

# El uso del Big Data en el análisis de variables económicas

Seminario de Investigación  
Banco Central de Reserva del Perú

Yalina Crispin  
Vanessa Belapatiño

25 de enero, 2017



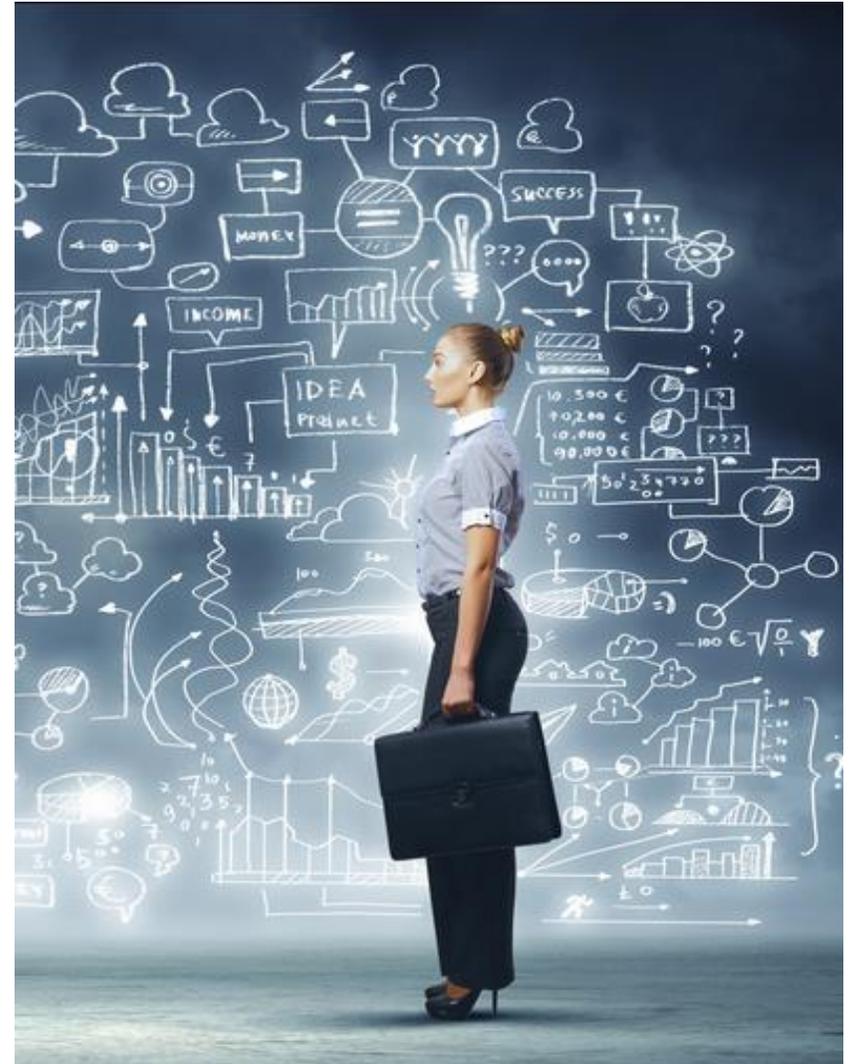
# Motivación

Utilizar herramientas tecnológicas de información que permitan realizar un mejor análisis del entorno económico.

Conocemos el pasado con retraso  
Ejemplos: PIB, retails, mercado inmobiliario, etc

Intuimos el presente (*nowcasting*)

Ignoramos el futuro (*forecasting*)



# Agenda

## 1. ¿Qué es Big Data?

- Fuentes de información.
- Sectores que están utilizando Big Data.
- Aplicaciones en la economía.

## 2. Aplicaciones para el análisis de variables económicas (con información de internet)

- Índice de intención de compra de departamentos en la ciudad de Lima.
- Índice de intención de compra de los hogares en Perú.
- Indicador de tensiones políticas.

## 3. Conclusiones

# Es hora de generar valor a los datos



¿Qué es **Big Data**? Es el uso de grandes cantidades de información, generadas a gran velocidad por diversas fuentes

Volumen

Variedad

Velocidad

# Fuentes de información

Registro de venta de empresas

Registro de pagos con tarjetas

Registro en los portales de búsqueda (Google, Facebook y Twitter)



# ¿En que sectores se está utilizando Big Data?



Comercio



Servicios financieros



Publicidad



Medios de  
Telecomunicación



Gobierno



# Aplicación: Comercio

## Cómo Netflix cambió la industria televisiva usando big data

## Cómo 'House of Cards' se convirtió en un éxito mundial gracias a un algoritmo

Algoritmos y 'Big Data'. Así aplica Netflix sus herramientas de análisis para garantizar el éxito de sus inversiones multimillonarias



Kevin Spacey posa ante un retrato de Francis Underwood, su personaje en 'House of Cards'. (Reuters)

## Sensores y 'big data': Amazon quiere 'matar' al supermercado (y saber todo sobre ti)

Una tienda "sin colas ni cajeros" en la que el cliente coge lo que quiere y se larga. Es la nueva propuesta con la que Amazon quiere hacer los 'súper' más cómodos- ¿Cómo?



# Aplicación: Banca – Producto BBVA Data & Analytics (España)

¿Quieres saber cómo activar tu negocio? ¿Quieres conocer tu situación respecto a la competencia? ¿Te gustaría entender mejor a tus clientes?



¿De dónde procede la información de Commerce360?



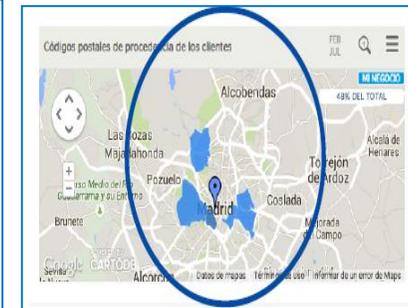
Los datos de Commerce360 se elaboran con una muestra de datos agregados y anonimizados de compras con tarjeta.



Los indicadores de tu negocio se basan en las compras de tu TPV BBVA. Sólo tú accedes a la información de tu comercio.



