

# Using GDELT Data to Evaluate the Confidence on the Spanish Government Energy Policy

Diego J. Bodas-Sagi, José M. Labeaga

*Universidad Francisco de Vitoria, Universidad Nacional de Educación a Distancia (UNED), Spain*

**Abstract** — The growing demand for affordable, reliable, domestically sourced, and low-carbon electricity is a matter of concern and it is driven by several causes including public policy priorities. Policy objectives and new technologies are changing wholesale market design. The analysis of different aspects of energy markets is increasingly on the agendas of academics, firms' managers or policy makers. Some concerns are global and are related to the evolution of climate change phenomena. Others are regional or national and they strongly appear in countries like Spain with a high dependence on foreign energy sources and high potential of domestic renewable energy sources. We can find a relevant case in Spanish solar energy policy. A series of regulatory reforms since 2010 reduce revenues to existing renewable power generators and they end up the previous system of support to new renewable generation. This policy change has altered the composition of the energy market affecting investment decisions. In this paper, we analyze the public opinion about energy policy of the Spanish Government using the Global Database of Events, Language, and Tone (GDELT). The GDELT Project consists of over a quarter-billion event records in over 300 categories covering the entire world from 1979 to present, along with a massive network diagram connecting every person, organization, location, and theme to this event database. Our aim is to build sentiment indicators arising from this source of information and, in a final step, evaluate if positive and negative indexes have any effect on the evolution of key market variables as prices and demand.

**Keywords** — Big Query, GDELT, Public Opinion, Energy, Electricity.

---

## I. INTRODUCTION

---

Public policy plays a critical role in regulating relationship between companies, investors and society. The importance of public policy for long-term investors has grown in recent years, due to [1]:

- Legislative reform of the financial sector in the wake of the global financial crisis.
- Governmental need for investors as a source of long-term growth.
- The increasing impact of environmental, social and governance factors on the ability of investors to deliver long-term returns.

One of the key case for studying the effects of public policy is the energy market. The growing demand for affordable, reliable, domestically sourced, and low-carbon electricity is on the rise. It “is driven in part by evolving public policy priorities, especially reducing the health and environmental impacts of electricity service... Well-designed markets encourage economically efficient solutions, promote innovation and minimize unintended consequences” [2].

Policy objectives and new technologies are changing wholesale

market design. A relevant case is the Spanish solar energy policy. A series of regulatory reforms since 2010 reduce revenues to existing renewable power generators and end the previous system of support to new renewable generation. This policy change has caused several claims by various organizations and altered the composition of the energy market. At the end, the Royal Decree of October 2015 strongly affected the solar energy market.

The analysis of the public opinion about specific government measures may be a useful component for some decisions of the agents on the long-term. The public opinion can play a role on influencing policy makers at several stages of their decision process and this public perception could affect the design of political programs or policy measures. Agents in any market can also be influenced by the positive and negative perceptions about the company. In any case, public representatives can be interested in analyzing the state of the public opinion when adopting measures which can distort the markets.

In this paper, we analyze the public opinion about energy policy of the Spanish Government using the Global Database of Events, Language, and Tone (GDELT). The GDELT Project [3] consists of over a quarter-billion event records in over 300 categories covering the entire world from 1979 to present. Our aim is to build sentiment indicators arising from this source of information and, in a final step, evaluate if positive and negative indexes have any effect on the evolution of key market variables as prices and demand. We do not try to evaluate whether there are causal effects from the sentiment indicators to the market variables but our purpose is to detect the existence of correlation among those variables.

The rest of the paper is structured as follows: First, we provide a brief summary of the GDELT Project and explain the relation with the Big Data paradigm. In section 3, we explain the tools, methods and techniques used in this study. The results obtained are discussed in section 4. The paper ends up providing some policy implications and ideas for future research.

---

## II. THE GDELT PROJECT

---

The GDELT Project, supported by Google Ideas, share real-time information and metadata with the world. This codified metadata (but not the text of the articles) is then released as an open data stream, updated every 15 minutes, providing a multilingual annotated index of the information. It includes broadcast, print, and online news sources. The project shares a database with trillions data points. Although, data is available as downloadable CSV files, few users have the storing capacity and processing power to download terabytes of data, and effectively query and analyze it. Google's BigQuery platform provides a way to interact with this huge information source. GDELT is a clear example of Big Data, while Google's BigQuery is an example of Infrastructure As a Service (IaaS) technology.

According to [4], Big Data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves

too fast, or doesn't fit the structures of our database architectures. To gain value from this data, one must choose an alternative way to process it. Big Data technologies have huge variety of sources, huge volume of information – so much less time is needed to process information thanks to parallel processing and clustering infrastructure.

