

# 100

## memorias

## y un índice de incertidumbre ECONÓMICA

MARCO VEGA\* Y VALENTINA PINELO\*\*

Este artículo presenta una medida de incertidumbre económica basada en el texto de las memorias del BCRP de los últimos 100 años. El indicador muestra una tendencia persistente de aumento de incertidumbre desde la década de los 50, fenómeno parecido al encontrado en medidas de incertidumbre estimadas para EE. UU. Además, el indicador estimado para Perú muestra cómo ha ido variando la contribución de diversos factores asociados con la incertidumbre económica en este centenario.



\* Subgerente de Investigación Económica del BCRP  
marco.vega@bcrp.gob.pe



\*\* Pontificia Universidad Católica del Perú  
valentina.pinel@pucp.edu.pe

El Banco Central de Reserva del Perú (BCRP) cumplió 100 años de fundación este 2022. Su creación se produjo el 9 de marzo de 1922, durante el segundo gobierno del presidente Augusto B. Leguía. A partir de esta fecha, el Banco Central ha publicado una memoria anual donde informa sobre los aspectos más resaltantes del ámbito económico y las políticas económicas llevadas a cabo. Las cien memorias publicadas hasta 2021 —que están disponibles en formato digital en el portal del BCRP— son una muestra viva de la evolución histórica de la economía peruana en las últimas diez décadas. En este artículo se utiliza el contenido textual de dichas memorias con el fin de extraer información sobre el grado de incertidumbre económica prevalente en el centenario transcurrido.

En un periodo largo, que ha sido testigo de varios ciclos económicos, es conocido que el nivel de incertidumbre afecta a las variables económicas agregadas como el consumo, la inversión y la actividad económica. La incertidumbre económica es una aproximación a la varianza de ciertas variables macroeconómicas como el PBI, la misma que varía en el tiempo de manera estocástica. Por su parte, los episodios de crisis o recesión económica son generalmente periodos de alta varianza o volatilidad macroeconómica. Estudios como los de Justiniano y Primiceri (2008), Bloom (2009) y Fernández-Villaverde et al. (2011) han explorado teóricamente cómo la incertidumbre afecta a dichas variables.

En términos empíricos, la incertidumbre macroeconómica se puede medir utilizando la volatilidad implícita de la bolsa de valores, la dispersión transversal de la productividad de las empresas o a través de la incorporación de la volatilidad estocástica a los choques estructurales de un modelo de vectores autorregresivos (VAR). En general, la medición de la incertidumbre a través del tiempo trata de capturar no simplemente el comportamiento promedio de las variables (primeros momentos), sino también la evolución de los segundos momentos o volatilidades.

Una forma empírica adicional de medir la incertidumbre es a través del análisis textual. El estudio pionero sobre este método es de Baker et al. (2016), que introduce la idea de contar términos asociados a incertidumbre económica y política a partir de los principales periódicos impresos. El estudio demostró que un simple conteo de estos términos a través del tiempo puede capturar muy bien el grado de incertidumbre; además, este indicador, así extraído, tiene efectos importantes sobre variables económicas.

El estudio de Baker et al. (2016) se basa en la frecuencia de artículos en 10 periódicos estadounidenses importantes que contienen al menos un término predefinido en las siguientes tres categorías



La incertidumbre económica es una aproximación a la varianza de ciertas variables macroeconómicas como el PBI, la misma que varía en el tiempo de manera estocástica.



(economía, incertidumbre y política): “económico” o “economía”; “incierto” o “incertidumbre”; y “Congreso”, “déficit”, “Reserva Federal”, “legislación”, “regulación” o “Casa Blanca”. El objetivo de este índice, que se conoce como Economic Policy Uncertainty (EPU), es capturar la incertidumbre sobre quién tomará las decisiones de política económica, qué acciones de política económica se emprenderán y cuándo, y los efectos económicos de dichas acciones (o inacciones).

