

Culturomics: una aplicación al caso colombiano

Juan Manuel Caicedo

Alejandro Gaviria

Javier Moreno

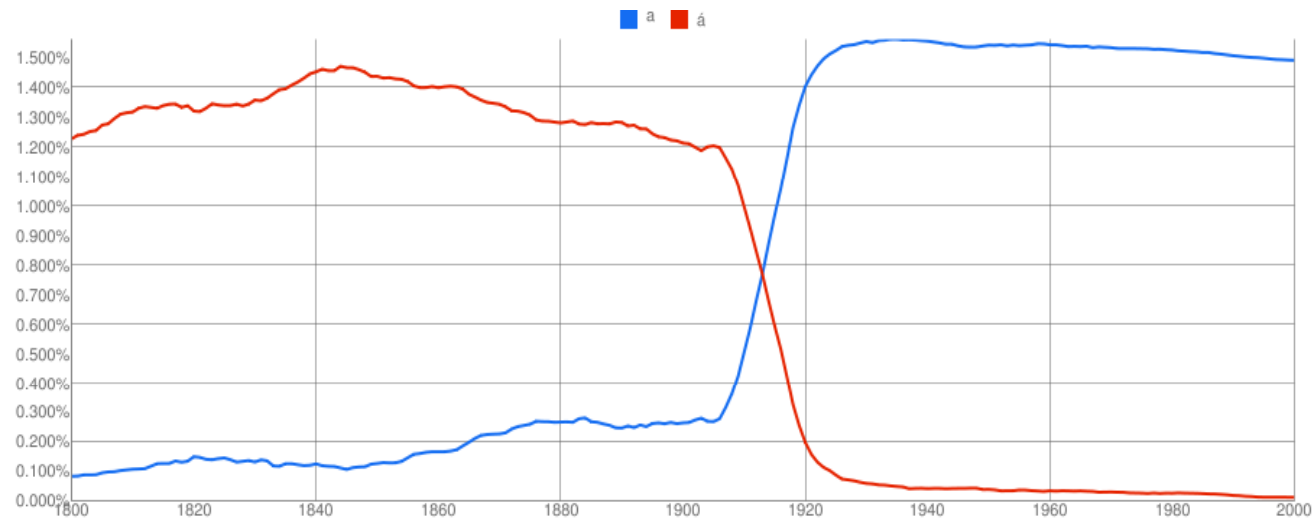
Agenda

- Cultoromics: definición y antecedentes
- Revisión de la literatura
- Metodología
- Ejemplos:
 - Realidad y percepción
 - Medir lo inmensurable

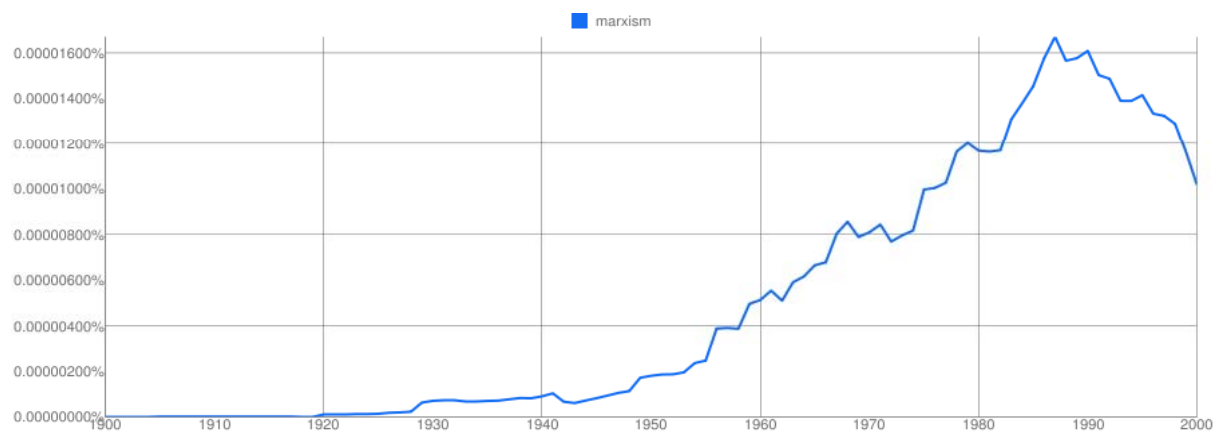
Antecedentes:

- *Culturomics* es el análisis cuantitativo de tendencias lingüísticas y culturales con base en libros, periódicos y otros textos virtuales disponibles en internet.
- Término acuñado por Michel et al. (2010) en artículo publicado en Science.
 - el análisis cuantitativo de textos virtuales constituye un nuevo (y en principio poderoso) método de análisis en las ciencias sociales.

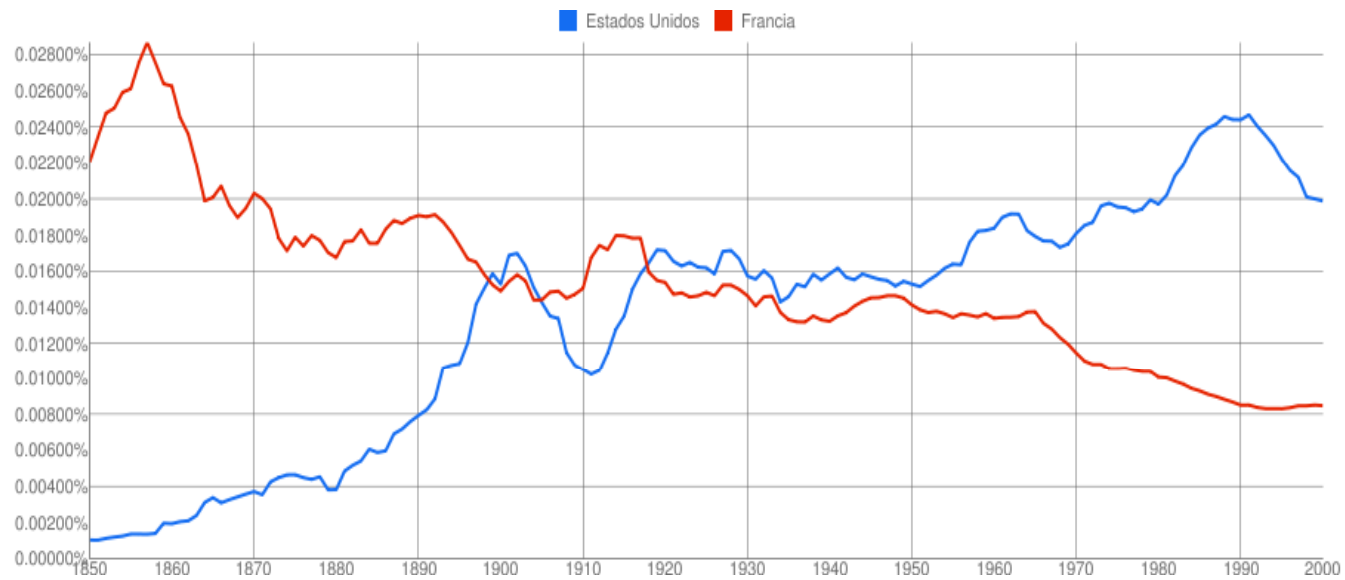
La evolución de la ortografía: “a” vs. “á”



La evolución de la ideología: “marxismo” en español e inglés



Francia vs. Estados Unidos



Revisión de la literatura (i):

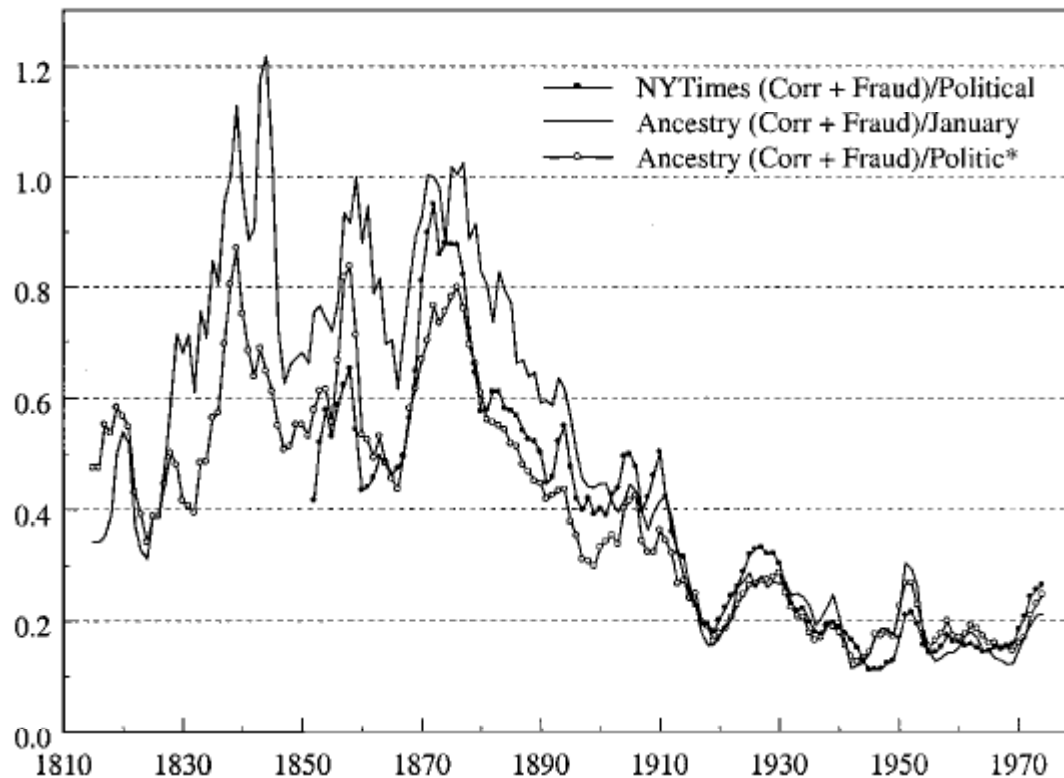
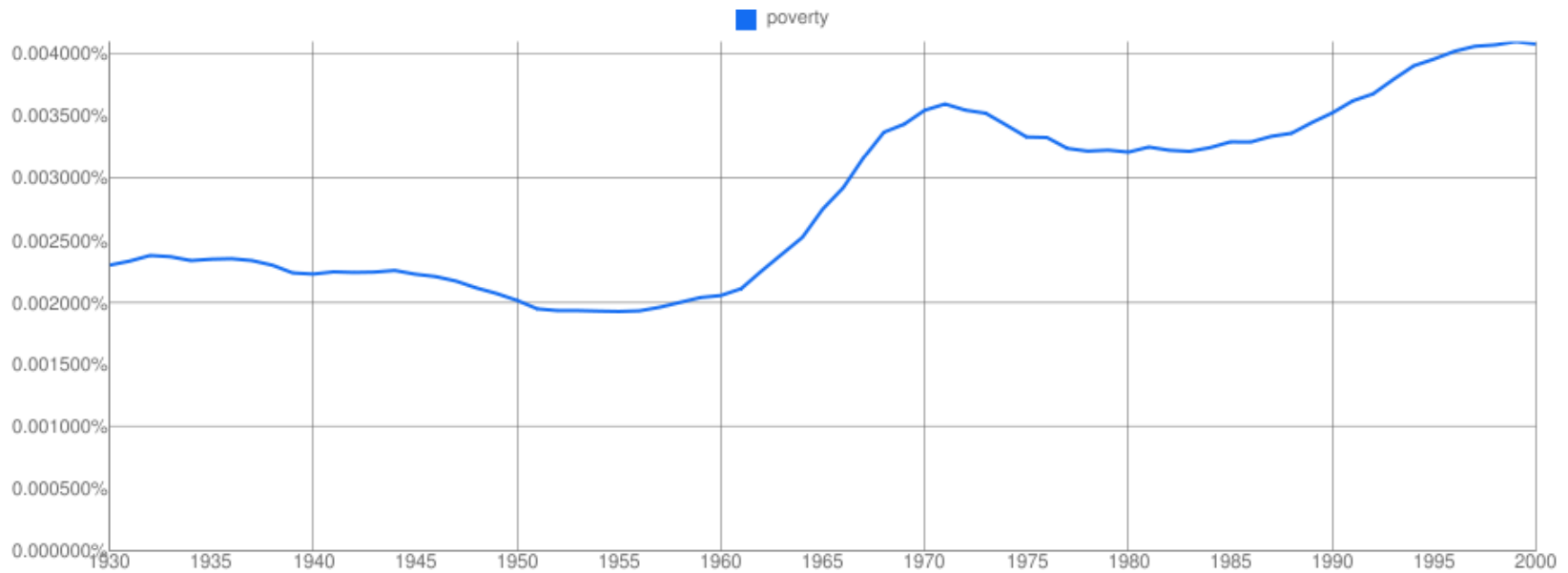