GDELT maintains the GDELT Event Database, and the GDELT Global Knowledge Graph (GKG). The GKG begins April 1, 2013 and "... attempts to connect every person, organization, location, count, theme, news source, and event across the planet into a single massive network that captures what's happening around the world, what its context is and who's involved, and how the world is feeling about it, every single day" [3]. The data files use Conflict and Mediation Event Observations (CAMEO) [5] coding for recording events. GKG also provides event identification (*EventIDs*) of each event found in the same article as the extracted information, allowing rich contextualization of events.

---

### III. METHODOLOGY

---

In this work, we have used GKG table on Google's BigQuery platform. GKG table provides the "Themes" attribute, the list of all themes found in the document. We want to filter documents related to, at least, one of these two themes: "ENV\_SOLAR" (which refers to solar power in general), and "FUELPRICES" (which refers to cost of fuel, energy and heating). The theme attribute is not available for the Event table. At the same time, we have looked for events that refer to Spain at some point using the "Locations" attribute (which contains a list of all locations found in the text). In summary, we are using GKG table to filter information about solar power or cost of fuel, energy and heating and related (in some way) with Spain. Attribute "V2Tone" allows us to analyze the average "tone" of the document as a whole. The score ranges from -100 (extremely negative) to +100 (extremely positive). Common values range between -10 and +10, with 0 indicating neutral. This is calculated as Positive Score minus Negative Score. Positive Score is the percentage of all words in the article that were found to have a positive emotional connotation. Negative Score is the percentage of all words in the article that were found to have a negative emotional connotation. Big Query allows interaction with the whole GDELT dataset using Structure Query Language (SQL). An account in Google Cloud Services and activate Google's Cloud Storage to export and download data is required.

R [6] has been used to analyze and process data. Downloaded data from Google's Cloud Storage have been imported into R. After that, we have done a basic and exploratory analysis of the downloaded data. The analysis shows that there are some references, documents or URL's that are not related to energy policy. For example, some entries refer to scientific news from Canary Institute of Astrophysics ("ENV\_SOLAR" theme). For this reason and for a more efficient measurement, we feel that it is necessary to analyze the text of the news and look for references to Spanish Government, council, ministry or ministers. This is a computation intensive task because it requires to deal with HTML tags and extract the text of the document. In the next section, we detail this process and explain different alternatives to improve execution time.

Next and for each theme, we have grouped by day all mentions and calculated the mean tone and typical deviation in tone per day. At this stage, we only have evaluated documents written in Spanish or English because we need to find mentions to Spanish Government (Spanish and English have been the chosen languages to process, other languages will be included in future versions). The results have been placed into context with the policies that the government of Spain has implemented.

Finally, we have applied a Correlated Topics Models (CTM) algorithm [7], based on Latent Dirichlet Allocation (LDA) algorithm [8], on the words contained in the mentions, news or documents written in Spanish. The rest of the dataset has not been included in this part of the study. We have not mixed languages, words with similar meaning but from different languages can be placed on different topics because writing is different. We leave the evaluation of other languages for future research. The "topicmodels" R package [9] allows us to execute LDA and CTM algorithms.

LDA allows to discover topics in large data collections described via topics. It does not require labeled data (unsupervised learning) and uses a stochastic procedure to generate the topic weight vector. LDA represents documents as mixtures of topics that spit out words with certain probabilities. It is a bag-of-words model. For this reason, LDA can be used for document modelling and classification. LDA fails to directly model correlation between the occurrence of topic and, sometimes, the presence of one topic may be correlated with the presence of another (for example: "economic" and "business"). CTM is very similar to LDA except that topic proportions are drawn from a logistic normal distribution rather than a Dirichlet distribution. Applying the "topicmodels" R package to our dataset, we can obtain a list of words for every topic and, also, check the correlation between the topics obtained.

#### A. Getting the text of the documents

As we have mentioned before, getting the text of the documents is a compute-intensive phase because it requires to deal with HTML tags and extract the text of the document. This task is done in the following steps:

1. First, documents are accessed through its URL.
2. We detect the text language using "textcat" R package [10]. If the document is written in Spanish or English we download the text and confirm that the text contains one of the following words: "gobierno", "government", "council", "ministers", "ministry", "ministro", "ministerio". In other case, we consider that the text does not mention Spanish Government. One should note that Spain location is referenced in the text according to GKG metadata.
3. For all downloaded texts, we clean the text removing HTML tags and stop-words (Spanish and English) in order to improve accuracy and performance. This task reduces the text size. Stop-words refer to the most common words in a language but given our aims they do not add value to the analysis of the topic.

Sequential and parallel execution modes have been tested here. A parallel algorithm, as opposed to a traditional sequential algorithm, is one which can be executed a piece at a time on many different processing  $p$  devices or processors, and then put back together again at the end to get the correct result.

