BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

The power of text-based indicators in forecasting the Italian economic activity

by Valentina Aprigliano, Simone Emiliozzi, Gabriele Guaitoli, Andrea Luciani, Juri Marcucci and Libero Monteforte

March 2021

Number 1321

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

The power of text-based indicators in forecasting the Italian economic activity

by Valentina Aprigliano, Simone Emiliozzi, Gabriele Guaitoli, Andrea Luciani, Juri Marcucci and Libero Monteforte

Number 1321 - March 2021

ISSN 1594-7939 (print)
ISSN 2281-3950 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

THE POWER OF TEXT-BASED INDICATORS IN FORECASTING THE ITALIAN ECONOMIC ACTIVITY

by Valentina Aprigliano*, Simone Emiliozzi*, Gabriele Guaitoli†, Andrea Luciani*, Juri Marcucci* and Libero Monteforte*‡

Abstract

Can we use newspaper articles to forecast economic activity? Our answer is yes and, to this aim, we propose a brand new economic dictionary in Italian with valence shifters, and we apply it on a corpus of about two million articles from four popular newspapers. We produce a set of high-frequency text-based sentiment and policy uncertainty indicators (TESI and TEPU, respectively), which are timely, not revised and computed both for the whole economy and for specific sectors or economic topics. To test the predictive power of our text-based indicators, we propose two forecasting exercises. First, using Bayesian Model Averaging (BMA) techniques, we show that our monthly text-based indicators greatly shrink the uncertainty surrounding the short-term forecasts of the main macroeconomic aggregates, especially during recessions. Secondly, we employ these indexes in a weekly GDP growth tracker, delivering sizeable gains in forecasting accuracy in both normal and turbulent times.

JEL Classification: C11, C32, C43, C52, C55, E52, E58.
Keywords: forecasting, text mining, sentiment, economic policy uncertainty, big data, BMA.
DOI: 10.32057/0.TD.2021.1321

Contents

1. Introduction 5
2. Data 7
   2.1 Why newspapers-based indicators 7
   2.2 The News Corpus 8
   2.3 The Italian Economic Dictionary 10
   2.4 Computing the sentiment 11
3. A text-based sentiment indicator for Italy 12
   3.1 Sentiment by Topic and by economic sector 13
   3.2 Comparison with other soft indicators of economic activity 15
   3.3 Sentiment heterogeneity across sources 16
4. Textual Economic Policy Uncertainty for Italy 17
   4.1 TEPU index for Italy 18
   4.2 TEPU indexes by topic and sector 19
5. Empirical applications 20
   5.1 Monthly BMA model 20
   5.1 A text-based weekly economic tracker 26
6. Conclusions 31
References 32
Appendix: Additional Figures and Tables 36

* Bank of Italy, Directorate General for Economics, Statistics and Research.
† University of Warwick.
‡ Ufficio Parlamentare di Bilancio.

1 Introduction [1]

Many radical transformations have been reshaping the structure of the economy in the last decades. Globalization fostered capital and financial linkages across the world, while the shift from manufacturing to services and digital technologies is structurally changing our economies. These radical transformations, together with the legacies of the Global Financial Crisis, before the Covid-19 outbreak, have been challenging the cornerstones of macroeconomic theory and traditional statistical tools for economic analysis and forecasting (Ng and Wright, 2013; Bok et al., 2018). Notwithstanding, analysts today can benefit from a proliferation of novel nontraditional data characterized by high frequency and high dimensionality and obtained from unconventional and previously unexplored sources. The big-data phenomenon is indeed spreading in economic analysis.

In this work, we explore one of the most peculiar types of this new information: text data, which differ from traditional data because of their unstructured nature. Text data - like this paragraph - represent a source of valuable qualitative information for (or about) economic agents, encoded through the complex rules of language. Recent advancements in statistical and computational methods have allowed researchers to access and process such data in order to extract quantitative information suitable for statistical models, such as forecasting models (Gentzkow et al., 2019a).

We use a multi-source database of Italian newspaper articles related to economic news to build a set of sentiment and uncertainty measures along the lines put forward by Soroka (2006), Tetlock (2007), Ardia et al. (2017), Loughran and McDonald (2011), Loughran and McDonald (2014), Baker et al. (2016) and then we use these indicators to nowcast and forecast the Italian real economic activity.

With respect to alternative sources of textual data, such as *tweets* from Twitter, newspaper articles have interesting properties that make their use particularly appealing: i) they have no (meaningful) space restriction, allowing for a more natural use of language and to explain more deeply or extensively about the news; ii) they are easily organized into topics, allowing us to explore topical sentiment and uncertainty regarding Monetary and Fiscal Policy, Economic Conditions, Domestic Policy and several other categories; iii) the

[1]We would like to thank Eleni Kalamara, Andrea Nobili, Dooruj Rambaccussing, Roberta Zizza and participants to the 31^{th} (EC)2 Conference on "High dimensional modeling in time series", the 2020 Bank of Italy and Federal Reserve Board Joint conference on "Nontraditional data & Statistical Learning with Applications to Macroeconomics", the Bank of England virtual conference on "Modelling with Big Data & Machine Learning: Measuring Economic Instability", the 21^{st} IWH-CIREQ-GW Macroeconometric Workshop on "Forecasting and Uncertainty", and the 2020 World Congress of the Econometric Society for their helpful comments. An earlier version of the paper circulated under the title "*Sentiment and Uncertainty indexes for forecasting Italian economic activity*". The views expressed herein are those of the authors and do not necessarily represent the views of the Bank of Italy or the Eurosystem. All remaining errors are our own. Corresponding author: Simone Emiliozzi, Bank of Italy, via Nazionale, 91, 00184, Rome, Italy. Email: simone.emiliozzibancaditalia.it.

daily number of articles is relatively more stable than that one of *tweets*, entailing better properties of the series and a correct interpretation of intensive and extensive margins. As for Italy, we have survey-based evidence of the importance of newspapers in shaping key firms' decisions regarding investment or employment. Using the Bank of Italy Survey on Inflation and Growth Expectations (SIGE), we show that firms put a great amount of confidence on news read on newspapers. In fact, news is the most important source of information after the contacts with customers and suppliers.

Unlike the existing literature, we employ Italian-language newspaper articles from multiple sources. Including multiple newspapers allows us to smooth the individual-newspaper bias. The choice of the language is due to the fact that the number of newspaper articles about the Italian economy written in Italian is much larger than the number of English articles about the same topic from international sources (such as Bloomberg). Our database is extracted from Dow Jones Factiva one of the largest archives used in the forecasting literature using textual based indicators (Thorsrud, 2018; Fraiberger, 2016; Bybee et al., 2019; Kelly et al., 2018; Shapiro et al., *forthcoming* 2020). Kalamara et al. (2020) have recently analyzed around half a million articles from three main British newspapers showing that simple text-based indicators can improve economic forecasts, in particular during downturns. Rogers and Xu (2019) have also recently analyzed the predictive performance of different measures of economic uncertainty in the US to forecast real and financial outcome variables, showing some additional predictive content of such measures. Similarly, Rambaccussing and Kwiatkowski (2020) use around 400 thousand UK newspapers articles from 1990 to 2018 to nowcast inflation, unemployment and output in the UK. They do not find evidence of superior performance of sentiment for inflation, while for output and unemployment forecasting the sentiment inferred from media seems useful.

One of our main contributions is the construction of an Italian dictionary of economics-specific terms, tailored for capturing news related to the Italian real economic activity. We enrich our dictionary with valence-shifting words[2] which help to capture the correct polarity of each sentence.

The sentiment and uncertainty indexes that we build display consistent signals. To the best of our knowledge, we are the first to document a negative correlation between sentiment and uncertainty measures for Italy derived from textual data, similar to the one found in survey data (Bachmann et al., 2013). Moreover, our indicators help in tracking the real-time evolution of the economic activity at different frequencies. Using a Bayesian Model Averaging approach, our monthly-aggregated indicators have high inclu-

[2]Valence-shifting terms are negations, conjunctions, adverbs that can change the general meaning of a sentence. In general valence-shifting terms can switch the polarity of close words or amplify or reduce their sentiment intensity.

sion probabilities compared with the standard soft and hard indicators commonly used in nowcasting. Furthermore, their inclusion tends to improve density nowcasts and forecasts, while the gains in terms of point-forecast are more limited and concentrated during recessions. We also explore the weekly properties of our text-based measures, building an high-frequency tracker of the Italian GDP as in Lewis et al. (2020) and Delle-Monache et al. (2020). These indexes provide large point forecast gains over a simple benchmark model when nowcasting the Italian GDP, in particular during recessions.

The rest of the paper is organized as follows. Section 2 contains the description of our textual database, the dictionary and the methodology put forward to compute the sentiment indexes. Sections 3 and 4 present the overall Sentiment and Economic Policy Uncertainty (EPU) indices and their sub topics (economic conditions, government, monetary and fiscal policy, domestic policy and labor markets, to cite just a few). Section 5 contains the empirical forecasting application. Section 6 concludes.

2 Data

2.1 Why newspapers-based indicators

Sentiment indicators and topic analysis are becoming popular tools for forecasting and structural analysis. Newspapers-based indicators can be a timely source of information to track the business cycle, as already documented by recent papers (Thorsrud, 2018; Kalamara et al., 2020; Bybee et al., 2019; Shapiro et al., *forthcoming* 2020; Rambaccussing and Kwiatkowski, 2020).

We provide novel evidence on why newspaper-based sentiment indicators are relevant for economic purposes by interviewing the firms participating in the Survey on Inflation and Growth Expectations conducted by the Bank of Italy (SIGE). The survey asks to rank up the three main sources of information out of a list of six (newspapers, TV news, institutional or trade associations' publications, analysis carried out by private companies, contacts with customers and/or suppliers, social networks) used for taking important business decisions such as investing and hiring.[3]

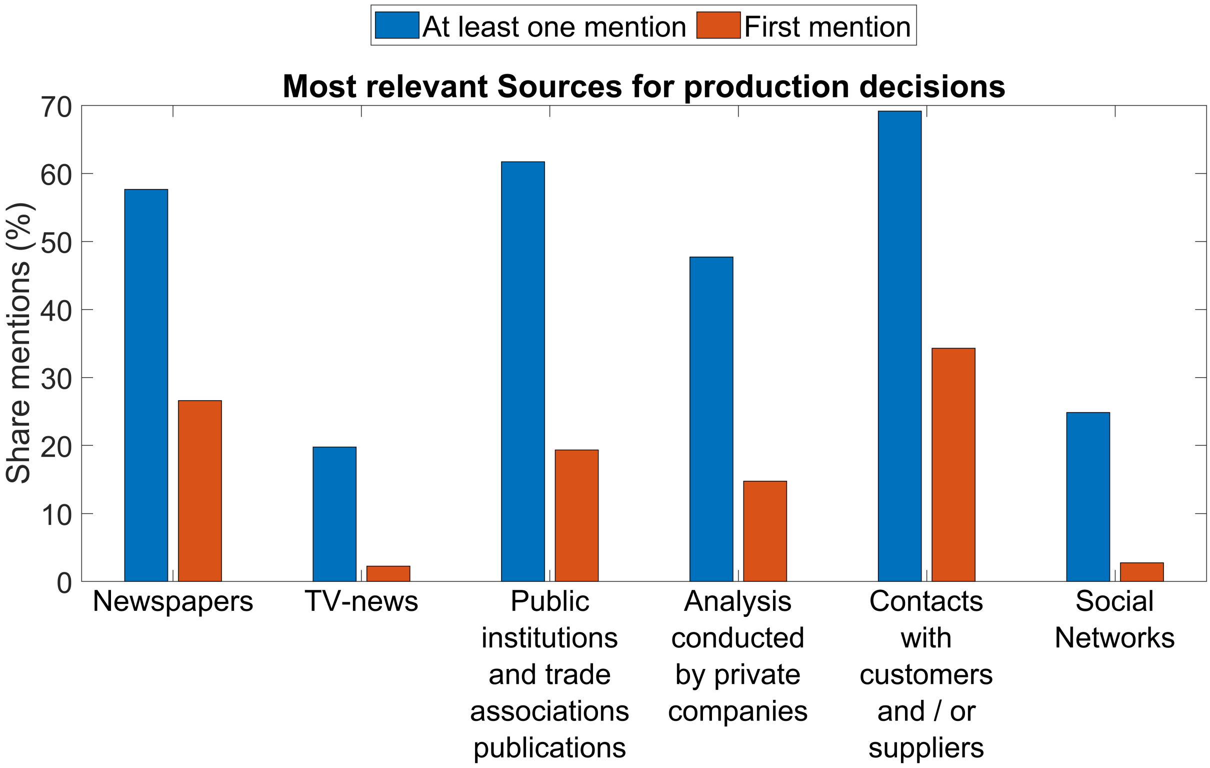
We find that firms consider newspapers a relevant and reliable source of information: they were ranked second as the most reliable information channel, after "Direct contact with clients and/or suppliers". A reasonable outcome, since supply-chain relations are crucial for business. Figure 1 shows that almost one third of the sample claims to use

[3]The questionnaire can be found at `https://www.bancaditalia.it/pubblicazioni/indagine-inflazione/2019-indagine-inflazione/12/index.html?com.dotmarketing.htmlpage.language=1`. The exact wording of the question and the available options have been reviewed and approved after a pilot, which provided feedback on the clarity and the effort necessary to answer the question. The questionnaire was filled by almost 1200 Italian firms during the last quarter of 2019.

newspapers as their main source of news to inform their business decisions, while almost 60% reported newspapers among the top three most reliable sources out of the six choices allowed in the survey.

Among all high-frequency publicly available sources, newspapers remain the most important one, while TV news and social networks are ranked first by less than 3% of the firms and in the the top-three by less than 25%.

Figure 1



Sample size: 1199 respondents. First mentions sum to 1. Blue bars do not sum to 3 despite three answers were available since some respondents (12.5%) provided only one answer (6.7%) or two (5.8%).

2.2 The News Corpus

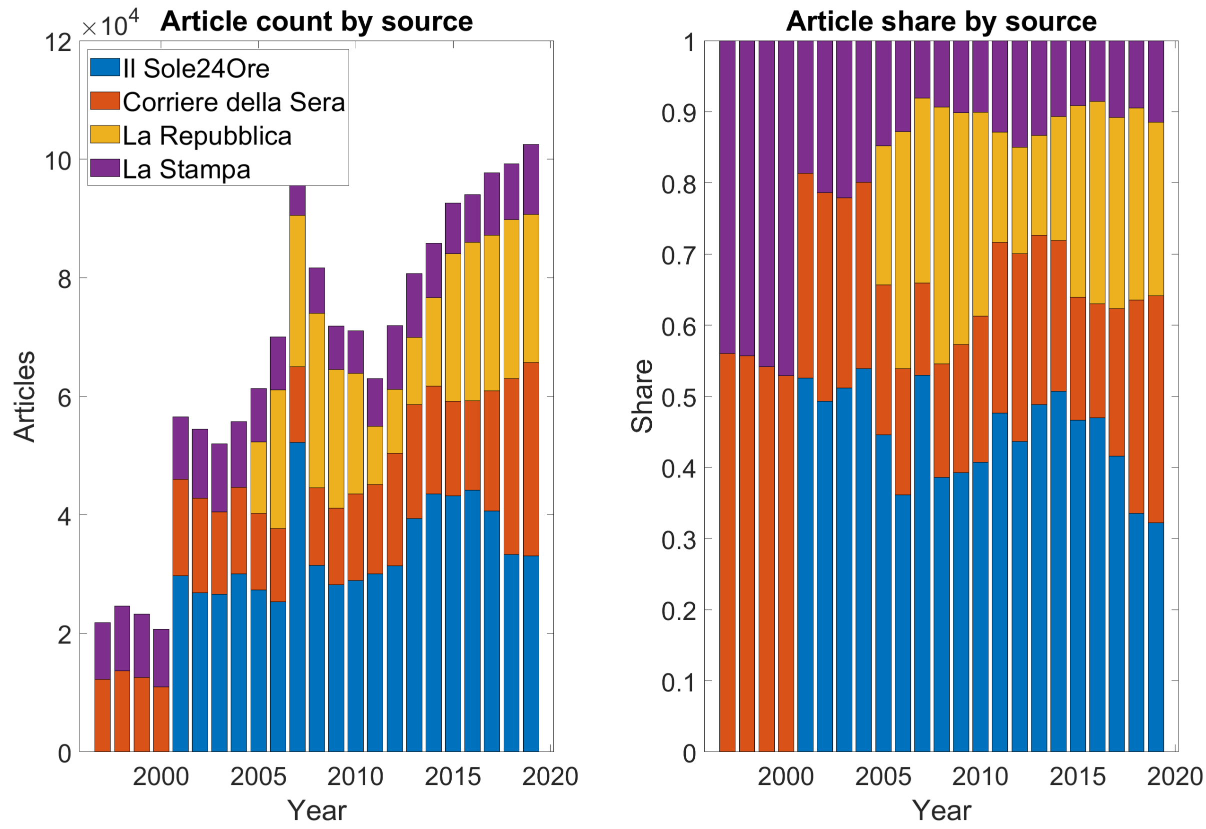
We use articles extracted from the repository Factiva.[4] Factiva's records include the title, text, date and source of each article, along with a number of other metadata and an automatically generated category tagging (e.g. topic, companies' name, geographical region; see figure A.1 in the Appendix for a FACTIVA article snippet).

We download all articles talking about the Italian economy from the four main national newspapers: *Il Corriere della Sera, Il Sole 24 Ore, La Repubblica* and *La Stampa* (both paper and online editions), which have a significant national circulation and are available

[4] https://professional.dowjones.com/factiva/.

for a reasonable number of years among all newspapers in the Italian language.[5] Figure 2 shows the number of articles by year and source and the share of articles for each source in each year.

Figure 2: Count of articles by year and source.



After the exclusion of articles including less than 100 words and of financial markets' purely technical news,[6] our final dataset contains over 1.6 million articles spanning from January 1997 to December 2019.

The text ("corpus") of the articles undergoes a preliminary screening ("pre-processing"), to get rid of words and characters that are not informative (i.e. articles, non-adversative conjunctions, non-meaningful punctuation).[7]

[5]According to *Accertamenti Diffusione Stampa* (ADS), i.e. the company that certifies the circulation of newspapers in Italy, *Il Corriere della Sera* and *La Repubblica* are the top two newspapers in Italy in terms of average circulation (with 216,149 and 165,748 copies sold in 2018, respectively), *La Stampa* is the fifth one (131,744 copies sold in 2018), while *Il Sole 24 Ore* is the tenth one (79,928 copies sold in 2018), but it is the most important Italian economic and financial newspaper. Data are available at this link `http://www.adsnotizie.it/_dati_DMS.asp`.

[6]We find all categories which refer to technical market communications or wraps of the daily market prices and exclude them from our database, as the quantitative correspondent of such news is already available, i.e. the stock market index time series itself.

[7]Our corpus has approximately 350,000 unique terms before stemming, too many to be directly utilized

2.3 The Italian Economic Dictionary

In order to extract quantitative information from textual data and to shrink the dimension of the available information we use a dictionary based approach. We compile an Italian economic dictionary focused on economic and financial terms that recur often in journalistic jargon (Soroka, 2006; Tetlock, 2007; Loughran and McDonald, 2011, 2014; Fraiberger, 2016). The dictionary is a set of words (unigrams) and phrases (n-grams) associated with a "polarity" (positive or negative meaning) and a "weight" (module that characterizes the relative importance of a word in a statement; see Table A.2 in the Appendix for a small sample of our dictionary). Differently from the literature, our dictionary is enriched with a set of valence-shifting words and specifically tailored for newspaper jargon.