Fig. 1 Indexes of corruption and fraud: 1815–1975

Sources: NYTimes: online, fully searchable edition of the *New York Times* from ProQuest Information and Learning Company, Ann Arbor, Michigan. Ancestry: Ancestry.com, part of MyFamily.com, Inc., Provo, Utah.

Fuente: Glaeser y Goldin (2002)

Revisión de la literatura (ii):

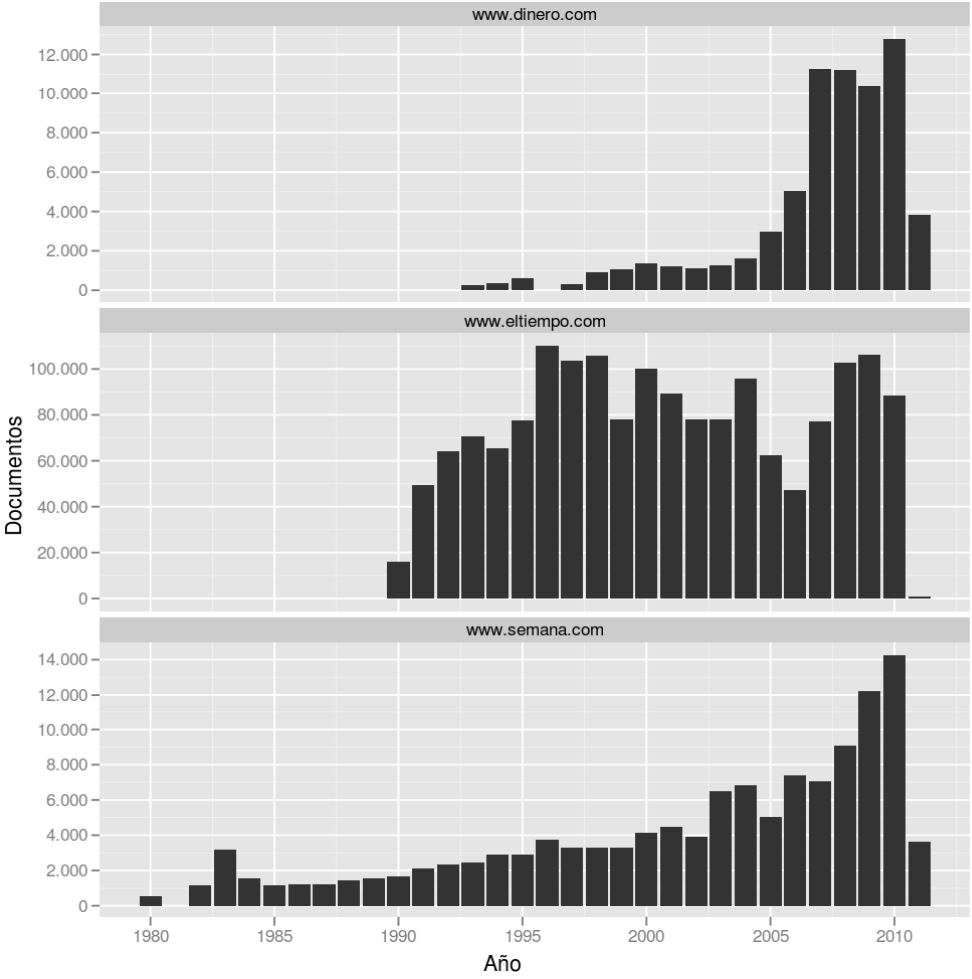


Fuente: Ravallion (2011)

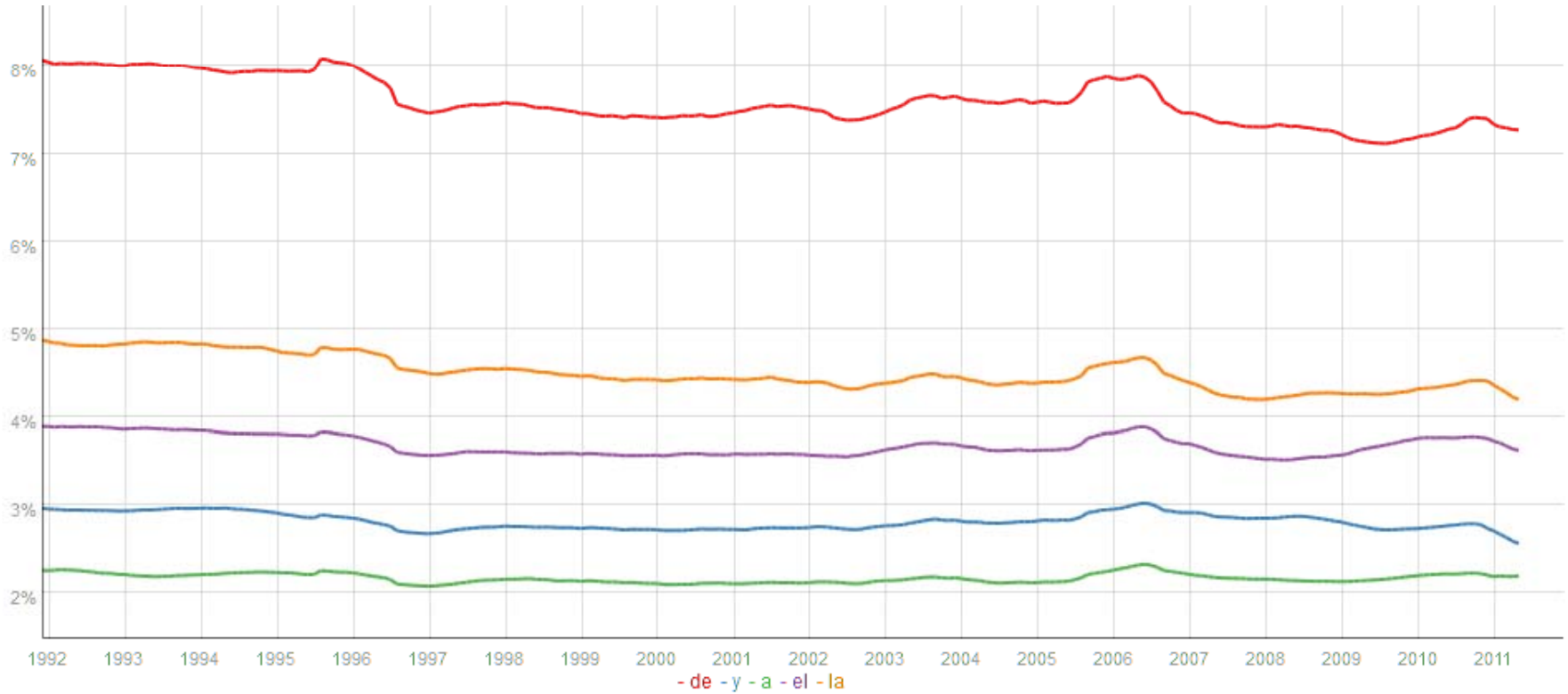
Metodología

- Corpus de varios millones de artículos publicados en El Tiempo, Semana y Dinero.
- Cálculo de frecuencias:
 - Cociente entre (i) el número de ocurrencias de un *n-grama* en todos los artículos publicados durante un mes dado en una de las tres publicaciones analizadas y (ii) el número total de *1-gramas* publicados durante el mismo mes en la misma publicación.

Cantidad de artículos por publicación

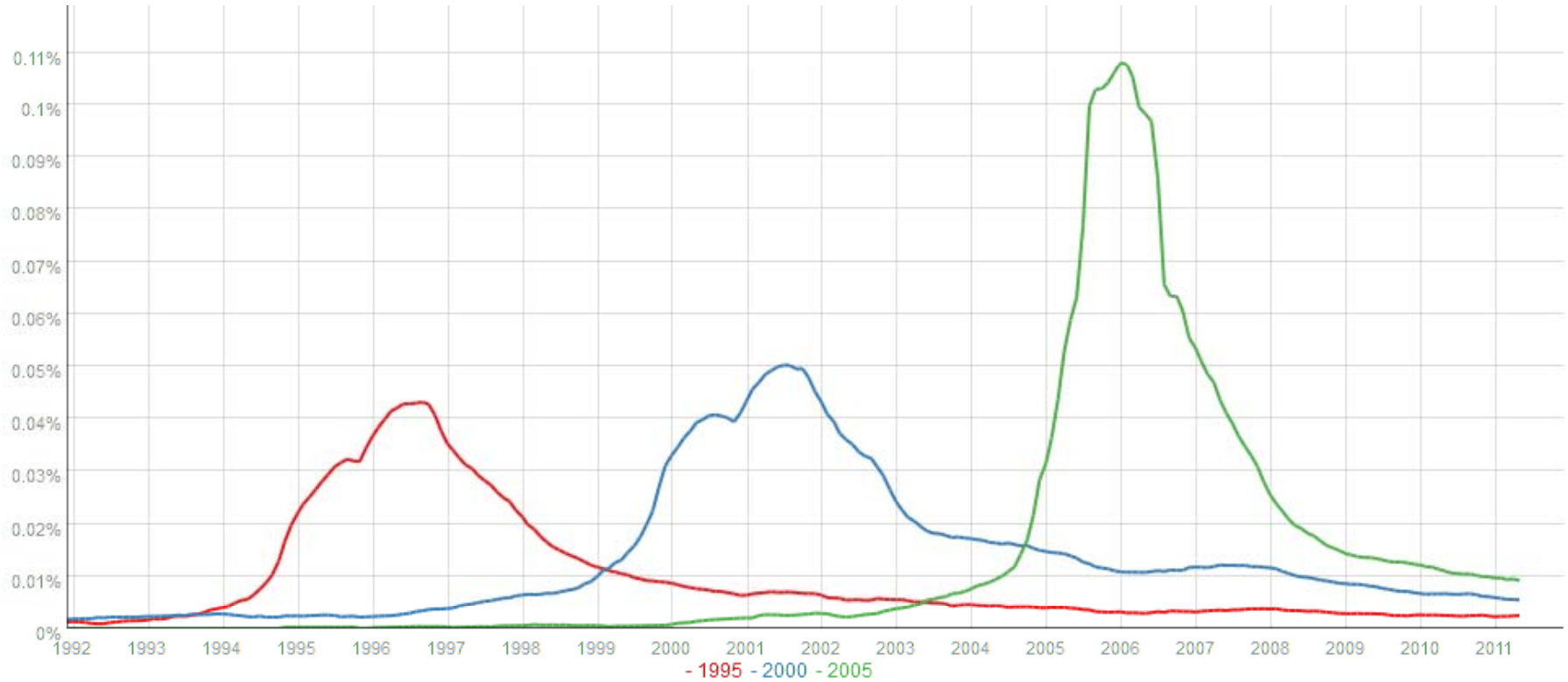


Prueba (1):



