

# Análisis exploratorio de las minutas del banco central de Inglaterra y su relación con la tasa de interés; 1997 al 2014.

Miranda Belmomte Hairo Ulises

**Presentación:** Ciencia de datos

11 de Junio del 2019

# Estructura

- 1 Introducción
- 2 Comunicados Banco Centrales de Inglaterra
  - Banco Centrales
  - Banco de Inglaterra
- 3 Datos
  - Minutas
  - Tasa de interés
- 4 Metodología empírica
  - Pre-procesamiento
  - Matriz de Documentos-Términos
  - TF-IDF
  - NMF; método multiplicativo (MM)
  - NMF y Text mining
- 5 Descriptivos
- 6 Resultados

# Introducción

- Mercados financieros analizan los comunicado del banco central para descifrar el significado de su postura
- El significado de las minutas varía dependiendo la condición económica
- Detectar temas en los comunicado deja en claro a los inversionistas las intenciones en sus ajustes

# Introducción

- Crayton, A. (2018) métodos más utilizados supuestos distribucionales
- Problemas de estos métodos es la estabilidad de los resultados
- Repetir varias veces el algoritmo genera distintos resultados

# Introducción

- Se basa en Crayton, A. (2018) y Moniz, A. y de Jong, F. (2011)
- Se utiliza Crayton utiliza modelos de factorización de tópicos,
- Crear factores de una matriz de documentos-terminos (pesada)

# Ventajas

- Estables, el resultado es el mismo para las veces que se estime el modelo
- O'Callaghan et al., (2015), generan tópicos más coherentes que métodos probabilistas.

# Objetivo

Resumir la información de los comunicados en tópicos y observar como cambios en la presencia de distintos temas influyen en las expectativas de la tasa de interés.

# Banco Centrales

- Los comunicados son conocidos como minutas
- Contienen decisiones del comité respecto al ajuste de la tasa de interés.
- Temas: crecimiento y estabilización financiera e inflación



# Banco de Inglaterra

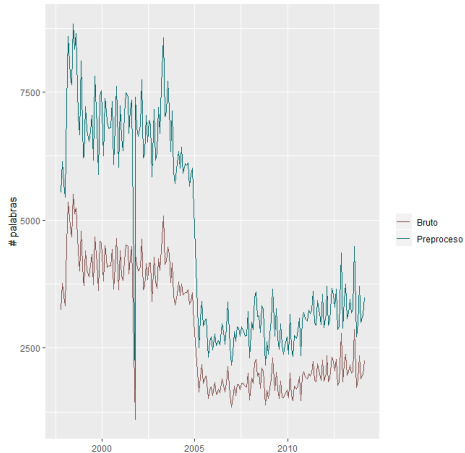
Los comunicados se publica ocho veces por año.

- Pre-MPC reunión: revisan datos de la economía
- Primera reunión: discuten lo encontrado de la reunión previa
- Segunda reunión: debaten que acción de política monetaria deben tomar
- Reunión final: Los miembros votan sobre la decisión
- Anuncio de la decisión: publican la decisión por medio de las minutas y explican la razón de la decisión en términos del contexto económico y su dirección.

# Minutas

- Minutas del banco central de Inglaterra fechas de Octubre de 1997 a Marzo del 2014
- Corpus de 195 documentos.

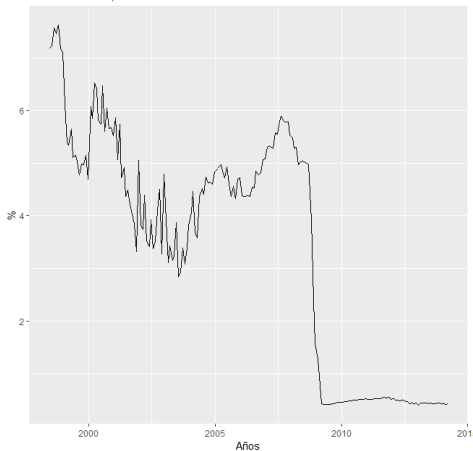
# Tamaño de Minutas corpus en bruto y pre-procesado



# Tasa de interés

- Tasa de interés *Sterling Overnight Interbank Average Rate (SONIA)*
- Se reporta a manera diaria.
- Gran liquidez y su gran relación respecto al tipo de cambio.