The usefulness of this type of parallelization is that once the program structure is known, a few changes must be made to the program to be executed by several processors, and not as a distributed algorithm in which, first, we must establish the optimal communication structure. This usually involves making substantial changes to the program.

Two parallelization schemes have been evaluated. In the first one, we take advantage of multiple cores in one computer. The second parallelization method employs a master-slave architecture [11]. This architecture features a single processor running the main algorithm (master) which delegates the mission of getting the text among a group of processors (slaves). Slaves are responsible for processing URLs and getting the text and communicating results to the central process.

In any case, if we have  $p$  processors, the original dataset is divided into  $p$  chunks, one processor processes only one of these chunks. In all cases, we have used a computer with Pentium V quad core and 8 GB RAM, managed by Operating System Centos 6.4. The Internet broadband speed is, roughly, 20 Mbps (download speed). Performance are usually measured in terms Speedup ( $S_p$ ) and Efficiency ( $E_p$ ):

$$S_p = T_1 / T_p \quad (1)$$

$$E_p = S_p / p \quad (2)$$

where  $p$  is the number of processors,  $T_1$  is the execution time of sequential algorithm and  $T_p$  is the execution time of the parallel implementation on  $p$  processors.

The Simple Network of Workstations (snow) package [12] allows executing parallel code in R. It requires loading the code, loading the snow library, create a snow cluster (or execute in local mode using multicore CPU) and running the code, maintaining this order. Snow library can be used to start new R processes (workers) in our machine. The snow package is a scatter/gather paradigm, which works as follows:

1. The manager partitions the data into chunks and parcels them out to the workers (scatter phase).
2. The workers process their chunks.
3. The manager collects the results from the workers (gather phase) and combines them as appropriate to the application.

Snow can be used with socket connections, Message Passing Interface (MPI), Parallel Virtual Machine (PVM), or NetWorkSpaces [12, 13]. The socket transport does not require any additional packages, and is very portable. We have used socket connections. Snow is a non-shared-memory system example, if we are using a network of workstations, each workstation has its own and independent memory. But, in the multicore and one-computer case, the memory is shared between all the running processes. In both cases, the cost of communications should be kept in mind. The cost of communication is dependent on a variety of features including the programming model semantics, the network topology, data handling and routing, and associated software protocols. Reducing the computation time by adding more processors would only improve marginally the overall execution time as the communication costs remains fixed.

#### IV. RESULTS

##### A. Results using data from GDELT GKG

We are analyzing tone in mentions from GKG database. All mentions have some common characteristic, they refer to Spain as location at some point, themes “ENV\_SOLAR” or “FUELPRICES” are detected in the text and, the lines contain one of the following words: “gobierno”, “government”, “council”, “ministers”, “ministry”, “ministro”, “ministerio” (we interpret it as the text mentions the Spanish Government). GDELT 2.0 and GKG new version are relatively recent. For this reason, we only have data filling these requisites from February 18th, 2015 to October 28th, 2015. Since is a project in constant update, in order to close our study, we refer here to data obtained in our last interaction with Google’s BigQuery on October 28th, 2015.

The following figures show the mean and typical deviation in mentions (tone). All mentions (mentions without filtering words nor languages) and only government mentions are displayed and compared. Blue bars represent mean values in tone, black ones represent error bars. According to Fig. 1 and its histogram in Fig. 3, the average index of the mentions due to fuel and energy prices are

negative indicating that the sentiment of news related to the solar energy policies is negative through this period. We must remember that the government introduced in October 2015 what was named as solar tax (“*impuesto al sol*”) regulating consumption made by consumers who produce their own energy through photovoltaic systems. The discussions at the media did not began at the time of publishing the Royal Decree on October the 9th, 2015 but several months before as soon as the agents knew government’s intention. It is not strange that the sentiment of the agents producing news is negative.

On the other hand, when we include the word government as a control to build the sentiment index, the average as presented in figures 2 and 4 is still more negative. So, the agents (producers, consumers, etc.) clearly express a negative reaction towards the fuel and energy prices and we associate it to the regulations in the energy market referred to these variables. The opinion expressed in other surrounding countries of Europe and also by the authorities of the UE was also negative towards the regulation.

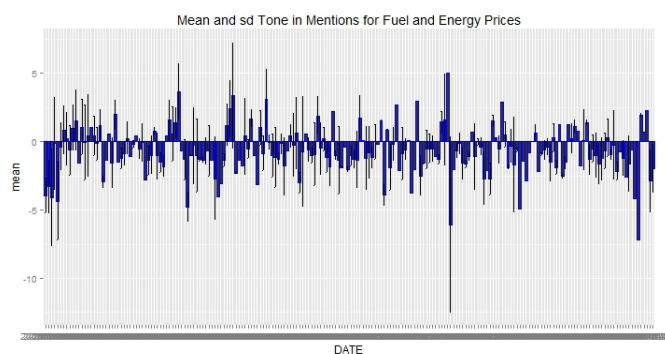


Fig. 1. All mentions for “FUELPRICES” theme in Spain.