Numerosos estudios han seguido este tipo de análisis. En algunos casos, la extracción y conteo de estas palabras es manual, mientras que en otros casos se realiza enteramente a través de programas de aprendizaje de máquina. Por ejemplo, Manela y Moreira (2017) construyeron un índice de incertidumbre a partir de 1890 basado en artículos de primera plana de *The Wall Street Journal*. El estudio emplea como medida de incertidumbre el comovimiento entre las coberturas de primera plana y el índice de volatilidad VIX<sup>1</sup> para construir un nuevo índice que se conoce como volatilidad implícita en noticias [*news implied volatility* (NVIX)]. El método de dichos autores supone que la variación temporal de los temas tratados por la prensa económica es una buena aproximación de la evolución de las preocupaciones de los inversionistas respecto a dichos temas.

El presente artículo sigue la línea de Baker et al. (2016) para construir un índice de incertidumbre a partir del conteo de palabras asociadas al concepto de incertidumbre. A diferencia de los trabajos mencionados, que utilizan medios de prensa escrita para extraer las palabras, el ejercicio que se realiza aquí utiliza las memorias del BCRP.

<sup>1</sup> El índice de volatilidad VIX —que es calculado por el Chicago Board Options Exchange (CBOE) en tiempo real— mide la volatilidad de los precios en el índice S&P 500.

Es plausible suponer que la comunicación de un banco central plasmada en las memorias sea más ponderada que la comunicación de noticias o eventos a través de la prensa. Por lo general, los bancos centrales buscan reducir cualquier exceso de incertidumbre que pueda surgir a partir de los eventos económicos. Por ello, como se verá líneas abajo, el conjunto de palabras que se tienen que analizar para poder extraer incertidumbre a partir de las memorias es más amplio que el que se seleccionaría cuando se usan periódicos, pues un banco central, en general, es cuidadoso con los términos que utiliza.

### DATOS Y METODOLOGÍA

Los datos utilizados en este artículo son los textos de las memorias del BCRP. Al conjunto total de 100 memorias se le denomina usualmente como corpus. Dado que un subconjunto de estas memorias corresponde a documentos físicos escaneados, primero se interpretaron las imágenes escaneadas como texto. Este procedimiento implicó dos pasos: i) la generación de archivos de imagen en formato JPG a partir de las memorias en formato PDF accesibles en el portal web del BCRP, procedimiento automatizado a través de la librería pdf2image en el *software* Python; y ii) la aplicación del motor Tesseract para el reconocimiento óptico de caracteres (*optical character recognition-OCR*) a través de Python. Este motor está basado en redes neurales del tipo LSTM (*long short-term memory*), que son especialmente idóneas para clasificar secuencias de imágenes.

Una vez extraído el texto de las memorias escaneadas, se completó todo el corpus de memorias en formato de base de datos de serie de tiempo, la misma que contiene los textos completos de cada publicación indexada por año de referencia.

### PROCESAMIENTO DEL TEXTO

Con la base de datos construida, se realizó un pre-procesamiento estándar de la parte textual que incluyó la limpieza, la tokenización<sup>2</sup> y la normalización de los textos. Por un lado, la limpieza del texto consistió en corregir errores de detección en encabezados, número de páginas, cuadros, entre otros, al momento de extraer los textos, tales como palabras incompletas o mal detectadas. Por otro lado, la tokenización consistió en dividir las cadenas de texto largas en listas de subcadenas denominadas tokens. Finalmente, la normalización se compuso de dos pasos: i) la eliminación de las puntuaciones, los símbolos y los *stop words*<sup>3</sup>; eliminación o conversión de los números a sus equivalentes en palabras; y conversión de mayúsculas a minúsculas;

y después ii) la reducción de las palabras conjugadas a sus raíces lingüísticas mediante la lematización, de modo que, por ejemplo, “encontraran” y “encuentras” se conviertan en “encontrar”.

### MÉTODO

Para poder aplicar los métodos del lenguaje de procesamiento natural es necesario transformar los textos a expresiones numéricas. Para ello, se empleó el método de ponderación de términos, denominado *tf-idf*<sup>4</sup>, que mide el grado de importancia de las palabras a lo largo del corpus al castigar tanto a los términos raros como a los muy frecuentes.