Tasa de Interés; 1998-06-06 al 2014-03-19



## Pre-procesamiento

- Se le retira la página inicial
- Los nombres de los integrantes del comité
- Palabras que no aporten al análisis del documento
- *stopwords*
- Espacios en blanco, puntuaciones, acentos, números y símbolos
- Cambia letras del corpus a minúsculas
- Se lematiza la palabra como proceso de estandarización del texto ( plurales a singular y verbos en pasado a presente)

# Matriz de Documentos-Términos

- Representación de texto en forma matricial
- *Bag of words*
- Matriz de tamaño  $n \times m$  con  $n$  documentos y  $m$  términos.
- Entrada  $c_{nm}$  de esta matriz conteo de los términos en cada documento

# Term frequency-inverse document frequency (TF-IDF)

Frecuencia de los terminos (tf), números de veces un termino aparece en el documento.

$$tf(n, m) = 1 + \log(c_{nm}) \quad (1)$$

La frecuencia inversa del documento (idf), número total de distintos documentos en los que aparece la palabra, penaliza términos que aparecen en un número largo de documentos.

$$w(n, m) = tf(n, m) \times \left( \log \frac{D}{df(m)} + 1 \right) \quad (2)$$

- En TF-IDF términos que aparecen con frecuencia en número pequeño de documentos tendrán mucho peso
- Capturar términos de importancia a específicos documentos y esenciales en su respectivo tópico.



## NMF; método multiplicativo (MM)

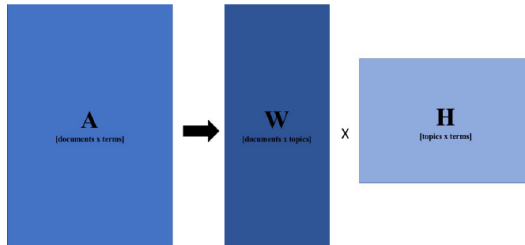


Figura: NMF

## NMF; método multiplicativo (MM)

$$\min ||A - WH||_F^2 \quad (3)$$

## Algoritmo del método multiplicativo (MM):

---

### Algorithm 1 NMF-MM MSE

---

```
1: Inicializa  $\mathbf{W}$ ,  $\mathbf{H}$ 
2: for  $i = 1 : \text{maxiter}$  do
3:    $\mathbf{H}_{\alpha\mu} \leftarrow \mathbf{H}_{\alpha\mu} \frac{\mathbf{W}'\mathbf{v}_{\alpha\mu}}{\mathbf{W}'\mathbf{H}\mathbf{H}_{\alpha\mu}}$ 
4:    $\mathbf{W}_{i\alpha} \leftarrow \mathbf{W}_{i\alpha} \frac{\mathbf{v}\mathbf{H}'_{i\alpha}}{\mathbf{W}\mathbf{H}\mathbf{H}'_{i\alpha}}$ 
5: end for
6:  $\mathbf{W}$ ,  $\mathbf{H}$ 
```