Los datos de tu mercado se obtienen a partir de una muestra anonimizada y agregada de compras con tarjetas BBVA.



# Aplicaciones en la economía, algunos ejemplos

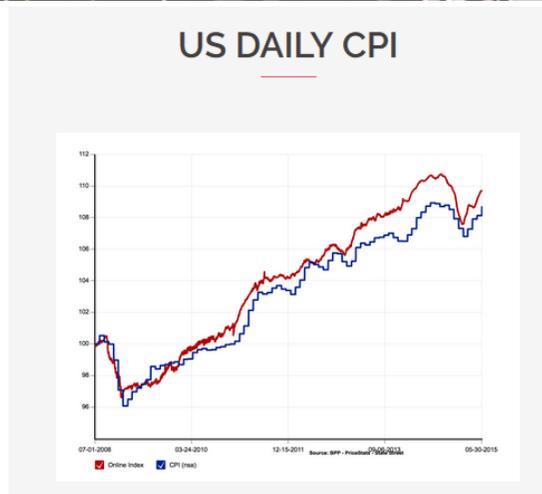
## Metodologías que hacen estimaciones en tiempo real de la evolución de los precios

Proyecto “ Billions Prices Projec”, MIT



Los datos scáner y los datos en línea son fuente importante en la construcción del IPC en Noruega

- Los precios recaudados por scanner representan el 30% de la muestra total de bienes y servicios, mientras que la data recaudada por internet representa el 19%.



### Cobertura:

- 96% de las cadenas minoristas.
- 85% de las cadenas de farmacias.

# Aplicaciones en la economía, algunos ejemplos

## Colombia (iniciativa pública – Ministerio de Hacienda, 2013)

Se construyeron indicadores sectoriales adelantados de la actividad económica con información de *Google* (para el análisis de tendencias y quiebres).

## España (iniciativa privada – BBVA, 2016 )

Se utiliza información de Google para la construcción de índices de tendencia de búsqueda relacionadas al turismo, los cuales incrementan la precisión de las estimaciones de corto plazo en este sector.

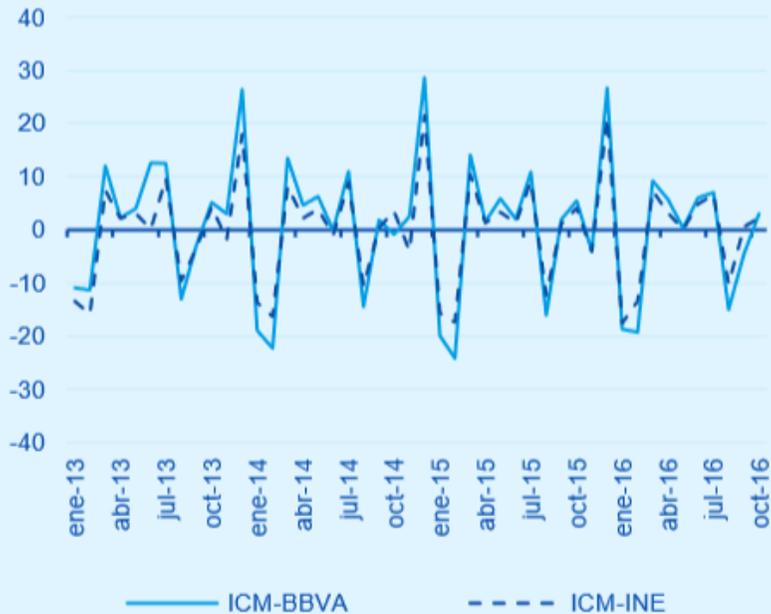
## España (iniciativa privada – BBVA, 2016)

Se estudia el comportamiento agregado de los consumidores mediante la información de las transacciones realizadas con tarjetas de crédito o débito por los clientes de BBVA con comercios minoristas. El indicador del gasto en comercios minoristas (ICM-BBVA) replica la estructura del índice de comercio al por menor publicado por el Instituto de Estadística. El ICM- BBVA ofrece un mayor detalle de información que el proporcionado por las fuentes oficiales. Este permite conocer el movimiento comercial a nivel provincial, por sectores (alimentos, salud, cuidado personal, equipamiento del hogar, tecnología, moda, etc).

# Aplicaciones en la economía, algunos ejemplos

## Índice BBVA de comercio al por menor

**Ventas al por menor vs ICM-BBVA  
(datos nominales brutos, % m/m)**



Fuente: BBVA a partir de INE

Permite obtener información comercial a nivel provincial ...

...por sectores ...

... y la estacionalidad del comercio  
(mensual y semanal)

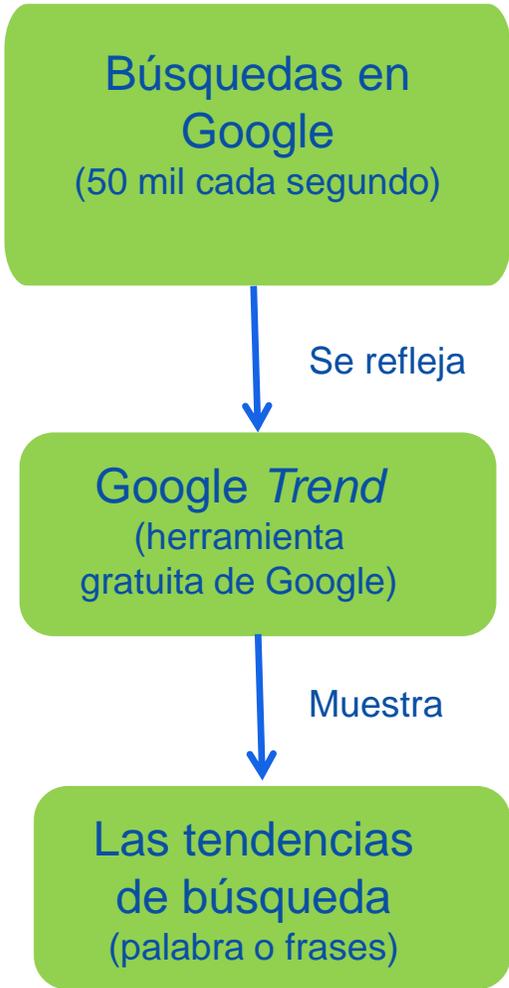
**ICM-BBVA diario: distribución semanal del consumo (% del gasto nominal en cada día de la semana, 1ª semana de junio)**



