

# Some new methods to evaluate key variables in energy markets

José M. Labeaga



VII Atlantic Workshop on Energy and Environmental Economics

A Toxa, June 27, 2016

# The problem to address (I)

- Extract information from everywhere to evaluate key variables in energy markets (special focus on demand and particularly price elasticities)
- How can we use the impressive amount of information (web, media, hundred of papers, social networks, etc.) to analyze correlation-causal effects relevant for economic decisions?
- Tools for accessing, storing, extracting, analyzing?
- Methods? Descriptives, graphics, correlation ...?

# The problem to address (II)

- Big Data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or does not fit the structures of our database architectures. To gain value from this data, one must choose alternative ways to process it
- Is it a good idea to storage data (trying to built long time series or long time series of cross-sections) for doing long-run analyses or should these data be analyzed for short-run purposes?
  - Academic vs non-academic analysis
- Is this kind of data useful? Can its use enrich current analyses?

# The problem to address (III)

- I do not have closed answers to these questions but it depends
  - On the data
  - On the questions to address
  - On the decisions to inform and their timing
  - ...
- So, my main purpose it to describe some (possibilities of) exercises using this kind of data

**Example:**

**Using GDELT data to evaluate the confidence on the  
Spanish government energy policy (joint work with  
D. Bodas-Sagi)**

# Example (I)

- Set-up of this exercise:
  - How do sentiment indicators affect energy prices and energy demand?
  - I do not know of any paper analyzing this issue in energy markets
  - I do not know of any economic paper using this database
  - Everything in the presentation constitutes “ad hoc” ideas
  - I will only give correlations

# Example (II)

- We use Global Database on Events Language and Tone (GDELT)
  - The GDELT Project is supported by Google Ideas and it shares real-time information and metadata with the world. This codified metadata (but not the text of the articles, which we will use later on) 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 (all) 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

# Example (III)

- GDELT ...
  - 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”
  - The data files use Conflict and Mediation Event Observations and we can identify each event (extracted from the original information), which allows rich contextualization of them. We can also extract information (words) on messages, news, etc. , related to a theme or subject

# Example (IV)

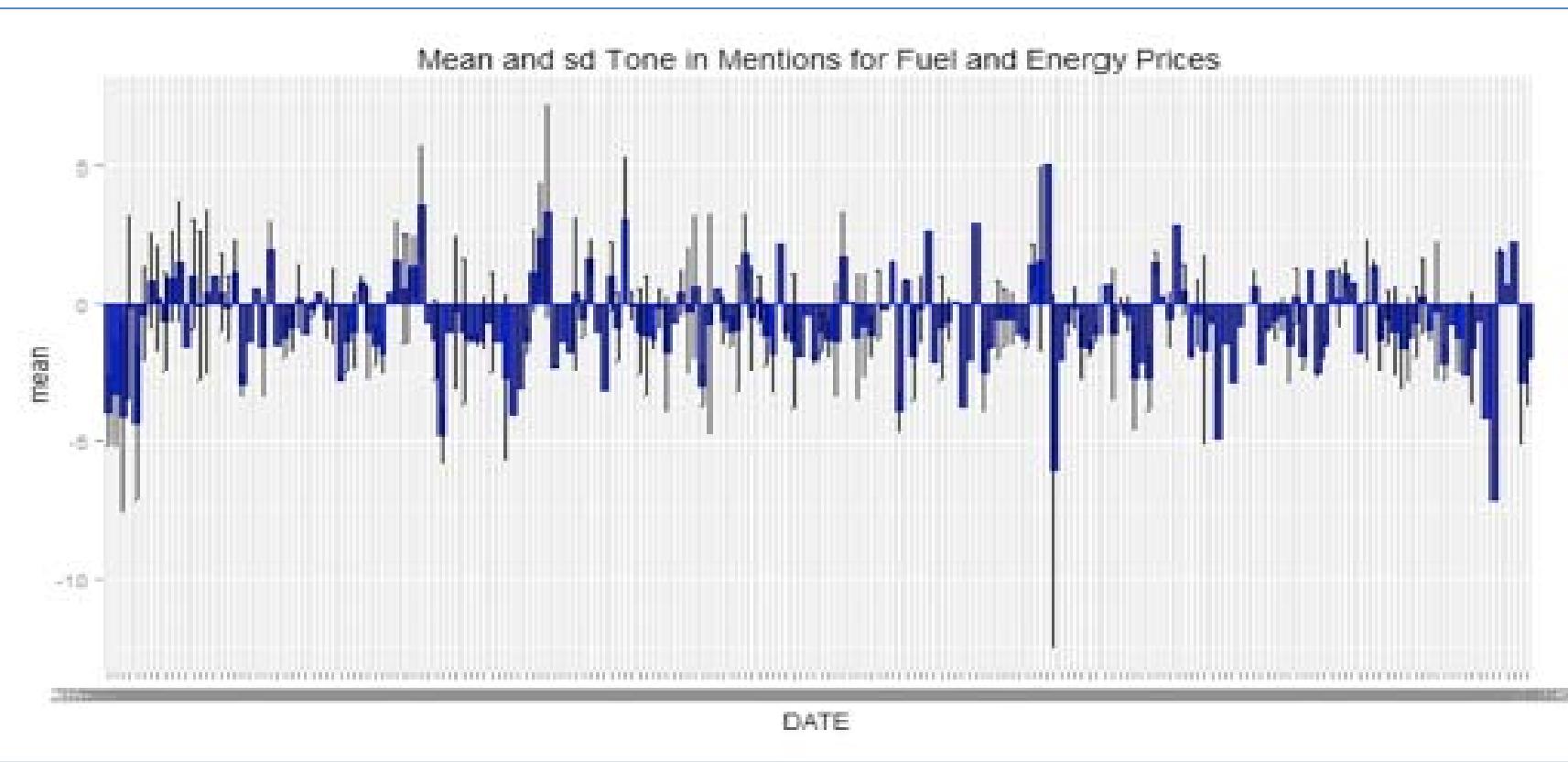


Figure 1. All mentions to fuel prices in Spain (February, the 18th 2015 to October, the 25th 2015)