A word or a "n-gram" carries its initial "polarity" that can be positive or negative according to its meaning. As an example, *debito* (debt) starts with a negative polarity while *fiducia* (trust) with a positive one. Some common collocations or phrases carry their own meanings, possibly different from the sum of the individual words they are made by. As an example, consider the "bi-gram" *debt growth*, which carries an overall negative meaning, despite the fact that "growth" is positive, as "debt" is negative.[8]

A *"valence shifter term"* is a word which does not have a meaning on its own, but affects both polarity term's sign and module. For example, one can think about negations (e.g. **fall of** *GDP*) and the amplification (deamplification) of a word in the vocabulary (e.g. **lot of** *uncertainty*).

Our final dictionary, after stemming, contains 433 and 190 unique polarity and valence-shifting terms, respectively.

For the compilation of our dictionary we do not use an unsupervised algorithm such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003), *Word2Vec* (Mikolov et al., 2013) or *GloVe* (Pennington et al., 2014). N-grams are included via a supervised process where the researcher includes relevant economic and financial terms exercising a great effort in balancing negative and positive polarity terms to balance the dictionary. Our methodology is however transparent since the n-grams included in the dictionary are publicly available and it is possible to debate on the opportunity of their inclusion (exclusion). Our dictionary based method is fast in computing sentiment for very large corpora as the one used in this work, while the unsupervised algorithm listed above can run into troubles

for statistical applications such as forecasting. This huge number of elementary words is far above the 160,000 terms commonly used in the Italian language and is due to the presence of English words (for example, Recovery Plan, Buy-Back, Next Generation EU, etc.), names, strictly financial terms appearing in short forms (for example, MRO, EAPP, TLTRO, etc.), but also typos.

[8]One could argue that "growth" is a neutral word unless associated to a meaningful name, but it is classified as a positive word by a number of dictionaries. In economics, talking about "growth" usually implies an increase in output and therefore economic development.

since they are computationally very demanding and suffer from multimodality problems that make topic identification challenging (Roberts et al., 2016).

2.4 Computing the sentiment

Our sentiment indicators are computed using the R package Sentometrics (Ardia et al., 2017), which allows fast sentiment computations (see figure A.2 in the Appendix for a schematic pipeline of the procedure).

Once the text is cleaned and processed, the sentiment is calculated at article-level as the sum of polarities divided by the number of words.

To give an example of how this approach works in practice, consider the following sentence extracted from an article appeared on the 10th December 2019 online edition of *Il Sole 24 Ore*: "*tra gennaio e ottobre il calo complessivo dell'output è dell'1,2%*" ("Production fell by 1.2% overall between January and October"). After our text processing routine, the sentence is collapsed in

gennaio ottobre calo complessivo output
(January October fall overall production)

to which our estimation procedure assigns a sentiment score of -0.2. In order to see how this is computed, we highlight and classify the words which find a match in our dictionary:

$$\text{gennaio ottobre } \underbrace{\text{calo}}_{\text{Shifter } -1} \text{ complessivo } \underbrace{\text{output}}_{\text{Polarity } 1}$$

The token "*output*" is assigned total polarity $1 \times (-1) = -1$ since a shifter with value -1 ("*fall*") appears within 5 words and it has not been used already together with another polarity term. Since there are 5 words in the trimmed sentence, the total sentiment score of article j (SS_j) is computed as

$$\mathrm{SS}_j = \frac{\sum_i^{\text{N words}} \text{Polarity}_i \times \text{Shifter}(i)}{\text{N words}} = \frac{1 \times (-1)}{5} = -0.2$$

Table 1 shows some examples of how our dictionary works on a selection of newspaper headlines.[9] While we almost always correctly classify the sign of the sentence, we are able to capture only partially the actual depth of the positivity/negativity. Our classification is helped by the inclusion of valence shifter terms, which allow a better signing of most polarity terms.

[9]The sentences have been translated from Italian to English in such a way to leave a sense of how certain common collocations are overridden by the newspaper headlines' jargon and style.

Table 1: Sentences and associated Sentiment score

Sentence	Sentiment
GDP has fallen	-1.000
Ex Ilva, does ArcelorMittal risk to leave Taranto?	-0.257
Oil: new cuts by OPEC, but the group argues about quotas	-0.143
Clash on plastic and sugar taxes. Renzi raises the bar: I don't want to vote, but [...]	-0.083
Censis: Italians, left with no confidence, leave even the BOT.	-0.029
More jobs and output. The government's plan for a "mixed-property" ILVA	0.257
China lends a helping hand on duties, financial indexes are positive	0.000
Istat, GDP at +0.2% in 2019. Growth expected to pick up in 2020	0.250
Ex Ilva, Conte against ArcelorMittal: "the plan is not yet ok, we will reject it"	0.000
Debt has grown	-0.500

Note: Sentences are translated from Italian, and the Sentiment is calculated using our Italian dictionary. Sentiment measures are the raw output and are not re-scaled.

The sentiment indicator in a given day t is obtained as a weighted average of the articles' scores SS_j, with the weights equal to the articles' length, that is defined as the total number of words in the text after the removal of stop-words.[10]

3 A text-based sentiment indicator for Italy

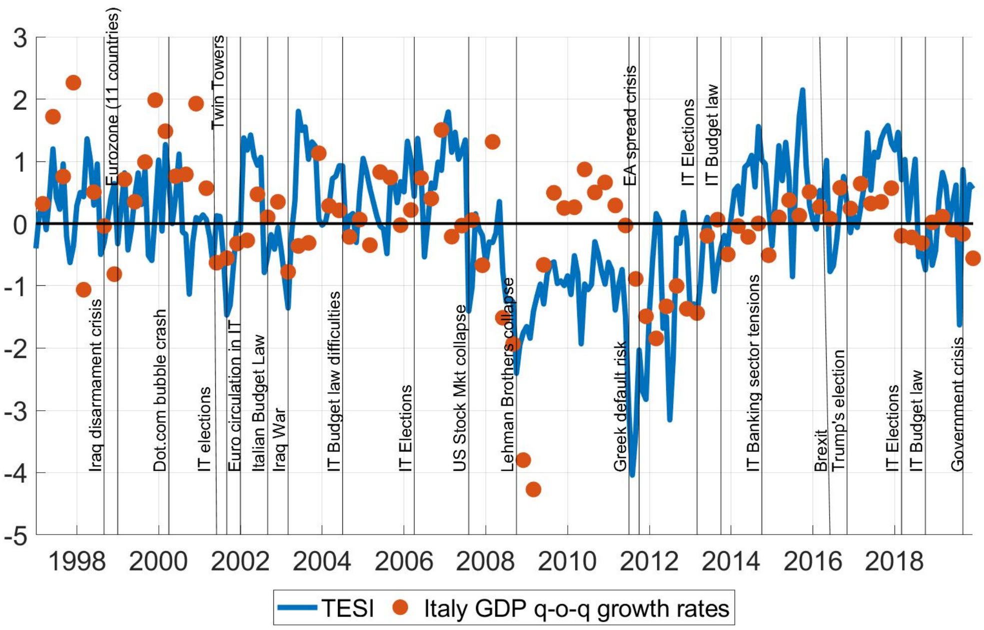
Figure 3 depicts our overall Text-based Sentiment Indicator (TESI), derived from newspaper articles from January 1997 to December 2019, along with the quarterly GDP growth rate (both standardized). Vertical bars represent key events for the Italian economy during this time.[11]

It is important to clarify that TESI is calculated in real time, while GDP data come from a vintage downloaded in January 2020. The contemporaneous comovement between the two series over the whole sample at the quarterly frequency is 0.45: it is low at the beginning of the sample, between 1997 ad 2001, while it starts to increase from 2002 onward.

[10]When taking sentiment averages among all articles published in a given day, we include in the computation of the index all articles with zero/almost zero score, so that our indicators - in particular the topical/sectoral ones - are not affected by large swings in the economic sentiment of a small fraction of articles.

[11]Since the source coverage of our database is limited until 2001, the measure before that year must be interpreted with caution.

Figure 3: Italian GDP (q-o-q growth rates) and TESI (daily data aggregated at monthly frequency). Vertical bars represent important events for the Italian economy.



Notes: TESI is calculated using daily articles in *real time*, while GDP growth rates are calculated using the vintage published by ISTAT in January 2020. Both series are standardized to have mean zero and variance one.

TESI tracks important episodes for the Italian economy. It decreases after negative events, such as the Iraq disarmament crisis in 1998, the Dot-com bubble in 2000, the burst of Iraq war in 2003, the US stock market collapse in 2007, the financial crisis after the burst of Lehman Brothers in 2008, the Greek risk of default in 2010, the start of the Sovereign Debt Crisis in 2011, the Brexit referendum in 2016, the 2019 Italian Budget Law discussion in October 2018 and finally the Government crisis in August 2019. It increases rapidly following the end of the Dot-com bubble in 2000, the period before the Great Recession, the recovery period in 2014 and 2015, and in 2017 when global trade was robustly growing.

3.1 Sentiment by topic and by economic sector

We use a Factiva's proprietary algorithm for articles tagging that assigns each article in our corpus to one or more granular subjects (e.g. economic news, sales figures, economic growth/recession), to estimate sentiment indexes for several economic topics and economic sectors. To exploit the richness of information provided by Factiva, we group their 300 detailed article categories into 15 topics (economic conditions, private sector, monetary policy to name but a few) and 21 sectors (see table A.1 in the Appendix for the details). These groupings are not mutually exclusive since the same article in the corpus can be used multiple times in the computation of the sentiment for each topic. This feature is due to the fact that newspapers articles deal with several topics. [12]

Figure A.3a in the Appendix displays the contemporaneous correlation between TESI and the 15 topic-indicators. The first row of the graph underlines the high correlation between TESI and the sentiment indicators related to economic conditions and to the private sector; those related to finance, government and monetary policy display a lower degree of association. Figure A.3b shows the contemporaneous correlations between TESI and the 21 sector-indicators. The former correlates strongly with the indicators calculated for the services, manufacturing, wholesale retail, business consumer services and real estate & construction, which are the most important sectors in the National Accounts.

Figure A.4 displays three of the most important sentiment indicators out of the 15 computed by topic: 1) government/fiscal policy; 2) monetary policy and 3) labor market.

The Sentiment about the topic "government/fiscal policy" (Figure A.4a) is calculated building on Factiva categories about government finance, budget and taxation, public debt, sales and income taxes and many others. It is highly correlated with TESI (0.7) and its dynamics matches up well key events of the Italian politics and fiscal policy. It tends to decrease sharply before national elections and to rebound thereafter (es. in 2001, 2006 and 2013). It reached historical low levels during the Sovereign Debt Crisis in 2011, with the fall of Berlusconi's Government in November 2011, during harsh political and public debates about Budget Laws (i.e. September 2002; July 2004 when Siniscalco took over Tremonti as Italian Treasury Minister, October 2018) and finally during the last Government crisis in August 2019.

Figure A.4b displays the sentiment about monetary policy conditions and it is calculated aggregating categories such as ECB, Central Bank interventions, interest rates and money supply among others. This index peaks with the introduction of the euro in January 1999 and with the change of the ECB monetary policy strategy in March 2003. It reaches minimal values during the episode of the stock market crash in August 2007, the Lehman collapse in September 2009, the Sovereign Debt Crisis and with the announced reductions of the Asset Purchase Programme. Further it sharply increased after impor-

[12] Suppose that a given article is composed of only two topics and that it is talking with a positive tone about monetary policy and with a negative one about labour market conditions in Italy. Suppose further that its overall score is positive. The article will be classified as positive in both topics. However, comparing the topical index with the overall sentiment TESI, it is still possible to see if the press is speaking about a given topic with a relatively better (or worse) tone, giving rise to interesting comparisons.

tant monetary policy interventions of the ECB such as the fixed rate full allotment tender procedure introduced in October 2008, the launch of the Security Market Programme in May 2010, the VLTRO in December 2011, Draghi's "Whatever it takes" in July 2012, the launch of OMT in September 2012, the launch of TLTRO in 2014, the start of the Asset Purchase Programme announced in January 2015 and its restart in September 2019 (Hartmann and Smets, 2018; Neri and Siviero, 2019).

Finally Figure A.4c shows the sentiment index for Labor Market conditions,[13] computed using articles tagged in categories such as: employment/unemployment, general labor issues, lay-offs/redundancies, job search. The index increases in months where important labor market reforms were passed in Italy, such as the Treu reform in June 1997, the Biagi's law in September 2003, the Fornero reform in June 2012, the first tranche of the Jobs Act in December 2014, its second tranche in June 2015 and the Dignity decree in July 2018.

Some of the sector-sentiment indicators used in the empirical application in section 5 are displayed in Figure A.5.[14]

3.2 Comparison with other soft indicators of economic activity

We compare TESI with measures of consumers' and firms' expectation about the economic conditions commonly used in nowcasting and forecasting Italian GDP: i) the Markit composite Purchasing Manager Index for Italy; ii) the Italian business (IESI) and Consumer Confidence indexes produced by ISTAT. Our sentiment indicator is a mixture of firms and consumers expectations, with more weight given to news regarding businesses.[15]

An advantage of our TESI over the standard sentiment indicators, such as the PMIs (published by Markit) or the the confidence indexes produced by Istat is that it can be computed (and released) at higher frequencies such as weekly or daily. Moreover, the Markit and Istat indicators are based on data collected in the first half a given month, while our sentiment can account for events occurring along the entire month. This means that relevant events happening in the second half of each month would be captured by the standard sentiment measures with a delay of one month, while our TESI would be able to capture them with 1-2 days of delay at most.

Figure A.6a shows TESI, the Italian composite PMI and the q-o-q variation of the Italian GDP. The contemporaneous correlation between our Sentiment Index and the

[13]One should be careful in interpreting the labor market sentiment index as an attempt to estimate an indicator of the state of this market. The index is a measure of how positive (negative) some labor market news are commented by journalists in the press.

[14]All the sentiment indicators not reported in the paper are available upon request.

[15]Topic indicators may be relevant for tracking firms or consumers beliefs. This is left for further analysis.

composite PMI is 0.5 at the monthly level.[16]

Figure A.6b displays TESI together with business' and consumers' confidence indices released by ISTAT. Looking at the contemporaneous correlation, the Sentiment aligns better with businesses' confidence rather than consumers' one.

Finally, Figure A.7 shows a comparison at monthly frequency between TESI and the Istat Social Mood on Economy (SME) index based on Twitter feeds.[17] Our measure has several advantages: i) it covers a longer time span, while the Istat index starts only in February 2016; ii) we provide a topical and sectoral decomposition of sentiment unlike the SME; iii) our index is based on major Italian newspapers articles while SME relies only on Twitter. The contemporaneous correlation between the two indices at monthly frequency is moderate (0.3). The association is elevated at the beginning of the sample, from February 2016 till June 2017. Instead, they tend to move in opposite directions during the first phase of the US-China trade dispute with the failed agreements in July 2017 and during the course of 2019. During the Italian Government crisis in August 2019 our indicator collapsed, while SME jumped up to high positive values. In this short time span TESI tracks better than SME the dynamics of Italian GDP growth.

3.3 Sentiment heterogeneity across sources

Since our sources may differ by political alignment, main focus (generalist news/ economic), geographical interest (national/local) and style, we compute complementary sentiment indicators by newspaper to look at the differences. The results can be seen in Figure A.8 for *Il Corriere della Sera*, *La Repubblica*, *Il Sole 24 Ore* and *La Stampa*.

We find a relevant level of heterogeneity in the sentiment computed across different newspapers. In particular, while all big events and trends show up in all the four newspapers and the overall trend seem similar, some differences arise in the monthly variation in some sub-periods and in the first two moments of the series. This last feature does not emerge in Figure A.8, where the series are standardized in order to focus on the first two differences.

For example, between 2015 and 2017 GDP growth reached its maximum (considering both q-o-q and y-o-y variations) since the Great Recession, but different sources show very different dynamics. The overall indicator has a slight negative trend between 2015 and late 2016, which seems inconsistent with the GDP growth figures. However, the change in the aggregate indicator is driven by a strong negative movement in the sentiment of *La*

[16]This high correlation is recorded also between the sentiment and the other PMI subcomponents such as PMI manufacturing, services and their forward looking components.

[17]The Istat Social Mood on Economy (SME) index based on Twitter feeds con be downloaded here `https://www.istat.it/en/archivio/219600`

Stampa and *La Repubblica*, while this fall is smaller for *Il Corriere della Sera* and *Il Sole 24 Ore*. On the contrary, all newspapers but *La Repubblica* seem to be overly pessimistic during 2018 and 2019. *La Repubblica* and *Il Sole 24 Ore* also seem to have had much worse news during the peak of the Sovereign Debt crisis than during the Great Recession, when GDP fell by more and faster. Overall, no newspaper seems to outperform the others in tracking the GDP strictly better over the full sample.

These differences in local trends can be related to two phenomena. First, newspapers can be generally biased against the current government in power. This is a well-known fact in the literature, as seen and measured in a number of works such as Groseclose and Milyo (2005) and Gentzkow and Shapiro (2010); Gentzkow et al. (2019b). Moreover, the four newspapers considered in our sample seem to have a different topical composition of their articles, suggesting a different reactivity of the sentiment series to different types of events.

4 Textual Economic Policy Uncertainty for Italy

Measuring Economic Policy Uncertainty (EPU henceforth) has become a key issue in recent years since the seminal work of Baker et al. (2016) (see also Rogers and Xu (2019) for an interesting application to forecasting). Following their methodology we propose several textual EPU indexes for Italy (TEPU henceforth) using newspaper articles at daily, weekly and monthly frequency. The availability of the topic tags allows us to construct granular topical-uncertainty indicators differently from Ardizzi et al. (2019) that propose an overall EPU index derived from articles in Bloomberg and *tweets* from Twitter.