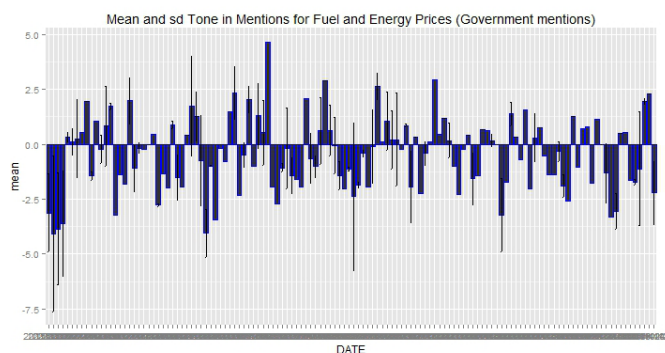


Fig. 2. Mentions for “FUELPRICES” theme in Spain filtering words (government mentions).

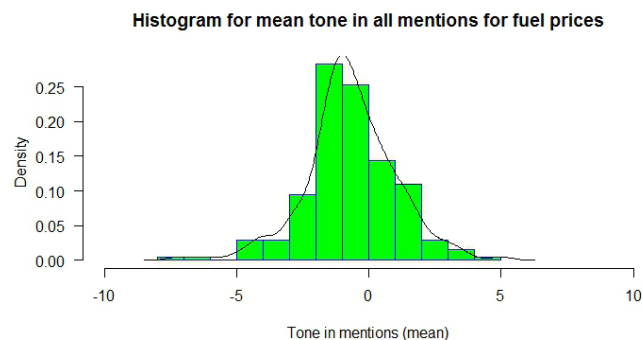


Fig. 3. Histogram - All mentions for “FUELPRICES” theme in Spain.

Histogram for mean tone in government mentions for fuel prices

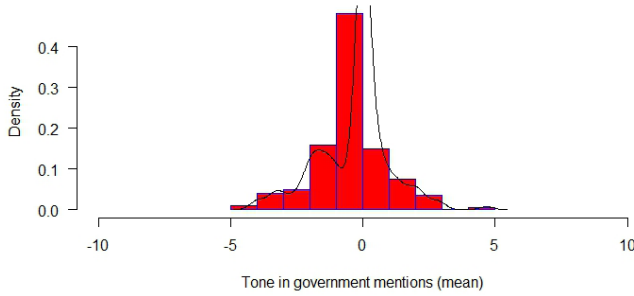


Fig. 4. Histogram - Mentions for “FUELPRICES” theme in Spain filtering words (government mentions).

In order to be able to use these data and conduct some test on them, we first check whether the indexes are normally distributed. figures 5 and 6 present Q-Q plots, which are probability plots, i.e., a graphical method for comparing two probability distributions by plotting their quantiles against each other. Here, we use Q-Q plot to compare data against Normal Distribution with mean and standard deviation according to the sample. Formally, the Shapiro-Wilk test [14] allows us to reject normality. For all data samples mean values are near to zero while typical deviation values are between 0.7 and 1.7. A normal distribution is symmetric about its mean, but this is not the case, and, taking into account the figures, we detect some extreme positive values which are balancing out a more frequent negative values and, for this reason, the mean tone is close to zero.

Normal Q-Q Plot

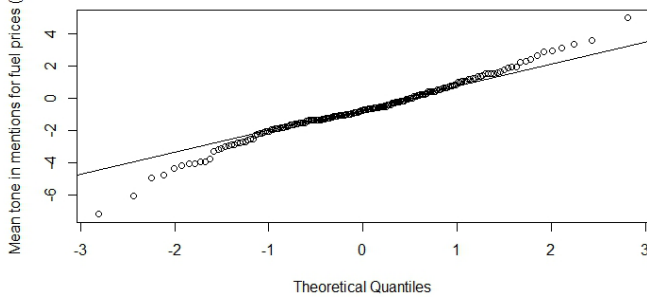


Fig. 5. Q-Q Plot - All mentions for “FUELPRICES” theme in Spain.

Normal Q-Q Plot

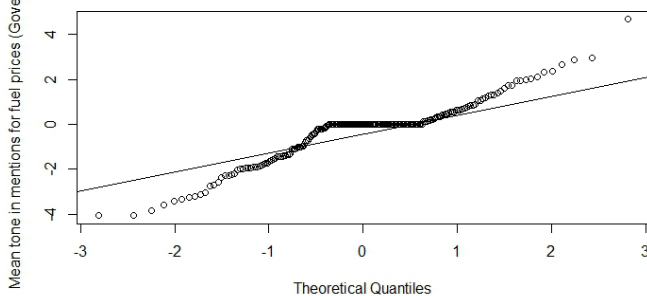


Fig. 6. Q-Q Plot - Mentions for “FUELPRICES” theme in Spain filtering words.