# Example (V)

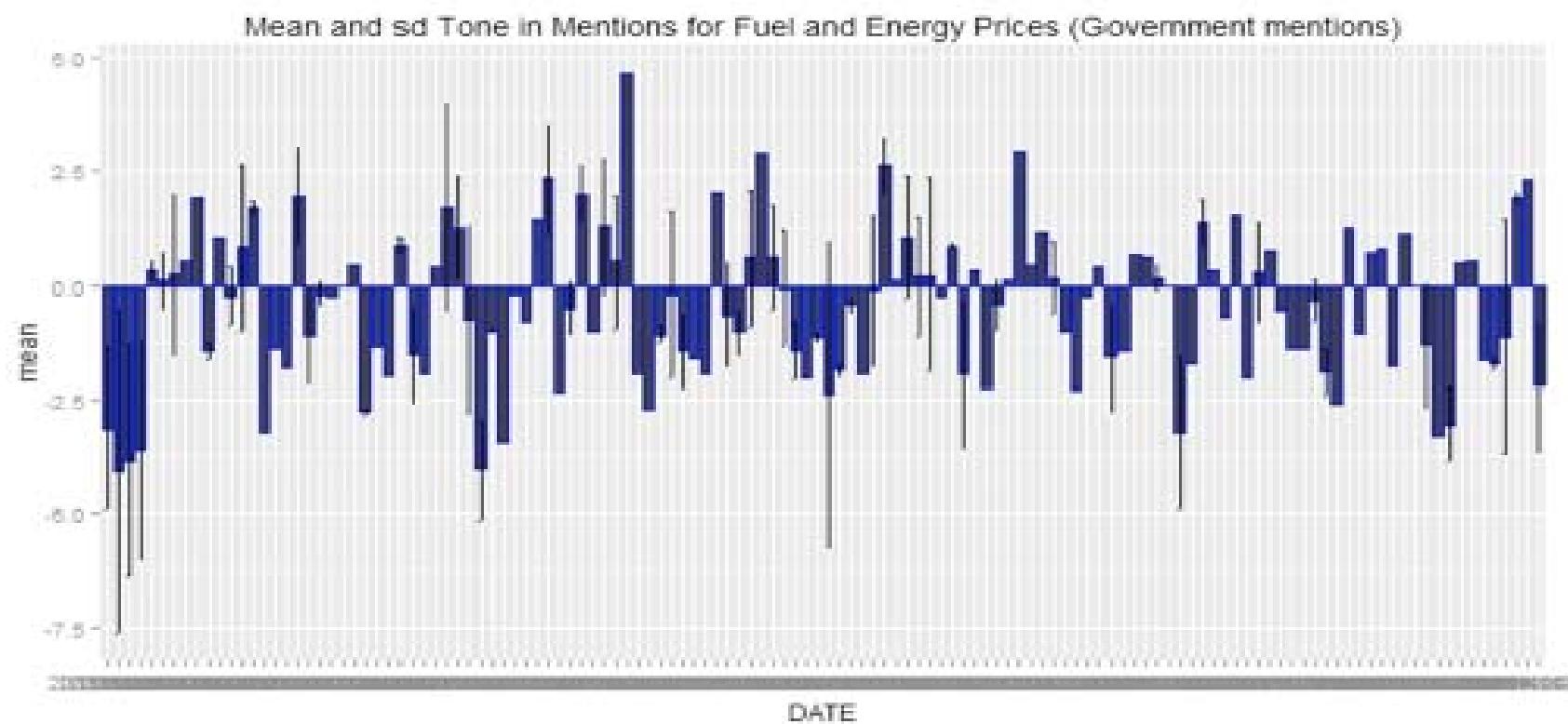


Figure 2. All mentions to fuel prices in Spain (February, the 18th 2015 to October, the 25th 2015)  
filtering words (government mentions)

# Example (VI)

Histogram for mean tone in all mentions for fuel prices

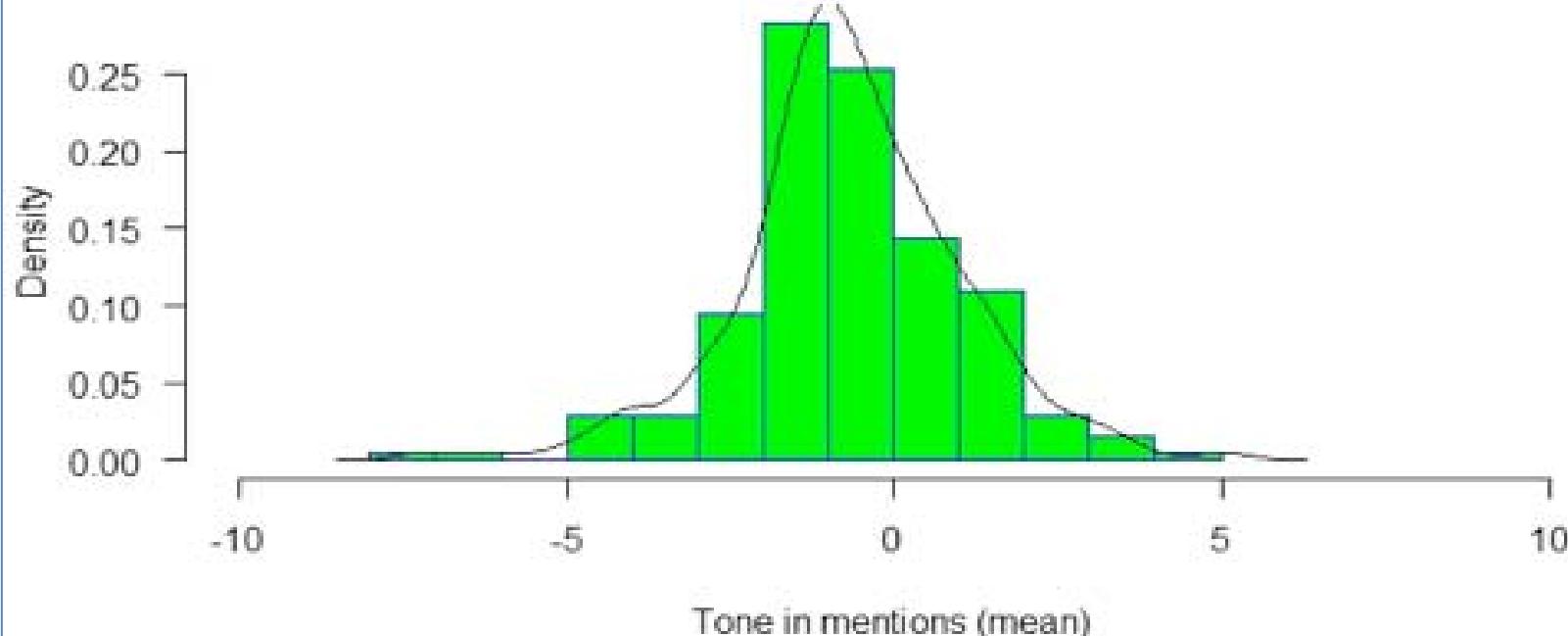


Figure 3. Histogram corresponding to all mentions to fuel prices in Spain (February, the 18th 2015 to October, the 25th 2015)