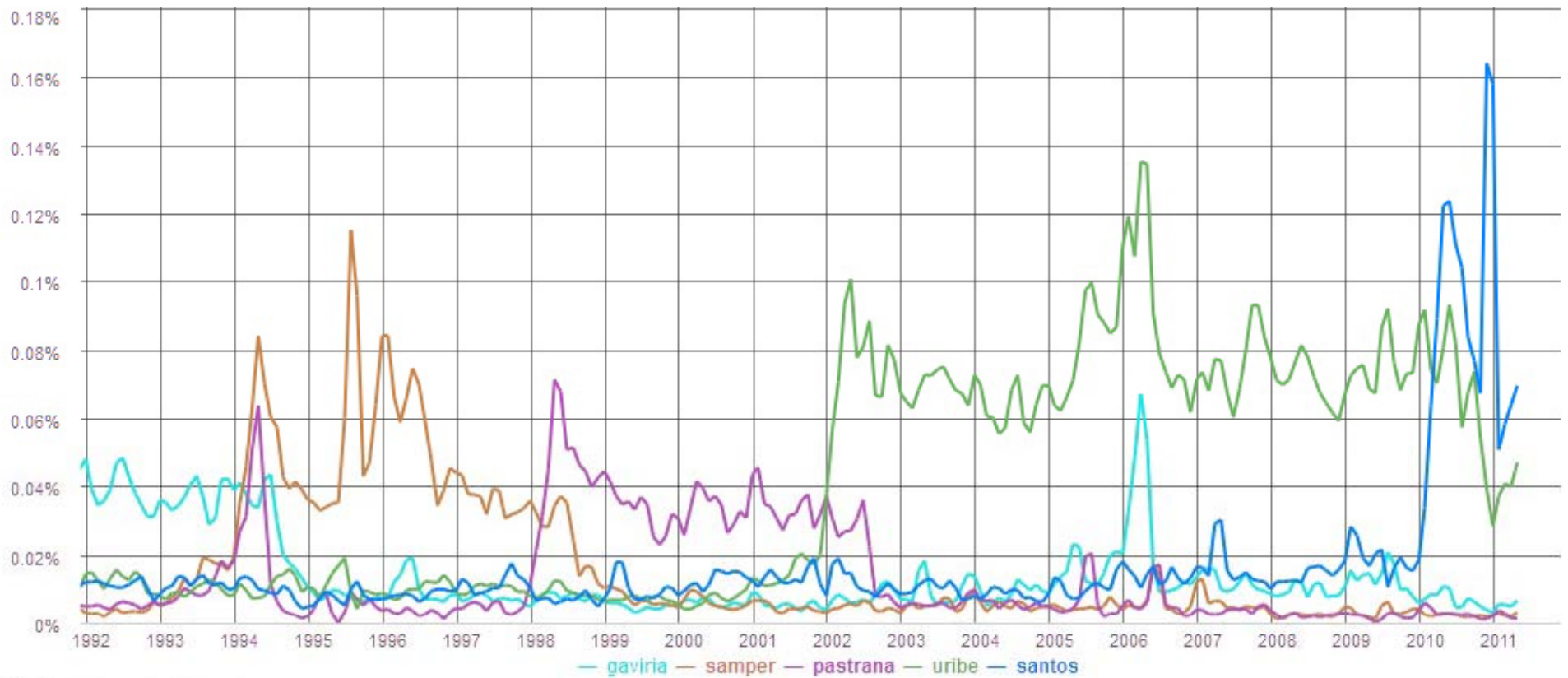
<http://ngrams.cavorite.com/>

Prueba (2):



<http://ngrams.cavorite.com/>

Ejemplo (1):

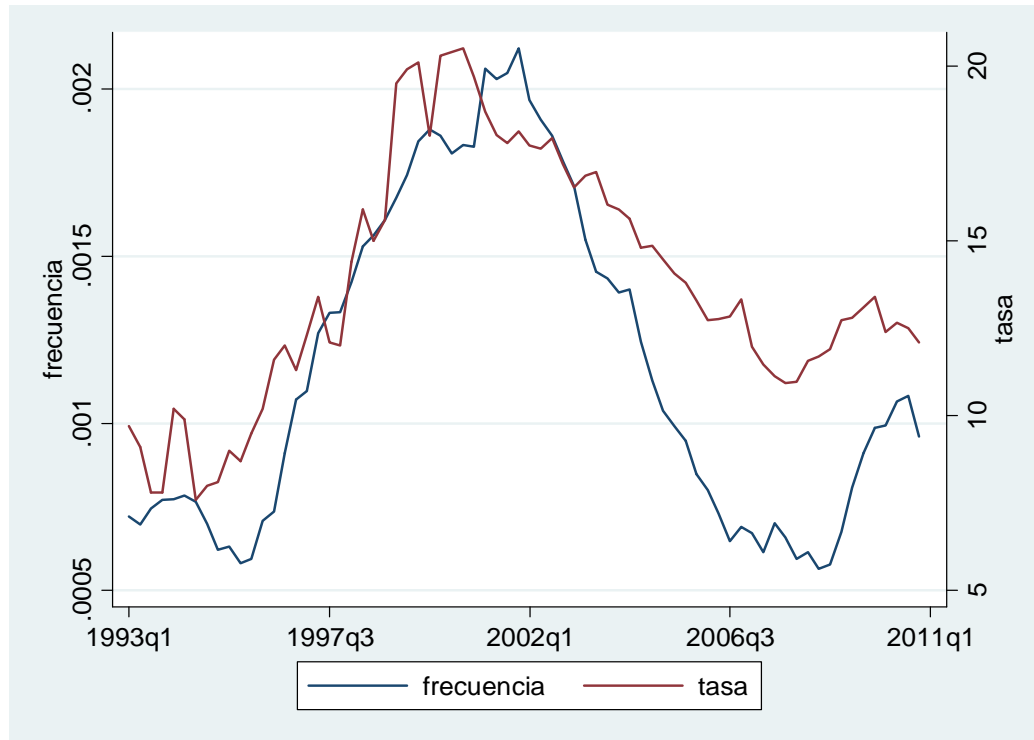


<http://ngrams.cavorite.com/>

<https://ngrams.cavorite.com/extra/>

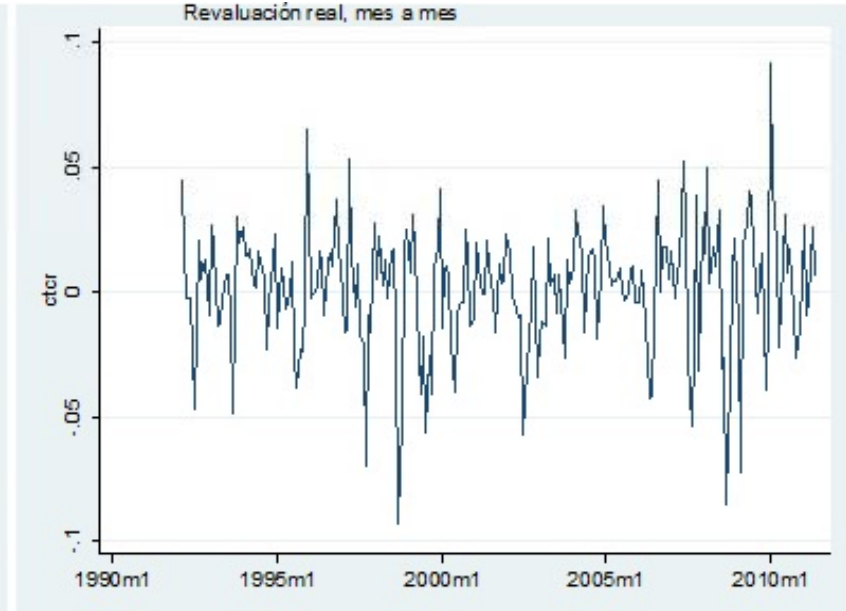
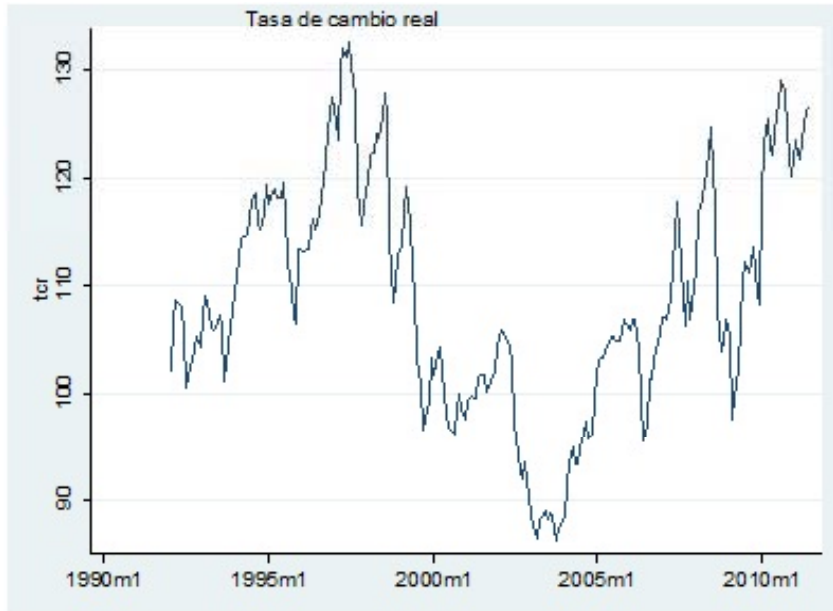
Aplicaciones

Tasa de desempleo y frecuencia de “desempleo”