Para obtener el *tf-idf* se calculó primero la frecuencia de cada palabra, llamada también *tf* (*term frequency*). Este indicador es la relación entre la frecuencia de un término o palabra en el documento correspondiente y el número total de palabras de dicho documento.

$$tf_{j,a} = \frac{\text{conteo de } j \text{ en } d}{\text{número de palabras en } d}$$

Donde *j* denota a cada palabra o término y *d* representa cada memoria. En segundo lugar, las frecuencias de términos o recuentos de términos se multiplicaron por la frecuencia de documento inversa de un término en la colección de documentos (*idf*). Esto se realizó para que las palabras que son comunes en la mayoría de los documentos del corpus tengan menos peso o no se incluyan en el vector, pues estas palabras no tendrían contenido relevante (Ashok & Mehram, 2009). Por lo tanto, el *idf<sub>j</sub>* de la palabra *j* se define de la siguiente manera:

$$idf_j = \log \left( \frac{N}{df_j + 1} \right);$$

donde *N* es el número total de documentos en el corpus, y *df<sub>j</sub>* es el número total de documentos que contienen la palabra *j* (ocurrencia de *j* en *N* documentos).

Por lo tanto, se obtiene la puntuación *tf-idf<sub>j,d</sub>* de la palabra *j* en el documento *d* de la siguiente manera:

$$tf-idf_{j,d} = tf_{j,a} \times \log \left( \frac{N}{df_j + 1} \right)$$

Como último paso en la construcción del índice, luego de haber calculado los *tf-idf<sub>j,d</sub>* de cada palabra *j*, se establecieron las palabras que podían estar asociadas al concepto de incertidumbre. La selección y clasificación que se realizó está basada en el criterio de los autores y se presenta en el Cuadro 1.

<sup>2</sup> La tokenización se refiere a la identificación tokens dentro del corpus. Los tokens son las unidades más pequeñas con significado dentro de un texto. Se refieren usualmente a palabras, pero pueden ser también caracteres, partes de una palabra y hasta signos de puntuación.

<sup>3</sup> Los *stop words* son palabras vacías que no tienen un significado por sí solas, sino que modifican o acompañan a otras. Estas palabras vacías son términos comunes, como ‘el’, ‘y’, ‘yo’, entre otros, que son muy frecuentes en el texto y, por lo tanto, no transmiten información sobre el tema específico de un documento.

<sup>4</sup> El término *tf-idf* representa las siglas de *term frequency-inverse document frequency*.

## CUADRO 1 ■ Términos utilizados por tipo de incertidumbre

Tipos de incertidumbre	Palabras
Central	Incertidumbre, riesgo, dispersión, variabilidad, volatilidad, temor, tensión.
Negatividad	Depresión, deterioro, recesión, desaceleración, dificultades, desfavorable, negativo, adverso, desconfianza, retroceso, contracción.
Alteración	Distorsión, perturbación, choque, trastorno, desbalance, desequilibrio, presiones, desajuste, alteración, fluctuación, disturbios, inestabilidad.
Geopolítico/político	Crisis, guerra, conflicto, paralización, paro, elección, huelga, electoral, geopolítico, protesta, inmovilización, bloqueo.
Naturaleza	Desastres, niño, niña, sequía, covid, pandemia, coronavirus, anomalía, sismo, plaga, inundación, lluvia, epidemia, ómicron, delta.

ELABORACIÓN: PROPIA.

Se establecieron cinco tipos de incertidumbre. La incertidumbre central estuvo vinculada directamente al término “incertidumbre” y palabras cercanas como “riesgo”, “variabilidad”, “volatilidad”, entre otras. El segundo tipo de incertidumbre tuvo que ver con eventos negativos en general, como “depresión”, “deterioro” o “desconfianza”, que ocurren principalmente en escenarios de baja actividad económica. Un tercer tipo de incertidumbre fue etiquetado como alteración al representar cambios, distorsiones o desajustes, respecto a una situación presumiblemente mejor. Finalmente, la cuarta y quinta forma de incertidumbre fueron las más exógenas respecto a la economía. Estas fueron la incertidumbre geopolítica y política (“guerra”, “crisis”, “conflicto”, entre otros) y la incertidumbre asociada a factores naturales (“desastres”, “niño”, “sequía”, “covid”, entre otros).