# Example (VII)

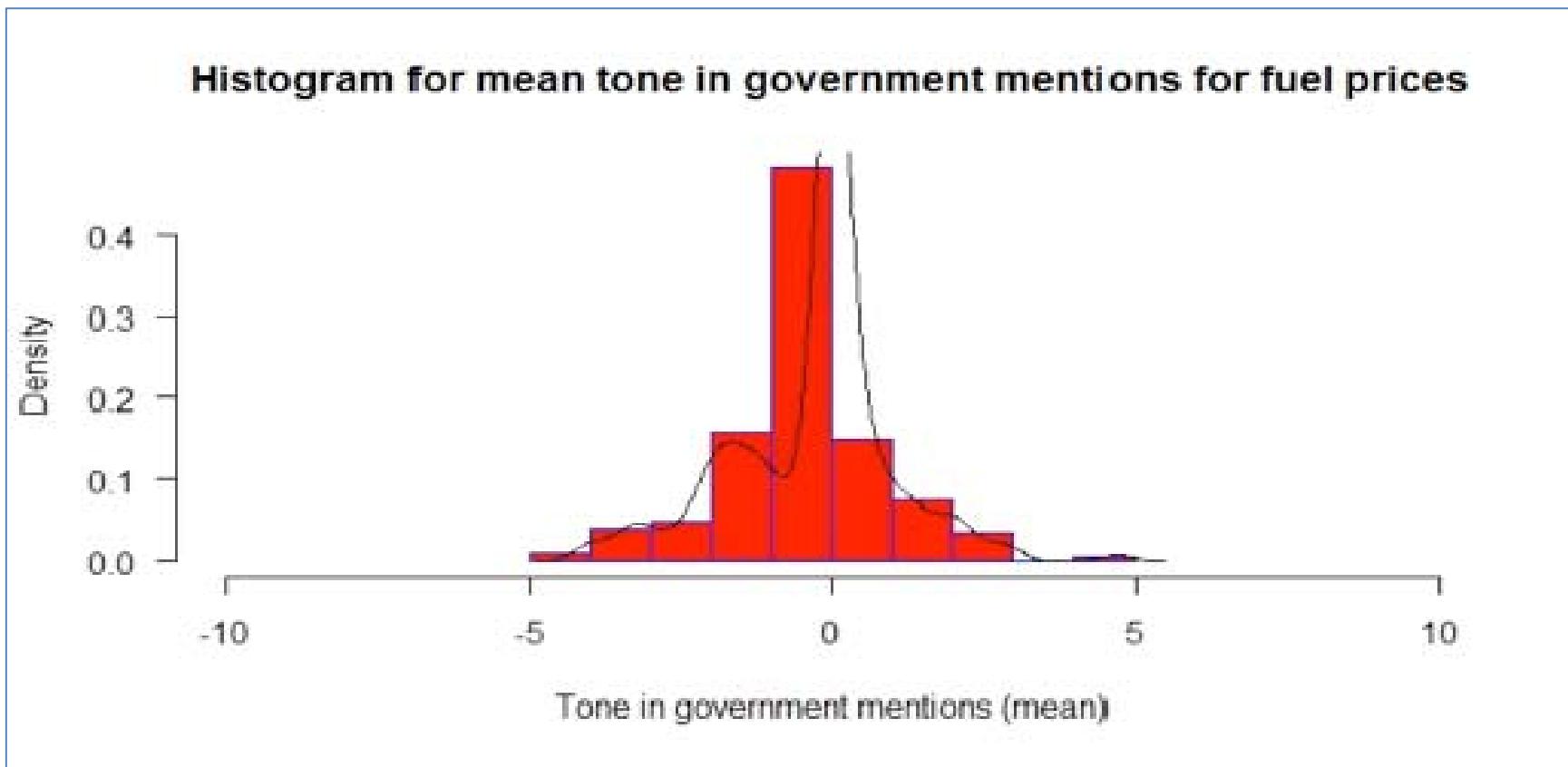


Figure 4. Histogram corresponding to all mentions to fuel prices in Spain (February, the 18th 2015 to October, the 25th 2015) filtering words (government mentions)