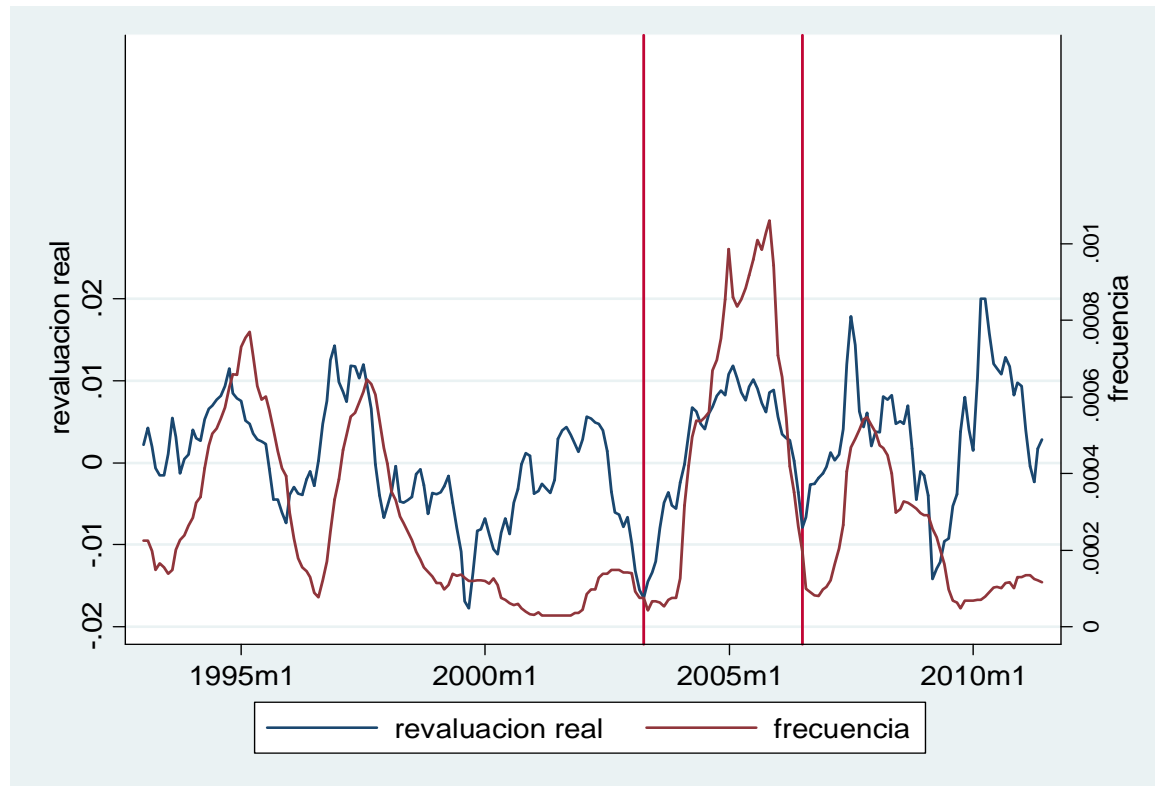


Pwcorr = 0,8897

Revaluación real



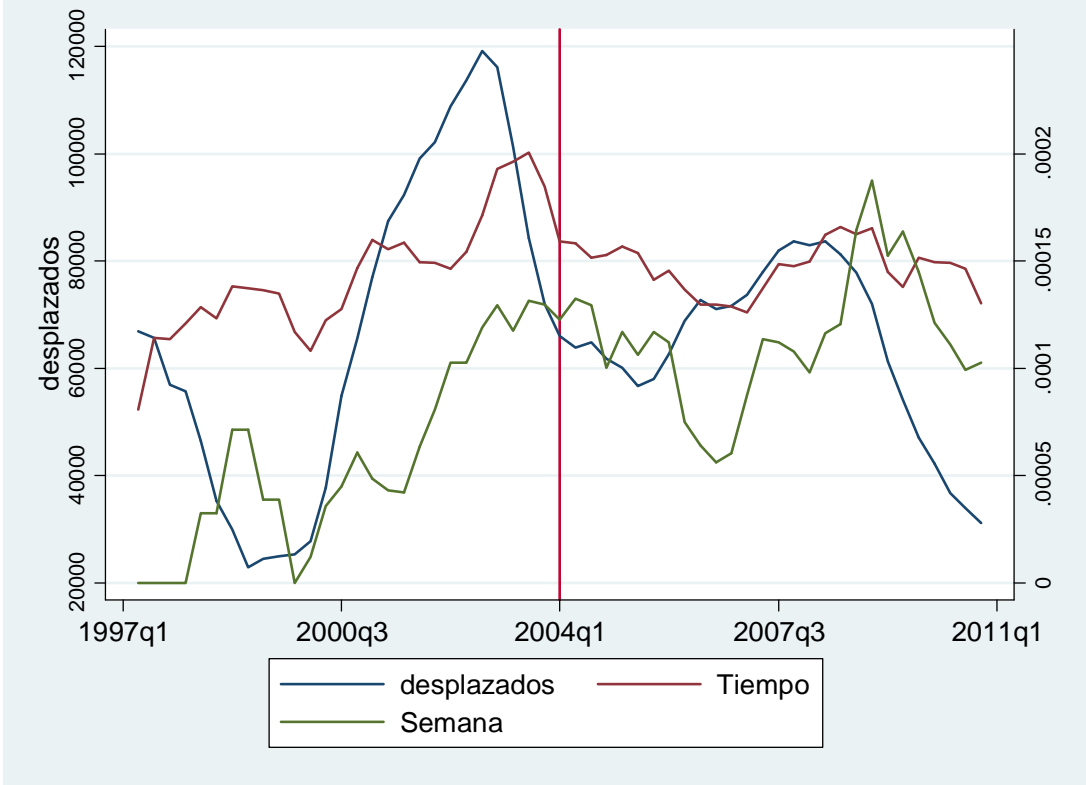
Revaluación y “revaluación”



Cambio en la correlación entre revaluación y “revaluación”:

	1993-2010	1993-2003	2003-2006	2006-2010
Revaluación real	0.4716	0.5016	0.9049	0.1063
Devaluación nominal	-0.4092	-0.3754	-0.8345	-0.1804

Desplazados en la calle y en la prensa



Correlaciones:

. pwcorr despl aza Ti empo Semana

	despl aza	Ti empo	Semana
despl aza	1.0000		
Ti empo	0.5254	1.0000	
Semana	0.3090	0.6928	1.0000

. pwcorr despl aza Ti empo Semana i f num<30

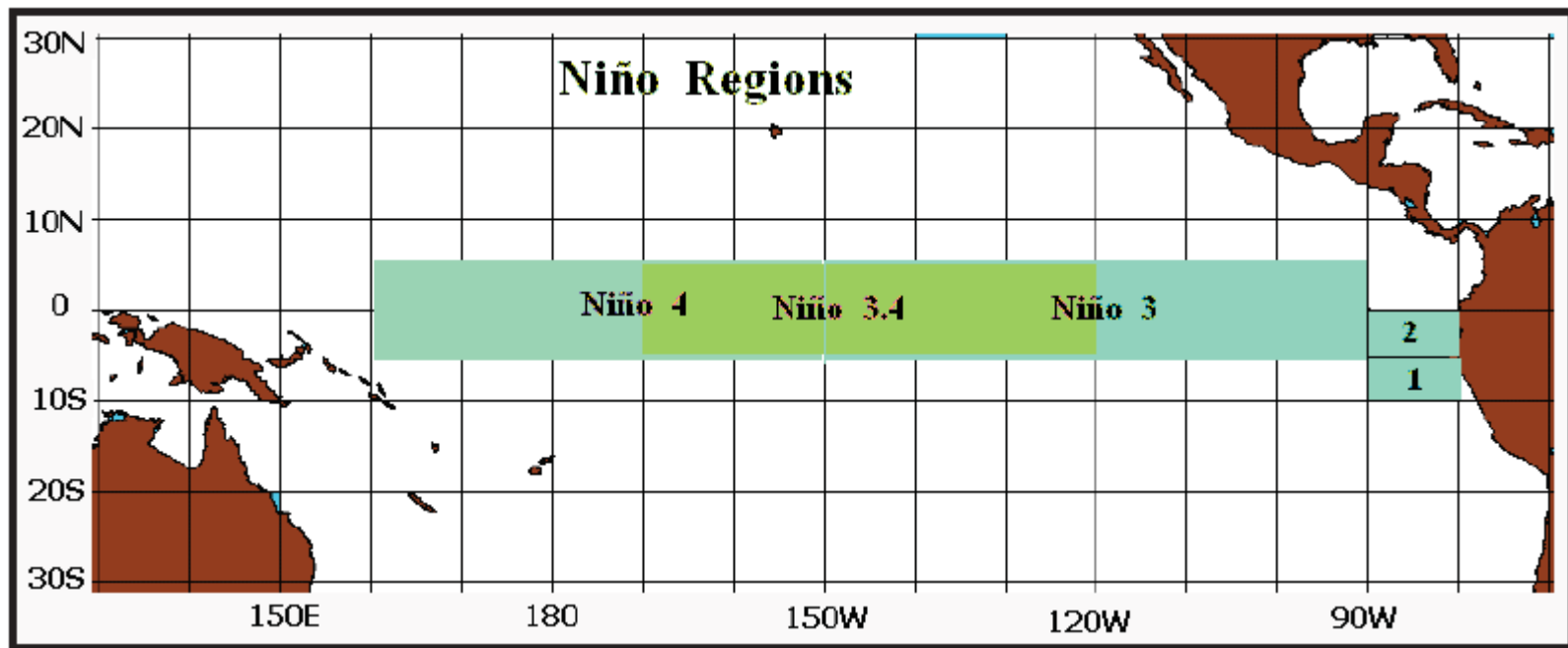
	despl aza	Ti empo	Semana
despl aza	1.0000		
Ti empo	0.6068	1.0000	
Semana	0.5944	0.8753	1.0000

. pwcorr despl aza Ti empo Semana i f num>29

	despl aza	Ti empo	Semana
despl aza	1.0000		
Ti empo	0.1488	1.0000	
Semana	-0.0818	0.6401	1.0000

