((Loughran and McDonald, 2011)). In the case of India however, while I cannot reject the hypothesis that lags of negative sentiment are jointly equal to zero, I can reject the hypothesis that lags of positive sentiment are jointly equal to zero. This suggests that positive sentiment measures might also be worth considering as a leading indicator of GDP growth.

Figure (2) shows that on average across countries, forecast errors of next quarter GDP growth diminish by 9.1% when news-based sentiment measures are included in the ADL(p,q) model described by equa-

	F [ $\pi$ ]	p [ $\pi$ ]	F [ $\nu$ ]	p [ $\nu$ ]	F [ $\pi, \nu$ ]	p [ $\pi, \nu$ ]	N
Argentina	8.76**	1.7E-06	4.67**	9.8E-04	4.87**	2.6E-05	86
Australia	2.34	1.1E-01	3.65**	3.2E-02	1.84	1.3E-01	64
Brazil	1.81	1.3E-01	4.87**	1.0E-03	3.04**	4.4E-03	66
India	2.55**	3.1E-02	1.01	4.3E-01	1.94*	5.1E-02	63
Indonesia	0.82	4.5E-01	5.94**	5.1E-03	3.03**	2.7E-02	57
Japan	8.25**	5.9E-03	7.61**	8.0E-03	5.27**	8.2E-03	61
New Zealand	3.69**	3.1E-02	3.33**	4.3E-02	2.13*	8.9E-02	64
South Africa	1.25	2.6E-01	1.46	1.5E-01	1.36	1.6E-01	101
South Korea	4.82**	5.9E-06	5.34**	1.4E-06	4.13**	1.4E-06	98
Sweden	5.76**	6.4E-04	3.61**	1.1E-02	3.60**	2.1E-03	63
Switzerland	2.00	1.6E-01	2.74	1.0E-01	1.50	2.3E-01	64
United Kingdom	0.14	7.1E-01	4.07**	4.9E-02	2.14	1.3E-01	55

Table 3: Granger Causality Tests. Significance tests of the coefficient estimates of the autoregressive distributed lag model described by equation (4) at a quarterly frequency and at the country level. The number of lags  $p$  and  $q$  are chosen using the AIC criterion. I test for the joint significance of lags of positive sentiment  $\pi$  (column 1 and 2), lags of negative sentiment  $\nu$  (column 3 and 4), and the union of lags of positive and negative sentiment ( $\pi, \nu$ ) (column 5 and 6). For each test of joint significance, I report F-statistics and p-values. \*\* and \* indicate that coefficients are jointly significantly different from zero at the 0.05 and 0.10 levels or better, respectively. News articles come from Factiva.com, GDP growth comes from the International Financial Statistics Database (IFS).

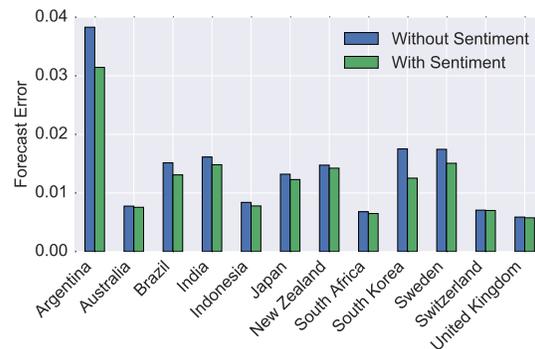


Figure 2: Reductions in GDP Growth Forecast Errors Using Sentiment. Forecast errors (measured by the regressions' Root Mean Square Error) of predictive regressions of GDP growth estimated with an AR(p) process (in blue) and when news-based measures of sentiment are also included in the regression (in green). News articles come from Factiva.com, GDP growth comes from the International Financial Statistics Database (IFS).