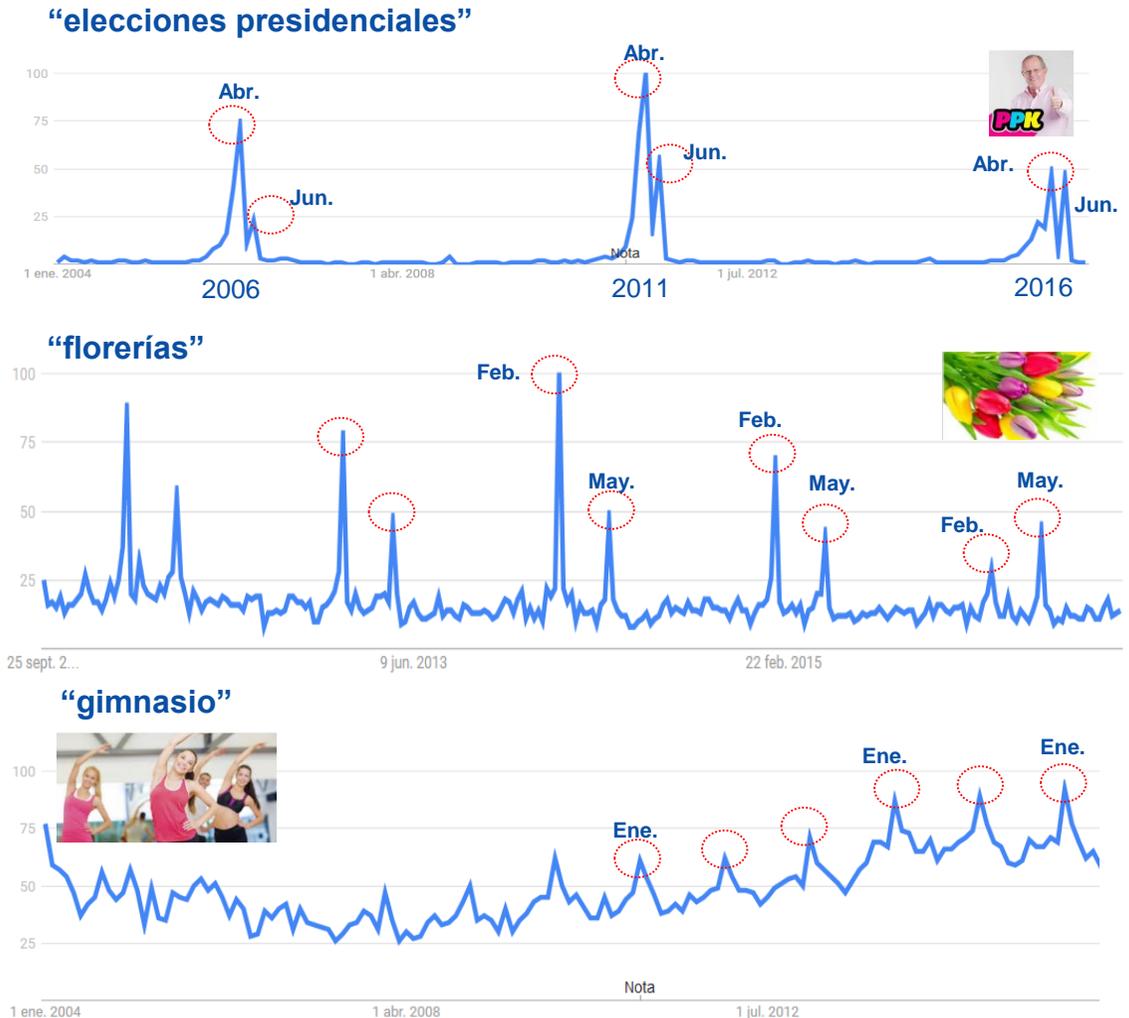
Fuente: BBVA a partir de INE



# Las búsquedas en internet contienen información relevante sobre las preferencias de los agentes económicos



## Ejemplos:



## Diversos estudios demuestran que los datos de búsqueda en internet (*google*) son útiles para realizar proyecciones de variables económicas

Autores	Sectores
Choi y Varian (2009, 2012) McLaren (2011) Chamberlin (2010) Carrire-Swallow y Labbé (2010)	Retail, autos, vivienda, turismo
Chang y Río (2013) Baker y Fradkim (2011) Askitas y Zimmerman (2010)	Empleo
Pena y Huan (2009)	Consumo privado

# Búsqueda en internet y las proyecciones de ventas de departamento en Lima

## Objetivo

Estimar un indicador de intención de compra de departamentos que permita proyectar las ventas en el corto plazo



Entre el deseo y la materialización de la compra de un departamento, hay un registro en internet de ese interés (que deseamos capturar)