# Example (VIII)

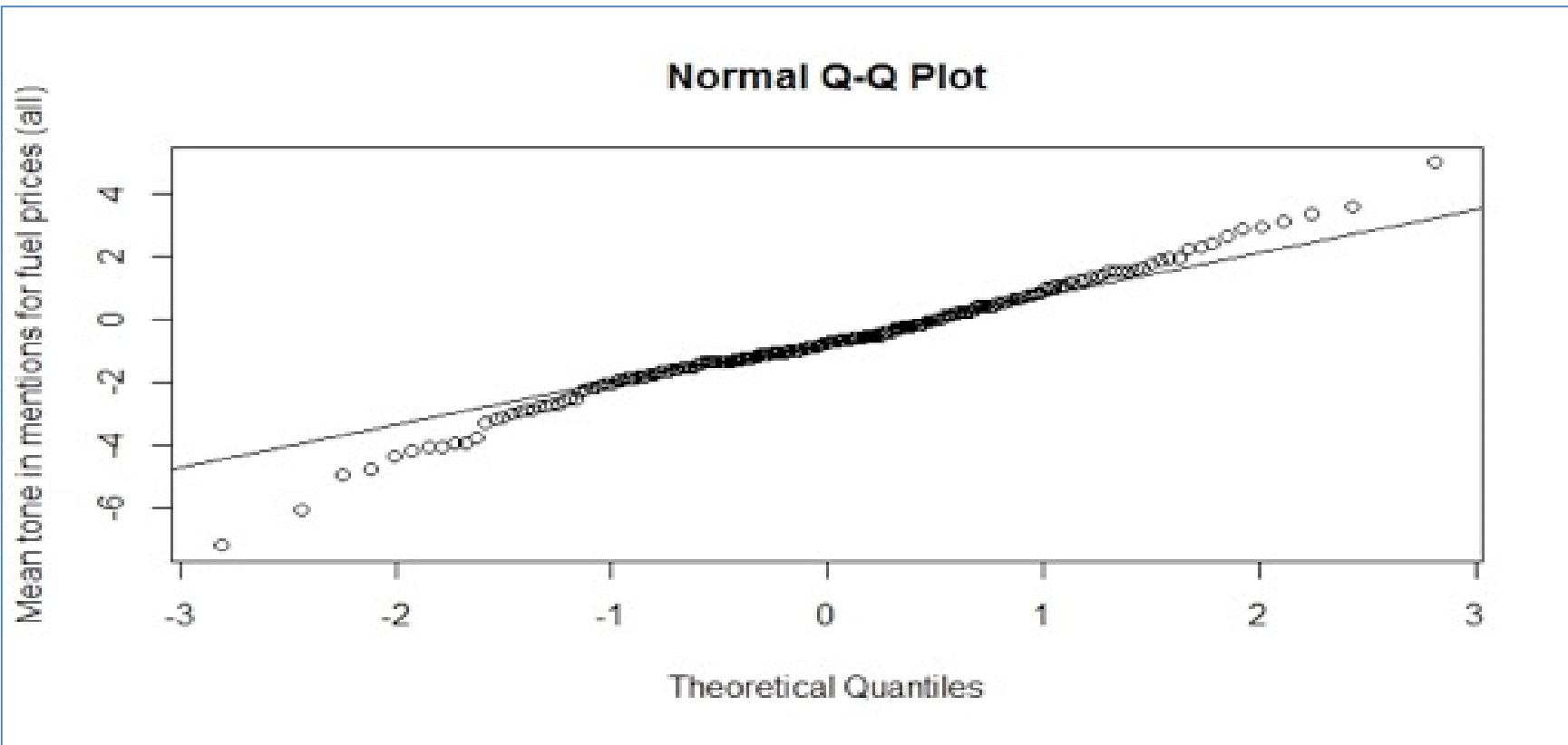


Figure 5. Q-Q plot for testing normality in mentions to fuel prices

# Example (IX)

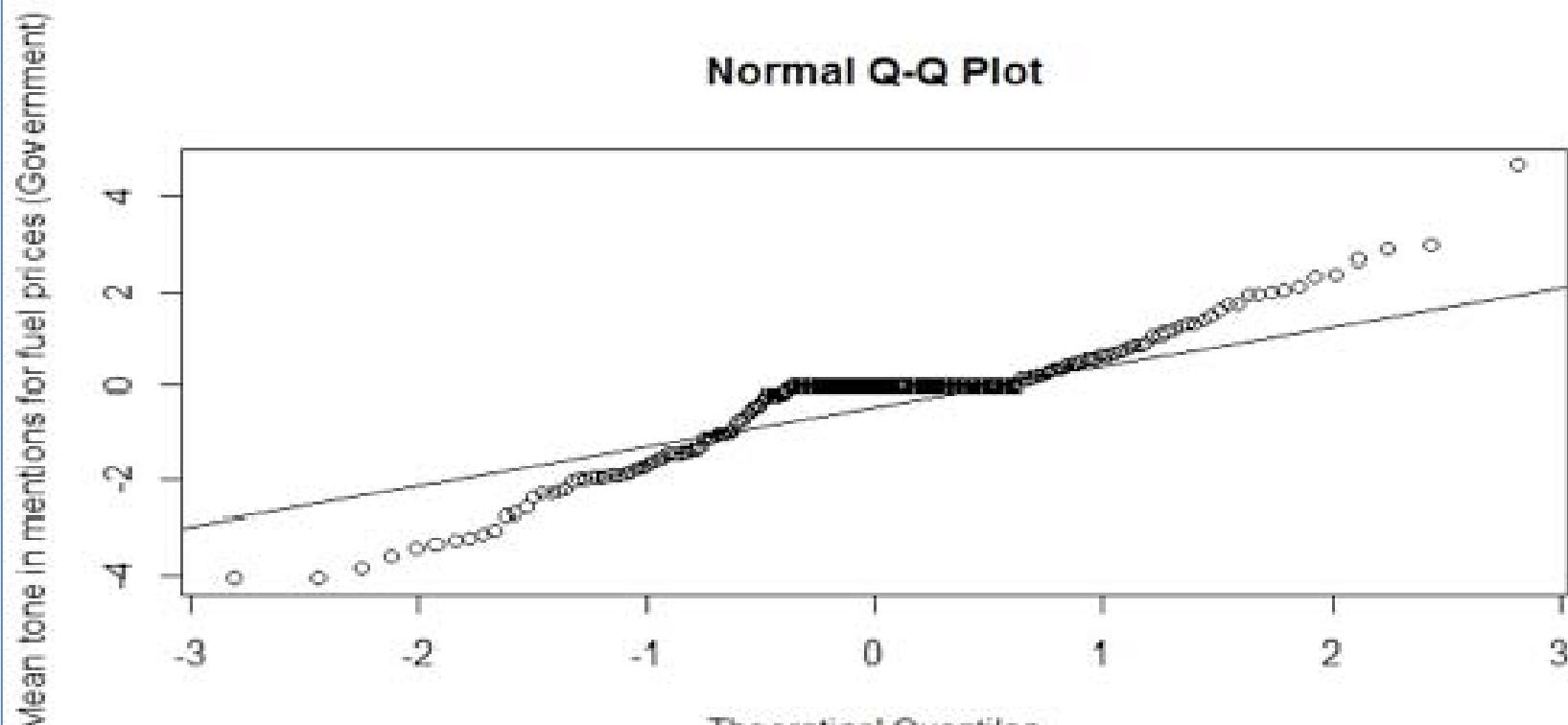


Figure 6. Q-Q plot for testing normality in mentions to fuel prices (filtering words)

# Example (X)

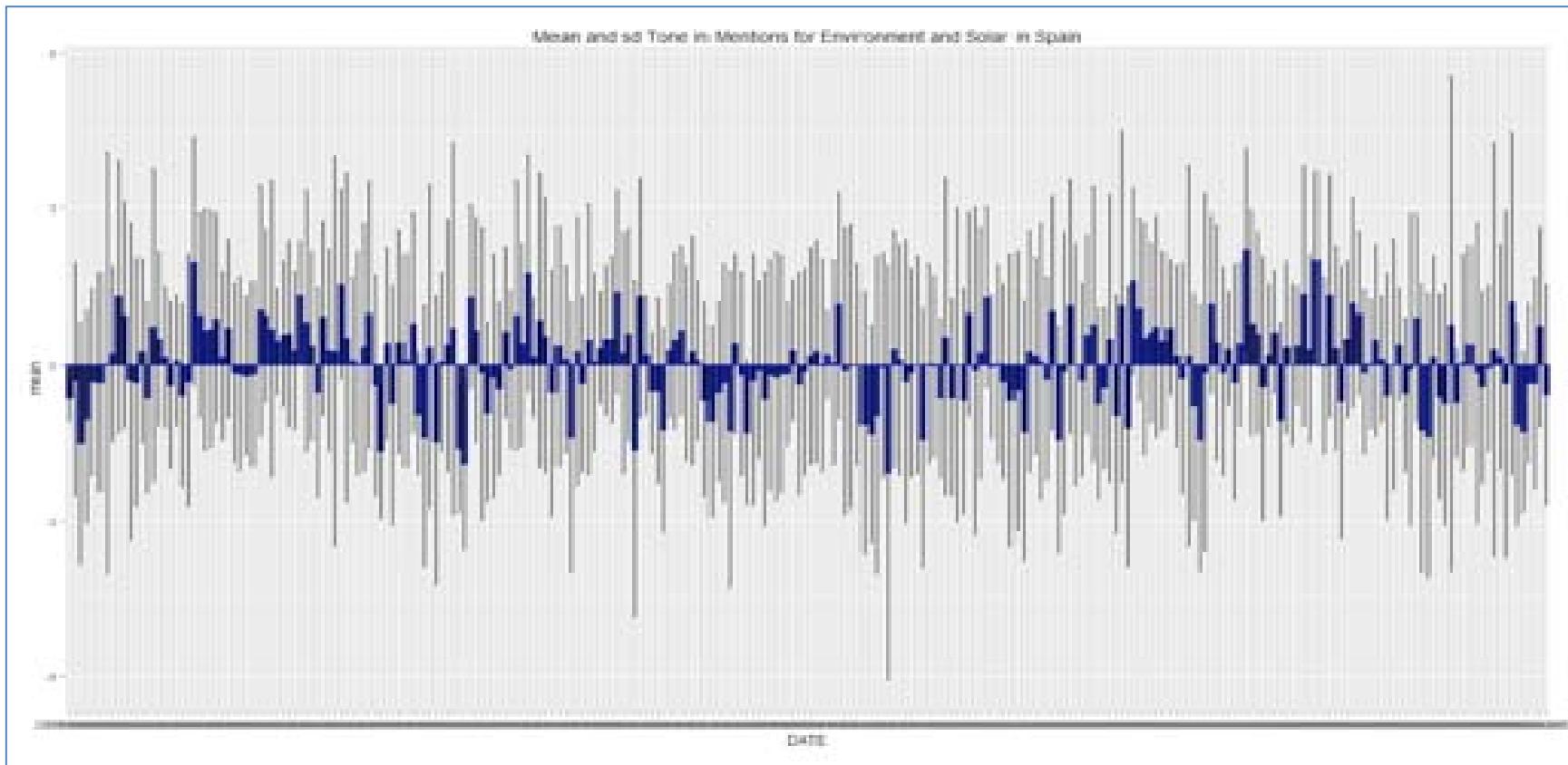


Figure 7. All mentions to environmental and solar energy in Spain (February, the 18th 2015 to October, the 25th 2015)