Our TEPU indexes are computed as:

$$\text{TEPU}_t = \frac{\sum_i^{N_t} \mathbf{1}\left[\text{TEPU Article}\right]_i}{N_t} \tag{1}$$

where N_t is the number of articles published at time t, while we define "TEPU Article" any piece of news which contains at least one "uncertainty word" (*incert** or *incertezz** in Italian, with * being a wild card) and one "policy word" as defined in Baker et al. (2016) for their Italian EPU indicator.[18] The TEPU index is the share of articles satisfying these criteria. Our TEPU indexes differ from those in Baker et al. (2016) for Italy since the denominator N_t in equation (1) is different for two reasons. First we use a larger set of newspapers; second while Baker et al. (2016) use all the articles published on a given day,

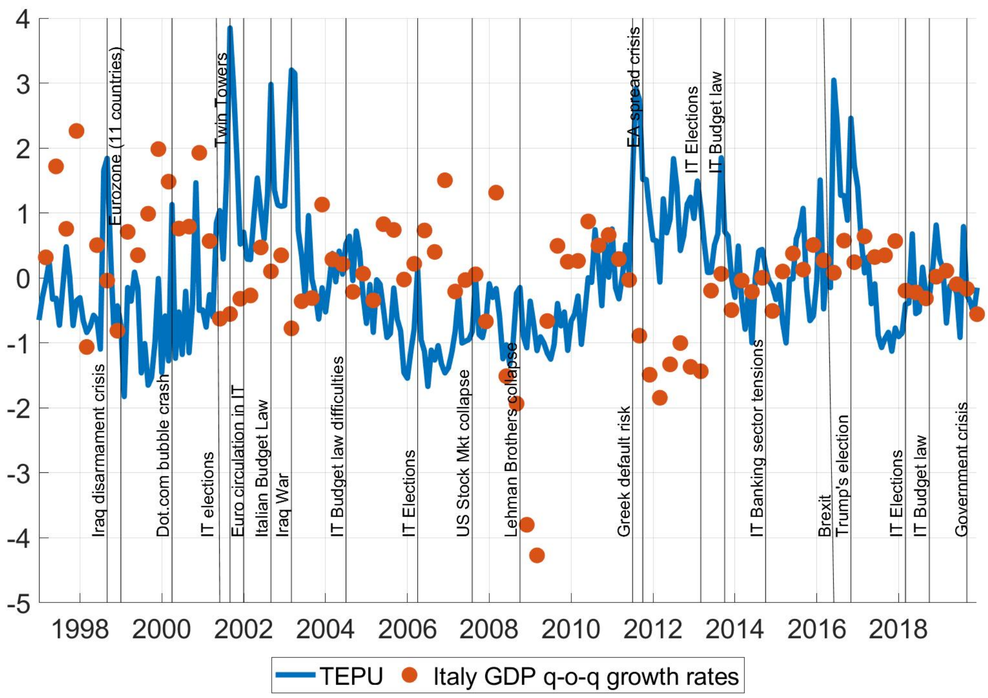
[18] See the Appendix in Ardizzi et al. (2019) for the adaptation of the "policy word" to the Italian case.

we use only those focusing on economics [19].

While variation in their EPU index can be driven by an increase in the coverage of economic news over the total amount of news in a given day, TEPU is more robust to such changes by construction. Despite these differences, our overall TEPU index for Italy is strongly correlated with the monthly Baker et al. (2016) EPU index as Figure A.10 shows. The monthly contemporaneous correlation between the two indicators is high and close to 0.8.

4.1 TEPU index for Italy

Figure 4: Textual Economic Policy Uncertainty index for Italy (TEPU) calculated on all available sources (Blue line; daily data are aggregated at the monthly frequency); Italian GDP (red dots) all sample September 1996 - May 2019; all sources



Notes: TEPU is calculated using daily articles in *real-time*, while GDP growth rates figures are calculated using the vintage published by ISTAT in November 2019.

The monthly Italian TEPU index is shown in Figure 4 together with the quarterly growth rates of Italian GDP and selected important events for Italy. The negative contemporaneous correlation between TEPU and GDP growth pointed out by the literature

[19]We use a broad filter to select the articles that speaks about economic activity and finance.

is evident. The tone of the newspapers articles gets more negative after bad episodes associated with a slowdown of the economic growth. This goes with strong spikes in our economic policy uncertainty index, as for example during the Dot.com bubble in 2000; after the struggling process to approve some Budget laws in Parliament in 2004 and, more recently, in 2018; during the Global Financial Crisis, where a big jump in EPU occurred after the *Great Moderation* period, and the Sovereign Debt Crisis in 2011; after Brexit and Trump's elections and during the the Government crisis in August 2019.

The negative correlation between Sentiment and Uncertainty indexes is a well consolidated result in the literature, and we prove it for our measures of sentiment and uncertainty derived from textual data (see Figure A.9 in the Appendix, where the contemporaneous correlation is -0.4).

4.2 TEPU indexes by topic and sector

Our TEPU measure is derived exploiting all the articles available in our dataset that are related to economic and financial news. Similarly to sentiment, we use Factiva's proprietary algorithm to calculate TEPU indexes by topic and sector (see the Table A.1 in the Appendix for the full list). Figures A.11a and A.11b show the time series of indicators calculated by topic and sector, together with a dating of important events having impacts on the Italian economy. Key events imply, according to their nature, different EPU specific indicators spiking up.

TEPU by topic is meant to capture the uncertainty in several sides of the economy such as monetary or fiscal policy, economic conditions and labor markets: as evident from Figure A.11a these topics seems to explain a lot of variation in our textual economic policy measures. TEPU calculated by sector is useful in detecting which sectors are the principal sources or are mostly affected by the business cycle uncertainty. Figure A.12a shows the contemporaneous correlation table of the overall TEPU index and its 15 components. The correlation patterns are similar to those previously described for the macro-topic components of our overall TESI: the highest correlation is reached by the topics on the top left part of the figure. Figure A.12b displays the contemporaneous correlation between the overall TEPU index and those built by economic sector. Also in this case results are similar to those described for sentiment in section 3.1: the correlation is higher for indexes calculates on articles about services, manufacturing, retail sales and construction sectors. Figure A.13 in the Appendix shows the three main TEPU macro topic indicators out of the 15 computed: 1) TEPU about government/fiscal policy; 2) TEPU about monetary policy, and 3) TEPU about labor markets.

Figure A.13a displays our TEPU index on government and fiscal policy articles. In the first part of the sample, the main peaks are around the Iraq's disarmament crisis towards the end of 1998, the *Dot-com* bubble in April 2000, Bush's election in November 2000. Afterwards, it moderately increases in periods when controversial Budget Laws had to pass the Parliament scrutiny and during national elections times. In the second part of the sample, high uncertainty episodes are the Sovereign Debt crisis, the Brexit referendum, the election of Trump, the fall of Renzi's government in December 2016 after the failed Constitutional Law referendum and the Government crisis in August 2019. TEPU reached its minimum when Italy adopted the euro at the beginning of 1999.

Figure A.13b shows TEPU stemming from monetary policy articles (Husted et al., 2019). Uncertainty regarding monetary measures was low when the euro was adopted. After that moment it stayed on low levels till the Global Financial crisis, when ECB started to adopt several unconventional measures such as the Fixed Rate Full Allotment Procedure after the Lehman collapse in October 2008 (Hartmann and Smets, 2018). The monetary TEPU decreased after the approval of important measures taken by the ECB such as the institution of the Security Market Programme in May 2010; the launch of the Very Long Term Refinancing Operations in December 2011; the Outright Monetary Transactions in September 2012; the announcement of the Expanded Asset Purchase Programme in January 2015 and the its restart after the announcement of APP purchases announced by the Governing Council in June 2019. On the contrary the index jumped up after the announcement of the reduction of the APP Programme in July 2017.

Finally, Figure A.13c shows TEPU for news concerning the Labor market. The index increased after the approval of important reforms for Italy such as the Treu's reform in June 1997 and the Biagi reform in September 2003.

TEPU indicators by sector, also used in the empirical application, are displayed in Figure A.14.[20]

5 Empirical applications

5.1 Monthly BMA model

We conduct a pseudo real-time forecasting experiment[21] to asses the ability of TESI and TEPU indices to track the economic activity. The Bayesian model averaging (BMA) described in Bencivelli et al. (2017) is employed to nowcast and forecast one step ahead

[20]EPU indexes not reported in the paper are available upon request.

[21]We use the last available vintage of data, i.e. December 2019, and we cut it recursively and we reproduce the missing values' pattern at the end of the sample.

the quarterly growth of both GDP and some of its main components (value added in service sector, VAS, which account for about 70% of the total economic activity in Italy; households consumption, HHC, and gross fixed investments, GFI).

BMA provides useful statistics to evaluate whether and to what extent our text-based indices contribute to improving both point-wise and density forecasts. The Monte Carlo Markov chain (MC^3) simulations are used to construct the confidence bands of the regression coefficients, the posterior inclusion probabilities (PIPs) of each regressor and to assess the ability of the text-based indices to shrink the uncertainty surrounding the forecasts.

BMA averages across the S randomly drawn models

$$y = \alpha_s \mathbf{1}_T + X_s \beta_s + \epsilon, \quad \epsilon \sim N(0, h^{-1} I_T) \tag{2}$$

where X_s is a $T \times K_s$ matrix with a subset of K_s regressors, weighting by their posterior probabilities (assuming an equal prior probability for all the models).[22] Model M^s is drawn among all the $R = 2^n$ possible ones, where n is the number of regressors. At each iteration, model M^{s+1} includes a variable in M^s or removes it. Based on M^s, the marginal likelihood $p(y|M^s)$ is estimated and the acceptance probability is computed as

$$p(M^s, M^{s+1}) = \left[\frac{p(y|M^{s+1})}{p(y|M^s)}, 1\right] \tag{3}$$

If M^s is accepted, its implied forecast is

$$\hat{y}^s_{t+h} = E(y_{t+h}|y_t, M^s) \tag{4}$$

and the final forecast is

$$\hat{y}_{t+h} = \frac{1}{S} \sum_{s=1}^{S} \hat{y}^s_{t+h} \tag{5}$$

The information set overall includes eight text-based indices distinguished by topic (sentiment and TEPU for economic conditions, service and manufacturing activity and sentiment for labor market and retail sales, see Table A.3 in the Appendix) and a selection of well-suited indicators extensively used to shape the short-term economic outlook, like the industrial production, Purchasing Managers' Indices (PMI), business and consumers' confidence indices, investors' expected earnings and, only for the HHC model, new cars registrations[23]. All the variables are demeaned (see Table A.3 for the detailed list of

[22]Refer to the extensive analysis in Bencivelli et al. (2017) for details.

[23]The HHC model does not include business confidence for service and manufacturing sectors and investors' expected earnings

variables included in each model and their pre-treatment). We train our model using data from January 2001 to December 2010, and we perform an out-of-sample forecast until December 2019 using a recursive window where the BMA is re-estimated as new data become available in real time.

The standard deviation of the investors earning expectations can be reasonably interpreted as an uncertainty measure and it has an appreciable negative correlation with the quarterly variation of the GDP. Therefore, it is particularly interesting comparing the latter to our uncertainty measure extracted from the textual data.

Figure 5 shows the nowcasts made with a model including our text-based indices (TB-model) with respect to a benchmark model, which excludes them from the information set. BMA's estimates are shown surrounded by the 25th and 75th percentiles bands, jointly with the historical data of the q-o-q growth of the targets variable. The TB-model turns out to grasp more effectively the depth of the downturn during the sovereign-debt crisis for all the targets. TB-model's nowcast is significantly negative for GFI in 2012.Q1, when the investments dropped by almost 6%, unlike the benchmark which provides a fairly brighter outlook. TB-model also tracks better the swing of the investments in 2017 and 2018, partly due to the uncertainty regarding the renewal of tax incentives. TB-model tracks the contraction of households' consumption between 2011 and 2014 significantly better than the benchmark but it is less effective in 2016, when it overestimates the slowing path of the dependent variable.

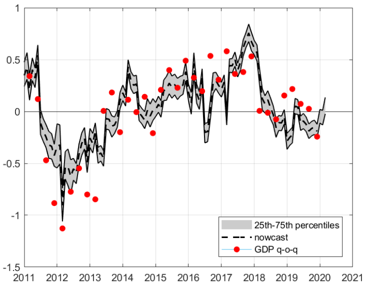
Table 2: Relative RMSFE for nowcasting (n) and forecasting (f)

	2011.1 - 2014.12		**2015.1 - 2019.12**		**2011.1 - 2019.12**	
	n	f	n	f	n	f
GDP	0.93	0.91	1.17	1.16	1.00	1.00
VAS	0.97	1.21	1.08	1.08	1.00	1.00
GFI	1.03	0.94	1.13	1.08	1.03	1.00
HHC	0.83	0.79	1.46	1.29	0.99	1.00

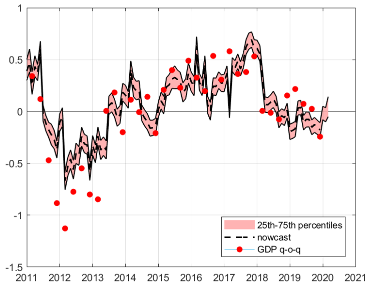
On the full sample the point forecasts do not show sizeable gains using the TB-model instead of its competitor, as measured by the relative RMSFE in Table 2 (the ratio between TB-model's and benchmark's RMSFE). In general, TB-model tends to lower the RMSFE during the most turbulent period in the sample, in particular for households' consumption.

The most interesting results derive from the density forecast assessment, based on the likelihood ratio test by Amisano and Giacomini (2007) (see Table 3). TB-model definitely

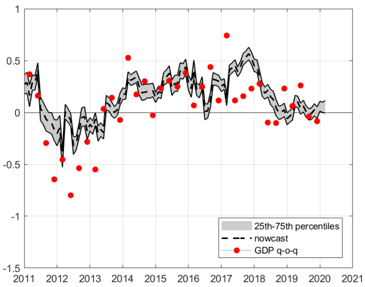
Figure 5: Nowcasts of the quarterly growth of GDP and its components (% changes)



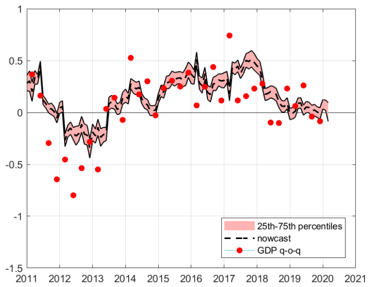
(a) TB-model (GDP)



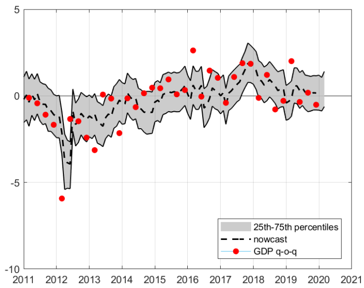
(b) Benchmark (GDP)



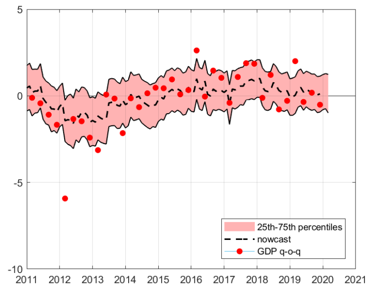
(c) TB-model (VAS)



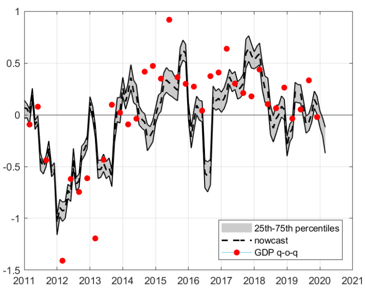
(d) Benchmark (VAS)



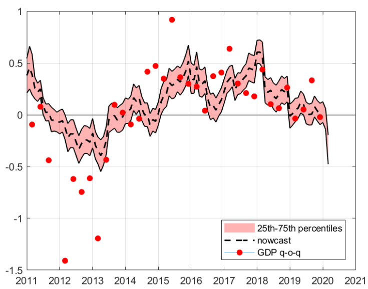
(e) TB-model (GFI)



(f) Benchmark (GFI)



(g) TB-model (HHC)



(h) Benchmark (HHC)

outperforms the benchmark, in particular during the sovereign-debt crisis, when the higher volatility challenges the forecasting heavily. Our TESI and TEPU indices squeeze the uncertainty around nowcasts and forecasts of the GFI evolution, both during crisis and in normal times.

BMA simulations provide the posterior inclusion probabilities as a measure of the

Table 3: Average log score

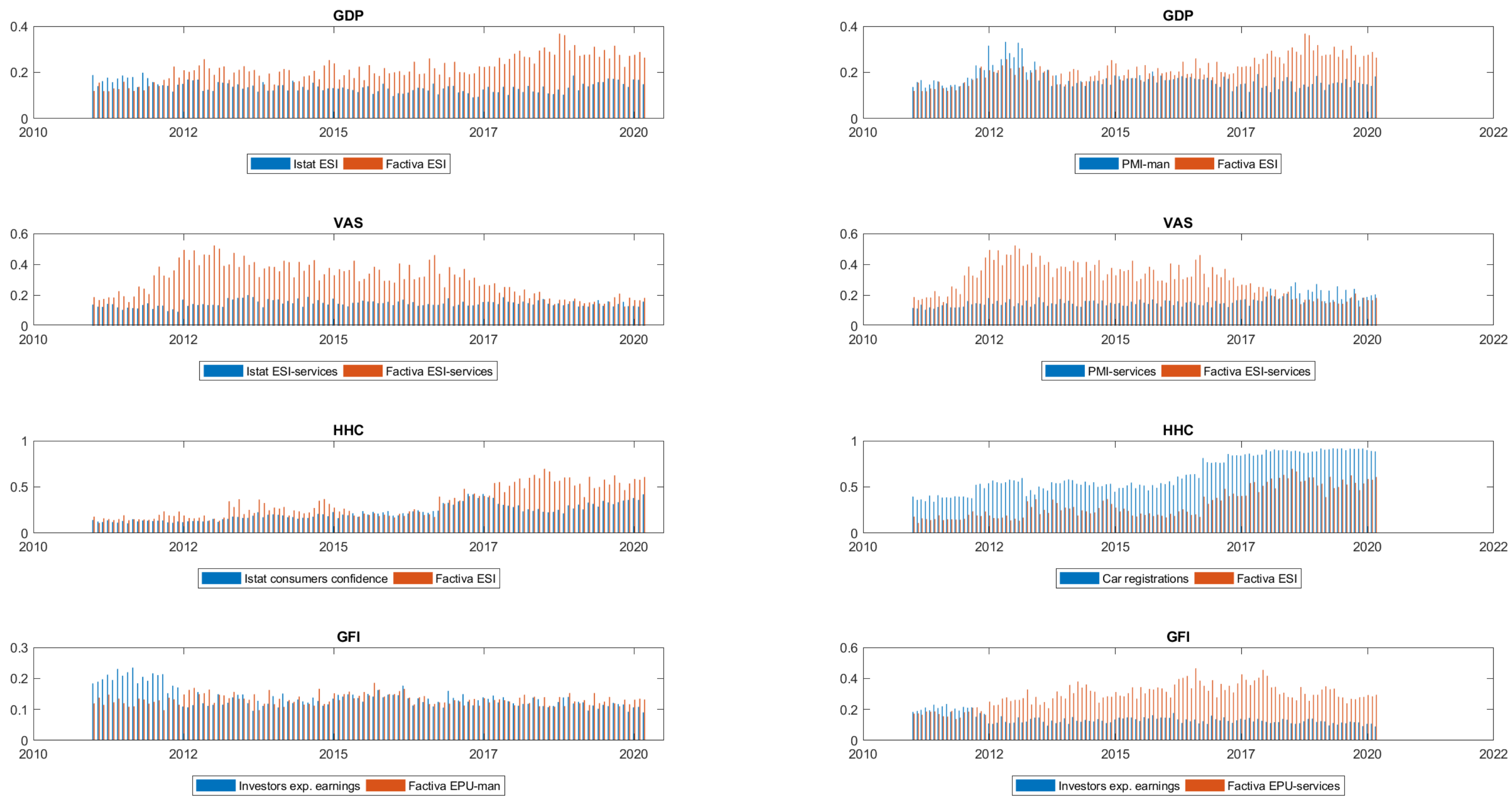
	2011.1 - 2014.12		**2015.1 - 2019.12**		**2011.1 - 2019.12**	
	n	f	n	f	n	f
GDP	9.1	8.6	-15.6	-22.4	6.1	6.6
VAS	5.3	6.6	-7.1	-10.3	3.7	6.4
GFI	3.6	23.7	6.9	10.5	4.6	24.5
HHC	14.6	12.2	-9.4	-11.8	11.4	11.8

relative importance of each regressor in explaining the variance of the dependent variable. By exploiting the pseudo real-time iterations, we can evaluate how PIPs evolve over time. The results, presented in Figure 6, refers to the nowcasting; TESI is an outstanding competitor of its popular counterparts produced by Markit (PMI) and by Istat (IESI). Our TEPU is as informative as the standard deviation of investors' expected earnings, which is a strong predictor for the short-term evolution of the Italian economic activity.