Una vez determinados los términos a incorporar en el índice de incertidumbre económica, agrupados en el conjunto de términos  $I$ , y calculadas también sus correspondientes frecuencias de términos ( $tf$ ) y la medida  $tf-idf$ , se procedió a calcular el índice de incertidumbre económica ( $IIE$ ) de cada memoria  $d$  según la siguiente expresión:

$$IIE_d = \sum_{j \in I} tf_{j,d}$$

Es decir, se sumaron todas las frecuencias de términos que pertenecen al conjunto de palabras que definen incertidumbre dentro de cada una de las memorias. Adicionalmente, dado que el número de páginas de las memorias se ha ido incrementando en los años, se realizó un ajuste de cada publicación de acuerdo con el promedio móvil a 10 años del tamaño de páginas de las memorias. En particular, se utilizó la siguiente fórmula:

$$IIE_d^{adj} = IIE_d \times f_d$$

Donde  $f_d$  es el factor de ajuste definido como la división entre el número de páginas del documento  $d$  y el promedio móvil del número de páginas correspondientes a 10 memorias centradas en  $d$ . Ambos índices, el simple y el ajustado, mostraron en general un comportamiento similar. Se reportaron en este documento solo los resultados del cálculo del índice de incertidumbre económica ajustado.

### RESULTADOS

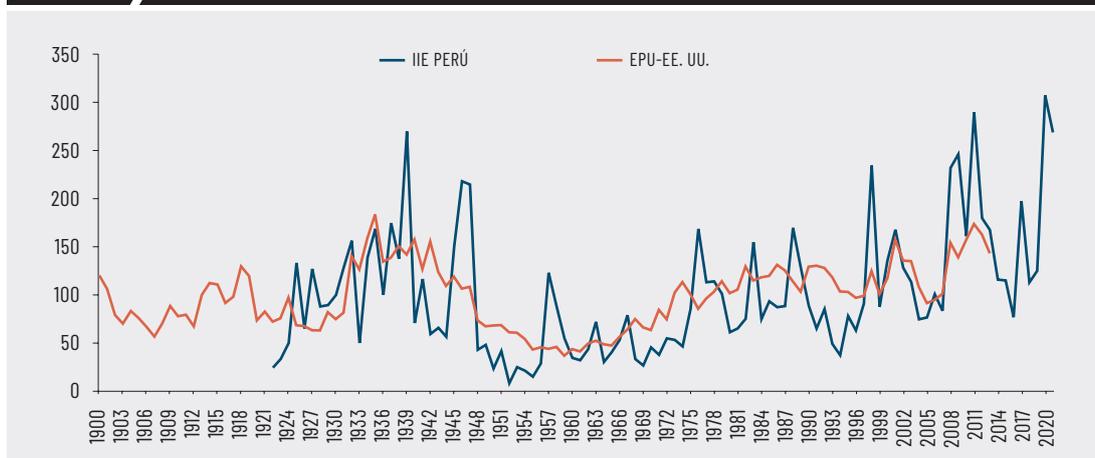
El Gráfico 1 (p. 8) presenta el  $IIE$  calculado para Perú de acuerdo con el método descrito líneas arriba y se compara con el Economic Policy Uncertainty (EPU) calculado por Baker et al. (2016). Se aprecian dos hechos saltantes. Primero, ambos



La medida de incertidumbre estimada tiene cierta semejanza con la medida de incertidumbre EPU estimada para EE. UU. **en cuanto a su comportamiento de largo plazo, pero tiene particularidades asociadas a la historia económica peruana en este centenario.**



**GRÁFICO 1** ■ Comparación de índices de incertidumbre IIE Perú vs. EPU EE. UU.



**FUENTE:** EL IIE-PERÚ ES CALCULADO CON EL MÉTODO EXPUESTO EN ESTE ESTUDIO. EL EPU-EE. UU. ES EXTRAÍDO DE BAKER ET AL. (2016). PARA GRAFICAR EL EPU, SE CALCULÓ EL PROMEDIO DE CADA AÑO DESDE 1900 HASTA 2013 PUESTO QUE EL ÍNDICE ESTÁ CONSTRUIDO A FRECUENCIA MENSUAL. PARA FINES DE COMPARACIÓN, LOS ÍNDICES ORIGINALES SE DIVIDIERON ENTRE EL PROMEDIO HISTÓRICO DE CADA ÍNDICE Y SE MULTIPLICARON POR 100.

índices alcanzaron valores mínimos en la década de los cincuenta del siglo pasado. Posterior a dicha década, se observó una tendencia creciente en el patrón de incertidumbre hasta los tiempos actuales. Es decir, la cantidad de incertidumbre tendió a aumentar desde dicha fecha, tanto en Perú como en EE. UU. Un segundo elemento resaltante de la comparación fue que el IIE de Perú es más volátil que el EPU de EE. UU. El IIE de Perú muestra picos —algo extremos— de incertidumbre cada cierto año.