# Example (XI)

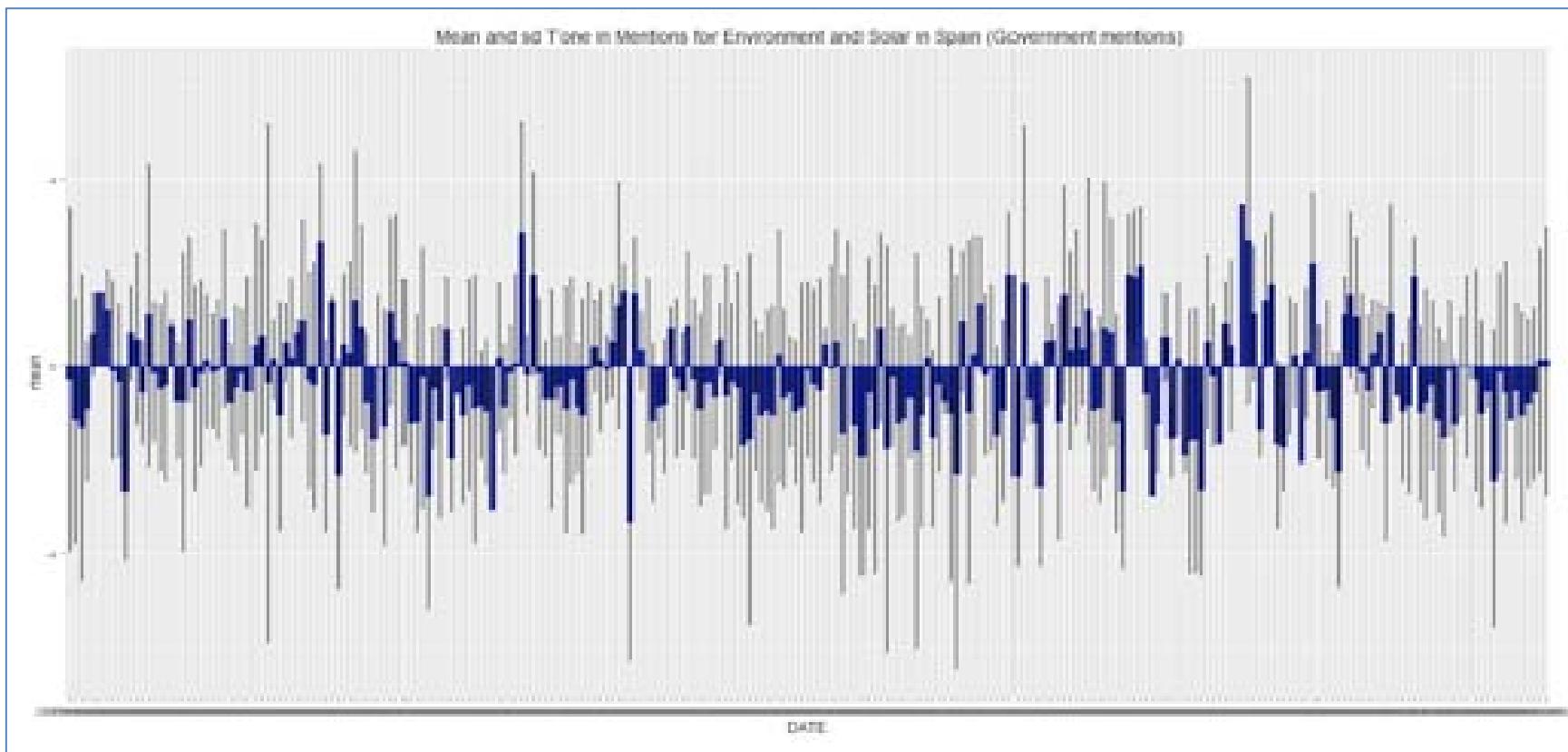


Figure 8. All mentions to environmental and solar energy in Spain (February, the 18th 2015 to October, the 25th 2015) filtering words (government mentions)