TESI is picked more frequently than IESI for nowcasting GDP's quarterly growth. The index referred to the topic "services activity" outperforms both the equivalent PMI and the Istat confidence throughout the sample and this result proves robust with respect to the target variable. The TESI for services is an important regressor p*er se*: it is selected up to 80% times for nowcasting GDP after the trough recorded in the sovereign-debt crisis. Further the PIPs of the TESI for manufacturing is comparable to those of PMI and Istat surveys indices. In the HHC model, TESI markedly overshadows the consumers' confidence index by Istat in the aftermath of the crisis and in the last two years, when it is persistently selected for nowcasting. Our EPU index for services is selected more frequently than the investors' expected earnings in both VAS and GFI models; for the latter, the PIP of the uncertainty measure extracted from the newspapers reaches almost 50% during the recovery phase, notwithstanding its high variability.

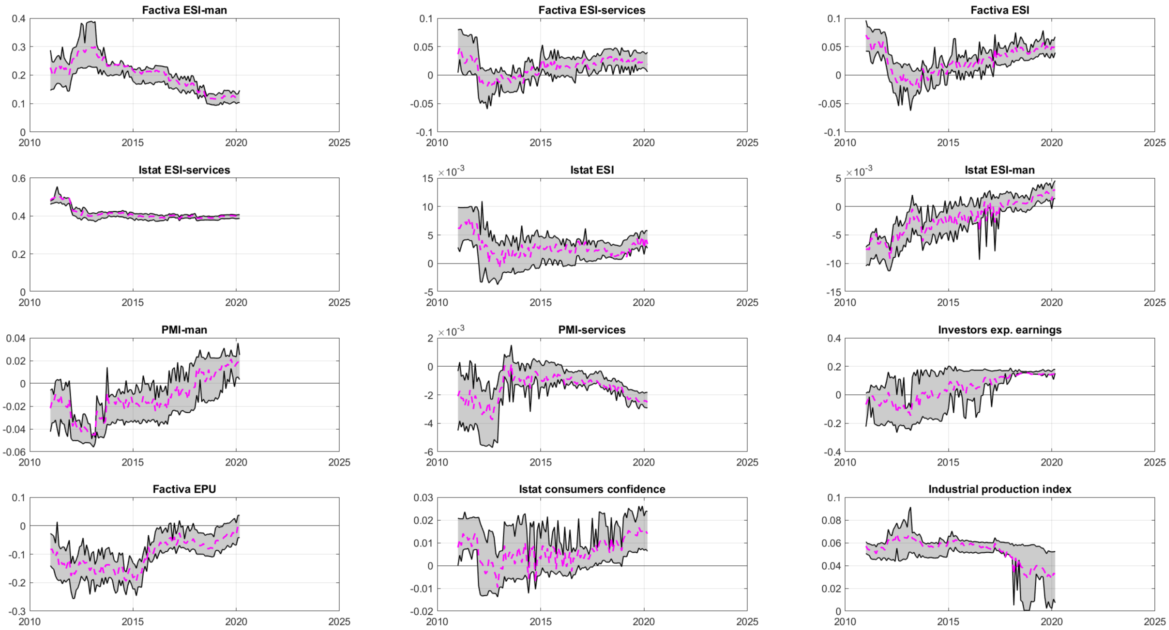
In order to measure the actual contribution of our sentiment and TEPU indices in the forecasting regressions, we focus on the coefficients. Figures from 7 to 10 show the results for the nowcasting exercise; the sentiment and TEPU indices extracted from the newspapers weigh more than many competitors such as official confidence indexes and PMIs. Text-based indices' coefficients display the expected sign and are larger in magnitude contrary to those attached to standard indicators, implying that they capture a reliable signal. TESI for manufacturing outperforms Istat's counterparts for manufacturing, services and consumers and it is outstanding even in comparison with the industrial production index. PMIs are forced on the wrong side: their contribution is shrunk toward

Figure 6: Posterior inclusion probabilities



zero and the sign of the betas is negative almost throughout the sample. TEPU gives a negative contribution as expected and its size is appreciable, particularly during the most turbulent period, unlike the investors' expected earnings, which is significantly close to zero in the same time interval and it becomes positive after 2015.

Figure 7: Regression coefficients and 25th-75th percentiles for GDP model



In the VAS model the coefficients associated to the Istat sentiment indices display a better pattern than that one of TESI, while PMI-services performs poorly. TEPU is again more suitable than investors' expected earnings, showing the correct sign. As for GFI, which is a volatile component of the GDP, TESI contributes more to explain its variance during the most critical period; the investors' expected earnings index has a negative and sizable coefficient up to 2015, when its contribution fades while TEPU plays a not negligible role during the whole sample. In HHC equation our sentiment index specific for the labor topic has a significantly positive contribution throughout the sample; moreover the coefficient of our TEPU is significantly negative during the first part of the sample during the Sovereign Debt Crisis in 2011, when both new cars registrations and the consumers' sentiment index by Istat lost their informative power glaringly.

5.2 A text-based weekly economic tracker

Following Lewis et al. (2020) and Delle-Monache et al. (2020) we build a weekly indicator of economic activity for Italy with two innovations: i) a different dataset, focused

Figure 8: Regression coefficients and 25th-75th percentiles for VAS model

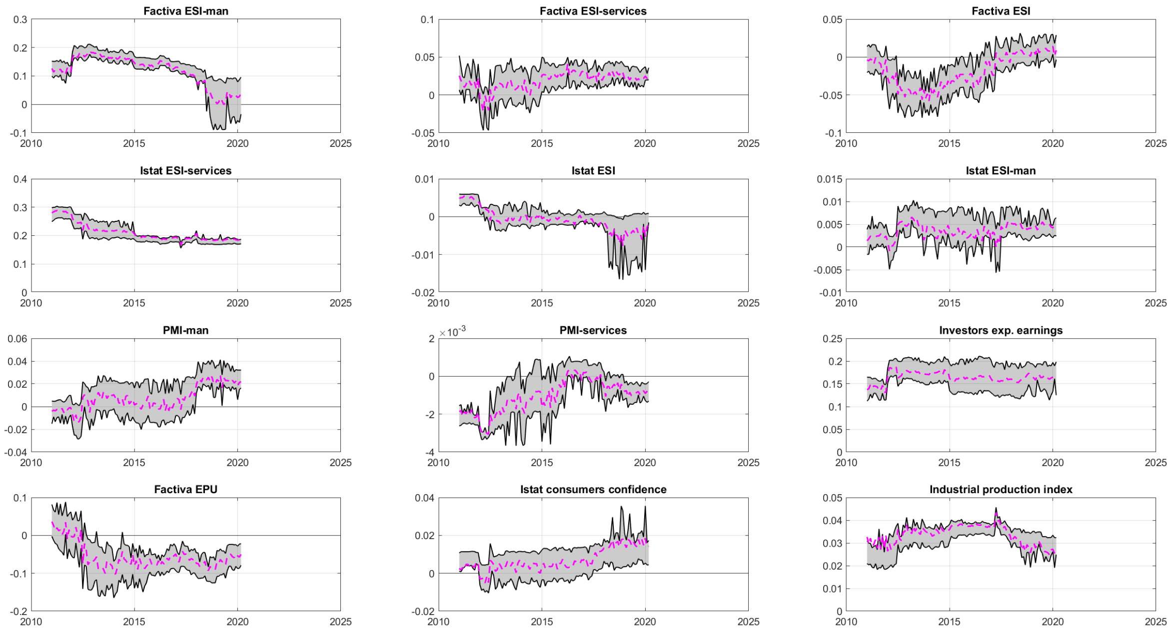


Figure 9: Regression coefficients and 25th-75th percentiles for GFI model

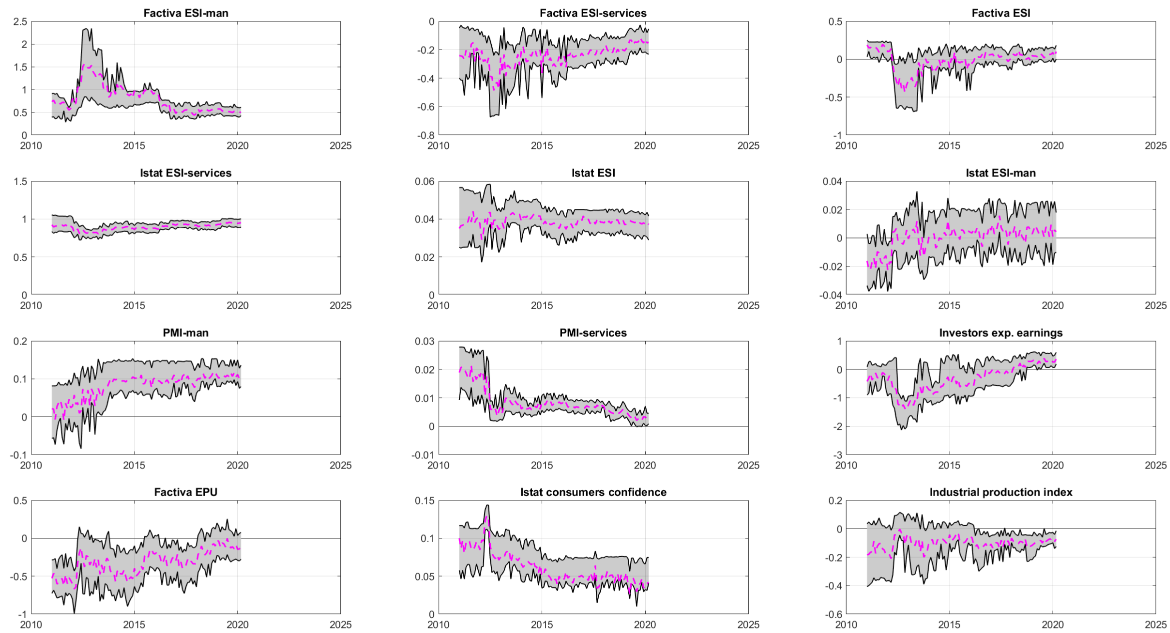
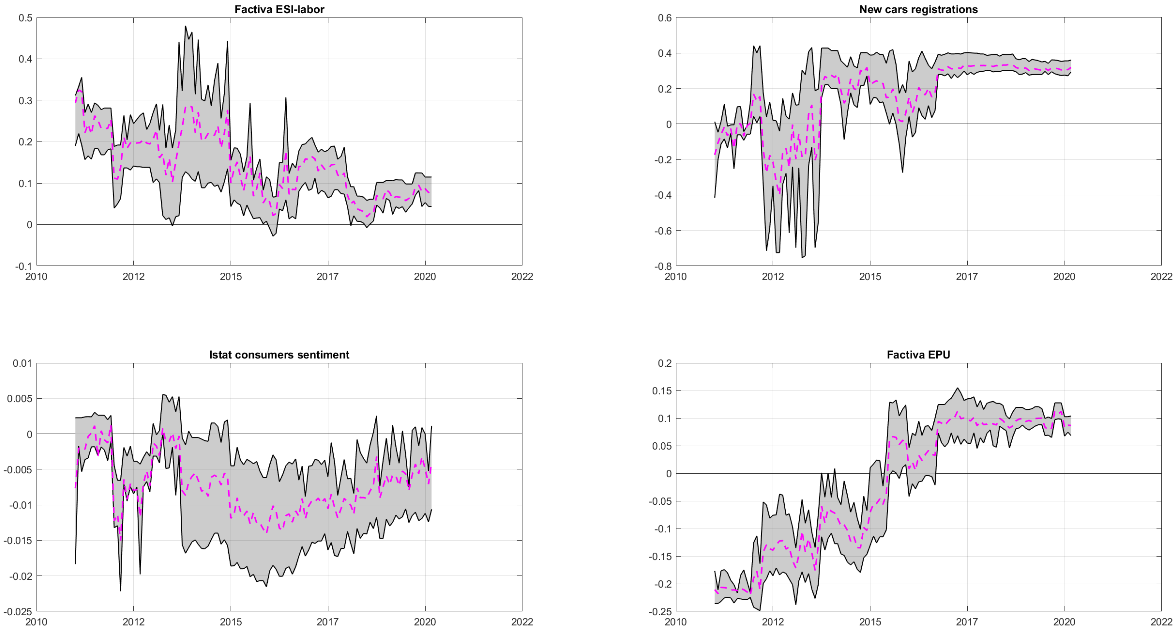


Figure 10: Regression coefficients and 25th-75th percentiles for HHC model



mainly on sentiment and uncertainty indexes; ii) a slightly diverse modeling approach that includes the Italian GDP y-o-y growth rate in the estimation of the weekly tracker. Adding the weekly TESI and TEPU series to a set of standard macroeconomic variables commonly used in nowcasting (see Table A.4 for the full list of variables), we obtain large point-forecast gains in weekly nowcasts of the quarterly Italian GDP growth. These improvements in forecast accuracy are large across the whole out-of-sample period (2009-2019) and every sub-period, with a 0.88 relative RMSFE in both expanding and rolling window for in-sample training, but are particularly strong (0.74 relative RMSFE with expanding window, 0.65 with rolling window) during periods of negative growth (2009, 2011-2013).

We build two weekly datasets where monthly variables have been disaggregated at weekly frequency via Chow-Lin methodology (Chow and Lin, 1971), using the higher frequency indicators available. The first dataset (X^{Bench}), used to calculate the benchmark weekly index , contains traditional data used in nowcasting such as electricity consumption and PMIs; the second dataset (X^{Text}) expands the first dataset to include sentiment (TESI) and uncertainty indicators (TEPU) for Italy. We make an horse race between two weekly economic activity indexes, $Z^{Bench}_{MA,t}$ and $Z^{Text}_{MA,t}$, derived from the two datasets as follows:

1. We extract for each dataset the first principal component;

2. We take a 13-periods ($\approx$ one quarter) moving average of the first principal component.

Since $Z^{Bench}_{\text{MA},t}$ and $Z^{Text}_{\text{MA},t}$ are standardized principal components, we scale them to the quarterly year-on-year GDP growth rate, $\Delta Y_{yoy,t}$, via the following regression:

$$\Delta Y_{yoy,t} = \alpha^i_t + \beta^i_t Z^i_{\text{MA},t} + \varepsilon_t \quad t = 1, \ldots, T \quad and \quad i = Bench, Text \tag{6}$$

Using the estimated coefficients in (6) we compute the value of the two weekly GDP trackers at week t, W^i_t for $i = Bench, Text$, as:

$$W^i_t = \widehat{\alpha}^i_t + \widehat{\beta}^i_t Z^i_{\text{MA},t} \quad t = 1, \ldots, T \quad and \quad i = Bench, Text$$

For training the model we use only available data for both the GDP and the principal component up to week t and we compute the corresponding nowcast error as:

$$error^i_t = \Delta Y_{yoy,t} - W^i_t.$$

We train our model using weekly data from January 2001 to December 2010; the out-of-sample goes from January 2011 till December 2019. Forecasting accuracy measures are computed on out of sample data using both an expanding and a rolling (335 weeks) window.

Table 4: Relative RMSFE of Weekly GDP trackers

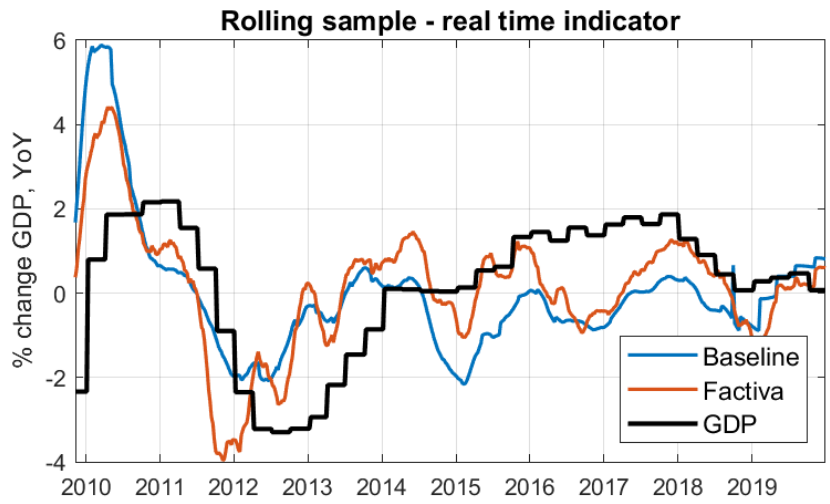
	Expanding	Rolling (335 weeks)
All sample	0.85	0.83
Negative GDP	0.88	0.96
Positive GDP	0.82	0.75

Table 4 shows the relative RMSFE as the ratio between the two weekly activity indexes taking the one computed with traditional data (W^{Bench}_t) as the benchmark. The one that includes TESI and TEPU series (W^{Text}_t), displays a better performance in nowcasting the year-on-year GDP growth since the relative RMSFE is smaller than 1. For the rolling sample, we run a Diebold-Mariano test [24] and find that the gains are significant at the 5% level.

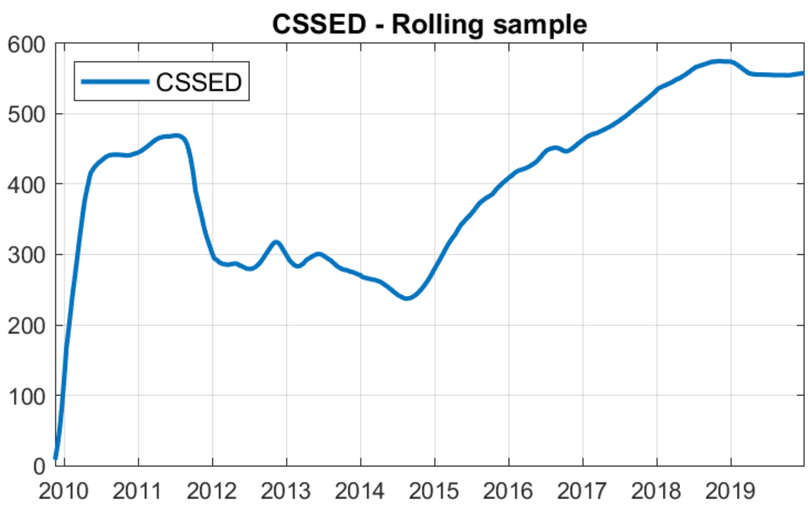
Figure 11a shows the point forecast gains documented in table 4 for the rolling case over the whole out of sample horizon. Similar results are obtained when an expanding scheme is implemented (Figure A.15a).

[24]The DM test is not consistent for nested models if the sample is expanding.

Figure 11: Rolling sample - Real-time weekly indicators of the GDP and Cumulative Sum of Squared Errors Difference (CSSED).



(a) Italian weekly real-time index of economic activity calculated using a rolling sample



(b) CSSED on the nowcasting errors comparing the models with and without our text based indexes on the rolling sample.