---

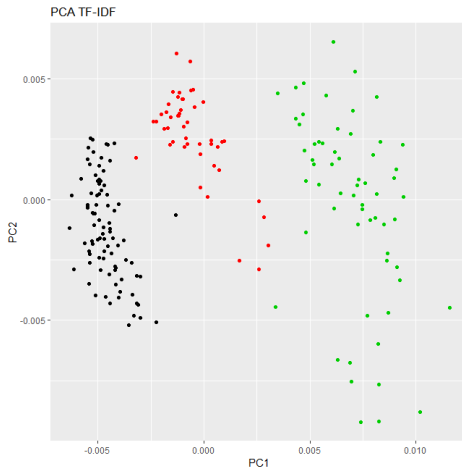
# NMF y Text mining

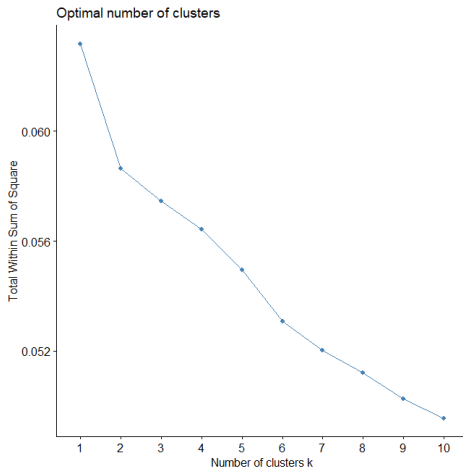
- Factorización no negativa para modelar tópicos
- Matriz base  $W$  los tópicos  $k$  en el documento  $n$
- Matriz de pesos  $H$  los pesos de los términos  $m$  en el tópico  $k$ .

## Características

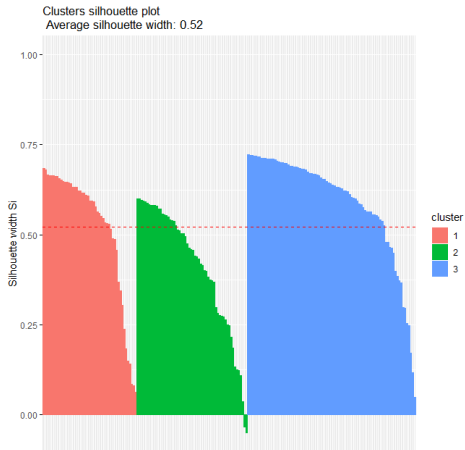
- Tópicos en  $W$  no son ortogonales pueden traslaparse
- Generación de *clustering*
- Facilidad en interpretación
- Factorización no es única
- Convergencia puede llegar a un mínimo local y no global

# PCA matriz TF-IDF





# Valores silhouette

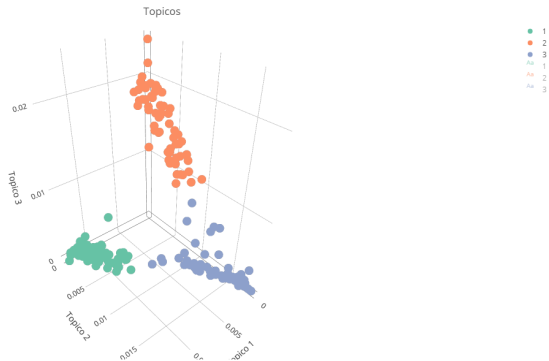




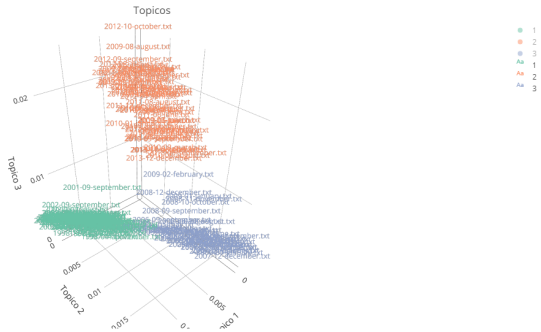
## Top 10 distintos tópicos

W1 K=3	W2 K=3	W3 K=3	W1 K=4	W2 K=4	W3 K=4	W4 K=4	W1 K=5	W2 K=5	W3 K=5	W4 K=5	W5 K=5
repo	gas	programm	gas	rpix	programm	repo	consolid	gas	programm	repo	rpix
rpix	cpi	stimulus	cpi	fell	stimulus	cip	slack	cpi	libor	cip	fell
cip	default	consolid	default	shortag	consolid	percentag	sovereign	default	scale	percentag	shortag
percentag	deterior	issuanc	slack	rose	issuanc	annual	ben	preview	asset	annual	rose
annual	slowdown	recess	upsid	skill	recess	securitis	programm	upsid	recess	securitis	skill
pnfcs	home	scheme	preview	bcc	scheme	rpix	judgement	slowdown	scheme	rpix	bcc
deflat	loss	asset	pose	budget	asset	pnfcs	issuanc	decad	stimulus	pnfcs	budget
securitis	upsid	precrisi	soften	cut	precrisi	ict	stimulus	home	commerci	ict	payment
bcc	preview	slack	home	billion	sovereign	wholeeconomi	peripheri	explan	sheet	wholeeconomi	cip
shortag	liquid	sovereign	slowdown	payment	slack	deflat	abovetarget	loss	billion	deflat	billion