**Interés de compra de un departamento**



**Búsqueda en internet**  
(inmobiliarias, precio, banco, etc)

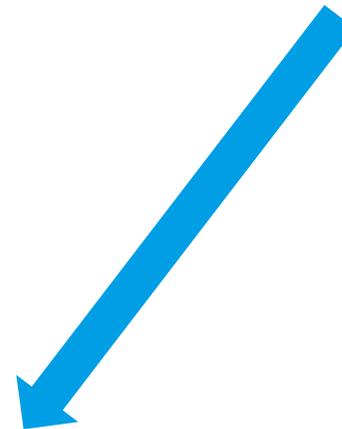


**Préstamo hipotecario**  
(evaluación crediticia y ahorro de la cuota inicial)



**Compra y venta**

**Índice de Intención de Compra de Departamentos (IICDG)**



# Construcción del IICDG

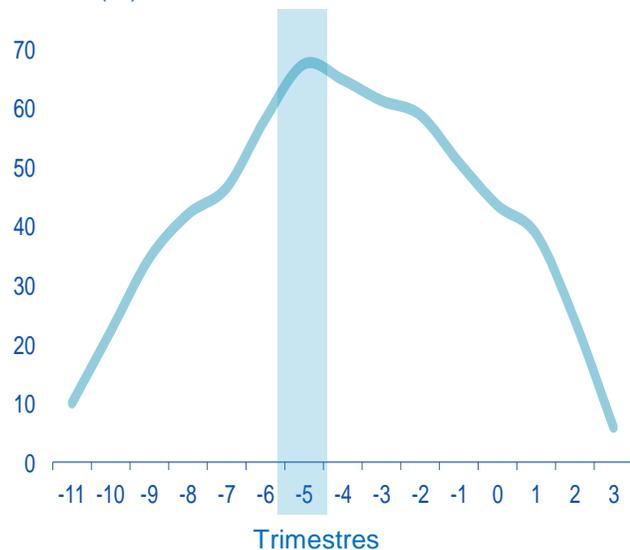
- 1 De una lista de palabras y frases relacionadas con el interés de compra de un departamento en la ciudad de Lima, se seleccionaron aquellas que mostraron una mayor relación con las ventas de departamento de la ciudad de Lima.
- 2 Se ponderan las series elegidas por la inversa de sus desviaciones estándar para reducir la importancia la volatilidad y la incertidumbre en el IICDG.

# IICDG y las ventas de departamento en Lima

Correlación positiva, alcanzando su mayor asociación cuando el IICDG se adelanta 5 trimestres\*

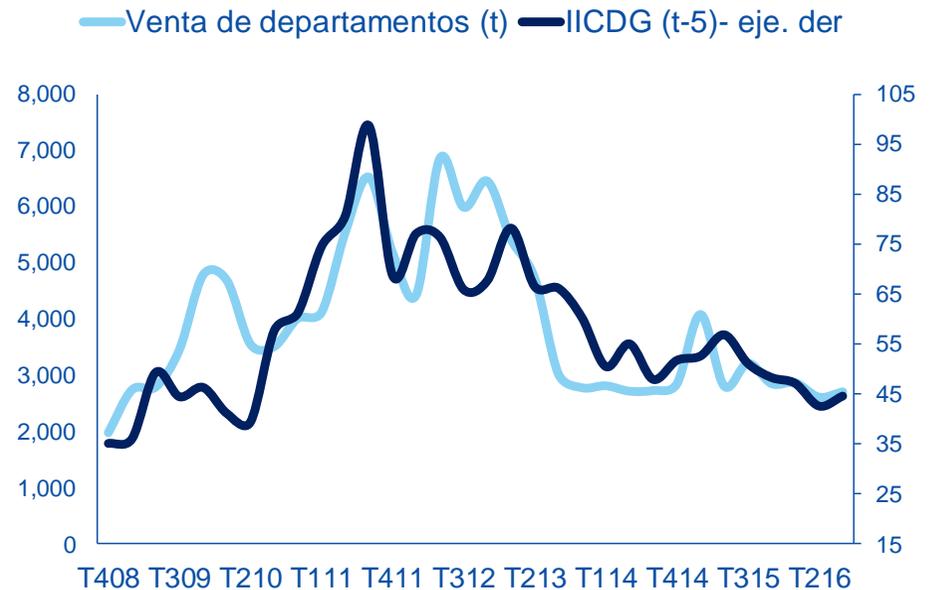
## Coeficiente de correlación entre las Ventas de departamentos en Lima y el IICDG con diferentes rezagos

Correlación (%)



Fuente: TINSA y Google.

## Venta de departamentos en Lima e IICDG (número de unidades e índice)



Fuente: TINSA, Google.  
Elaboración: BBVA Research.

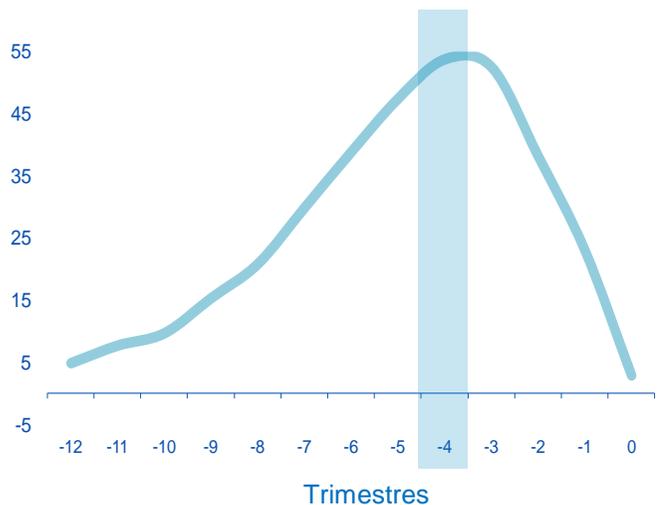
\*Se realizó prueba de precedencia estadística y se encontró que el IICDG anticipa (causa) el comportamiento de las ventas de departamentos en Lima.

# IICDG y el crecimiento del crédito hipotecario

Correlación positiva, alcanzando su mayor asociación cuando el IICDG se adelanta 4 trimestres\*

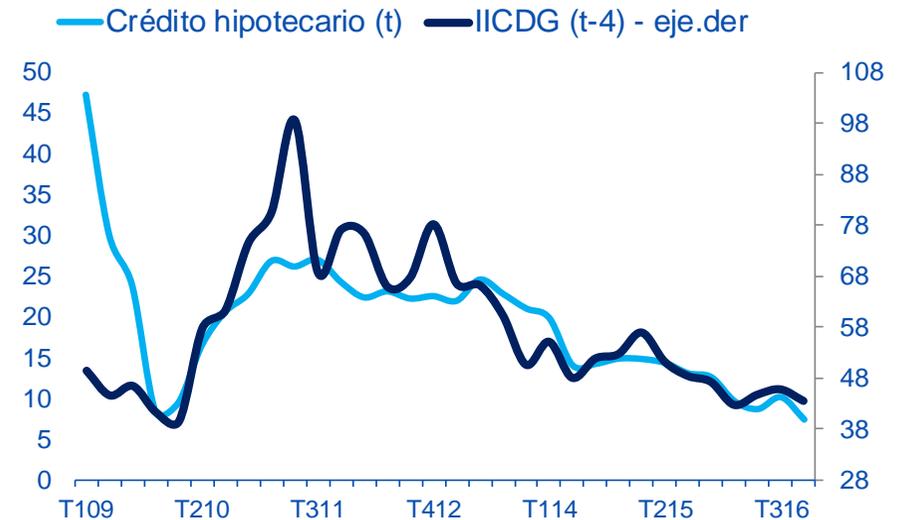
## Coeficiente de correlación entre el crecimiento del crédito hipotecario y el IICDG con diferentes rezagos

Correlación (%)



Fuente: ASBANC y Google.