Using the the Cumulated Sum of Squared Errors Difference (CSSED) [25] proposed by Welch and Goyal (2007), Figure 11b shows the magnitude and stability of the forecasting gains of the W_t^{Text} weekly index throughout the whole out of sample period when a rolling scheme is used. Similar results apply when estimation is carried out with an expanding sample (Figure A.15b).

6 Conclusions

We use daily newspapers article and text-mining techniques to construct sentiment and economic policy uncertainty indicators. They match up well with the Italian business cycle and with important events shaping it; their timeliness and high-frequency availability make them suitable candidates for tracking short-run developments of the Italian economic activity.

The use of text data from newspapers articles for nowcasting and forecasting is also motivated by the survey evidence on the role played by newspapers as an important source of information for firms, affecting their decisions on investments and the workforce.

We compute overall sentiment and TEPU indicators, together with 36 topic- and sector-specific ones, by using a dictionary-based approach to extract quantitative information from the text of the articles. We compile a brand new Italian economic dictionary tailored for newspaper economic news. Differently from most of the literature, our dictionary contains valence-shifting words, which help to interpret the overall meaning of each sentence better.

Our indicators prove effective in nowcasting and forecasting quarterly growth of the economy, both at monthly and weekly frequency. A monthly BMA application shows that our text-based sentiment (TESI) and uncertainty indicators (TEPU) help to improve the accuracy of the point-projections during recessions, despite they provide little gains across the whole sample; interesting results come out in terms of density forecasts. When used at the weekly frequency, our indicators provide sizeable and statistically significant gains in nowcasting the Italian GDP growth.

A number of questions still have to be addressed. The measures presented in this work can be further tuned for specific applications: separating news about the past and future can allow to use different types of information for different applications. Further exploring the observed source-specific heterogeneity in news volumes, polarity and what specific "stories" are reported - instead of simply averaging - may provide additional information useful to track the business cycle. We plan to evaluate in the near future

[25] The CSSED (Cumulated Sum of Squared Errors Difference) is calculated as $CSSED_\tau = \sum_{\tau=R}^{T}(e^2_{\tau,(\text{baseline})} - e^2_{\tau,(\text{factiva})})$, so that positive values show that the model with textual-variables outperformed the benchmark between time R and T.

the ability of these textual measures in forecasting economic activity during the Covid-19 pandemic crisis. When the economy is subject to strong and abrupt regime shifts, these measures can be really important since they provide timely, high-frequency and non-revised information.

References

Amisano, Gianni and Raffaella Giacomini, "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business & Economic Statistics*, 2007, *25* (2), 177–190.

Ardia, David, Keven Bluteau, Samuel Borms, and Kris Boudt, "The R package sentometrics to compute, aggregate and predict with textual sentiment," *Aggregate and Predict with Textual Sentiment (November 9, 2017)*, 2017.

Ardizzi, Guerino, Simone Emiliozzi, Juri Marcucci, and Libero Monteforte, "News and consumer card payments," *Bank of Italy Temi di Discussione (Working Paper)*, 2019, *No. 1233.*

Bachmann, Rüdiger, Steffen Elstner, and Eric R Sims, "Uncertainty and economic activity: Evidence from business survey data," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 217–49.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, "Measuring economic policy uncertainty," *The quarterly journal of economics*, 2016, *131* (4), 1593–1636.

Bencivelli, Lorenzo, Massimiliano Marcellino, and Gianluca Moretti, "Forecasting economic activity by Bayesian bridge model averaging," *Empirical Economics*, 2017, *53* (1), 21–40.

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

Bok, Brandyn, Daniele Caratelli, Domenico Giannone, Argia M Sbordone, and Andrea Tambalotti, "Macroeconomic nowcasting and forecasting with big data," *Annual Review of Economics*, 2018, *10*, 615–643.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, "The Structure of Economic News," *Available at SSRN 3446225*, 2019.

Chow, Gregory C and An loh Lin, "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series," *The Review of Economics and Statistics*, 1971, pp. 372–375.

Delle-Monache, Davide, Simone Emiliozzi, and Andrea Nobili, "Tracking economic growth during the Covid-19: a weekly indicator for Italy," *Bank of Italy, Mimeo*, 2020.

Fraiberger, Samuel P, "News Sentiment and Cross-Country Fluctuations," *Available at SSRN 2730429*, 2016.

Gentzkow, Matthew and Jesse M Shapiro, "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 2010, *78* (1), 35–71.

_ **, Bryan Kelly, and Matt Taddy**, "Text as Data," *Journal of Economic Literature*, September 2019, *57* (3), 535–74.

_ **, Jesse M Shapiro, and Matt Taddy**, "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 2019, *87* (4), 1307–1340.

Groseclose, Tim and Jeffrey Milyo, "A measure of media bias," *The Quarterly Journal of Economics*, 2005, *120* (4), 1191–1237.

Hartmann, Philipp and Frank Smets, "The first 20 years of the European Central Bank: Monetary policy," in "Brookings Papers on Economic Activity Conference Draft, Fall" 2018.

Husted, Lucas, John Rogers, and Bo Sun, "Monetary policy uncertainty," *Journal of Monetary Economics*, 2019.

Kalamara, Eleni, Arthur Turrell, Chris Redl, George Kapetanios, and Sujit Kapadia, "Making Text Count: Economic Forecasting Using Newspaper Text," Staff Working Paper, number=865, Bank of England May 2020.

Kelly, Bryan, Asaf Manela, and Alan Moreira, "Text Selection," Working paper 2018.

Lewis, Daniel J, Karel Mertens, James H Stock et al., "Monitoring Real Activity in Real Time: The Weekly Economic Index," Technical Report, Federal Reserve Bank of New York 2020.

Loughran, Tim and Bill McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of Finance*, 2011, *66* (1), 35–65.

_ **and** _ , "Measuring readability in financial disclosures," *The Journal of Finance*, 2014, *69* (4), 1643–1671.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, 2013, *26*, 3111–3119.

Neri, Stefano and Stefano Siviero, "The non-standard monetary policy measures of the ECB: motivations, effectiveness and risks," *Bank of Italy Occasional Paper*, 2019, (486).

Ng, Serena and Jonathan H. Wright, "Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling," *Journal of Economic Literature*, December 2013, *51* (4), 1120–54.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning, "Glove: Global vectors for word representation," in "Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)" 2014, pp. 1532–1543.

Rambaccussing, Dooruj and Andrzej Kwiatkowski, "Forecasting with news sentiment: Evidence with UK newspapers," *International Journal of Forecasting*, 2020, *36* (4), 1501 – 1516.

Roberts, Margaret E, Brandon M Stewart, and Dustin Tingley, "Navigating the local modes of big data," *Computational social science*, 2016, *51.*

Rogers, John and Jiawen Xu, "How Well Does Economic Uncertainty Forecast Economic Activity?," Finance and Economics Discussion Series 2019-085, Board of Governors of the Federal Reserve System 2019.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel Wilson, "Measuring News Sentiment," *Journal of Econometrics*, *forthcoming* 2020.

Soroka, Stuart N, "Good news and bad news: Asymmetric responses to economic information," *The journal of Politics*, 2006, *68* (2), 372–385.

Tetlock, Paul C, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, 2007, *62* (3), 1139–1168.

Thorsrud, Leif Anders, "Words are the new numbers: A newsy coincident index of the business cycle," *Journal of Business & Economic Statistics*, 2018, pp. 1–17.

Welch, Ivo and Amit Goyal, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *The Review of Financial Studies*, 03 2007, *21* (4), 1455–1508.

Appendix

Additional figures and tables

Table A.1: Sentiment indexes calculated by Topic and Sector

N.	Sentiment by Topic	N.	Sentiment by Sector
1)	Economic Condition	1)	Automation
2)	Finance	2)	Agriculture - Mining
3)	Domestic Policy	3)	Manufacture
4)	Foreign Policy	4)	Services
5)	Government	5)	Agriculture
6)	Institutions	6)	Automotive
7)	Labor	7)	Basic Materials & Resources
8)	Fraud	8)	Business Consumer Services
9)	Migration	9)	Consumer Goods
10)	Monetary Policy	10)	Energy
11)	Natural Disasters	11)	Financial Services
12)	Prices	12)	Health Care & Life Sciences
13)	Private Sector	13)	Industrial Goods
14)	Terrorism	14)	Leisure, Arts & Hospitality
15)	Pandemic	15)	Media & Entertainment
		16)	Real Estate & Construction
		17)	Retail & Wholesale
		18)	Technology
		19)	Telecommunication Services
		20)	Transportation Logistics
		21)	Utilities

Table A.2: Sample taken from the Italian Dictionary used to calculate the Italian Sentiment indexes. For valence shifters, the "type" column is an equivalent classification to the "value" column.

Dictionary terms	**polarity**	**Valence shifters**	**type**	**value**
ristagneranno	-1	una crescita di	2	2
ristagnerà	-1	una crescita del	2	2
ristagnare	-1	una crescita delle	2	2
ristagnano	-1	una crescita della	2	2
risorse	1	una contrazione delle	1	-1
risorse	1	una contrazione della	1	-1
rischio	-1	una contrazione	1	-1
rischia di	-1	una contrazione del	1	-1
risanamento dei conti pubblici	1	un rialzo	2	2
risanamento	1	un rallentamento	1	-1
ripresa economica	1	un incremento	2	2
ripresa	1	un aumento	2	2
rilassamento	1	sufficientemente	3	0.5
riforme	1	specialmente	2	2
riformare	1	sotto le attese	1	-1
riforma	1	sotto le aspettative	1	-1
rialzo dei tassi di interesse	-1	sopratutto	2	2
rendimento del decennale italiano	-1	soprattutto per	2	2
rendimento del btp decennale	-1	sono peggiorati	1	-2
rendimento dei titoli di stato	-1	sono peggiorate	1	-2
rendimento	-1	sono migliorati	2	2
rendimenti dei titoli di stato	-1	sono migliorate	2	2
rendimenti	-1	sicurezza	2	2
record	1	si sono ridotti gli	3	0.5
recessione	-1	si sono ridotte le	3	0.5
realismo	1	si è ridotto il	3	0.5
rapporto tra deficit e pil	-1	si è ridotta la	3	0.5
rapporto deficit pil	-1	senza successo	1	-1
pil	1	senza	1	-1
rapidità	1	sempre più	2	2
rally delle borse	1	secondario	3	0.5
rallentamento generale	-1	secondari	3	0.5
rallentamento economico	-1	sarà peggiore	1	-1
cuneo fiscale	-1	sarà migliore	2	2
crisi di liquidità	-1	sarà meglio	2	2
crisi	-1	rilevanti	2	2
crimini	-1	rilevante	2	2
crimine	-1	riduzione di	1	-1
crescita del prodotto interno lordo	1	riduzione delle	1	-1
crescita del PIL	1	riduzione della	1	-1

Notes: The first column contains the Dictionary words while the second contains their polarity. The third column shows *Valence shifter terms* together with the associated type (1 = negation, 2 = amplifiers, 3 = deamplifiers) and the associated multiplicative value. The "type" is the input taken by Sentometrics to compute by how much a polarity word should be "shifted".

Figure A.1: Article information in Factiva dataset


HTML Screenshot

Metadata

SE PRIMO PIANO
HD Rallenta la crescita, plausibile una manovra da 9 miliardi
BY Nicoletta Picchio
WC 472 words
PD 28 June 2018
SN Il Sole 24 Ore Digital Replica Edition of Print Edition
SC SOLE
PG 02
LA Italian
CY © Copyright Il Sole 24 Ore- Tutti i diritti riservati.
LP ROMA

Article text

L'**economia** italiana rallenta nel biennio 2018-2019, un calo «anticipato e più ampio» rispetto alle stime di dicembre dell'anno scorso. Frenata che renderebbe «plausibile una manovra correttiva in corso d'anno da 0,5 punti di Pil, pari a 9 miliardi». Il Centro studi Confindustria ieri ha diffuso i nuovi dati: il pil quest'anno salirà dell'1,3% (-0,2 rispetto alle previsioni precedenti) e dell'1,1 nel 2019 (-0,1). Andamento dovuto ad una serie di fattori: a livello internazionale le nuove politiche protezionistiche degli Stati Uniti creano **incertezza** sul futuro degli scambi mondiali e a ciò si aggiungono le tensioni geopolitiche. Già si osserva un rallentamento degli scambi mondiali che si riflette sull'export italiano. Le esportazioni, ha spiegato Andrea Montanino, direttore del Centro studi, aumenteranno meno della domanda mondiale nel 2018 per la prima volta dal 2013. L'Italia, quindi tornerà a perdere quote di mercato.

TD Inoltre si va esaurendo il ciclo degli investimenti a livello nazionale, anche per l'avvicinarsi della fine degli incentivi. In uno scenario in cui cresce il costo del finanziamento: +100 punti base ad oggi rispetto alla media dei primi quattro mesi pesa sul finanziamento dell'**economia** reale, oltre al fatto che l'aumento dello spread rende l'Italia un rischio per l'area euro. L'occupazione, pur continuando a crescere, ha perso slancio: aumenterà dello 0,8% nel 2018 e dello 0,7 nel 2019 contro la media del +1,2 nel 2017. Quella dipendente torna ad essere trainata dal lavoro temporaneo. Il costo del lavoro per unità di prodotto tornerà a crescere nel 2018, +0,4% e balzerà dell'1% nel 2019.

La crescita che rallenta, sottolinea il Csc, si riflette sui conti pubblici: «ci sono pochi spazi di **bilancio** per l'Italia», ha detto Montanino, anche perché il percorso di risanamento negli anni passati è stato debole, a differenza di gran parte dei paesi Ue. L'indebitamento della Pa è previsto all'1,9 nel 2018 e all'1,4 nel 2019, al di sopra dei target di governo e condivisi con l'Europa. È «palusibile» quindi, dice il Csc, la richiesta di una manovra correttiva di 0,5 punti di pil nel 2018 (9 miliardi), che non è stata calcolata nelle previsioni. Nel 2019 la correzione dovrebbe essere di 0,6 punti (quasi 11 miliardi). È stata richiesta e ottenuta molta flessibilità in Europa, quasi 30 miliardi, e le clausole di salvaguardia sono state disinnescate per tre quarti in **deficit**. Ora molto dipenderà, dice il Centro studi, dalle scelte di **politica economica** che saranno adottate sia sulle clausole di salvaguardia sia sull'attuazione del contratto di governo. Pesa anche la gradualità di come saranno realizzate le misure.

© RIPRODUZIONE RISERVATA

Topics

NS e11 : Economic Performance/Indicators | e1108 : Budget Account | e1101 : Economic Growth/Recession | e211 : Government Budget/Taxation | ccat : Corporate/Industrial News | gcat : Political/General News | e21 : Government Finance | ecat : Economic News

RE italy : Italy | eecz : European Union Countries | eurz : Europe | medz : Mediterranean | weurz : Western Europe

PUB Il Sole 24 Ore SpA

Identifier

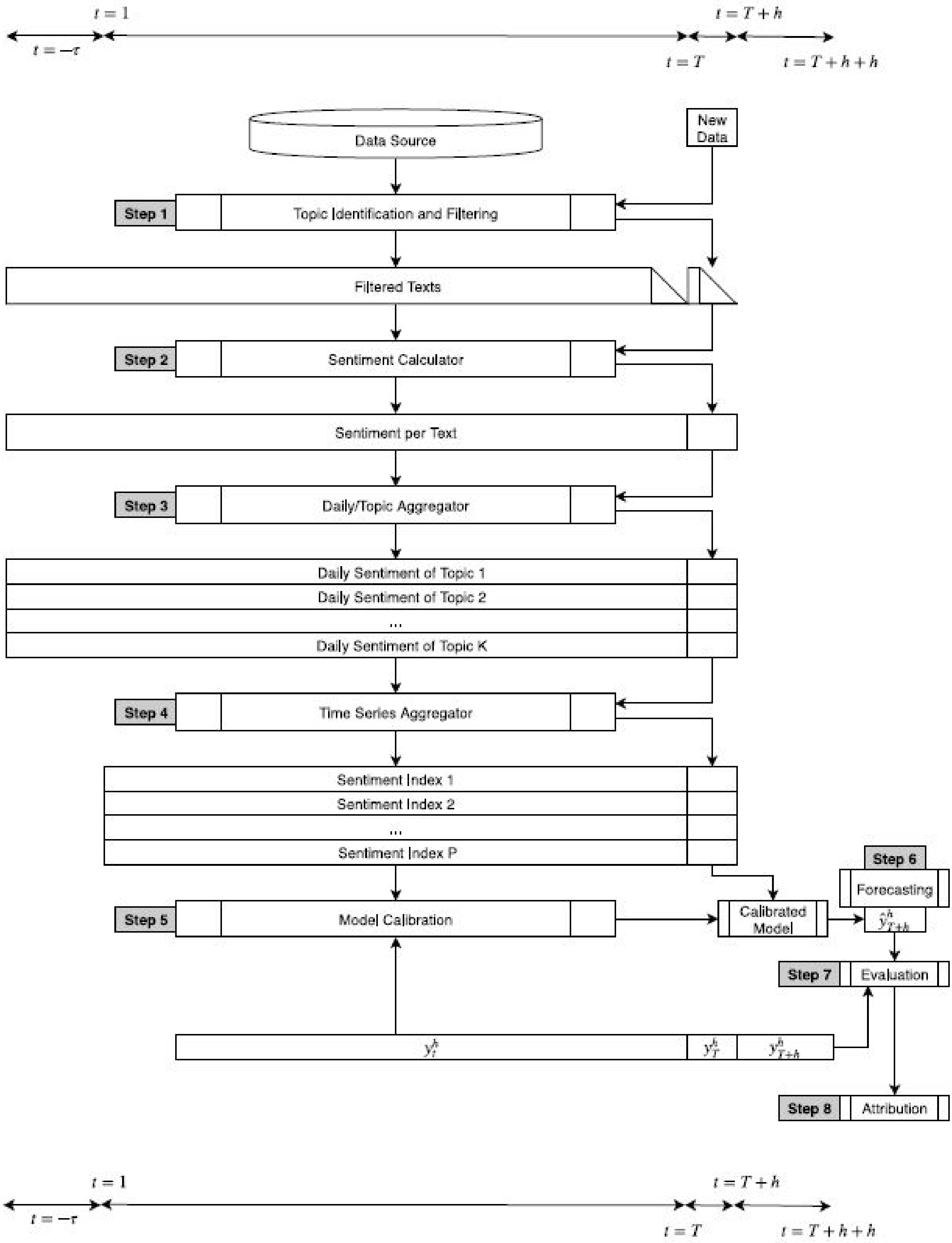
AN Document SOLE000020180628ee6s0002r

Table A.3: Variables included in the BMA exercise: Baseline vs Factiva specification.