Next, we conduct a similar exercise but filtering in GDELT a different theme than before. So, we include environment and solar (“ENV\_SOLAR” theme) to the previous exercises and analyze tone in the same way. We can see that the sentiment index does provide some negative tone messages when we do not filter using words related

to the government. However, once we filter for words related to the government, the negative tone appears much more clearly.

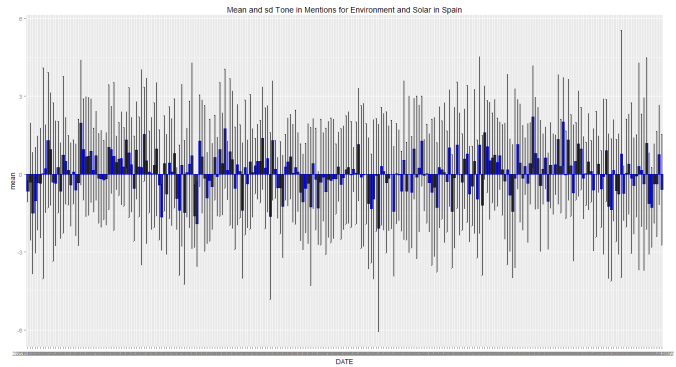


Fig. 7. All mentions for “ENV\_SOLAR” theme in Spain.

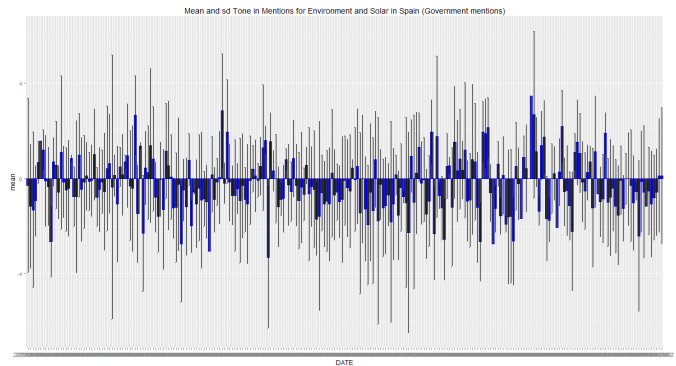


Fig. 8. Mentions for “ENV\_SOLAR” theme in Spain filtering words (government mentions).

Histogram for mean tone in all mentions for Environment and Solar

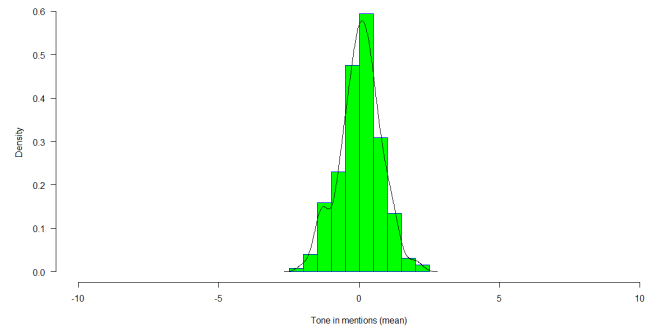


Fig. 9. Histogram – All mentions for “ENV\_SOLAR” theme in Spain.

Histogram for mean tone in government mentions for Environment and Solar

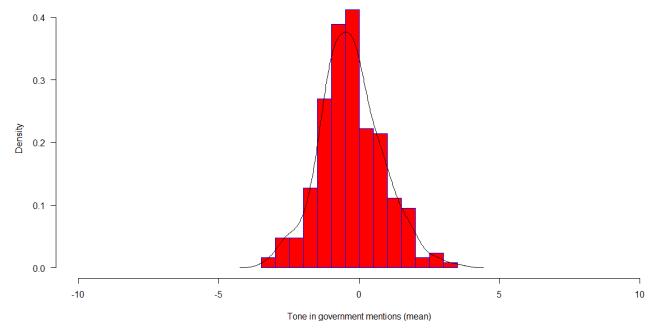


Fig. 10. Histogram – Mentions for “ENV\_SOLAR” theme in Spain filtering words (government mentions).