## Crédito hipotecario del Sistema bancario y el IICDG (var.% interanual e Índice)



Fuente: ASBANC, Google.  
Elaboración: BBVA Research.

\*Se realizó prueba de precedencia estadística y se encontró que el IICDG anticipa (causa) el comportamiento de las ventas de departamentos en Lima. Para más detalle descargue el informe [aquí](#)

# IICDG mejora la capacidad para predecir las ventas de departamento en Lima

## Indicadores de capacidad predictiva de los dos modelos de proyección

Indicadores	Modelo de referencia *	Modelo ** aumentado
<b>R-squared</b>	81	85
<b>Errores de la muestra</b>		
ECM	1,542	<b>958</b>
EMA	1,318	<b>779</b>

\*Se considera un modelo ARMA: Las ventas de departamento en Lima se encuentran en función al comportamiento pasado y a las perturbaciones económicas y financieras que afectan esta variable.\*\* Al modelo de referencia se le agrega el IICDG.

ECM: error cuadrático medio. EMA: error medio absoluto.

Elaboración: BBVA Research.

## Venta de departamentos en Lima (número de unidades)



Fuente: TINSA

Elaboración: BBVA Research.

# Búsqueda en internet y las proyecciones del gasto de las familias

## Objetivo

Estimar un indicador de intención de compra que permita proyectar el consumo de los hogares en el corto plazo





**Interés de compra**



**Búsqueda en internet**  
(precios, modelos, descuentos, ofertas, compras *online*, etc)



**Índice de intención de compra de los hogares (IICG)**



**Compra**  
(consumo privado)



# Construcción del IICG

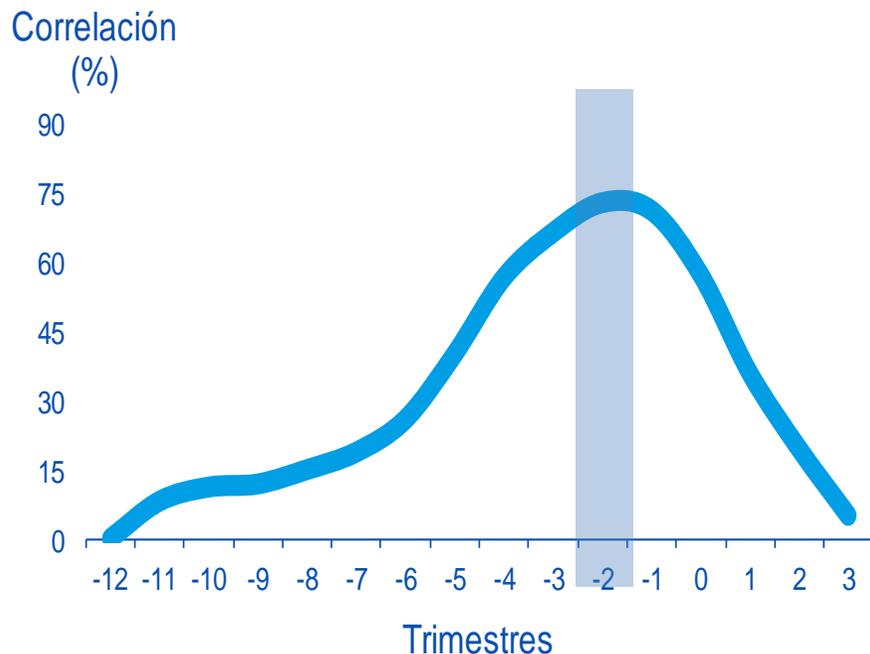
- 1 De una lista de palabras y frases relacionadas con la compra de bienes y servicios, se seleccionaron aquellas que mostraron una mayor relación con la dinámica del consumo privado.
- 2 Se ponderan las series elegidas por la inversa de sus desviaciones estándar para reducir la importancia la volatilidad y la incertidumbre en el IICG.

# IICG y el consumo privado

Correlación positiva, alcanzando su mayor asociación cuando el IICG se adelanta 2 trimestres

## COEFICIENTE DE CORRELACIÓN ENTRE EL CRECIMIENTO DEL CONSUMO PRIVADO Y EL IICG CON DIFERENTES REZAGOS

(%)