				GDP		HHC		GFI		VAS	
N	label	Description	Treatment	Base	Text	Base	Text	Base	Text	Base	Text
1	ITCNFCONR	ITA household confidence index	none	x	x	x	x	x	x	x	x
2	ITCNFBUSQ	ITA business confidence indicator	none	x	x	x	x	x	x	x	x
3	ITTOTPRDR	ITA business svy.: production level	none	x	x			x	x	x	x
4	ITEUSVCIQ	ITA services: confidence sadj	none	x	x			x	x	x	x
5	ITIPMAN.G	ITA industrial prod.- manufacturig	deltaog(1)	x	x	x	x	x	x	x	x
6	EMPMIM..Q	PMI Manufacturing - EA	none	x	x	x	x	x	x	x	x
7	ITPMIM..Q	PMI Manufacturing - IT	none	x	x	x	x	x	x	x	x
8	ITPMIS..Q	PMI Services - IT	none	x	x	x	x	x	x	x	x
9	@:ITMSCIP	Weighted ave. Std. Dev. of the EPS forecast for the t+1 Fiscal Year	deltalog(12)	x	x			x	x	x	x
10	AUTOD	Car registrations in Italy	deltaog(1)			x	x				
11	fct_sent	TESI	zscore MA(3)		x		x		x		x
12	fct_epu	TEPU	zscore MA(3)		x		x		x		x
13	fct_sent_man	TESI - manufacturing	zscore MA(3)		x		x		x		x
14	fct_sent_ser	TESI - Services	zscore MA(3)		x		x		x		x
15	fct_epu_man	TEPU - manufacturing	zscore MA(3)		x		x		x		x
16	fct_epu_ser	TEPU - Services	zscore MA(3)		x		x		x		x
17	fct_sent_lab	TESI - Labor	zscore MA(3)		x		x		x		x
18	fct_sent_ret	TESI - Retail	zscore MA(3)		x		x		x		x

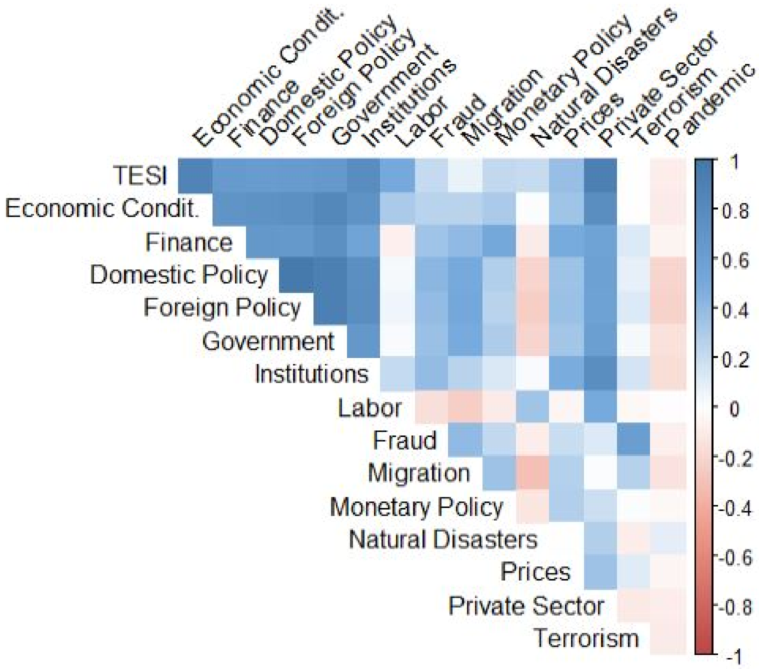
Notes: Base is the Baseline model and Text is the Text-based model

Figure A.2: Pipeline followed for calculating the Sentiment

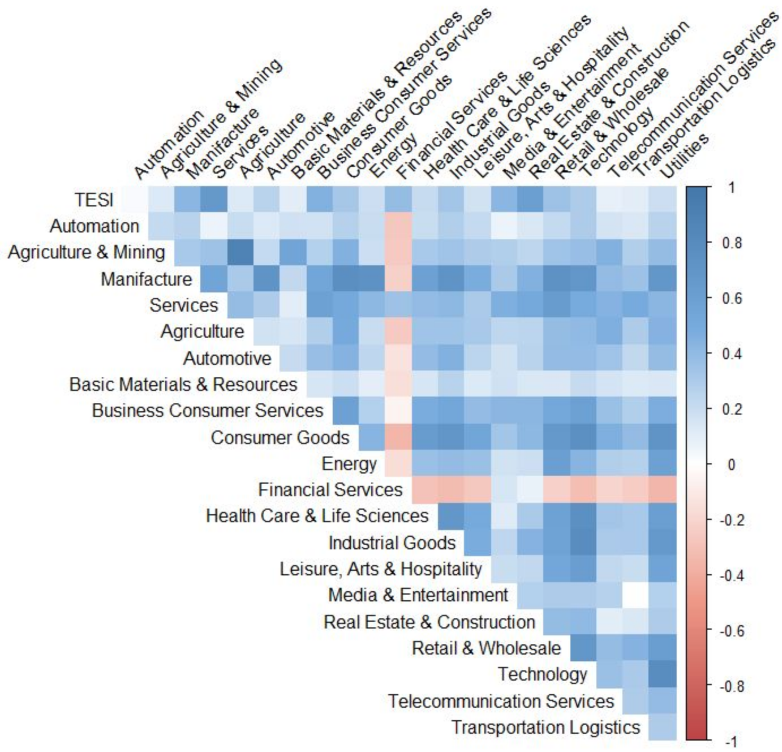


Source: this figure is taken from Ardia et al. (2017)

Figure A.3: Correlation between TESI and a) the the 15 topics derived from Italian newspapers; b) the 21 sentiment indexes derived by sector. Series are aggregated at monthly frequency



(a) Correlation between TESI and the 15 topic specific indexes; monthly frequency.

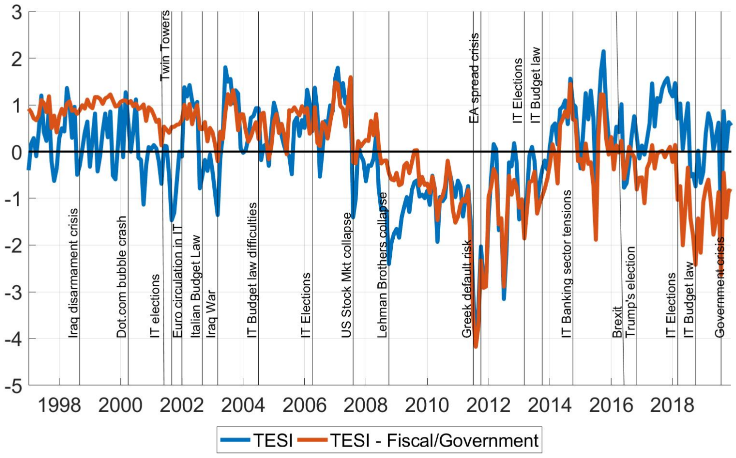


(b) Correlation between TESI and the sectoral indexes; monthly frequency.

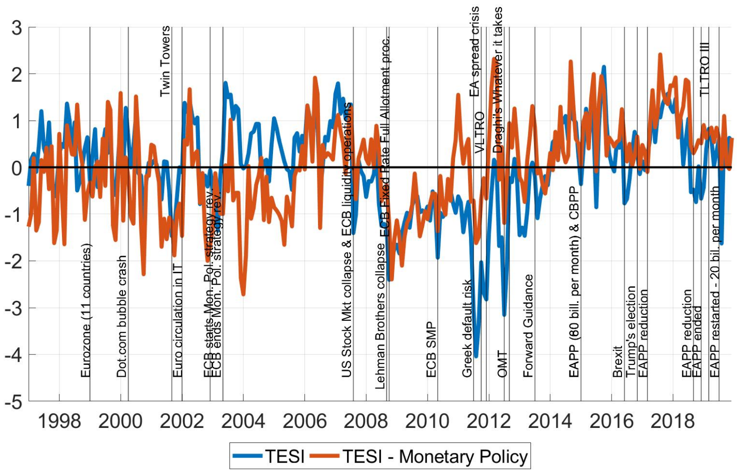
Table A.4: Variables used in the estimation of the weekly index of economic activity

Variable	Frequency	Baseline	Factiva
Electricity consumption	Weekly	x	x
Earning forecast std	Weekly	x	x
ISTAT Confidence	Monthly	x	x
PMI indices	Monthly	x	x
TESI	Weekly		x
TEPU	Weekly		x
TESI manufacturing	Weekly		x
TEPU manufacturing	Weekly		x
TESI services	Weekly		x
TEPU services	Weekly		x
TESI labor	Weekly		x
TESI retail	Weekly		x
TESI leisure	Weekly		x

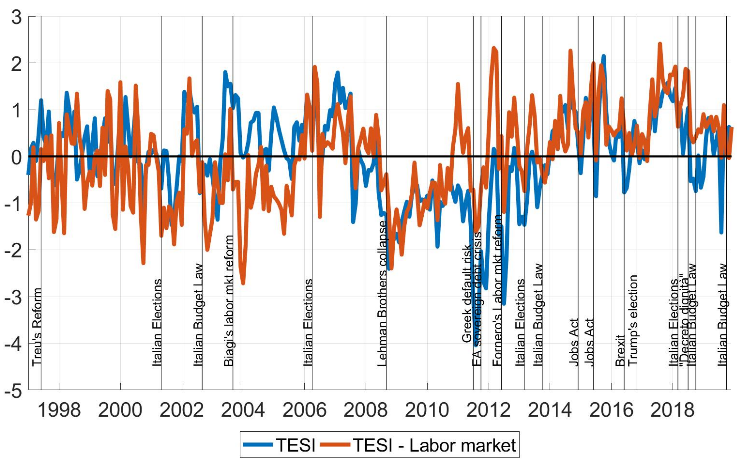
Figure A.4: Sentiment (TESI) for three major macro topics: 1) Government; 2) Monetary Policy ; 3) Labor Markets



(a) TESI on Government macro-topic

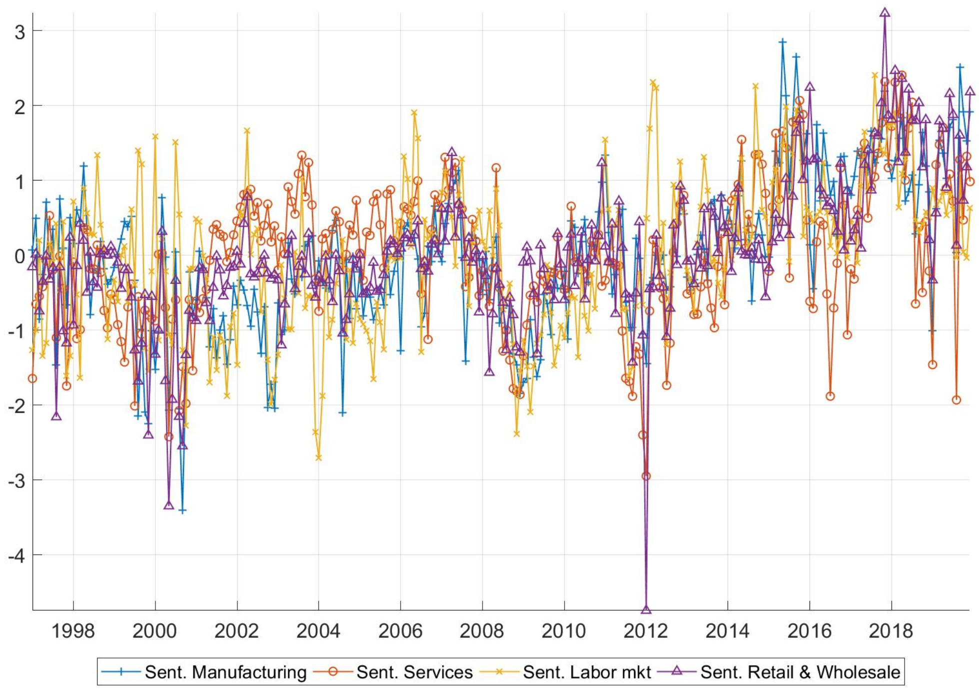


(b) TESI on Monetary Policy macro-topic



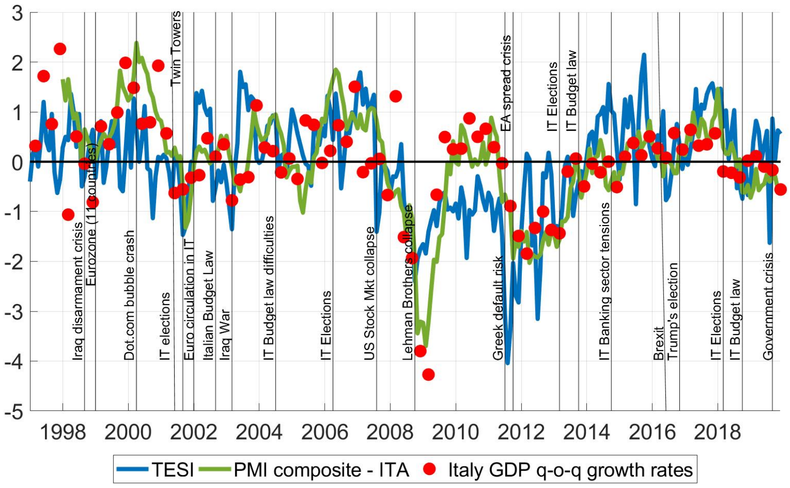
(c) TESI on Labor markets macro-topic

Figure A.5: Sentiment indicators (TESI) by sector

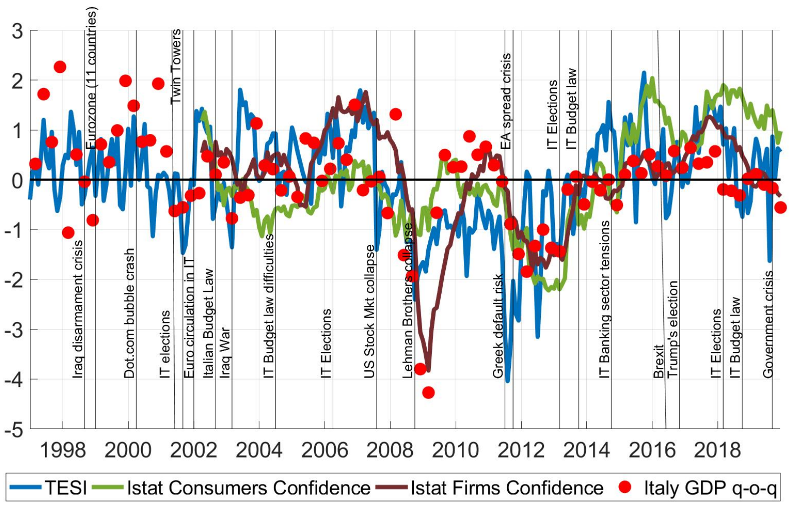


Notes: the series are standardized to have zero mean and unit variance.

Figure A.6: Comparison between TESI, the PMI composite and the confidence indexes provided by Istat



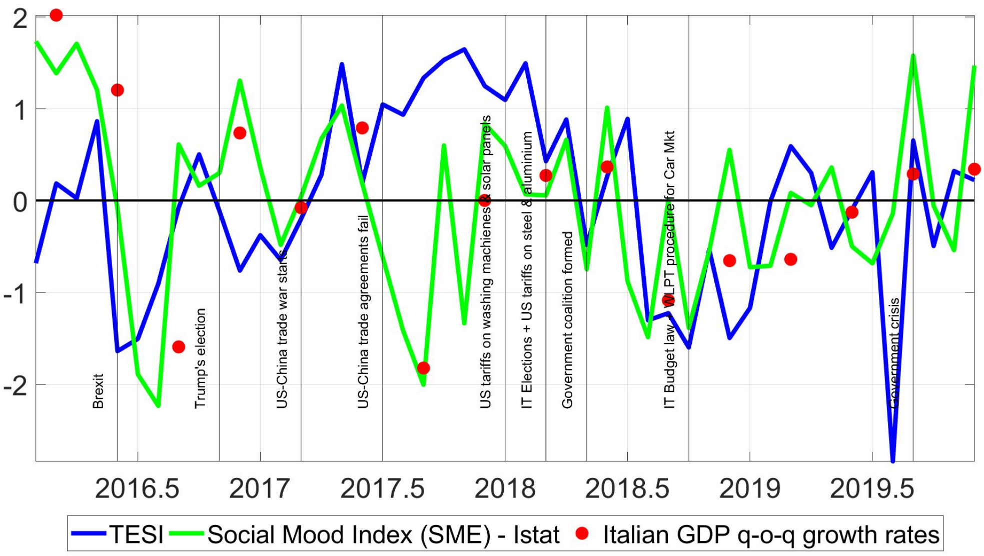
(a) Comparison between TESI and the PMI composite



(b) Comparison between TESI and the official Business confidence indicator (IESI) and Consumers confidence indicator

Notes: the series in the graphs are standardized to have zero mean and unit variance.

Figure A.7: Comparison between TESI and Social Mood Index (SMI) derived from Twitter (Istat)



Notes: the series in the graphs are standardized to have zero mean and unit variance. SMI is available from 2016, hence the comparison is over the period January 2016 - December 2019.