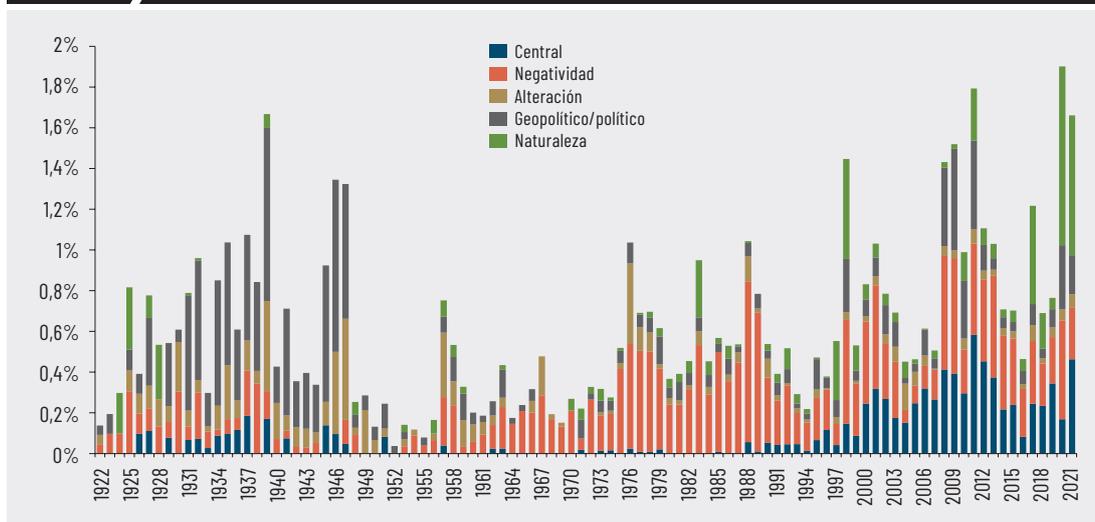
El Gráfico 2 presenta el IIE y la participación de cada tipo de incertidumbre definido en el Cuadro 1. Se puede observar algunos patrones interesantes. Por ejemplo, la incertidumbre central —después de un periodo de calma de cuatro décadas— empezó a subir fuertemente a partir de la década de los 90 y alcanzó su máximo en

2011. Sin embargo, el porcentaje de participación de esta incertidumbre alcanzó su pico en el periodo 2005-2007 (Gráfico 3), periodo asociado a una alta volatilidad de los mercados internacionales.

De acuerdo con los resultados, la incertidumbre asociada a negatividad fue muy importante entre las décadas del 60 y 80, importancia que disminuyó un poco a partir de 1990. En los 50 años que van desde 1947 hasta 1996, la incertidumbre asociada a negatividad explica el 54 por ciento del IIE (Gráfico 3) y, en algunos años, llega a explicar hasta más del 80 por ciento del IIE (años 1965, 1968, 1969, 1972, 1975, 1985, 1987, 1989).

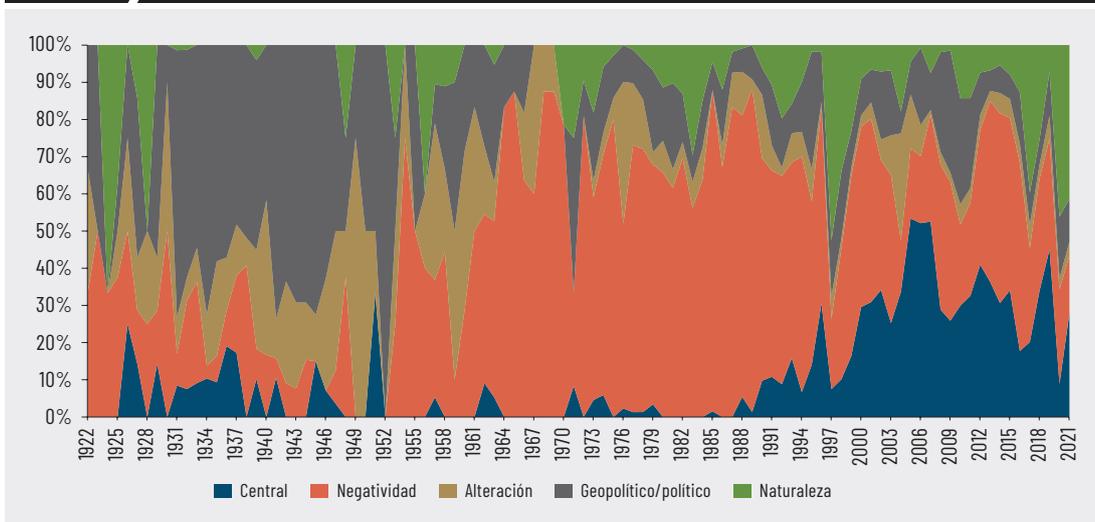
La incertidumbre asociada al concepto de alteración del estado normal de la economía tuvo una presencia fuerte en el periodo inicial; luego