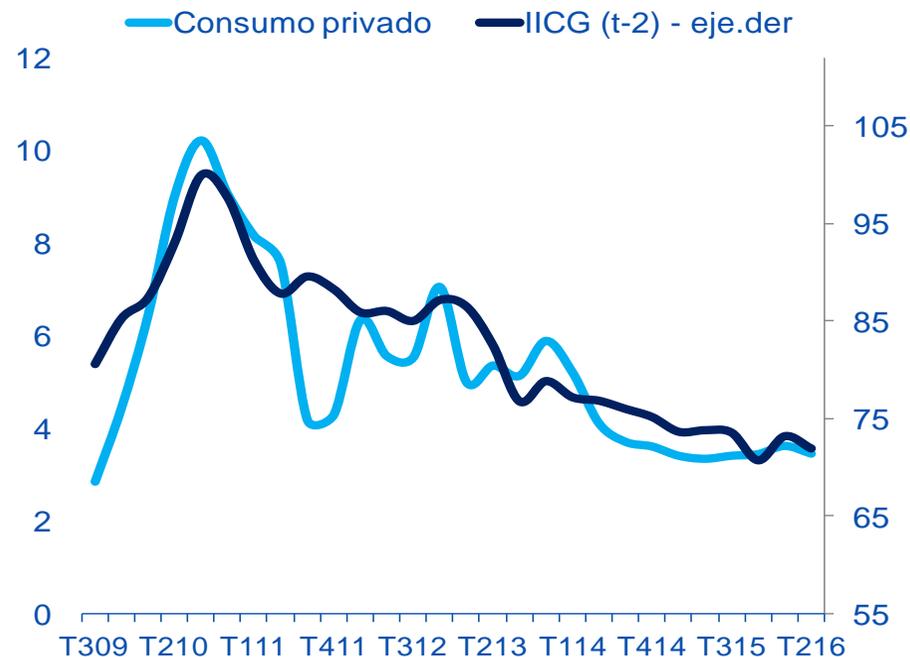


Fuente: BCRP, Google  
Elaboración: BBVA Research

Para más detalle descargue el informe [aquí](#)

## CONSUMO PRIVADO Y EL IICG

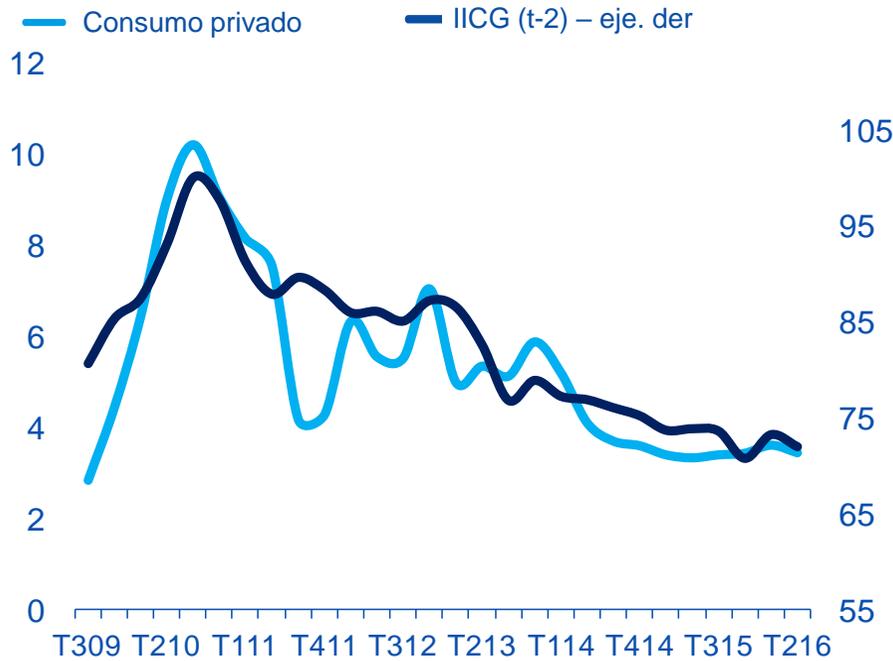
(var.% interanual e índice)



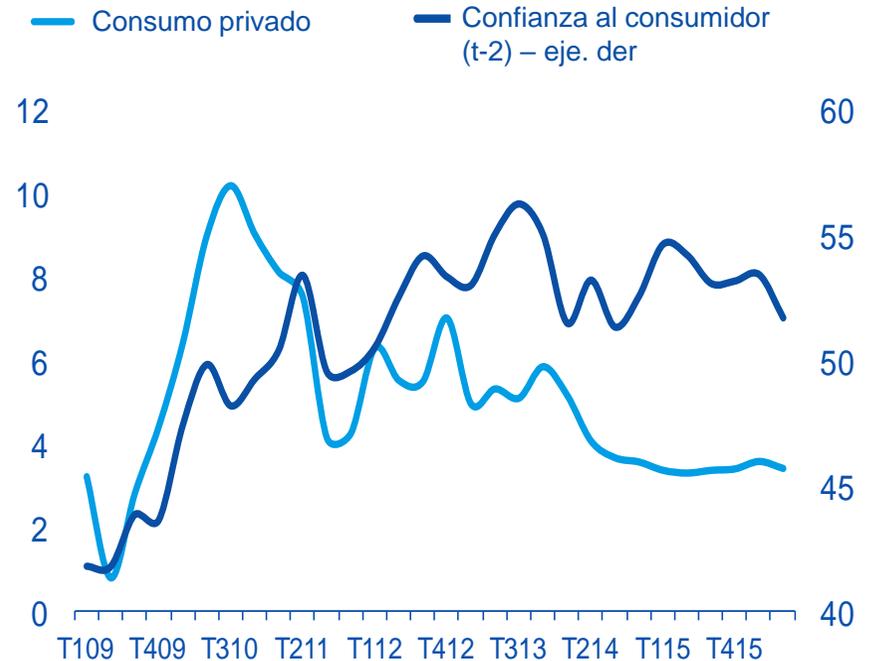
Fuente: BCRP, Google  
Elaboración: BBVA Research

# El IICG muestra una mejor relación con el consumo privado que la confianza al consumidor

**CONSUMO PRIVADO Y EL IICG**  
(var.% interanual e índice)



**CONSUMO PRIVADO Y EL INDICE DE CONFIANZA DEL CONSUMIDOR \***  
(var.% interanual y puntos)



\*Valores por encima de 50 se ubica en el tramo optimista. El análisis de correlación dinámica (para el periodo 2T2003-2T2016) revela que existe una relación positiva entre el consumo privado y el Índice de confianza del consumidor, alcanzando su máxima asociación (33%) cuando el índice se adelanta dos trimestres.