Figure A.8: Newspaper-specific sentiment

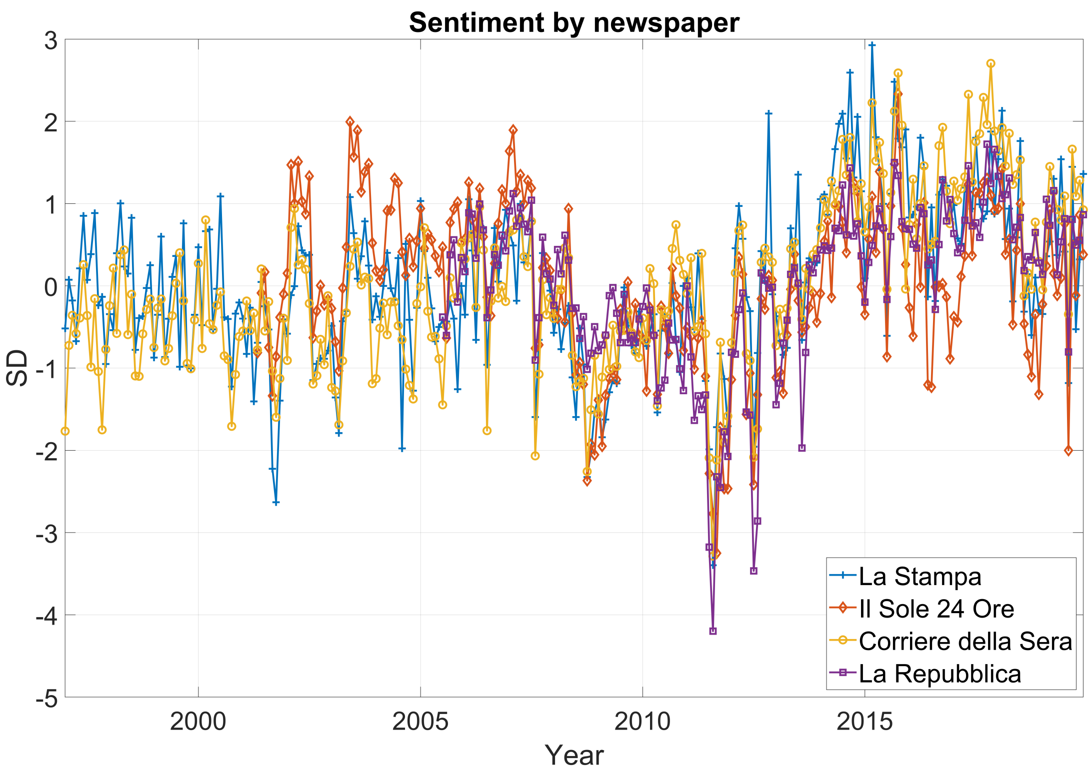


Figure A.9: TESI and TEPU for Italy - overall indexes

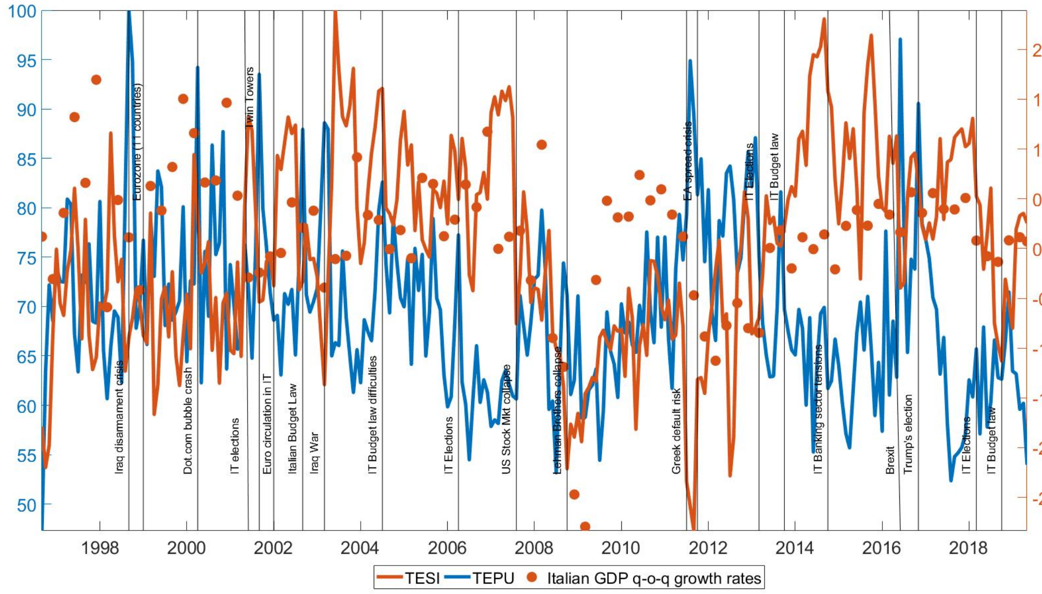
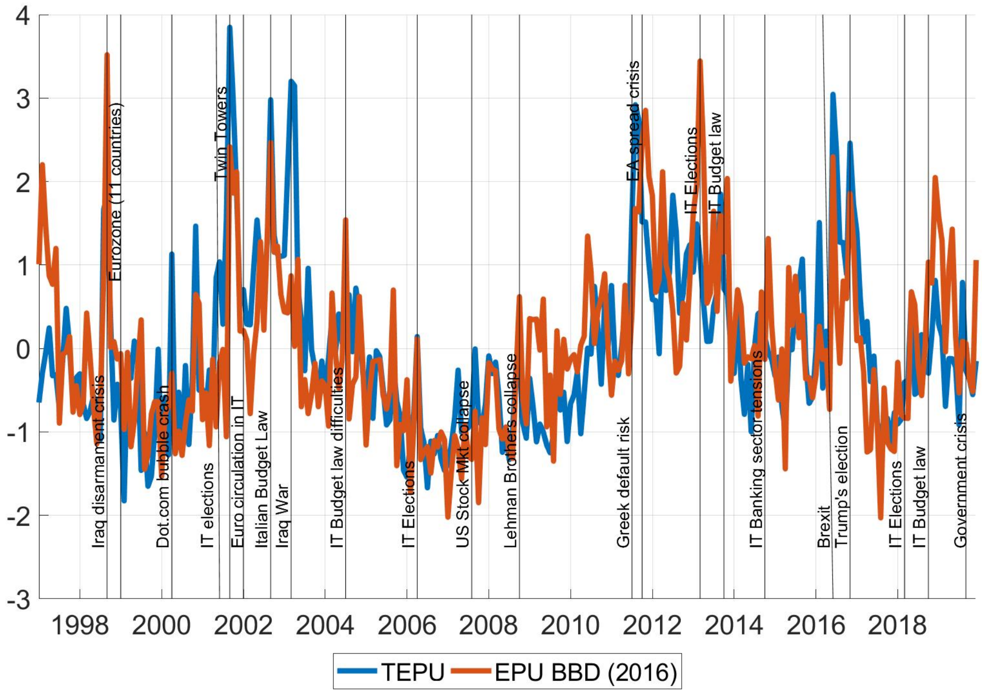
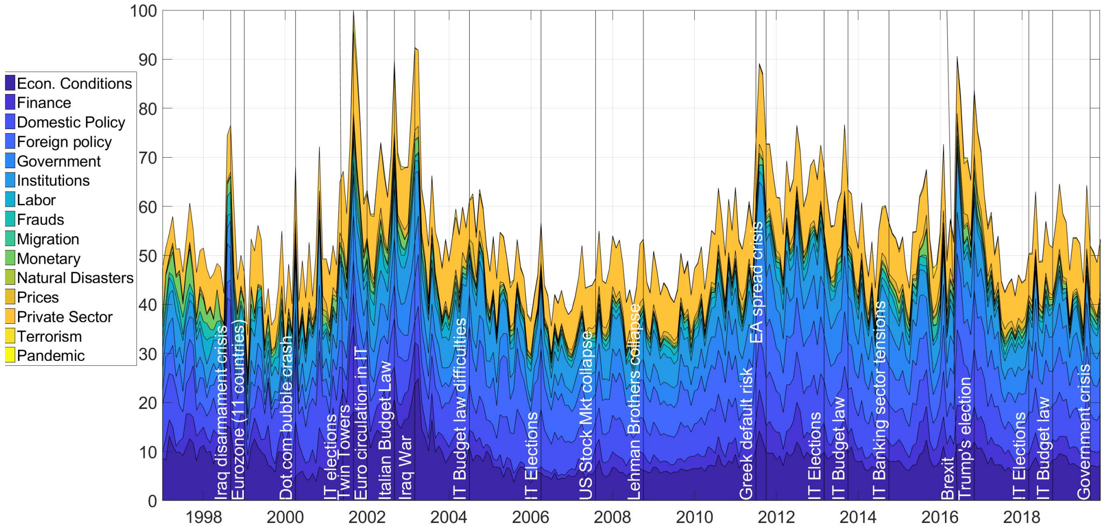


Figure A.10: Comparison between our TEPU and the the Baker et al. (2016) EPU index for Italy.

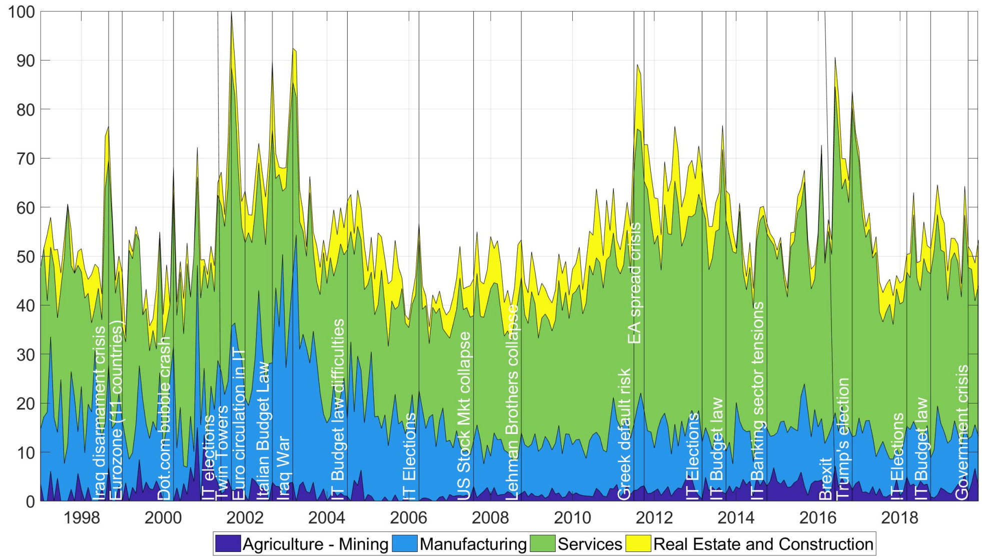


Notes: The EPU by Baker et al. (2016) correlates 0.76 with TEPU measure for Italy and 0.83 with the TEPU measure for economic conditions. TEPU and TEPU for economic conditions have a correlation of 0.92

Figure A.11: TEPU by topic and by sector for Italy (monthly shares of articles)

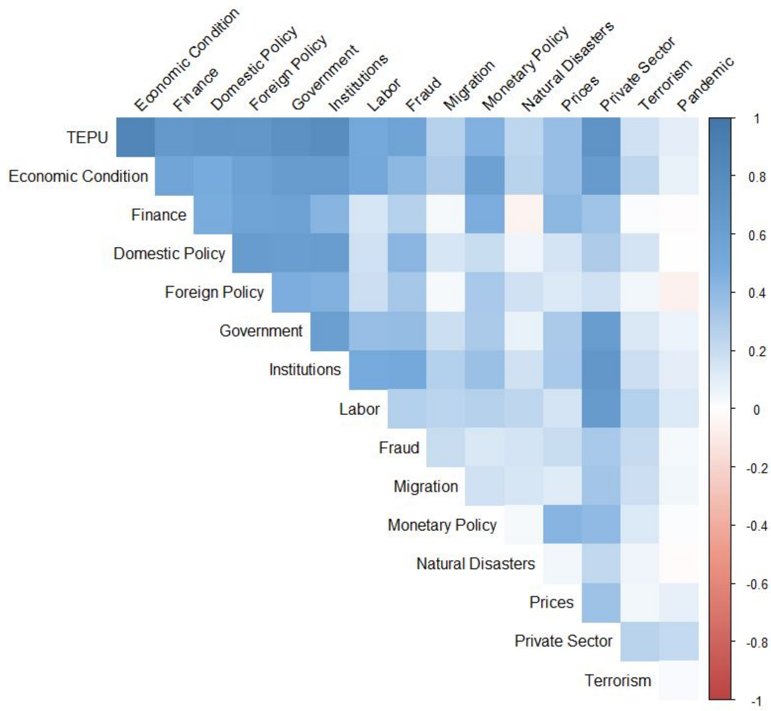


(a) TEPU by topic

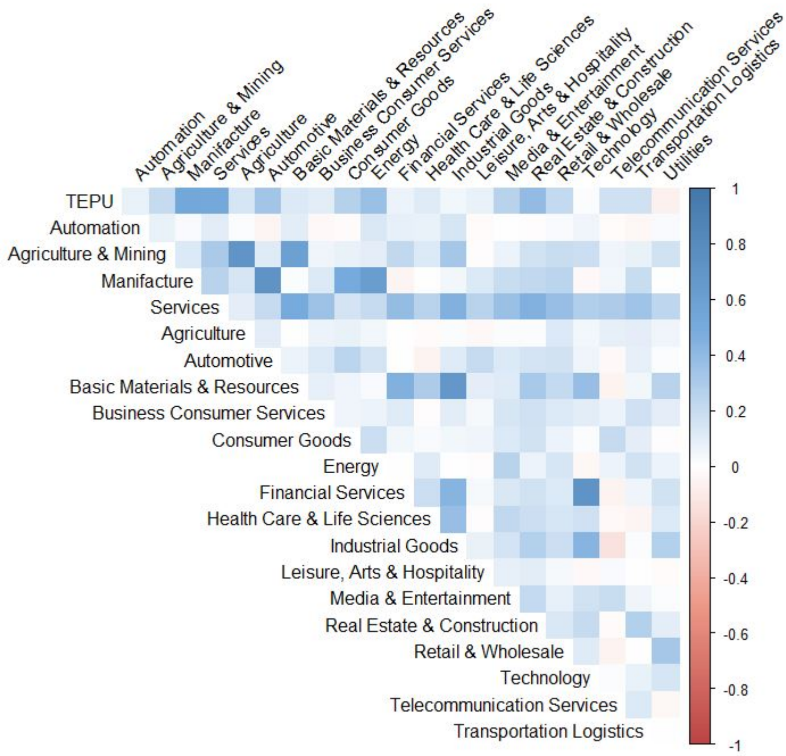


(b) TEPU by sector

Figure A.12: Correlation between Textual EPU (TEPU) indexes for Italy

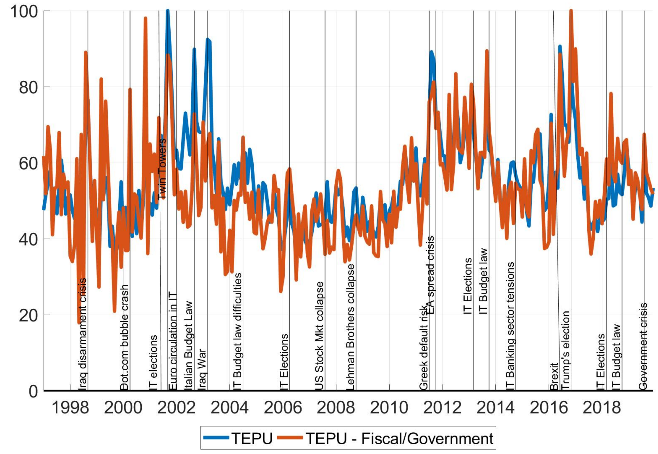


(a) Correlation between TEPU (overall) and TEPU calculated by topic

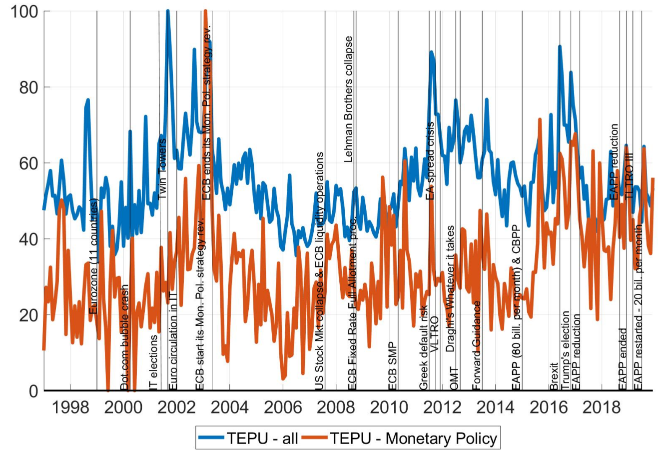


(b) Correlation between TEPU (overall) and TEPU calculated by sector

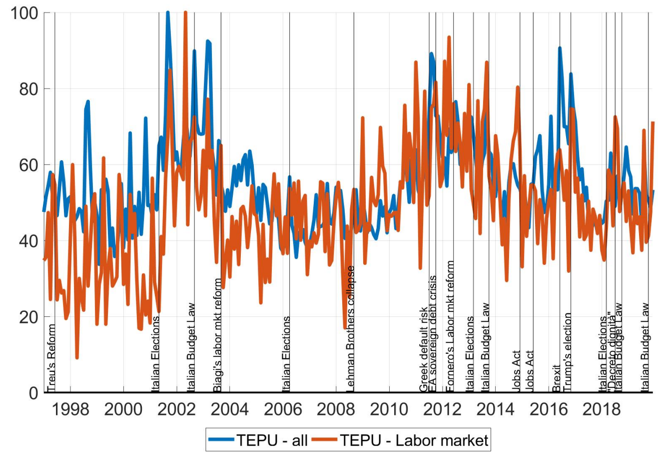
Figure A.13: TEPU indexes for three major macro topics: 1) Government; 2) Monetary Policy ; 3) Labor Markets



(a) TEPU on Fiscal/Government topic compared with the overall TEPU index

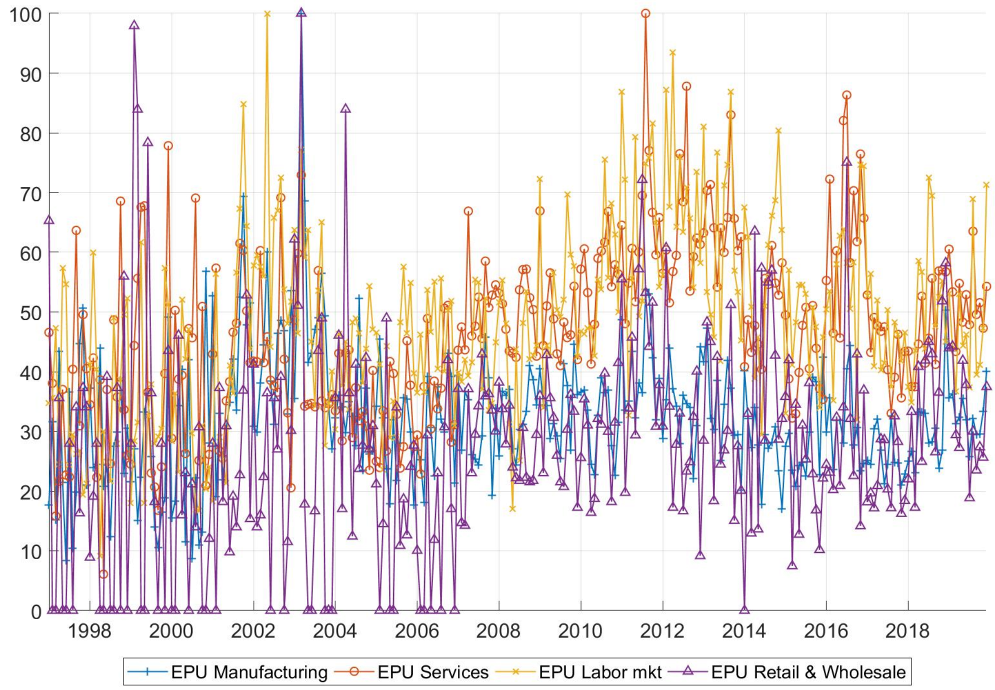


(b) TEPU on Monetary Policy compared with the overall TEPU index



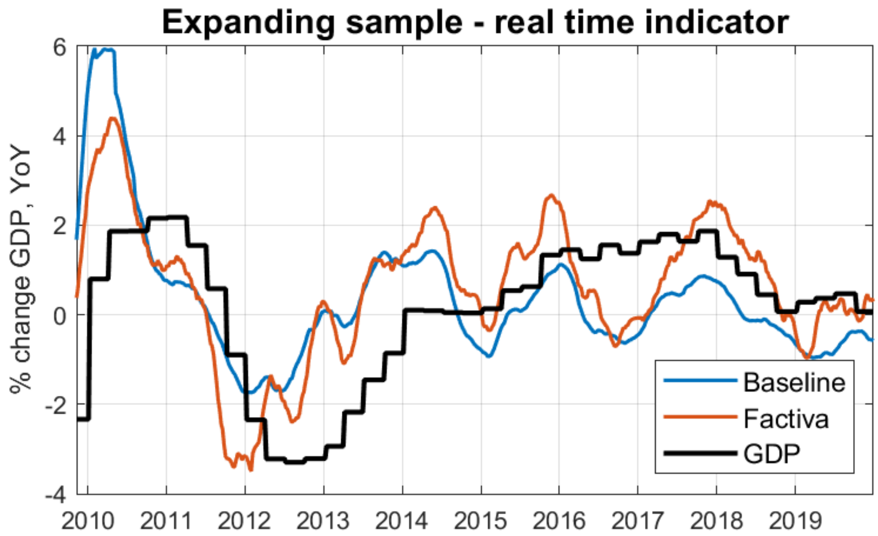
(c) TEPU on Labor markets topic compared with the overall TEPU index

Figure A.14: EPU indexes by sector

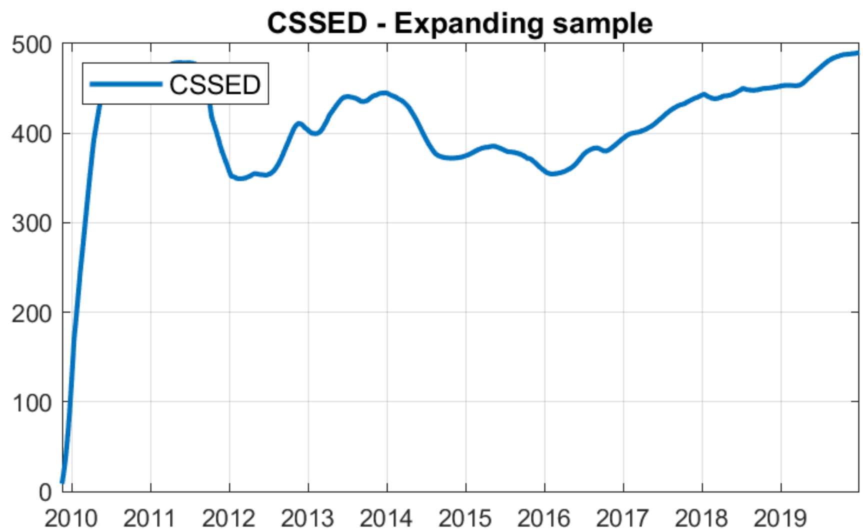


Notes: the series are standardized to have zero mean and unit variance.