**GRÁFICO 2** ■ Evolución de la IIE por componentes



ELABORACIÓN: PROPIA.

**GRÁFICO 3** ■ Participación porcentual de cada tipo de incertidumbre en el *IIE*



ELABORACIÓN: PROPIA.

pasa a tener una participación mediana; y finalmente pasar a ser un componente muy pequeño en los últimos 20 años. Este concepto está asociado a la existencia de desequilibrios, desajustes o alteraciones macroeconómicas que han sido reducidos al mínimo por la institucionalización de buenas prácticas de política macroeconómicas en las dos últimas décadas.

La incertidumbre geopolítica y política fue más importante hasta fines de la Segunda Guerra Mundial. El peso de esta incertidumbre como porcentaje de la *IIE* total también se puede apreciar en el Gráfico 3. En el periodo entre 1922 y 1946, la incertidumbre geopolítica y política explica el 48 por ciento del *IIE*.

Finalmente, la incertidumbre asociada a la naturaleza tuvo picos importantes en años específicos. Por ejemplo, en 1924 hubo una sequía extrema que afectó a la agricultura y en 1925 un Niño Costero extremo. También, en el período 82-83, se elevó este tipo de incertidumbre asociada al Fenómeno El Niño, y lo mismo ocurrió en 1997 y 2017. Los años 2020 y 2021 están marcados por la pandemia del COVID-19.

## CONCLUSIÓN

En este artículo se muestra que es posible capturar una medida de incertidumbre a partir del texto de las memorias del BCRP de estos últimos 100 años. La medida de incertidumbre estimada tiene cierta semejanza con la medida de incertidumbre EPU estimada para EE. UU. en cuanto a su comportamiento de largo plazo, pero tiene particularidades asociadas a la historia económica peruana en este centenario. Entre las especificidades del *IIE* de Perú se tiene, por ejemplo, la casi eliminación de la incertidumbre asociada a desequilibrios, desbalances o desajustes macroeconómicos en los últimos 20 años. También, la fuerte presencia de incertidumbre que denota negatividad macroeconómica en el periodo que va desde la década de 1950 hasta la de 1990 y, finalmente, el aumento de la incertidumbre asociada principalmente al concepto de volatilidad de mercados financieros a partir de 1990.

Cabe mencionar que la evolución del *IIE* de Perú por tipo de incertidumbre también permite conocer cómo ha ido cambiando la narrativa sobre la incertidumbre económica.

## REFERENCIAS:

- Ashok, S. & Mehram, S. (2009). *Text Mining: Classification, Clustering, and Applications*. Chapman and Hall/CRC.
- Baker, S., Bloom, N. & Davis, S. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593-1636. <https://doi.org/10.1093/qje/qjw024>
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623-685. <https://doi.org/10.3982/ECTA6248>
- Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J. F., & Uribe, M. (2011). Risk matters: The real effects of volatility shocks. *American Economic Review*, 101(6), 2530-61. <https://doi.org/10.1257/aer.101.6.2530>
- Justiniano, A. & Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, 98(3), 604-41. <https://doi.org/10.1257/aer.98.3.604>
- Manela, A. & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1), 137-162. <https://doi.org/10.1016/j.jfineco.2016.01.032>
- Martín, F. & Ruiz, J. (2013). *Procesamiento del lenguaje natural*. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad de Sevilla.