# IICG mejora la capacidad para predecir el consumo privado

## INDICADORES DE CAPACIDAD PREDICTIVA DE LOS MODELOS DE PROYECCIÓN DEL CONSUMO DE LOS HOGARES

Indicadores	Modelo de referencia <sup>1/</sup>	Modelo Estructurado <sup>2/</sup>	Modelo con IICG <sup>3/</sup>
R-squared	75	67	<b>81</b>
<b>Errores de proyección <sup>4/</sup></b>			
ECM	1.38	1.05	<b>1.00</b>
EMA	1.12	0.77	<b>0.68</b>

1/ Se considera un modelo ARMA: el crecimiento del consumo privado se encuentran en función al comportamiento pasado y a las perturbaciones económicas y financieras que afectan esta variable.  
 2/ El consumo privado en función al ingreso disponible.  
 3/ Al modelo estructurado se le agregó el IICG.  
 4/ ECM: error cuadrático medio. EMA: error medio absoluto

## CONSUMO PRIVADO (var.% interanual)



Fuente: BCRP  
 Elaboración: BBVA Research

# Aplicaciones Big Data: *GDELT* y el análisis de variables económicas

1. ¿Qué es GDELT y qué se puede encontrar?
2. Aplicaciones
3. Indicador de tensiones políticas



# GDELT: **G**lobal **D**atabase on **E**vents Location and **T**one

Monitorea los **medios de comunicación** a nivel mundial

Información desde 1979, incluye más de 100 lenguajes

Se actualiza cada 15 minutos

Usa **algoritmos para detectar emociones** de las noticias

**Geo-referencia los eventos**

**GDELT** proporciona información para análisis geopolítico, social y económico

# GDELT y su procesamiento de datos



- Cada 15 minutos

- Traduce las noticias al inglés (98% no están en inglés)
- Analiza el texto
  - Analiza cada palabra y la sintaxis.
  - Separa verbos, adjetivos nombres propios.
  - Reconoce patrones
- Ordena
- Permite:
  - Clasifica por categoría
  - Relacionar países, organizaciones y personas
  - Encuentra el tono de la noticia.

# La cobertura de GDELT y procesamiento de datos

BigQuery

The screenshot displays the Google BigQuery web interface. On the left, there is a sidebar with navigation options: 'COMPOSE QUERY', 'Query History', 'Job History', a search box for 'Filter by ID or label', and a project selection dropdown currently set to 'Mi primer proyecto'. Below this, there are links for 'Mapa\_conflictos', 'gdelt-bq', and 'Public Datasets'. The main area is titled 'New Query' and contains a SQL query editor with the following code:

```

1 SELECT a.date date, a.count/b.counttotal ratio, a.tone from
2 (select substr(string (DATE),1,8) date, COUNT (*) as count,
3 avg(FLOAT(REGEXP_EXTRACT(V2Tone, '^(.*)')))) tone
4 from [gdelt-bq:gdeltv2.gkg_partitioned] WHERE
5 _PARTITION_LOAD_TIME BETWEEN TIMESTAMP ('2016-12-31')AND
6 TIMESTAMP ('2017-01-31') and V2Themes like '%POLITICS%' and
7 V2Locations like '%Peru%' and TranslationInfo like
8 '%src:spa%' group by 1) a JOIN EACH
9 ( select substr (string (DATE), 1, 8) date, COUNT(*)
10 as counttotal, avg(FLOAT(REGEXP_EXTRACT(V2Tone, '^(.*)'))))
11 tone from [gdelt-bq:gdeltv2.gkg_partitioned]
    
```

Below the query editor are buttons for 'RUN QUERY', 'Save Query', 'Save View', and 'Format Query'. A 'Show Options' button is also present. The results section shows a table with the following data:

Row	date	ratio	a_tone
1	20161231	0.2839657282741738	-1.901596266386774
2	20170101	0.3153296266878475	-1.3461178841440216
3	20170102	0.2922044398554466	-0.6488735990940866

# GDELT: Base de datos

## Base de datos de eventos

Catálogo de eventos en más de 300 categorías (protestas, ataques militares, intercambios diplomáticos, etc). Desde 1979

## Gráfico global de conocimiento (GKG)

Extrae nombre de personas, organizaciones, ubicación, millones de temas y miles de emociones. Desde 2013

## Televisión americana (GKG)

Monitorea más de 100 estaciones de televisión de EEUU para conocer al usuario. Se actualiza cada 2 días.

## Archivo de libros de historia americana

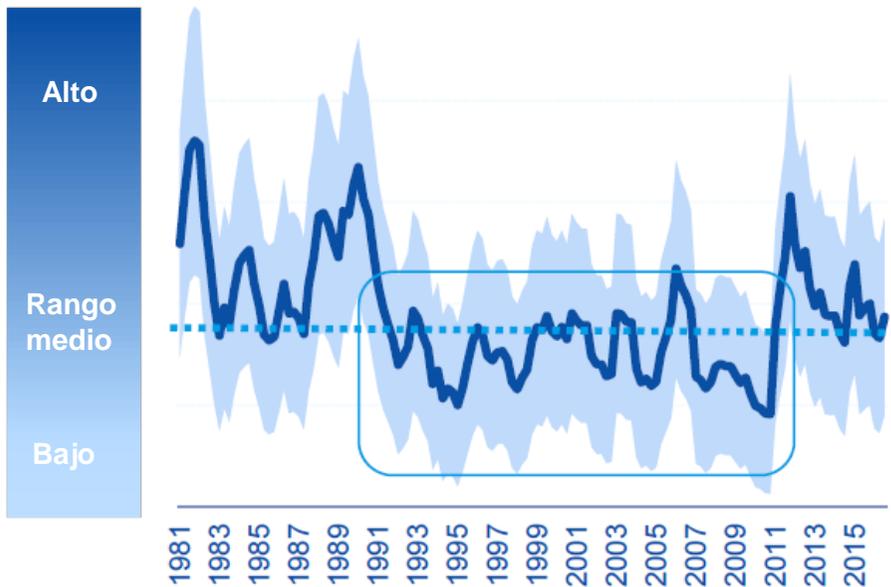
Procesó más de 3.5 millones de volúmenes de archivos de internet y documentos digitales.