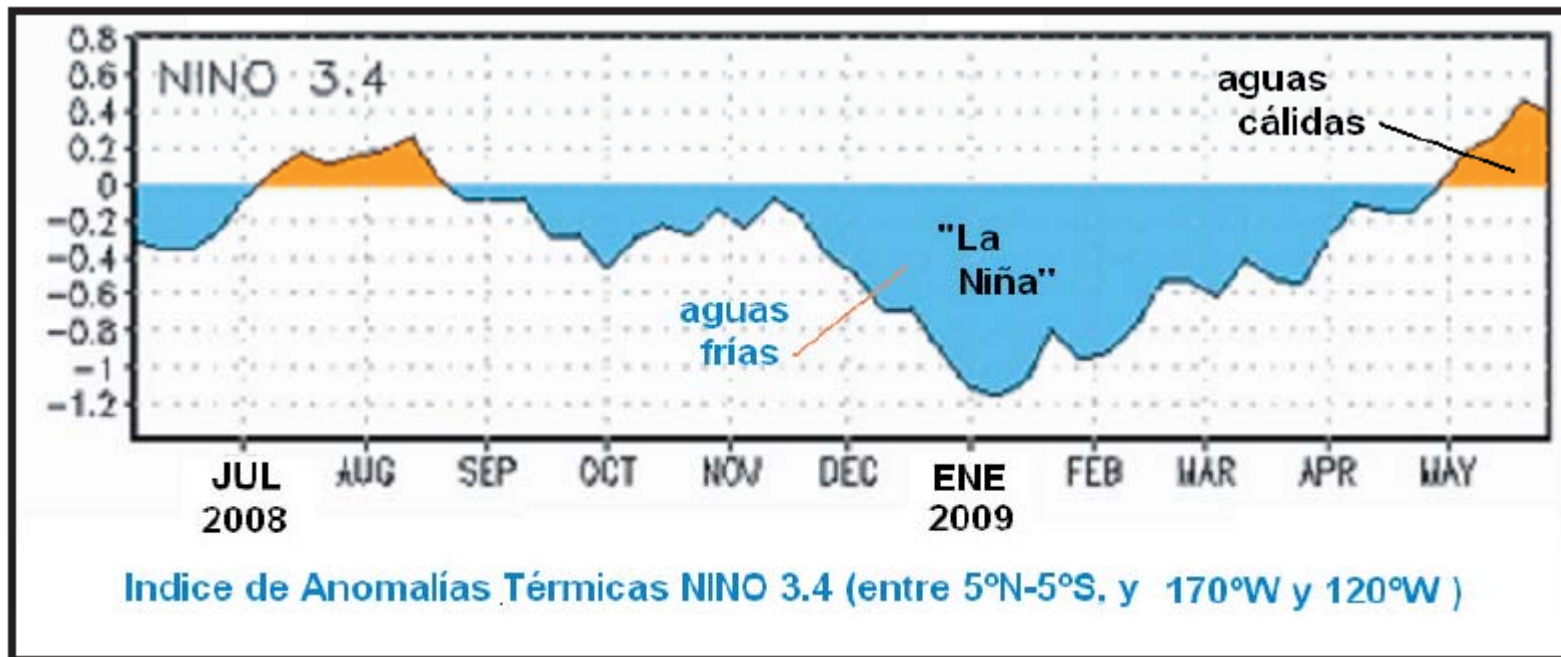
La Niña y el Niño en la prensa:

Un seguimiento del Índice Niño 3.4

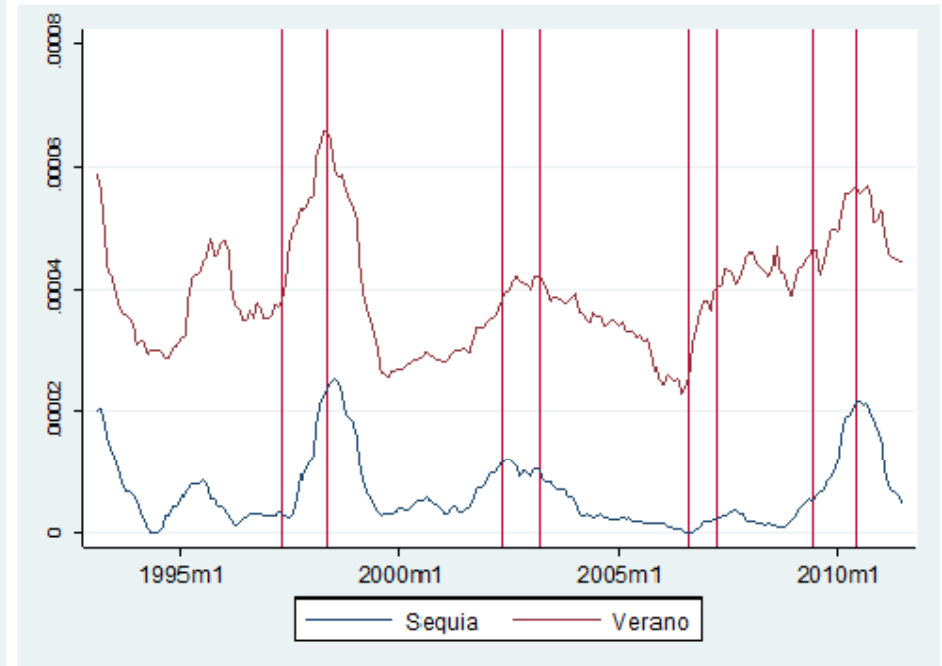
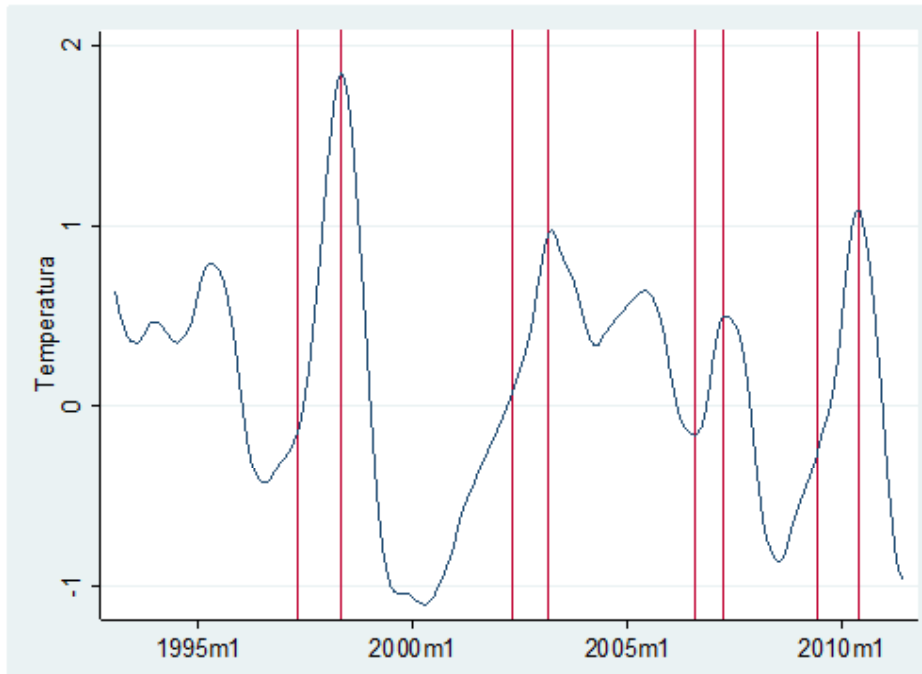


La Región Niño 3.4, se encuentra en el Océano Pacífico Ecuatorial.
Comprendida entre la Latitud 5°N y 5° Sur,
y Longitud Oeste, entre el meridiano 170° y el 120°.

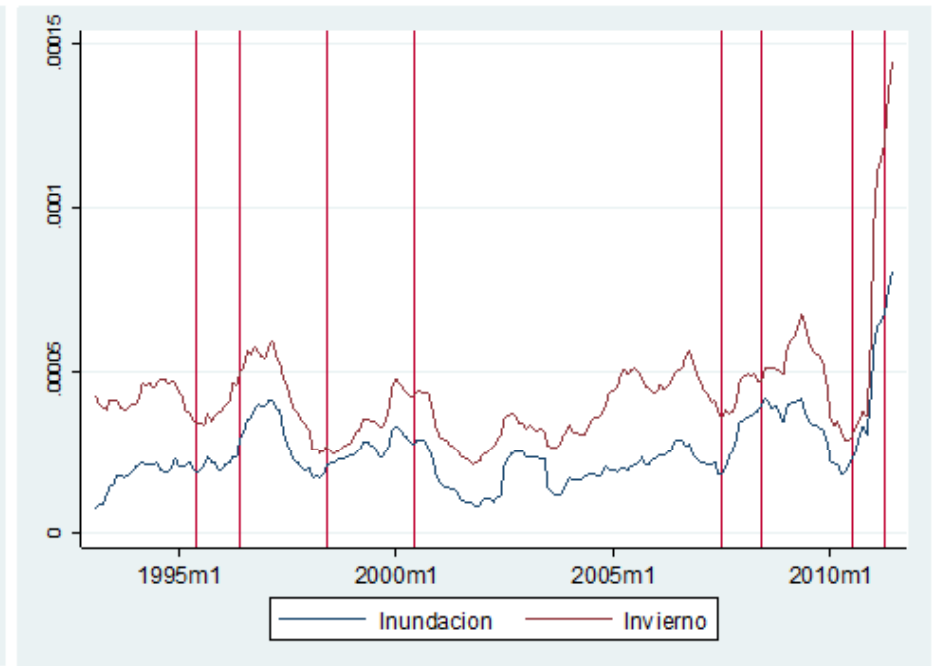
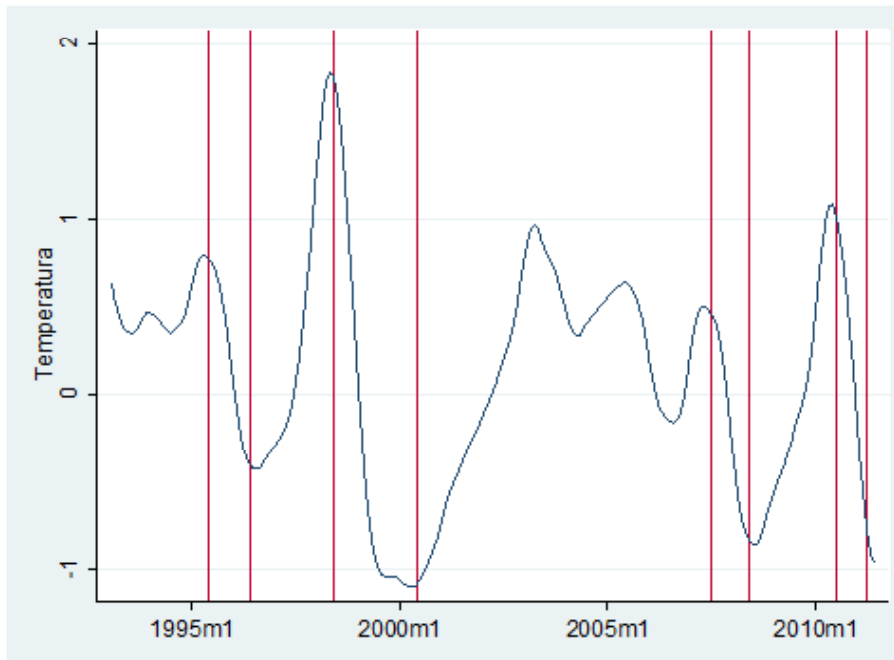
La Niña y el Niño en la prensa:



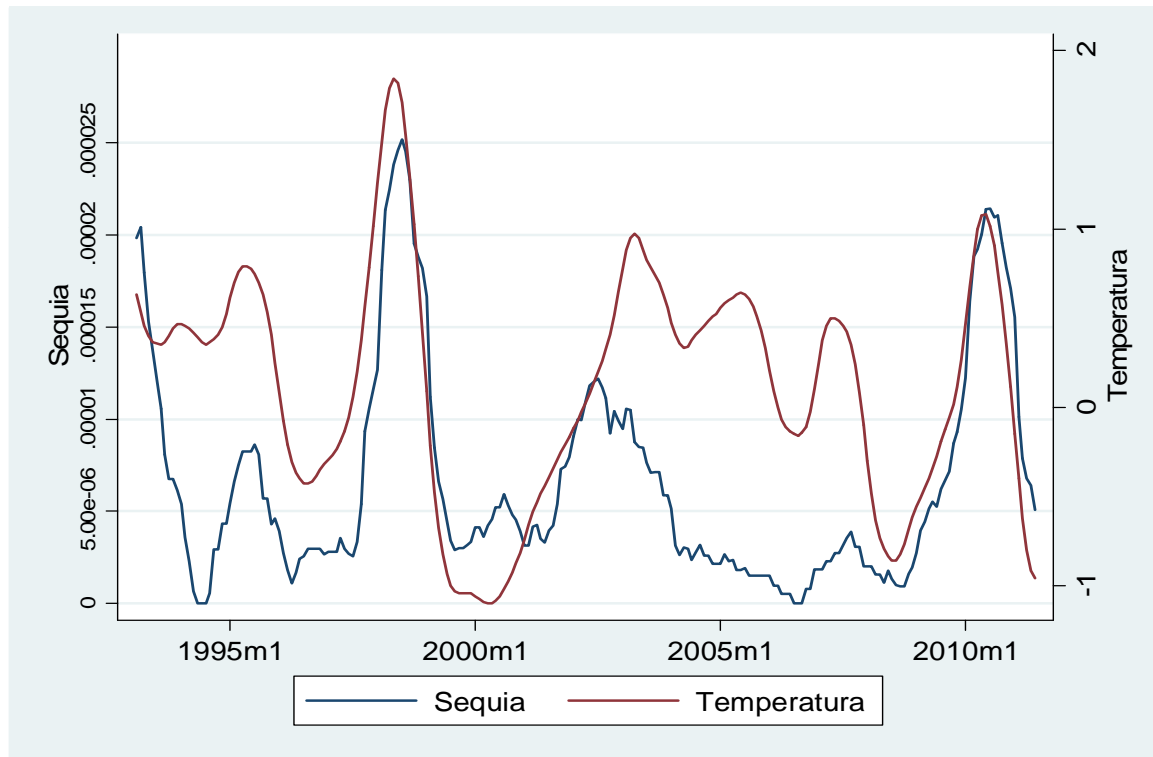
Niño vs. “verano” y “sequía”



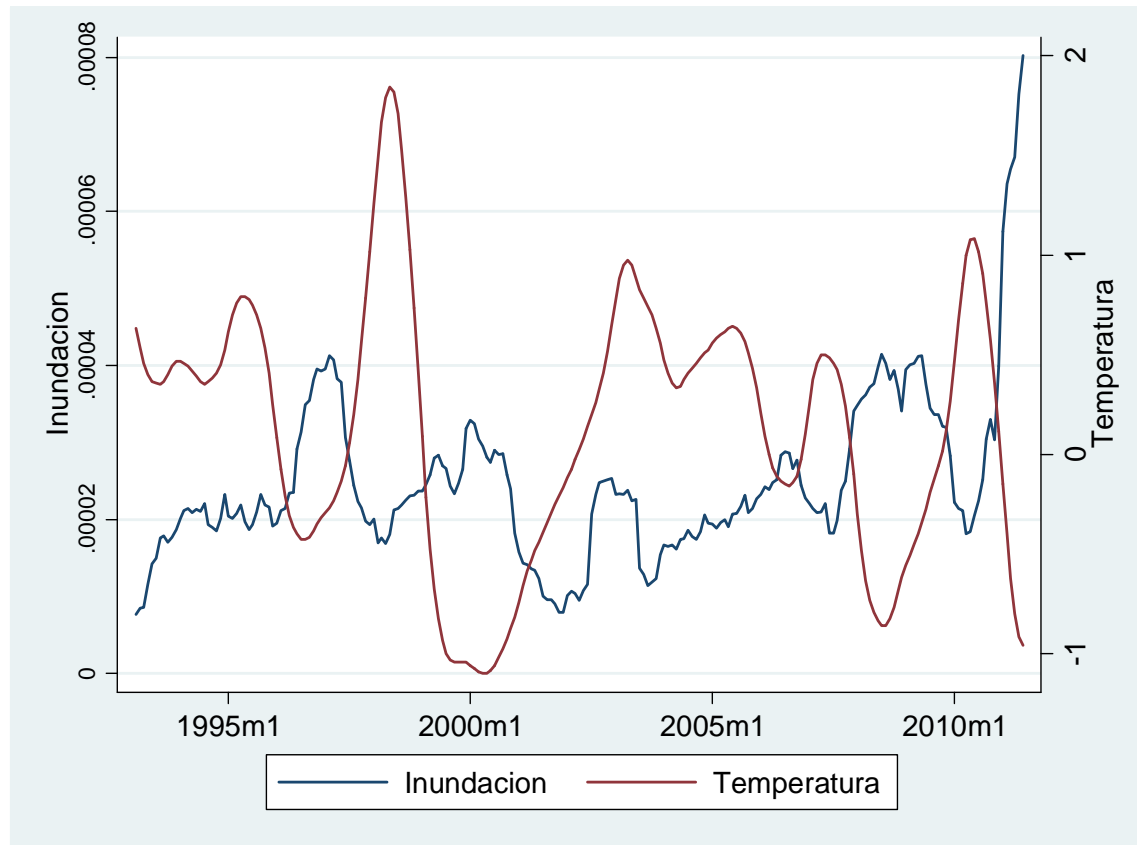
Niña vs. “invierno” e “inundación”



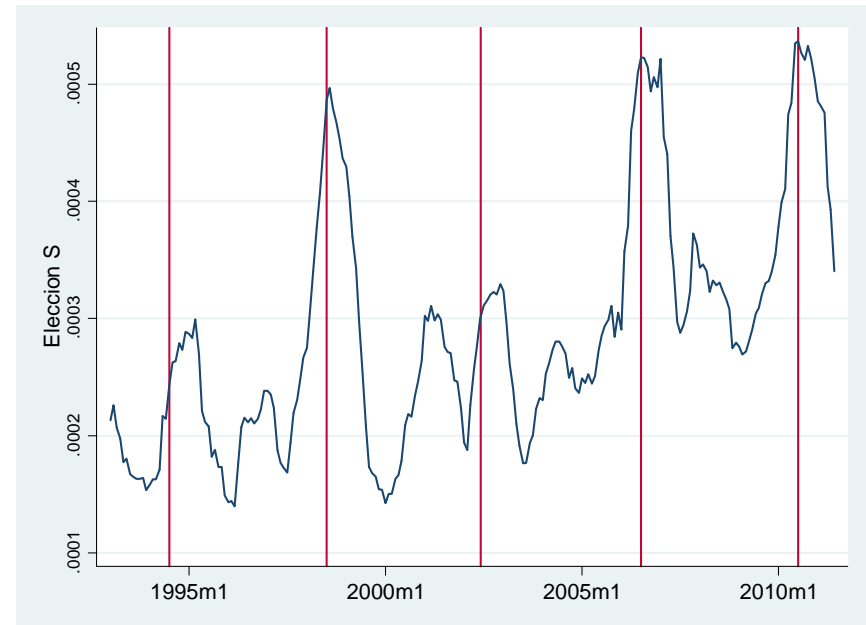
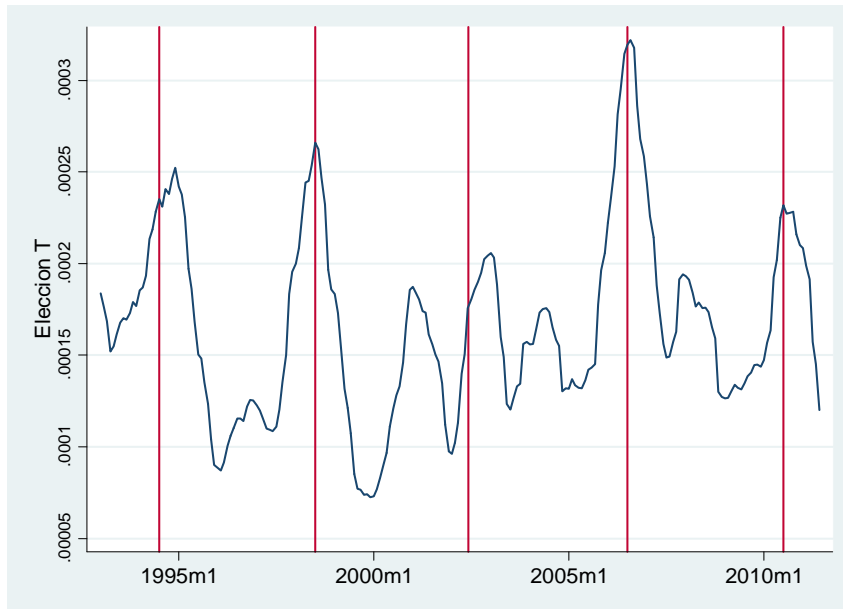
Temperatura y “sequía”