# Example (XII)

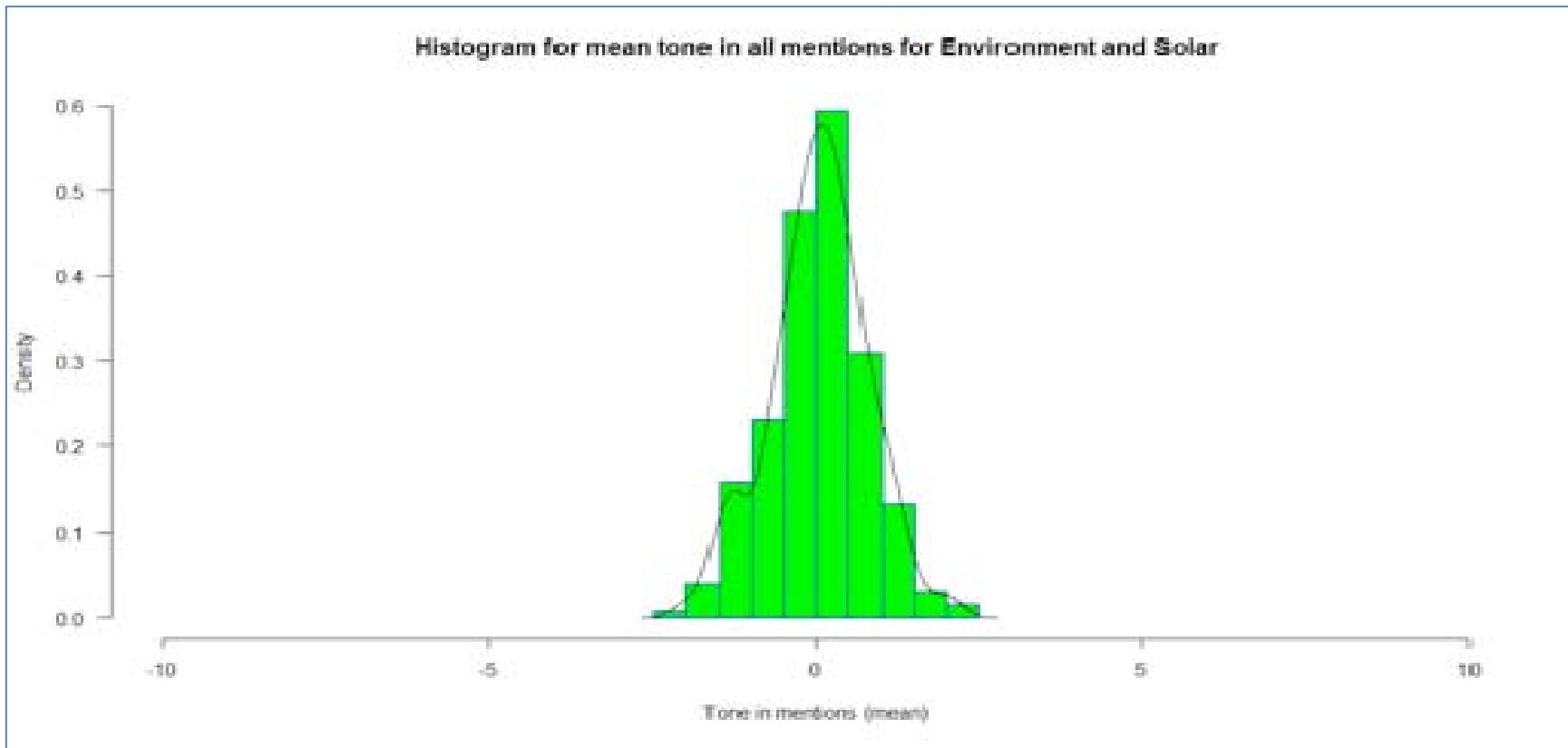


Figure 9. Histogram corresponding to all mentions to environmental and solar energy in Spain  
(February, the 18th 2015 to October, the 25th 2015)

# Example (XIII)

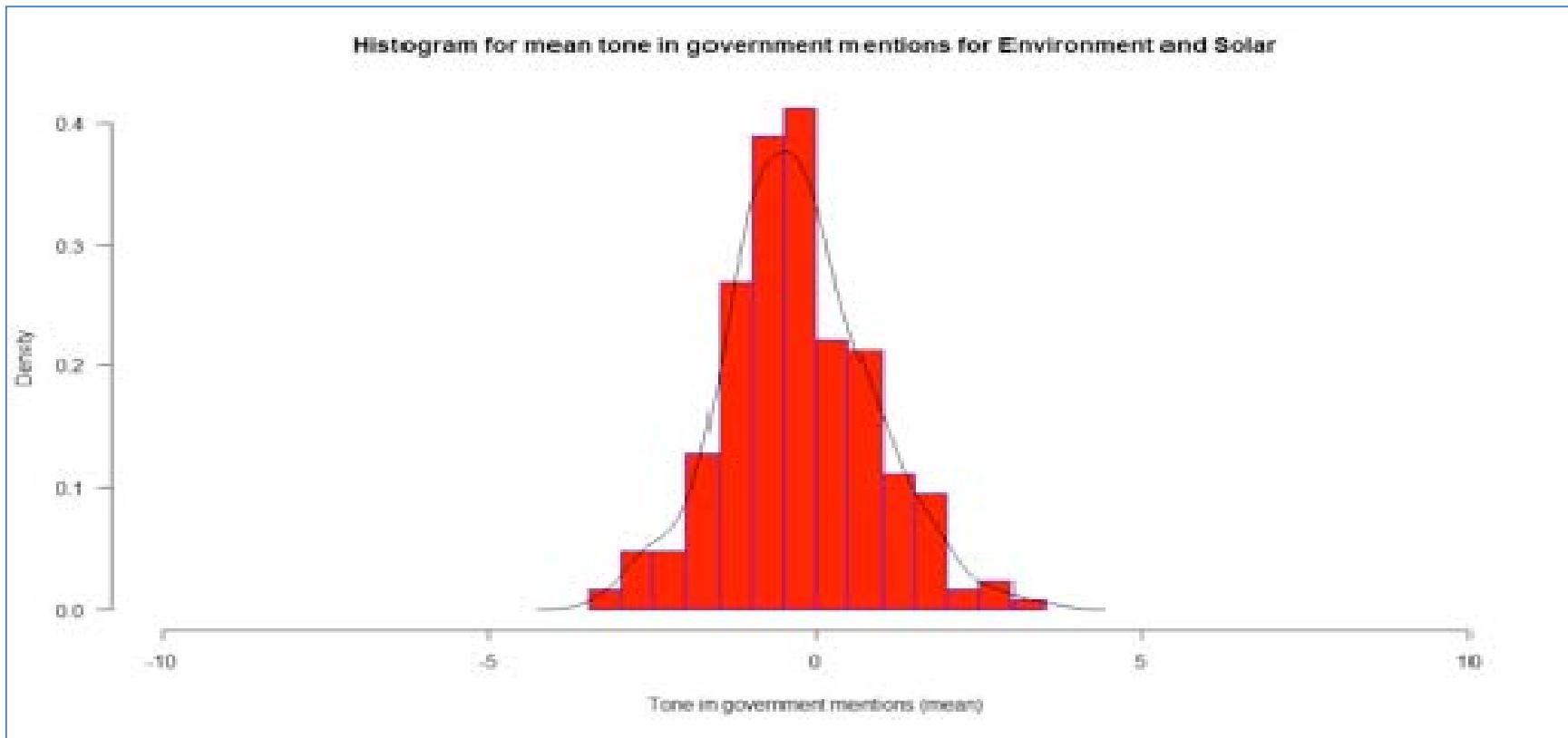


Figure 10. Histogram corresponding to all mentions to environmental and solar energy in Spain (February, the 18th 2015 to October, the 25th 2015) filtering words (government mentions)