tion (4) compared to an AR(p) process.<sup>3</sup>

### 3.2 News Sentiment and Consensus Forecast

Several aggregate time series (e.g. weekly initial jobless claims, monthly payroll employment, etc...) are well known for containing information to help measure current economic conditions ((Aruoba et al., 2009)). Does my sentiment index simply combine information already contained in these well-known leading indicators? Obtaining data on leading indicators of GDP growth across countries is challenging, but these leading indicators should presumably be included in professional forecasters' information set. Since 1989, Consensus Economics Inc. provides a monthly survey of professional forecasters who are asked to forecast annual GDP growth across countries. For each realization of yearly GDP growth, the dataset contains GDP growth forecasts made by public and private economic institutions for each horizon  $h=1, \dots, 24$  months. (Fildes and Stekler, 2002) show that survey-based consensus forecasts are most of the time more accurate than those generated by time series models. The other advantage of forecasts produced by Consensus Economics is its common format for a large cross section of emerging market countries. If professional forecasters use all available information in producing their forecasts, the information contained in my news-based sentiment measures should not reduce the forecast errors of predictive regressions of GDP growth using consensus forecasts.

Predictive regressions of GDP growth using consensus forecasts and news-based sentiment measures are described by equation (5):

$$y_{t,c} = \alpha + \beta y_{t-h,c} + \gamma \sum_{i=1}^q \pi_{t-h-i,c} + \zeta \sum_{i=1}^q \nu_{t-h-i,c} + \epsilon_{t,c}, \quad (5)$$

where  $y_{t,c}$  is the log GDP growth between  $t$  and  $t + 12$  months in country  $c$  and  $\epsilon_{t,c}$  is the error term. First, I estimate predictive regressions of GDP growth on consensus forecasts at the country level

<sup>3</sup>All the regressions' forecast errors are measured in sample by computing the regressions' root mean square errors (RMSE).

for each horizon  $h = 1, \dots, 24$ . Because sample sizes are small, estimating coefficients for each lagged measure of sentiment would lead to large standard errors. I instead include moving averages of my positive and negative sentiment measures (averaged at a monthly frequency); the moving average horizon  $q$  is chosen by minimizing regressions' forecast errors.<sup>4</sup>

On average across countries and horizons, forecast errors diminish by 19% when news-based sentiment measures are included in predictive regressions of GDP growth on consensus forecasts. The top right panel of figure (3) shows that, on average across horizons, forecast errors diminish for each country in my sample. The top left panel shows that this reduction is larger for longer horizon: the average reduction in forecast error goes from 12% for horizons up to 12 months, to 25% for horizons longer than 12 months. This evidence supports a model of information frictions where forecasters slowly incorporate textual information in forming their forecasts.

It is well established that forecast errors tend to be larger during bad times. Does the reduction in forecast errors resulting from the inclusion of sentiment measures differentially improve forecasts of good and bad times? I fit an H-P filter to quarterly GDP growth times series at the country level ((Hodrick and Prescott, 1997)). Good (bad) times are defined to be the periods when the realized annual GDP growth is above (below) the trend produced by the H-P filter. I use the estimates of the model defined by equation (5) and I separately compute the forecast errors measured during good and bad times. The middle column of figure (3) presents forecast errors of good times and the right column presents forecast errors of bad times.

Forecast errors of good times diminish by 13% on average as a result of the inclusion sentiment measures in equation (5). The improvement in forecast error goes up to 21% during bad times. The middle left and bottom left panel of figure (3) shows that the same result holds if I restrict the sample to short (long) forecasting horizons: during good times, forecast errors diminish by 10% (15%) on av-

<sup>4</sup>To be clear, sentiment variables are lagged to only include information released prior to the forecasts.