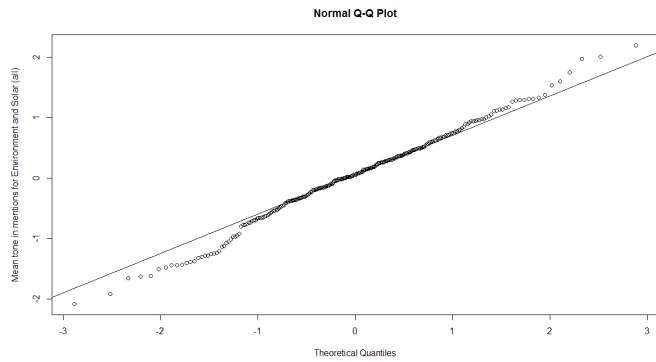


Fig. 11. Q-Q Plot – All mentions for “ENV\_SOLAR” theme in Spain.

Despite the graphic, Shapiro-Wilk normality test produces no suspicion about normality. But the mean tone, although is close to zero, is negative.

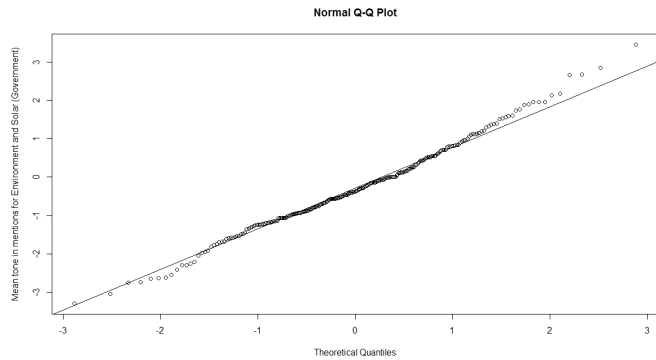


Fig. 12. Q-Q Plot – Mentions for “ENV\_SOLAR” theme in Spain filtering words.

**B. Correlation analysis: Prices and Demand**

We can obtain historical data about electricity prices and demand from OMIE [15]. OMIE manages the electrical market for Spain and Portugal (MIBEL Market). We have used data from February 18th, 2015 to October 28th, 2015 (similar to GDELT data). Our aim is to evaluate whether there is any type of correlation between prices or demand in the MIBEL market and the mean tone of public opinion evaluated thanks to GKG data. For prices and demand, we have taken natural logs. Variables in Fig. 13 must be interpreted in the following way: *MeanALLSolar* (*ENV\_SOLAR* theme for Spain) does not include references to Spanish Government, *MeanGovSolar* does include it. Interpretation is similar for *MeanALLFuel* and *MeanGovFuel* (for *FUELPRICES* theme in this case). *LogPrices* and *LogDemand* refer to logarithm mean and daily values for prices and demand (respectively) from MIBEL historical data and the same period.

	MeanALLSolar	MeanGovSolar	MeanALLFuel	MeanGovFuel	LogPrices	LogEnergy
MeanALLSolar	1.0000	0.4132	0.0170	-0.0161	-0.0318	0.0047
MeanGovSolar	0.4132	1.0000	-0.0023	-0.0442	-0.0421	0.0302
MeanALLFuel	0.0170	-0.0023	1.0000	0.4228	-0.0708	-0.0446
MeanGovFuel	-0.0161	-0.0442	0.4228	1.0000	-0.0149	-0.0062
LogPrices	-0.0318	-0.0421	-0.0708	-0.0149	1.0000	0.4234
LogEnergy	0.0047	0.0302	-0.0446	-0.0062	0.4234	1.0000

Fig. 13. Correlations results.

There is some week evidence of correlation between *LogPrices* and mean tone collected by *MeanALLFuel*. A test of the null that the coefficient (-0.0708) is equal to zero gives a normal value of -1.65, which is significantly different from zero at 9.9 percent of significance. The negative coefficient of correlation between *LogPrices* and mean tone collected by *MeanGovSolar* has a p-value of 0.32 for testing the same assumption. Finally, the correlation between *LogPrices* and mean

tone collected by *MeanALLFuel* is not significantly different from zero at standard levels. Public opinion could to some extent weakly affect fuel prices and, indirectly, fuel demand in the short-term.

**C. CTM results using CTM algorithm**

In this subsection we present the results obtained using a CTM algorithm to discover and correlate topics. We must note that we only have applied the algorithm to Spanish texts. The R package “topicmodels” allow us to display the graphs collected in Fig. 14. For the theme named “FUELPRICES” and text written in Spanish, the cluster Group 1 corresponds to HTML tags or other words that have not been properly removed or English words that appear in texts that “textcat” R package has been classified as written in Spanish. On the other hand, clusters named Group 3 and 6 refer to words like (translated from Spanish) “stock exchange”, “market”, “government”, “income”, “Europe”, “state”, “congress”... “gas”, “price”, “growth” and other Spanish locations as “Madrid” or “Barcelona” are words contained in the rest of the groups.