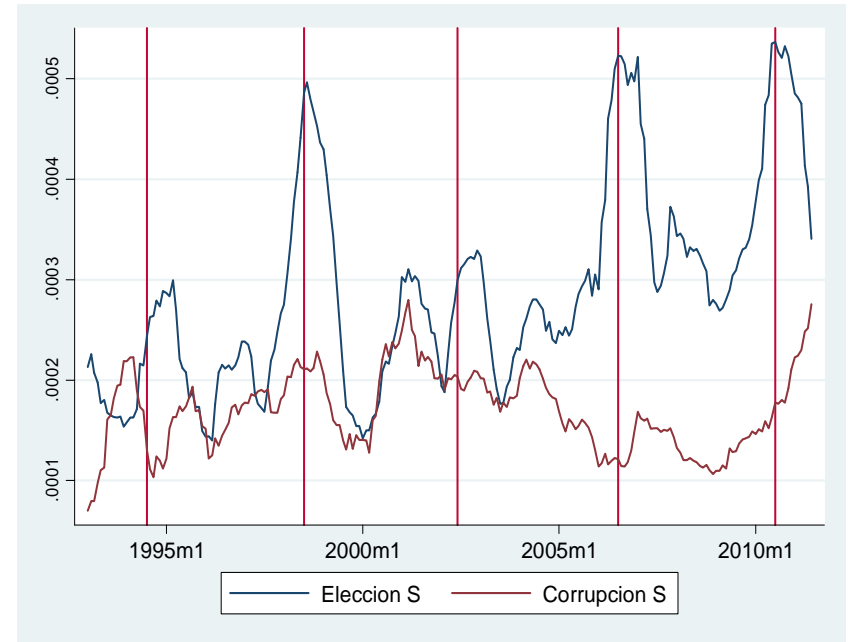
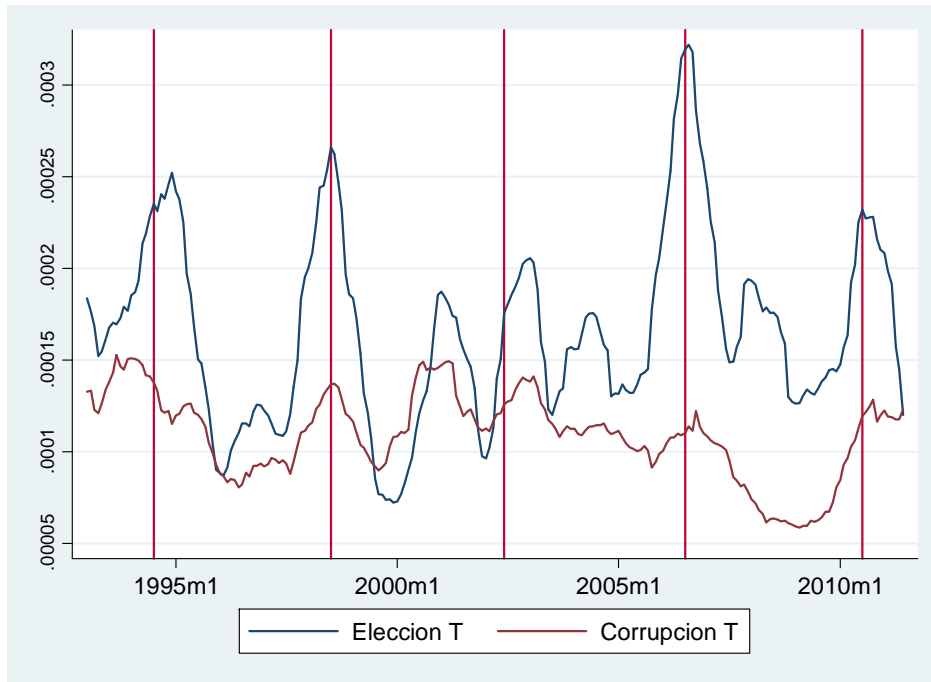
Temperatura e “inundación”



Elecciones: los ciclos de la democracia



“elecciones” y “corrupción”



Correlaciones:

```
. pwcorr electm eleesm corruptm corrupsm if num>12
```

	electm	eleesm	corruptm	corrupsm
electm	1.0000			
eleesm	0.7138	1.0000		
corruptm	0.3403	-0.0252	1.0000	
corrupsm	-0.0338	0.0309	0.5817	1.0000

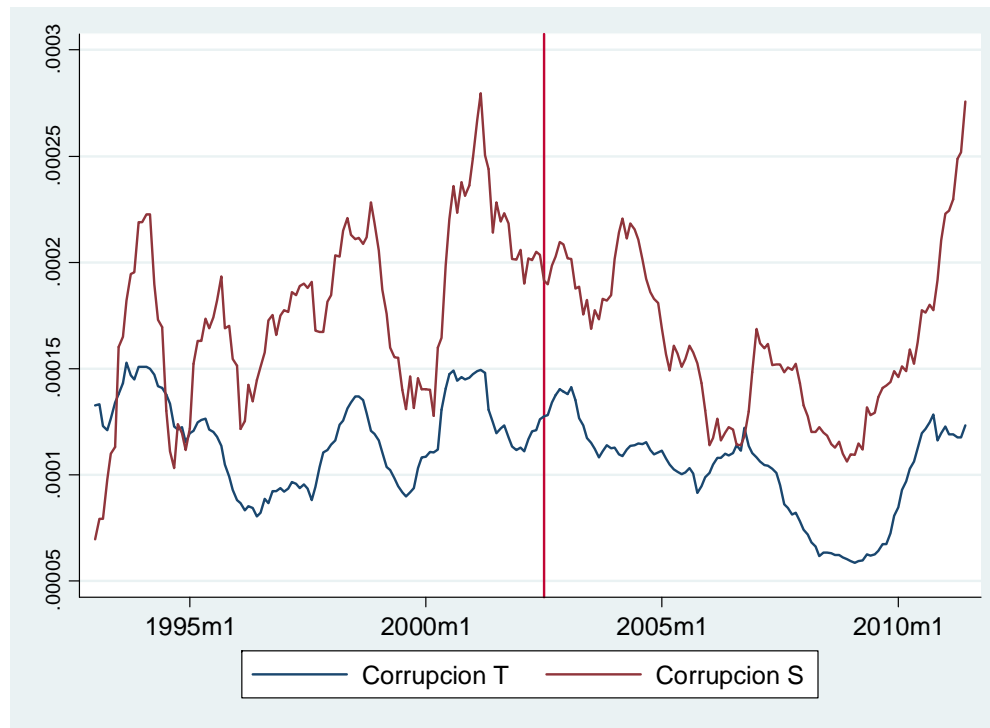
```
. pwcorr electm eleesm corruptm corrupsm if num>12 & num<126
```

	electm	eleesm	corruptm	corrupsm
electm	1.0000			
eleesm	0.6416	1.0000		
corruptm	0.6035	0.2064	1.0000	
corrupsm	0.1562	0.3648	0.4214	1.0000

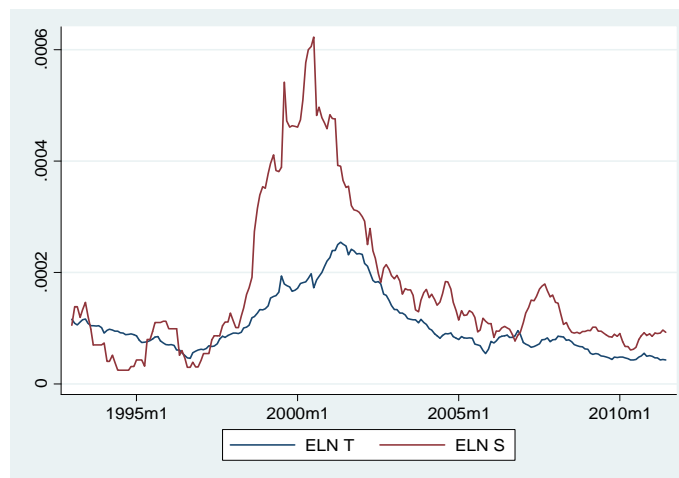
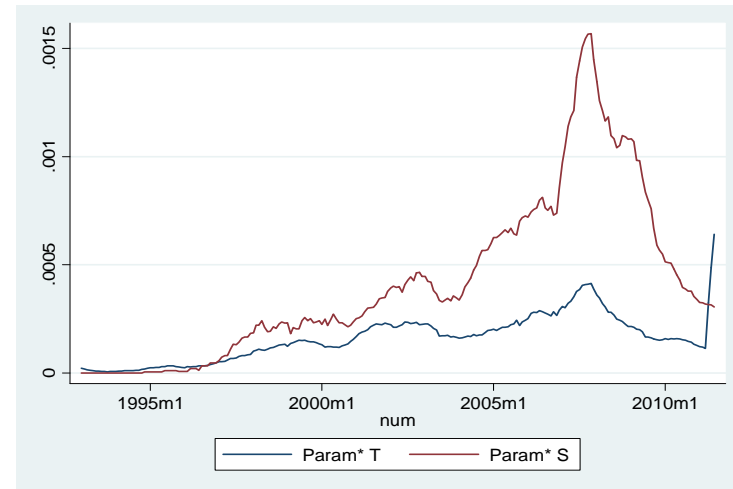
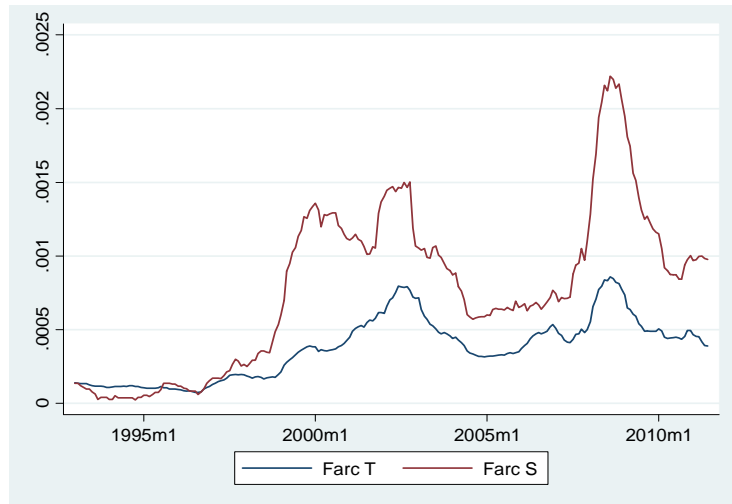
```
. pwcorr electm eleesm corruptm corrupsm if num>126
```

	electm	eleesm	corruptm	corrupsm
electm	1.0000			
eleesm	0.7751	1.0000		
corruptm	0.3512	0.1757	1.0000	
corrupsm	-0.1693	-0.0705	0.7058	1.0000

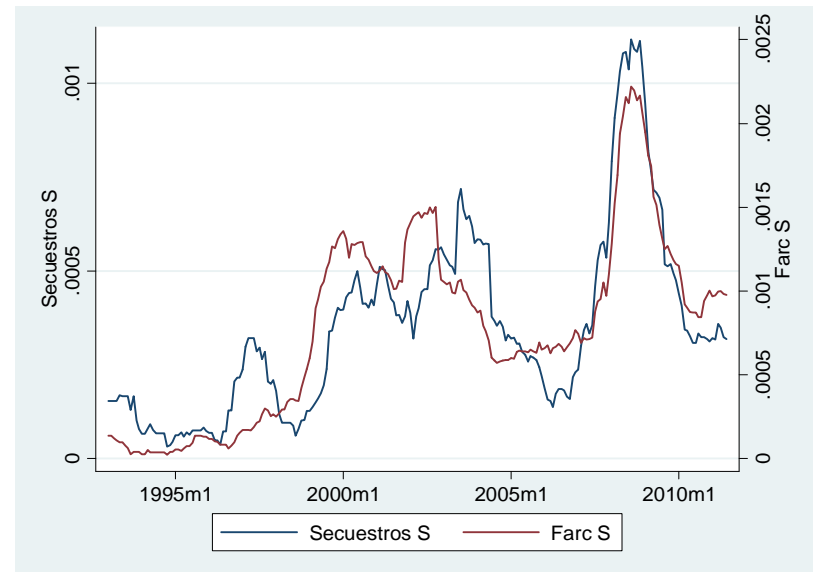
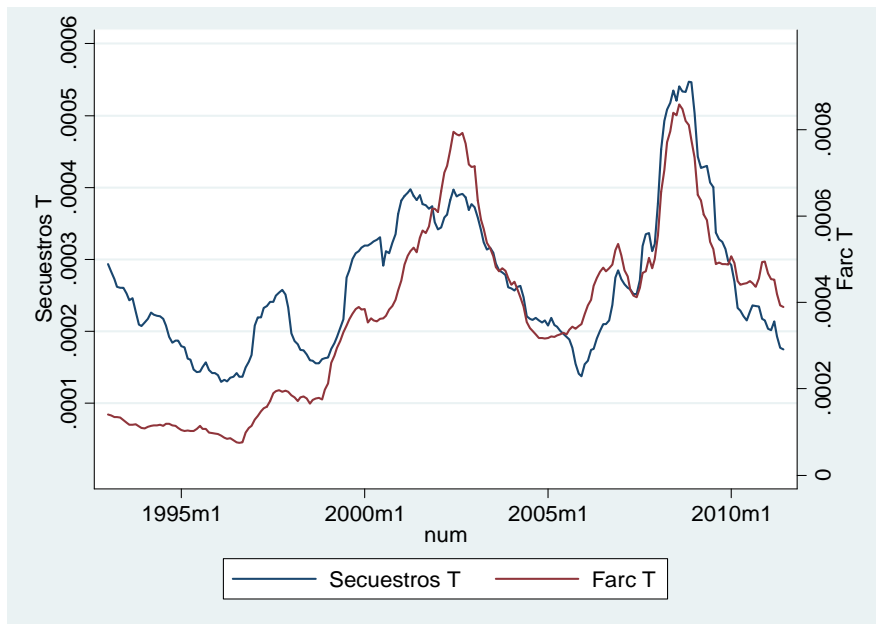
¿Puede la prensa ayudar a cuantificar lo incuantificable?