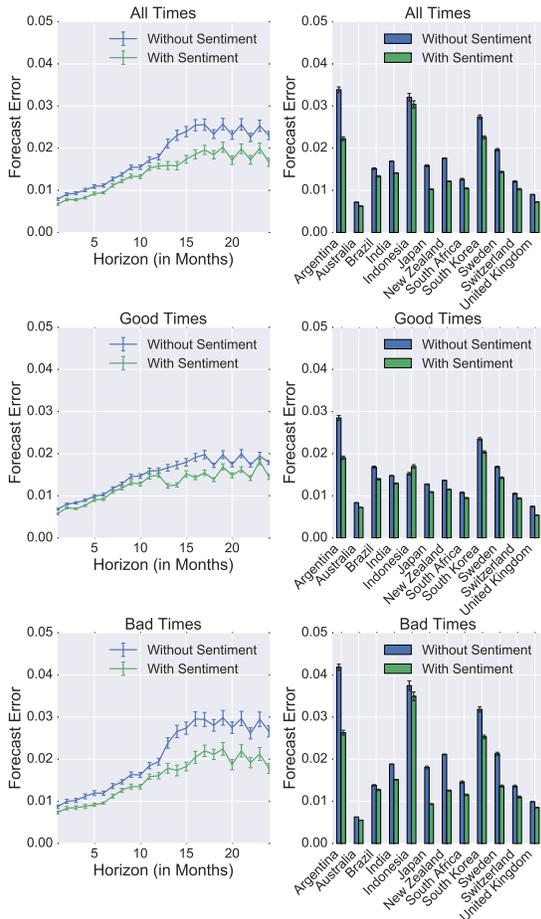


Figure 3: Reduction in GDP Growth Forecast Errors. Changes in forecast errors (measured in RMSE) across countries and horizons when news-based sentiment measures are included in predictive regressions of GDP growth on consensus forecasts for GDP growth (see equation (5)). In the left column, forecast errors are averaged across countries. In the right column they are averaged across horizons. The left panel shows forecast errors during both good and bad times; the middle panel shows forecast errors during good times; the right panel shows forecast errors during bad times. Good and bad times are determined with respect to an HP filter estimated on quarterly GDP growth data. A period is considered to be a good (bad) time if annual GDP growth is above (below) the trend estimated by the HP filter. Errors bars represent standard errors. News articles come from Factiva.com, GDP growth comes from the International Financial Statistics Database (IFS), consensus forecasts come from Consensus Economics, Inc.

erage; during bad times they diminish by 16% (28%) on average.

If forecasters were simply slowly incorporating information but correctly assigning weights when updating their forecasts, I should not observe a difference in changes in predictive accuracy between good and bad times. The fact that reductions in forecast error are larger in bad times than in good times suggests that forecasters tend to underreact to negative information.

#### 4 Conclusion and Future Work

This paper describes the information content of news-based measures of sentiment and their relationship to fluctuations in GDP growth. Sentiment measures tracks fluctuations in GDP and we show that they are a leading indicator of GDP growth at the country level for 10 out of 12 countries in our sample. Sentiment measures contain information which is not accounted for by professional forecasters. News-based sentiment measures lead to a 19% average reduction in forecast error of GDP growth relative to consensus forecasts. Reductions in forecast errors are larger for longer forecasting horizons which suggests that forecasters slowly incorporate textual information into their forecasts. Reductions in forecast errors are also larger during bad times which indicates that forecasters tend to underreact to bad news.

From a policy perspective, news-based measures of sentiment provide a direct, real-time, automated and inexpensive measures of aggregate sentiment about current and future economic conditions, especially for countries for which official statistics might be sparse, inaccurate or noisy. As a result, it could help policy makers react in a more efficient manner to changes in economic conditions.

#### References

[Aruoba et al.2009] S. Boragan Aruoba, Francis X. Diebold, and Chiara Scotti. 2009. Real-time measurement of business conditions. *Journal of Business and Economic Statistics*, 27:417–427.

- [Baker et al.2016] Scott Baker, Nicolas Bloom, and S.J. Davis. 2016. Measuring economic policy uncertainty. *Quarterly Journal of Economics*, *Forthcoming*.
- [Choi and Varian2012] Hyunyoung Choi and Hal Varian. 2012. Predicting the present with google trends. *The Economic Records*, 88:2–9.
- [Fildes and Stekler2002] Robert Fildes and Herman Stekler. 2002. The state of macroeconomic forecasting. *Journal of Macroeconomics*, 24(4):435–468.
- [Garcia2013] Diego Garcia. 2013. Sentiment during recession. *Journal of Finance*, 68(3):1267–1300.
- [Hodrick and Prescott1997] Robert Hodrick and Edward C. Prescott. 1997. Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, 29(1):1–16.
- [Jurafsky and Martin2009] Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing*. Prentice Hall, Upper Saddle River, NJ.
- [Loughran and McDonald2011] Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance*, 66(1):35–65.
- [Nordhaus1987] William Nordhaus. 1987. Forecasting efficiency: Concepts and applications. *Review of Economics and Statistics*, 69(4).
- [Tetlock2007] Paul C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.
- [Young and Soroka2012] Lori Young and Stuart Soroka. 2012. Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29:205–231.