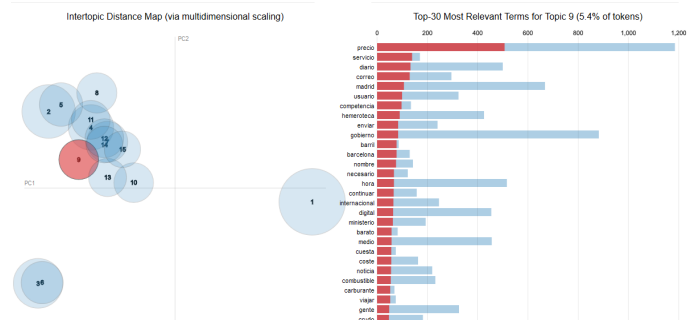


Fig. 14. Using “topicmodels” R package. “FUELPRICES” theme, text written in Spanish.

For “ENV\_SOLAR” and text written in Spanish, Group 1 is similar to the last case. Group 5, 12 and 15 refer to words like “solar”, “system”, “electricity”, “law”, “change”, “tax”, and months (written in Spanish).

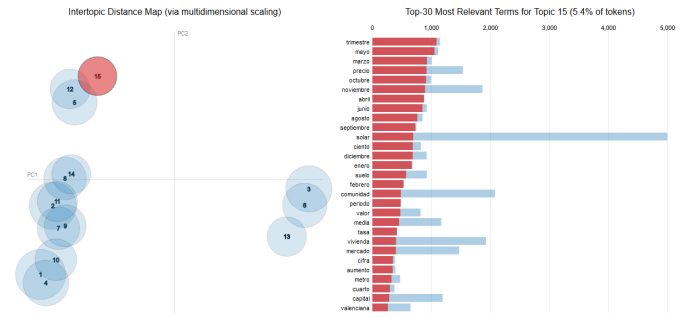


Fig. 15. Using “topicmodels” R package. “ENV\_SOLAR” theme, text written in Spanish.

The CTM model allows to classify documents or news in media. We also think that we will be able to use it in the future to do further correlation analysis. Of course our final aim will be to use this technique in future research to do causal analysis from the information collected (the indexes built on it) and the movement of key variables in energy markets.

**D. Speedup and Efficiency analysis (getting the text)**

As we have explained in the methodology section, getting the text from URLs is a compute-intensive phase. We present two figures for analyzing sequential and parallel execution modes. They summarize the performance in terms of Speedup and Efficiency according to equations (1) and (2). Fig. 16 shows that speedup improves when using

a network of workstations. Although efficiency (in terms of reducing execution time) increases with multicore execution a network of workstations is still preferred:

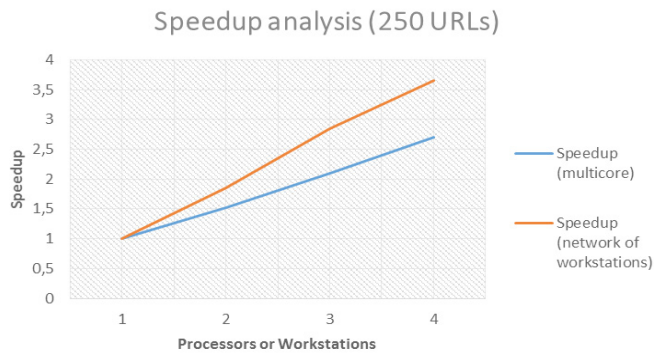


Fig. 16. Speedup comparison.

Our code does not require the dispatch of regular data between processes. Therefore, when we are using a network of workstations, the communication cost is not excessive and efficiency can be maintained at a constant level. However, in the multicore case the computer memory has to be shared and this issue impacts in the efficiency values.

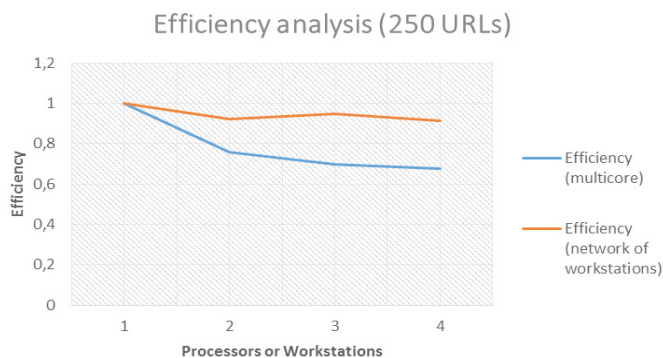


Fig. 17. Efficiency analysis

A multicore execution can be used to reduce execution time. Nevertheless, a network of workstations is preferred.

## V. CONCLUSIONS AND FUTURE WORK

In this paper we have used extensive data from several sources to analyze two issues related to energy markets. First, we analyze the public opinion about energy policy of the Spanish government using GDELT. Second, we conduct a correlation analysis between sentiment variables about the public policy and real prices and demand taken from the MIBEL energy market for the same period. Two results are worth emphasizing. On one hand, we detect negative feelings about the solar energy policy introduced by the Spanish government in 2015. On the other, hand, we find weak correlation between the indexes (tone) in mentions from GKG database and average daily log prices of energy. We do not find any correlation to average daily energy demand.