# Tópicos



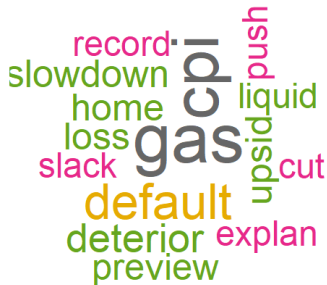
# Tópicos



# Tópico 1



## Tópico 2

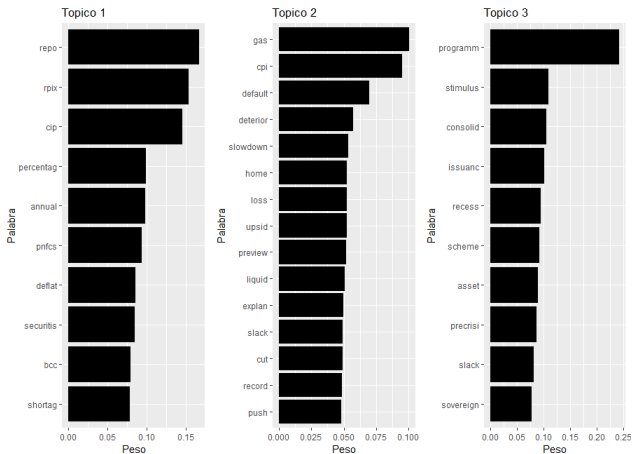


## Tópico 3

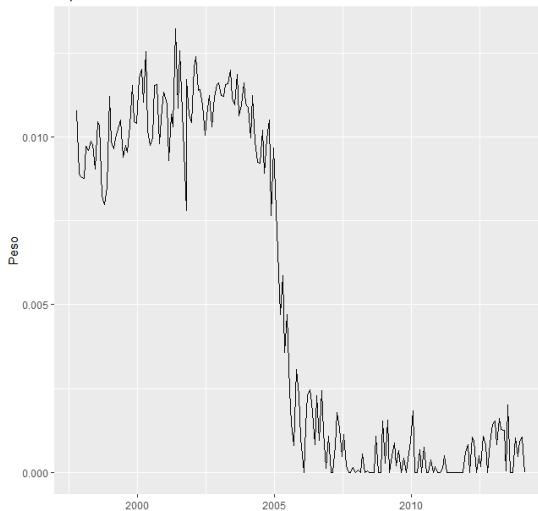
A word cloud visualization of financial terms. The word 'programm' is the largest and most central. Other prominent words include 'stimulus', 'recess', 'scheme', 'asset', 'billion', 'fund', 'precrisi', 'consolid', 'issuanc', 'ben', 'pmis', 'slack', 'crisi', and 'sovereign'. The words are arranged in a circular pattern around the central 'programm'.

billion fund precrisi  
scheme recess  
asset stimulus  
programm  
consolid ben  
issuanc pmis  
slack crisi  
sovereign