# Example (XIV)

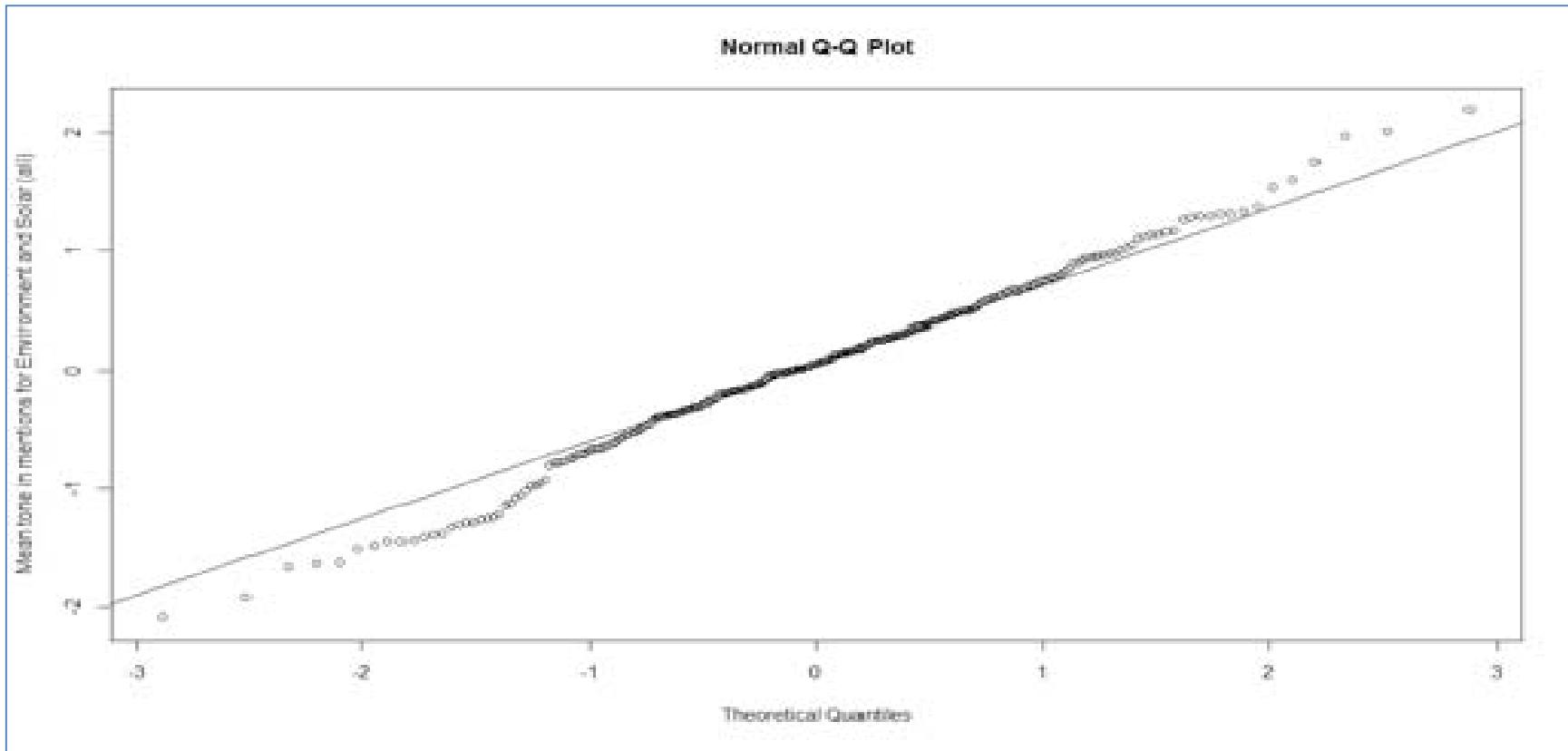


Figure 11. Q-Q plot for testing normality in mentions to environmental and solar energy

# Example (XV)

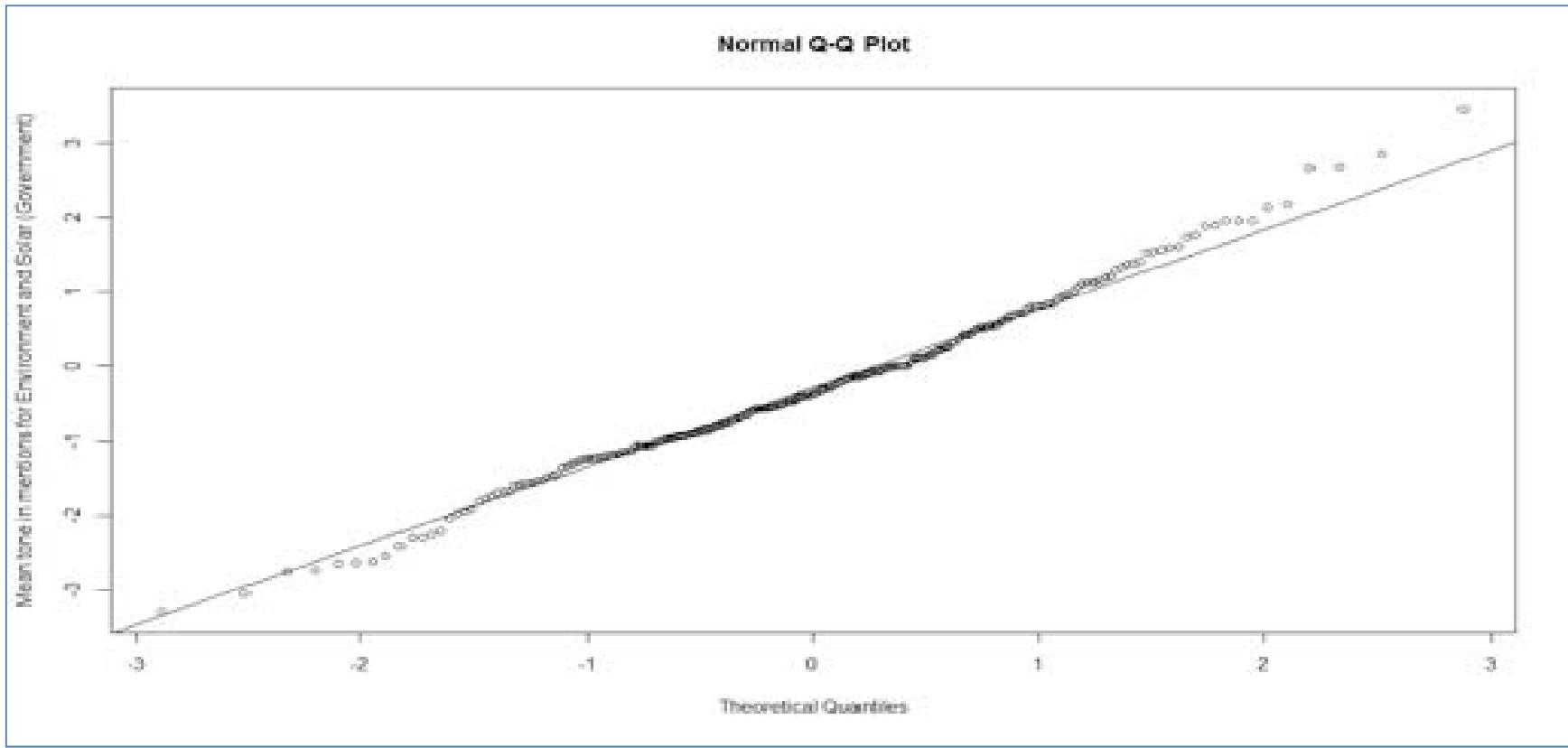


Figure 12. Q-Q plot for testing normality in mentions to environmental and solar energy (filtering words)

# Example (XVI)

- From all these analysis, I will only present at this moment some correlation (but I have also done some more formal correlation analysis through regression, i.e., demand models including sentiment indicators)
  - Coefficient of correlation log average prices of energy and meangovsolar index = -0.042 (significant at 0.32)
  - Coefficient of correlation average log prices of energy and meanallfuel index = -0.071 (significant at 0.09)
  - Coefficient of correlation log demand of energy and meanallfuel index = -0.045 (significant at 0.30)