There are many extensions using extensive databases like the one in this paper or similar to follow different research lines in the future. We only quote two possibilities closely related to this exercise. First, we have only taken into account Spanish or English while alternative languages could be important to build sentiment indexes. Second, we have only presented correlation analysis between the indexes and average prices and demand but some formal demand model where to include these indexes as explanatory variables is necessary to accurately

measure the potential of these variables to explain the evolution of key energy market variables.

## ACKNOWLEDGMENT

We acknowledge very useful comment from an editor of the journal.

## REFERENCES

- [1] R. Sullivan, W. Martindale, N. Robins, and Winch H. (2014, November). *Principles for Responsible Investment. Policy Frameworks for Long-Term Responsible Investment: The Case for Investor Engagement in Public Policy*. Available: [http://www.unpri.org/wp-content/uploads/PRI\\_Case-for-Investor-Engagement.pdf](http://www.unpri.org/wp-content/uploads/PRI_Case-for-Investor-Engagement.pdf)
- [2] J. Cochran, et al. (2013, October). *Market evolution: Wholesale electricity market design for 21st century power systems*. Technical Report. Contract. Available: <http://www.nrel.gov/docs/fy14osti/57477.pdf>
- [3] The GDELT Project: Watching Our World Unfold. <http://gdeltproject.org/>
- [4] E. Dumbill, *Planning for big data*. O'Reilly Media, Inc. 2012.
- [5] D. J. Gerner, R. Abu-Jabr, P. A. Schrodt, and Ö. Yilmaz. "Conflict and Mediation Event Observations (CAMEO): A new event data framework for the analysis of foreign policy interactions." *International Studies Association*, New Orleans (2002).
- [6] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: <https://www.R-project.org/>.
- [7] D. Blei, and J. Lafferty. "Correlated topic models." *Advances in neural information processing systems*, vol. 18, pp. 147. 2006.
- [8] D. Blei, Y. Andrew, and M. I. Jordan. "Latent dirichlet allocation." *The Journal of machine Learning research*, vol 3, pp. 993-1022. 2003.
- [9] K. Hornik, and B. Grün. "topicmodels: An R package for fitting topic models." *Journal of Statistical Software*, vol 40.13, pp: 1-30. 2011.
- [10] K. Hornik, P. Mair, J. Rauch, W. Geiger, C. Buchta, and I. Feinerer. "The textcat package for n-gram based text categorization in R." *Journal of Statistical Software* vol 52.6, pp. 1-17. 2013.
- [11] E. Cantú-Paz, Erick. *Designing Efficient and accurate Parallel Genetic Algorithms*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign, Champaign, IL, USA. Advisor(s) David E. Goldberg. 1998.
- [12] L. Tierney, A. J. Rossini, N. Li, H. Sevcikova. (2008). *Snow: simple network of workstations*. R package version 0.3-3, Available: <https://cran.r-project.org/web/packages/snow/index.html>
- [13] N. Matloff, *Parallel Computing for Data Science: With Examples in R, C++ and CUDA*. Chapman and Hall/CRC. 2015.
- [14] S. S. Shapiro, and M. B. Wilk. "An analysis of variance test for normality (complete samples)." *Biometrika*, vol. 52, pp. 591-611. 1965.
- [15] OMI-Polo Español S.A. (OMIE): Market Results. Available: <http://www.omie.es/files/flash/ResultadosMercado.swf>



**Diego J. Bodas-Sagi** is an Associate Professor in the Francisco de Vitoria University (Polytechnic School). He has a PhD from the Complutense University of Madrid in computer science. His research interests include Big Data, Data Science, Computational Economics, modelling and e-Health. Contact address is: Universidad Francisco de Vitoria. Carretera Pozuelo a Majadahonda, Km 1.800, 28223 Pozuelo de Alarcón, Madrid (Spain).



**José M. Labeaga** is Professor of Economics at the Open University in Madrid and Research Affiliated at UNU-MERIT (Maastricht University) and Economics for Energy. He is Ms. and PhD. in Economics by Universitat Autònoma de Barcelona. He has served for the Spanish Government as General Director of the Institute for Fiscal Studies during the period 2008-2012. His main research interests rely on applied microeconomic models, microsimulation and ex-ante evaluation of programs as well as ex-post or impact evaluation of public policies in several fields as health, energy or taxation. Contact address is: Universidad Nacional de Educación a Distancia. Departamento de Análisis Económico II. C/ Senda del Rey, 11 28040, Madrid (Spain).