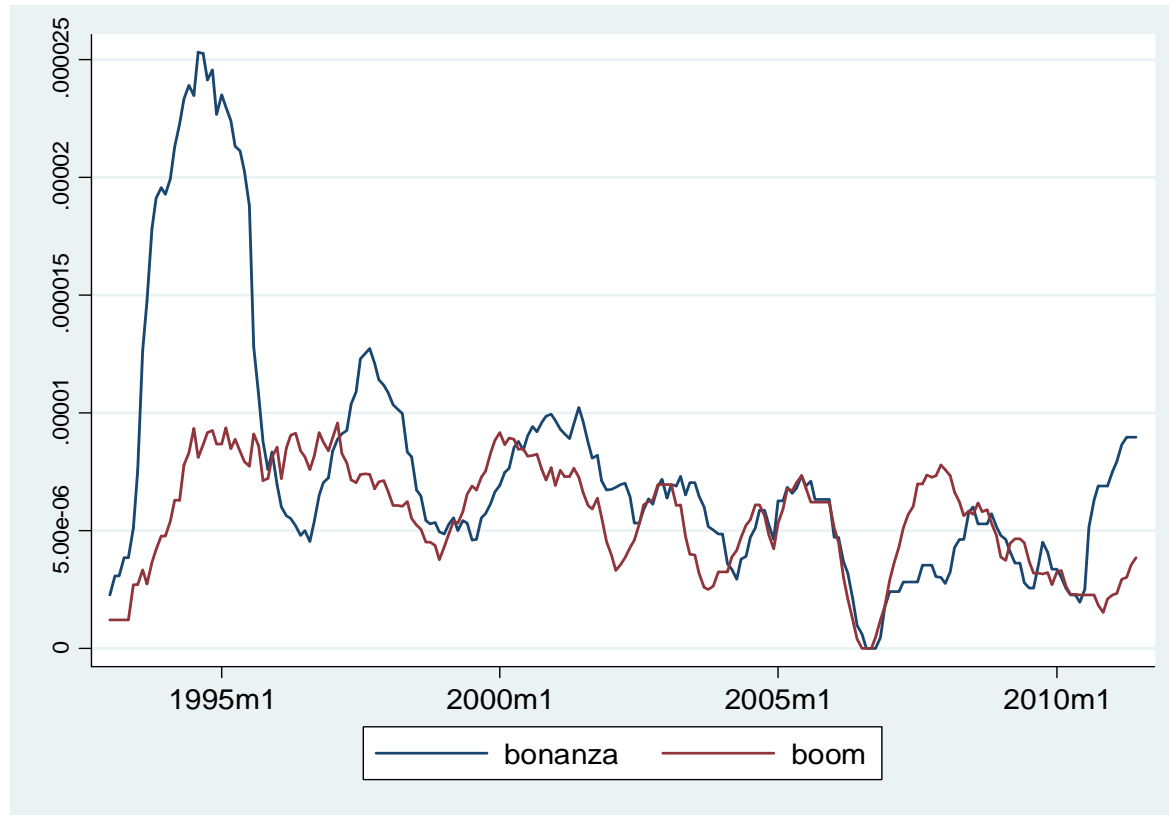
El conflicto en la prensa



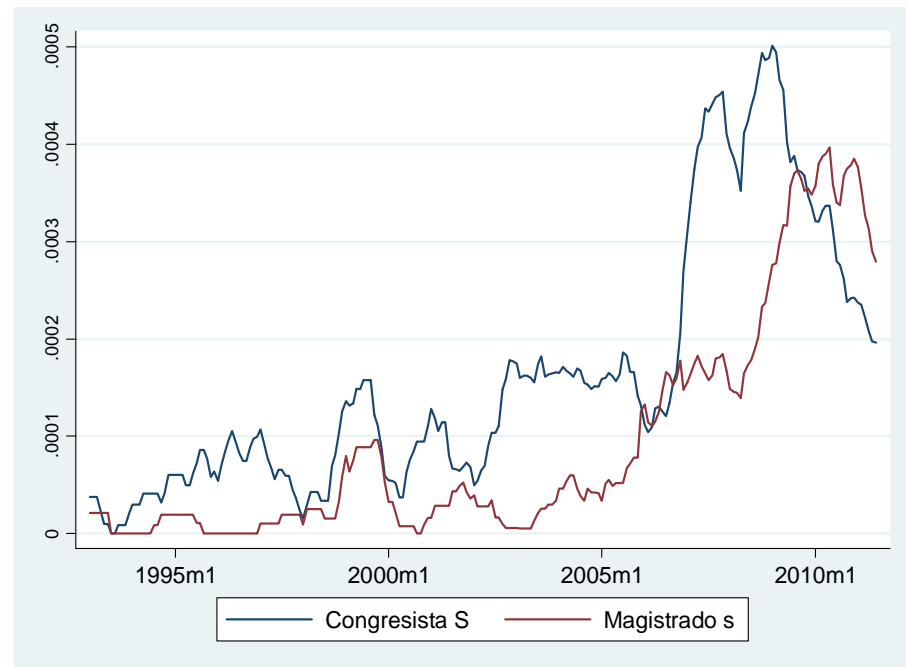
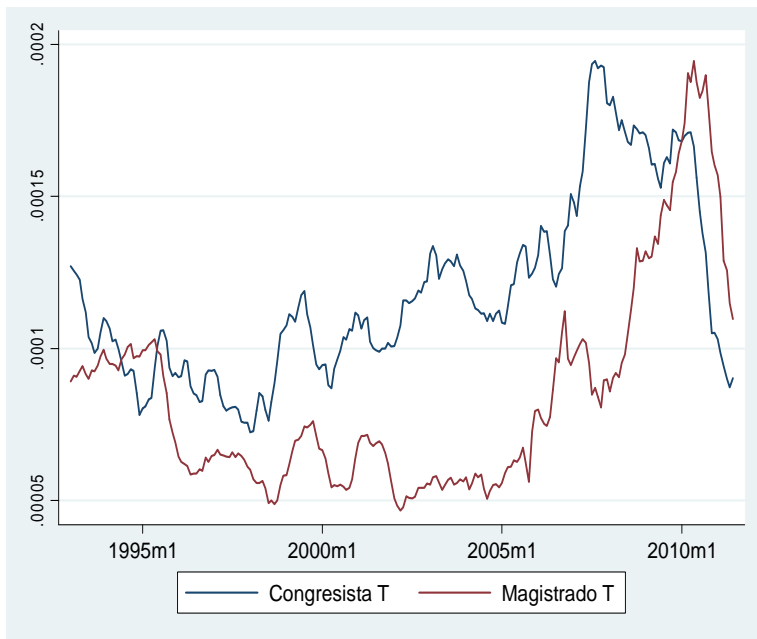
“Farc” y “secuestros”



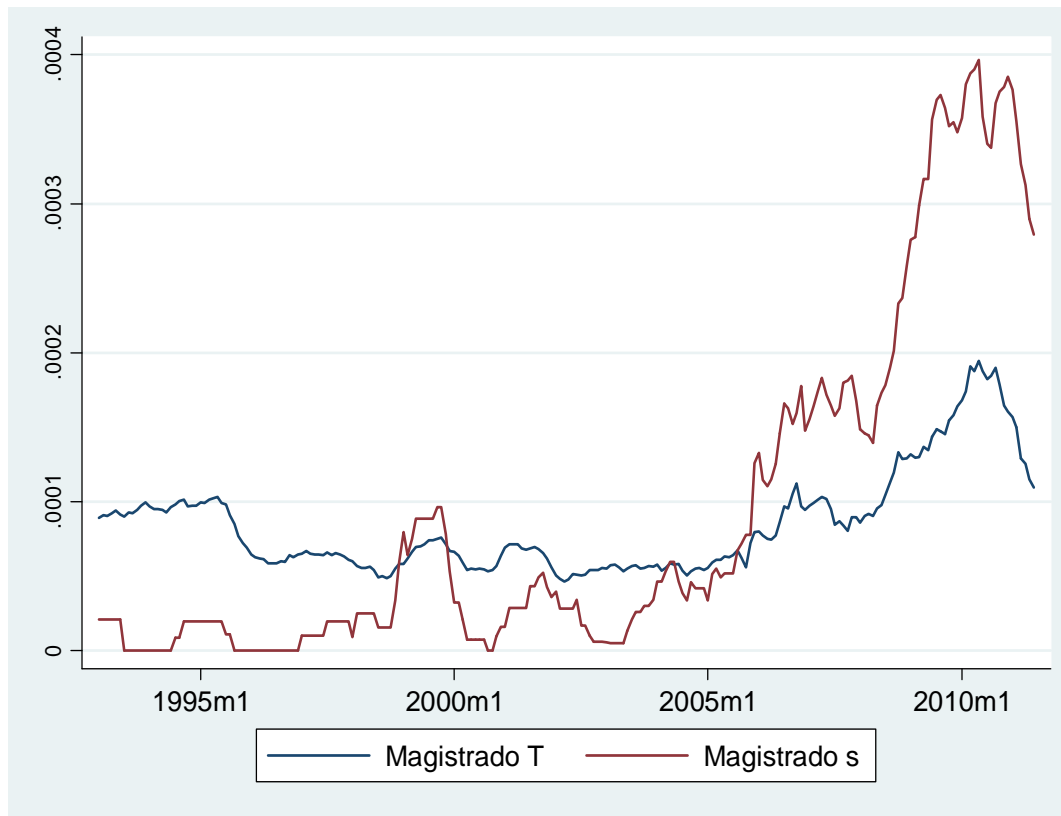
“Bonanza” y “boom”



“Congresistas” vs. “Magistrados”



El Tiempo vs. Semana



“alcaldía”, “gubernación” y “presidencia”