Figure A.15: Expanding sample - Real-time weekly indicators of the GDP and Cumulative Sum of Squared Errors Difference[a]



(a) Italian weekly real-time index of economic activity calculated using an expanding sample



(b) CSSED on the nowcasting errors comparing the models with and without our text based indexes on the expanding sample.

[a]The CSSED (Cumulative Sum of Squared Errors Difference) is calculated as $CSSED_\tau = \sum_{\tau=R}^{T}(e^2_{\tau,(\text{baseline})} - e^2_{\tau,(\text{factiva})})$

RECENTLY PUBLISHED “TEMI” (*)

N. 1301 – *A quantitative analysis of distortions in managerial forecasts*, by Yueran Ma, Tiziano Ropele, David Sraer and David Thesmar (November 2020).

N. 1302 – *Bargaining power and the Phillips curve: a micro-macro analysis*, by Marco J. Lombardi, Marianna Riggi and Eliana Viviano (November 2020).

N. 1303 – *The effects of structural reforms: Evidence from Italy*, by Emanuela Ciapanna, Sauro Mocetti and Alessandro Notarpietro (November 2020).

N. 1304 – *Consumption and Wealth: New Evidence from Italy*, by Riccardo De Bonis, Danilo Liberati, John Muellbauer and Concetta Rondinelli (November 2020).

N. 1305 – *Do details matter? An analysis of the Italian personal income tax*, by Martino Tasso (November 2020).

N. 1306 – *Effects of eligibility for central bank purchases on corporate bond spreads*, by Taneli Mäkinen, Fan Li, Andrea Mercatanti and Andrea Silvestrini (November 2020).

N. 1307 – *On the design of grant assignment rules*, by Francesca Modena, Santiago Pereda Fernandez and Giulia Martina Tanzi (December 2020).

N. 1308 – *Monetary policy strategies in the New Normal: a model-based analysis for the euro area*, by Fabio Busetti, Stefano Neri, Alessandro Notarpietro and Massimiliano Pisani (December 2020).

N. 1309 – *Rare disasters, the natural interest rate and monetary policy*, by Alessandro Cantelmo (December 2020).

N. 1310 – *The market stabilization role of central bank asset purchases: high-frequency evidence from the COVID-19 crisis*, by Marco Bernardini and Annalisa De Nicola (December 2020).

N. 1311 – *Equity tail risk in the treasury bond market*, by Dario Ruzzi and Mirco Rubin (December 2020).

N. 1312 – *Scars of youth non-employment and labour market conditions*, by Giulia Martina Tanzi (December 2020).

N. 1313 – *The COVID-19 shock and a fiscal-monetary policy mix in a monetary union*, by Anna Bartocci, Alessandro Notarpietro and Massimiliano Pisani (December 2020).

N. 1314 – *Working horizon and labour supply: the effect of raising the full retirement age on middle-aged individuals*, by Francesca Carta and Marta De Philippis (February 2021).

N. 1315 – *Bank credit and market-based finance for corporations: the effects of minibond issuances*, by Steven Ongena, Sara Pinoli, Paola Rossi and Alessandro Scopelliti (February 2021).

N. 1316 – *Is inflation targeting a strategy past its sell-by date?*, by Alberto Locarno and Alessandra Locarno (February 2021).

N. 1317 – *Declining natural interest rate in the US: the pension system matters*, by Jacopo Bonchi and Giacomo Caracciolo (February 2021).

N. 1318 – *Can we measure inflation expectations using Twitter?*, by Cristina Angelico, Juri Marcucci, Marcello Miccoli and Filippo Quarta (February 2021).

N. 1319 – *Identifying deposits’ outflows in real-time*, by Edoardo Rainone (February 2021).

N. 1320 – *Whatever it takes to save the planet? Central banks and unconventional green policy*, by Alessandro Ferrari and Valerio Nispi Landi (February 2021).

(*) Requests for copies should be sent to:
Banca d’Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2019

ALBANESE G., M. CIOFFI and P. TOMMASINO, *Legislators' behaviour and electoral rules: evidence from an Italian reform,* European Journal of Political Economy, v. 59, pp. 423-444, **WP 1135 (September 2017).**

APRIGLIANO V., G. ARDIZZI and L. MONTEFORTE, *Using the payment system data to forecast the economic activity,* International Journal of Central Banking, v. 15, 4, pp. 55-80, **WP 1098 (February 2017).**

ARNAUDO D., G. MICUCCI, M. RIGON and P. ROSSI, *Should I stay or should I go? Firms' mobility across banks in the aftermath of the financial crisis,* Italian Economic Journal / Rivista italiana degli economisti, v. 5, 1, pp. 17-37, **WP 1086 (October 2016).**

BASSO G., F. D'AMURI and G. PERI, *Immigrants, labor market dynamics and adjustment to shocks in the euro area,* IMF Economic Review, v. 67, 3, pp. 528-572, **WP 1195 (November 2018).**

BATINI N., G. MELINA and S. VILLA, *Fiscal buffers, private debt, and recession: the good, the bad and the ugly,* Journal of Macroeconomics, v. 62, **WP 1186 (July 2018).**

BURLON L., A. NOTARPIETRO and M. PISANI, *Macroeconomic effects of an open-ended asset purchase programme,* Journal of Policy Modeling, v. 41, 6, pp. 1144-1159, **WP 1185 (July 2018).**

BUSETTI F. and M. CAIVANO, *Low frequency drivers of the real interest rate: empirical evidence for advanced economies,* International Finance, v. 22, 2, pp. 171-185, **WP 1132 (September 2017).**

CAPPELLETTI G., G. GUAZZAROTTI and P. TOMMASINO, *Tax deferral and mutual fund inflows: evidence from a quasi-natural experiment,* Fiscal Studies, v. 40, 2, pp. 211-237, **WP 938 (November 2013).**

CARDANI R., A. PACCAGNINI and S. VILLA, *Forecasting with instabilities: an application to DSGE models with financial frictions,* Journal of Macroeconomics, v. 61, **WP 1234 (September 2019).**

CHIADES P., L. GRECO, V. MENGOTTO, L. MORETTI and P. VALBONESI, *Fiscal consolidation by intergovernmental transfers cuts? The unpleasant effect on expenditure arrears,* Economic Modelling, v. 77, pp. 266-275, **WP 1076 (July 2016).**

CIANI E., F. DAVID and G. DE BLASIO, *Local responses to labor demand shocks: a re-assessment of the case of Italy,* Regional Science and Urban Economics, v. 75, pp. 1-21, **WP 1112 (April 2017).**

CIANI E. and P. FISHER, *Dif-in-dif estimators of multiplicative treatment effects,* Journal of Econometric Methods, v. 8. 1, pp. 1-10, **WP 985 (November 2014).**

CIAPANNA E. and M. TABOGA, *Bayesian analysis of coefficient instability in dynamic regressions,* Econometrics, MDPI, Open Access Journal, v. 7, 3, pp.1-32, **WP 836 (November 2011).**

COLETTA M., R. DE BONIS and S. PIERMATTEI, *Household debt in OECD countries: the role of supply-side and demand-side factors,* Social Indicators Research, v. 143, 3, pp. 1185–1217, **WP 989 (November 2014).**

COVA P., P. PAGANO and M. PISANI, *Domestic and international effects of the Eurosystem Expanded Asset Purchase Programme,* IMF Economic Review, v. 67, 2, pp. 315-348, **WP 1036 (October 2015).**

ERCOLANI V. and J. VALLE E AZEVEDO, *How can the government spending multiplier be small at the zero lower bound?,* Macroeconomic Dynamics, v. 23, 8. pp. 3457-2482, **WP 1174 (April 2018).**

FERRERO G., M. GROSS and S. NERI, *On secular stagnation and low interest rates: demography matters,* International Finance, v. 22, 3, pp. 262-278, **WP 1137 (September 2017).**

FOA G., L. GAMBACORTA, L. GUISO and P. E. MISTRULLI, *The supply side of household finance,* Review of Financial Studies, v.32, 10, pp. 3762-3798, **WP 1044 (November 2015).**

GERALI A. and S. NERI, *Natural rates across the Atlantic,* Journal of Macroeconomics, v. 62, article 103019, **WP 1140 (September 2017).**

GIORDANO C., M. MARINUCCI and A. SILVESTRINI, *The macro determinants of firms' and households' investment: evidence from Italy,* Economic Modelling, v. 78, pp. 118-133, **WP 1167 (March 2018).**

GOMELLINI M., D. PELLEGRINO and F. GIFFONI, *Human capital and urban growth in Italy,1981-2001,* Review of Urban & Regional Development Studies, v. 31, 2, pp. 77-101, **WP 1127 (July 2017).**

LIBERATI D. and M. LOBERTO, *Taxation and housing markets with search frictions,* Journal of Housing Economics, v. 46, article 101632, **WP 1105 (March 2017).**

MAGRI S., *Are lenders using risk-based pricing in the Italian consumer loan market? The effect of the 2008 crisis,* Journal of Credit Risk, v. 15, 1, pp. 27-65, **WP 1164 (January 2018).**

MERCATANTI A., T. MAKINEN and A. SILVESTRINI, *The role of financial factors for european corporate investment,* Journal of International Money and Finance, v. 96, pp. 246-258, **WP 1148 (October 2017).**

MIGLIETTA A., C. PICILLO and M. PIETRUNTI, *The impact of margin policies on the Italian repo market,* The North American Journal of Economics and Finance, v. 50, **WP 1028 (October 2015).**

MONTEFORTE L. and V. RAPONI, *Short-term forecasts of economic activity: are fortnightly factors useful?,* Journal of Forecasting, v. 38, 3, pp. 207-221, **WP 1177 (June 2018).**

NERI S. and A. NOTARPIETRO, *Collateral constraints, the zero lower bound, and the debt–deflation mechanism,* Economics Letters, v. 174, pp. 144-148, **WP 1040 (November 2015).**

PANCRAZI R. and M. PIETRUNTI, *Natural expectations and home equity extraction,* Journal of Housing Economics, v. 46, 4, **WP 984 (November 2014).**

PEREDA FERNANDEZ S., *Teachers and cheaters. Just an anagram?*, Journal of Human Capital, v. 13, 4, pp. 635-669, **WP 1047 (January 2016).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment,* Macroeconomic Dynamics, v. 23, 2, pp. 590-624, **WP 871 (July 2012).**

2020

ALESSANDRI P. and M. BOTTERO, *Bank lending in uncertain times,* R European Economic Review, V. 128, **WP 1109 (April 2017).**

ANTUNES A. and V. ERCOLANI, *Public debt expansions and the dynamics of the household borrowing constraint,* Review of Economic Dynamics, v. 37, pp. 1-32, **WP 1268 (March 2020).**

ARDUINI T., E. PATACCHINI and E. RAINONE, *Treatment effects with heterogeneous externalities,* Journal of Business & Economic Statistics, , v. 38, 4, pp. 826-838, **WP 974 (October 2014).**

BOTTERO M., F. MEZZANOTTI and S. LENZU, *Sovereign debt exposure and the Bank Lending Channel: impact on credit supply and the real economy,* Journal of International Economics, v. 126, article 103328, **WP 1032 (October 2015).**

BRIPI F., D. LOSCHIAVO and D. REVELLI, *Services trade and credit frictions: evidence with matched bank – firm data,* The World Economy, v. 43, 5, pp. 1216-1252, **WP 1110 (April 2017).**

BRONZINI R., G. CARAMELLINO and S. MAGRI, *Venture capitalists at work: a Diff-in-Diff approach at late-stages of the screening process,* Journal of Business Venturing, v. 35, 3, **WP 1131 (September 2017).**

BRONZINI R., S. MOCETTI and M. MONGARDINI, *The economic effects of big events: evidence from the Great Jubilee 2000 in Rome,* Journal of Regional Science, v. 60, 4, pp. 801-822, **WP 1208 (February 2019).**

COIBION O., Y. GORODNICHENKO and T. ROPELE, *Inflation expectations and firms' decisions: new causal evidence,* Quarterly Journal of Economics, v. 135, 1, pp. 165-219, **WP 1219 (April 2019).**

CORSELLO F. and V. NISPI LANDI, *Labor market and financial shocks: a time-varying analysis,* Journal of Money, Credit and Banking, v. 52, 4, pp. 777-801, **WP 1179 (June 2018).**

COVA P. and F. NATOLI, *The risk-taking channel of international financial flows,* Journal of International Money and Finance, v. 102, **WP 1152 (December 2017).**

D'ALESSIO G., *Measurement errors in survey data and the estimation of poverty and inequality indices,* Statistica Applicata - Italian Journal of Applied Statistics, v. 32, 3, **WP 1116 (June 2017).**

DEL PRETE S. and S. FEDERICO, *Do links between banks matter for bilateral trade? Evidence from financial crises,* Review of World Economic, v. 156, 4, pp. 859 - 885, **WP 1217 (April 2019).**

D'IGNAZIO A. and C. MENON, *The causal effect of credit Guarantees for SMEs: evidence from Italy,* The Scandinavian Journal of Economics, v. 122, 1, pp. 191-218, **WP 900 (February 2013).**

ERCOLANI V. and F. NATOLI, *Forecasting US recessions: the role of economic uncertainty,* Economics Letters, v. 193, **WP 1299 (October 2020).**

MAKINEN T., L. SARNO and G. ZINNA, *Risky bank guarantees,* Journal of Financial Economics, v. 136, 2, pp. 490-522, **WP 1232 (July 2019).**

MODENA F., E. RETTORE and G. M. TANZI, *The effect of grants on university dropout rates: evidence from the Italian case,* Journal of Human Capital, v. 14, 3, pp. 343-370, **WP 1193 (September 2018).**

NISPI LANDI V., *Capital controls spillovers,* Journal of International Money and Finance, v. 109, **WP 1184 (July 2018).**

PERICOLI M., *On risk factors of the stock–bond correlation,* International Finance, v. 23, 3, pp. 392-416, **WP 1198 (November 2018).**

RAINONE E., *The network nature of OTC interest rates,* Journal of Financial Markets, v.47, article 100525, **WP 1022 (July 2015).**

RAINONE E. and F. VACIRCA, *Estimating the money market microstructure with negative and zero interest rates,* Quantitative Finance, v. 20, 2, pp. 207-234, **WP 1059 (March 2016).**

RIZZICA L., *Raising aspirations and higher education. Evidence from the UK's widening participation policy,* Journal of Labor Economics, v. 38, 1, pp. 183-214, **WP 1188 (September 2018).**

SANTIONI, R., F. SCHIANTARELLI and P. STRAHAN, *Internal capital markets in times of crisis: the benefit of group affiliation*, Review of Finance, v. 24, 4, pp. 773-811, **WP 1146 (October 2017).**

SCHIANTARELLI F., M. STACCHINI and P. STRAHAN, *Bank Quality, judicial efficiency and loan repayment delays in Italy,* Journal of Finance , v. 75, 4, pp. 2139-2178, **WP 1072 (July 2016).**

FORTHCOMING

ACCETTURO A., A. LAMORGESE, S. MOCETTI and D. PELLEGRINO, *Housing Price elasticity and growth: evidence from Italian cities,* Journal of Economic Geography, **WP 1267 (March 2020).**

ALBANESE G., E. CIANI and G. DE BLASIO, *Anything new in town? The local effects of urban regeneration policies in Italy,* Regional Science and Urban Economics, **WP 1214 (April 2019).**

ALBANESE G., G. DE BLASIO and A. LOCATELLI, *Does EU regional policy promote local TFP growth? Evidence from the Italian Mezzogiorno,* Papers in Regional Science, **WP 1253 (December 2019).**

ANZUINI A. and L. ROSSI, *Fiscal policy in the US: a new measure of uncertainty and its effects on the American economy*, Empirical Economics, **WP 1197 (November 2018).**

BOLOGNA P., A. MIGLIETTA and A. SEGURA, *Contagion in the CoCos market? A case study of two stress events,* International Journal of Central Banking, **WP 1201 (November 2018).**

COVA P., P. PAGANO, A. NOTARPIETRO and M. PISANI, *Secular stagnation, R&D, public investment and monetary policy: a global-model perspective,* Macroeconomic Dynamics, **WP 1156 (December 2017).**

DE PHILIPPIS M., *Multitask agents and incentives: the case of teaching and research for university professors,* Economic Journal, **WP 1156 (December 2015).**

DEL PRETE S. and M. L. STEFANI, *Women as "Gold Dust": gender diversity in top boards and the performance of Italian banks,* Economic Notes, Monte dei Paschi di Siena, **WP 1014 (June 2015).**

FIDORA M., C. GIORDANO and M. SCHMITZ, *Real exchange rate misalignments in the Euro Area,* Open Economies Review, **WP 1042 (January 2018).**

HERTWECK M., V. LEWIS and S. VILLA, *Going the extra mile: effort by workers and job-seekers,* Journal of Money, Credit and Banking, **WP 1277 (June 2020).**

LI F., A. MERCATANTI, T. MAKINEN and A. SILVESTRINI, *A regression discontinuity design for ordinal running variables: evaluating central bank purchases of corporate bonds,* Annals of Applied Statistics, **WP 1213 (March 2019).**

LOSCHIAVO D., *Household debt and income inequality: evidence from Italian survey data,* Review of Income and Wealth, **WP 1095 (January 2017).**

METELLI L. and F. NATOLI, *The international transmission of US tax shocks: a proxy-SVAR approach,* IMF Economic Review, **WP 1223 (June 2019).**

MOCETTI S., G. ROMA and E. RUBOLINO, *Knocking on parents' doors: regulation and intergenerational mobility,* Journal of Human Resources, **WP 1182 (July 2018).**

NISPI LANDI V. and A. SCHIAVONE, *The effectiveness of capital controls,* Open Economies Review, **WP 1200 (November 2018).**

PEREDA FERNANDEZ S., *Copula-based random effects models for clustered data,* Journal of Business & Economic Statistics, **WP 1092 (January 2017).**

PERICOLI M. and M. TABOGA, *Nearly exact Bayesian estimation of non-linear no-arbitrage term-structure models,* Journal of Financial Econometrics, **WP 1189 (September 2018).**