## Top 10 de términos en los tópicos

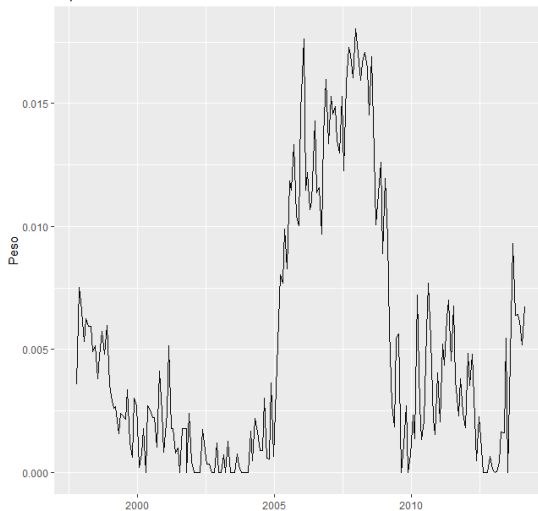


Topico 1

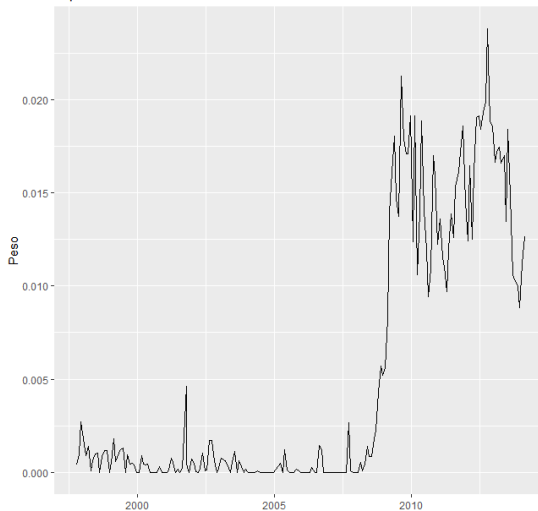


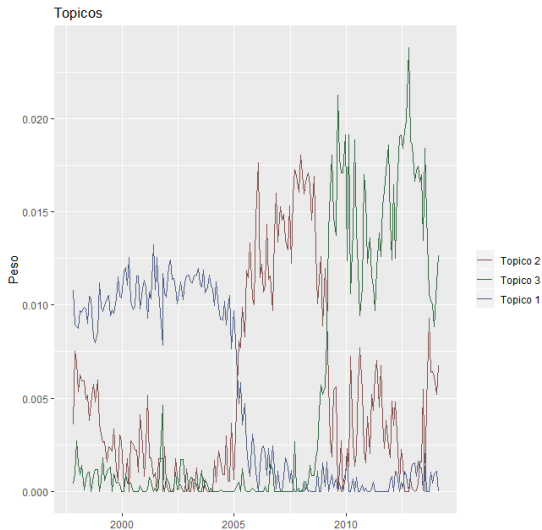


Topico 2

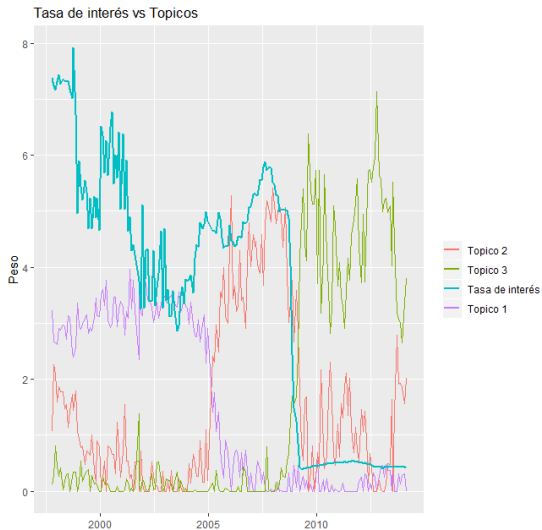


Topico 3





- **Tópico 1:** Objetivos Intermedios y periodo pre-crisis
- **Tópico 2:** Mercados Financieros y Crisis Financieras
- **Tópico 3:** Condiciones y crecimiento económico y periodo post-crisis



## Tasa de interés vs Tópicos

1998-2004

2004-2009



2009-2014

1998-2014

## Conclusión

- A trabajos a futuros se pretende utilizar análisis de polaridad para detectar sentimientos positivos y negativos de los comunicados del banco central
- Hacer uso de resultados para la clasificación en el efecto que tienen en las expectativas de los inversionistas y los cambios en la tasa de interés y clasificar
- Aplicar esto para México

## Referencias

- Crayton Ancil (2018). Central Bank Communication and the Yield Curve: A Semi-Automatic Approach using Non-Negative Matrix Factorization
- Moniz1, A. y de Jong, F. (2011). Predicting the impact of central bank communications on financial market investors' interest rate expectations. Springer-Verlag Berlin Heidelberg
- O'Callaghan, D., Greene, D., Carthy, J., y Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. Expert Systems with Applications.