## Example (XVII)

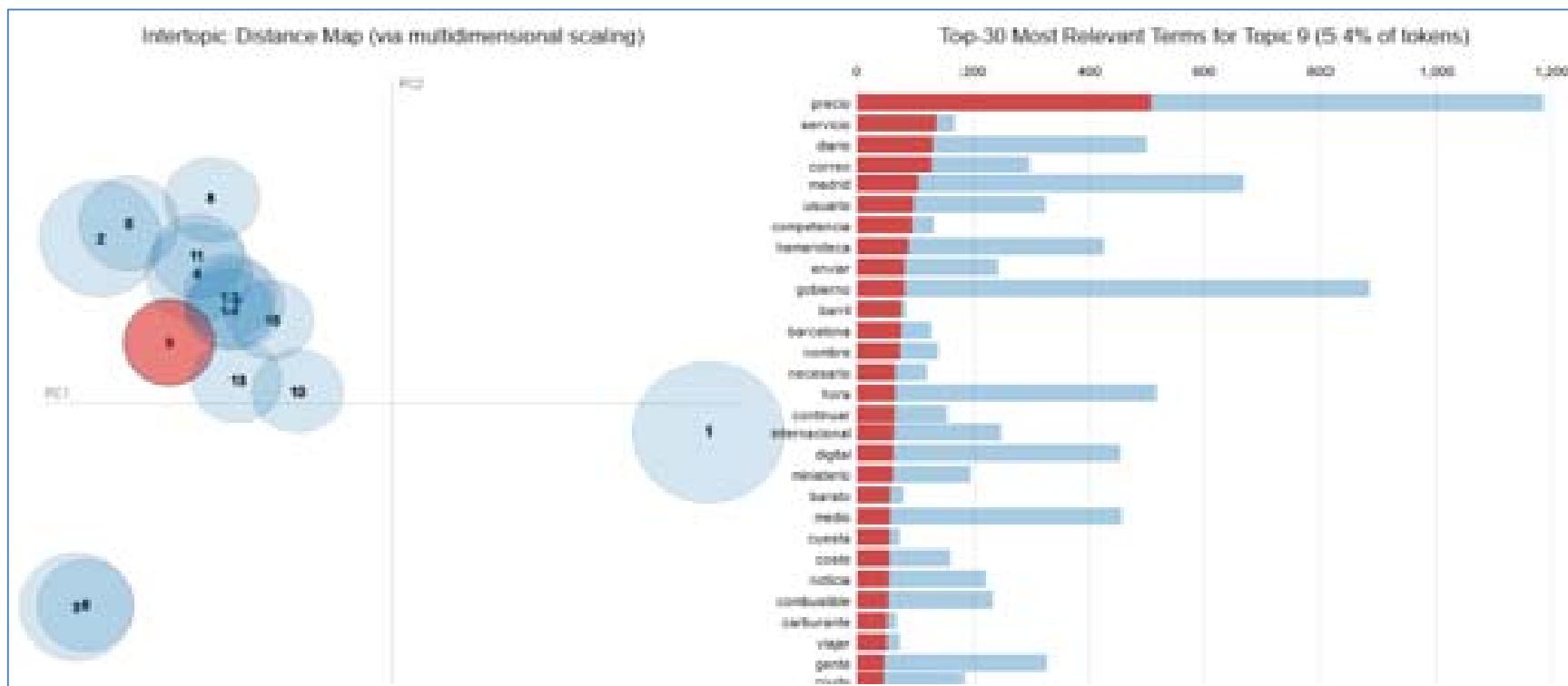


Figure 13. Analyzing “FUELPRICES” theme using topics model, text written in Spanish

## Example (XVIII)

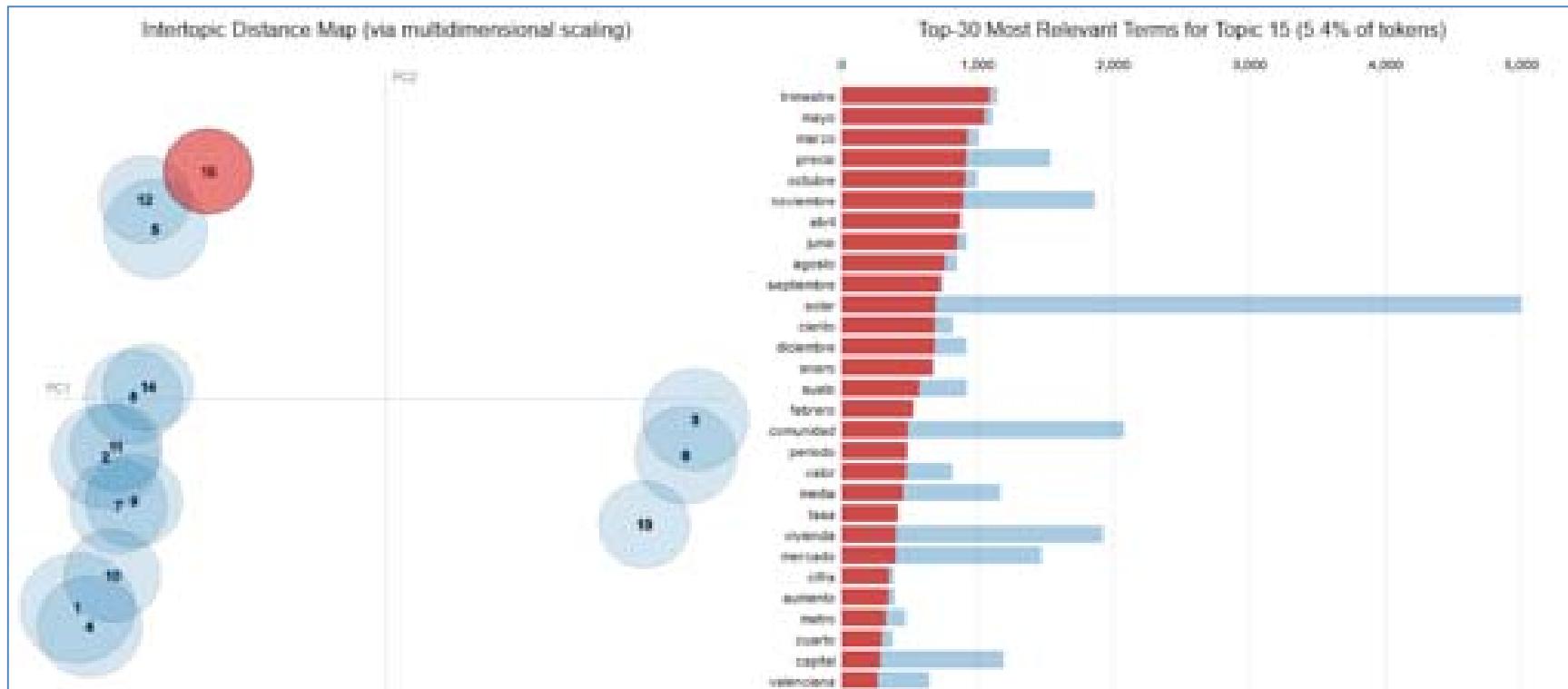


Figure 14. Analyzing “ENVSOLAR” theme using topics model, text written in Spanish

# Example (XIX)

- Our final aim:
  - To estimate demand models with daily data (coming from OMIE) including sentiment indicators on a daily base and see how news, media, opinion, etc. do affect energy demand (directly or indirectly through prices) to be able to:
    - Inform about the effects of policy decisions on demand through “non-market variables”
    - Analyze whether price-elasticities (or income elasticities are affected)
    - Etc.
- Preliminary results:
  - Some negative effects on demand at short run and they influence market variables at specific dates

# Many thanks



VII Atlantic Workshop on Energy and Environmental Economics  
A Toxa, June 27, 2016