## Literatura Académica de África y Medio Oriente

Información sociocultural, menciones de grupos étnicos y religiosos

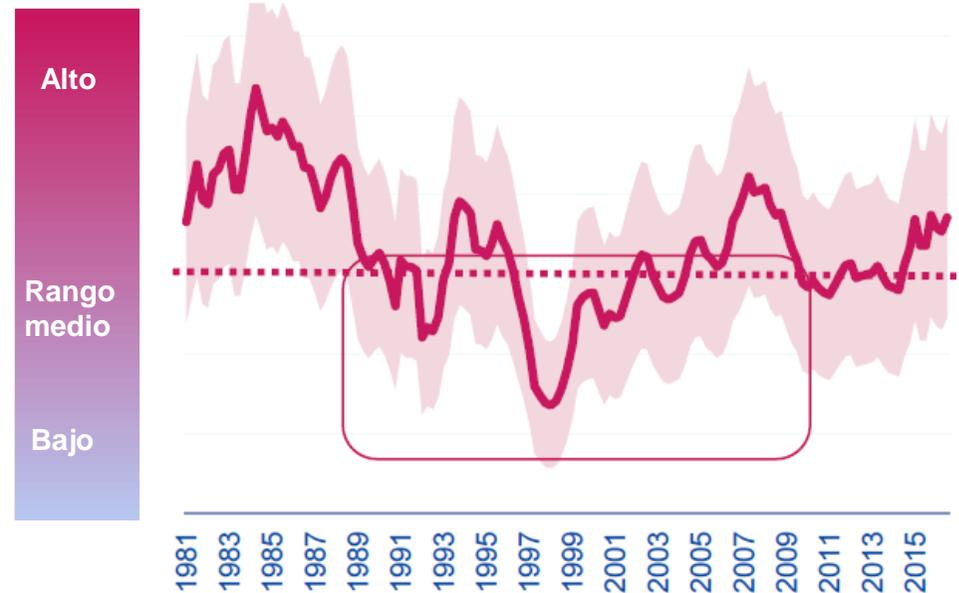
# Algunas aplicaciones

**ÍNDICE DE INTENSIDAD DE PROTESTAS 1981-2016**  
 (Número de protestas en el mundo/ total de eventos en el mundo)



Fuente: BBVA Research y GDELT  
 Elaboración: BBVA Research

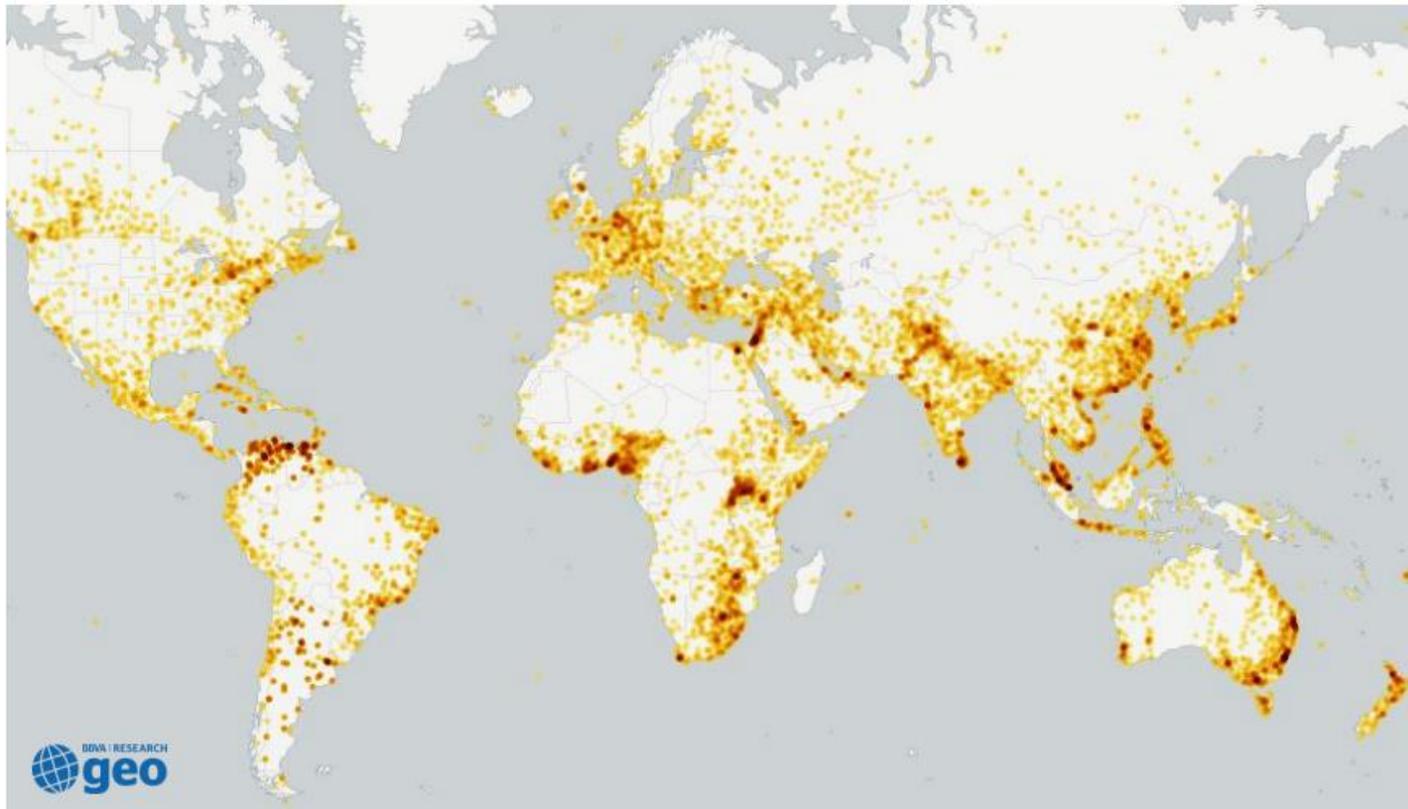
**ÍNDICE DE INTENSIDAD DE CONFLICTOS 1981-2016**  
 (Número de conflictos en el mundo/ total de eventos en el mundo)



Fuente: BBVA Research y GDELT  
 Elaboración: BBVA Research

# Algunas aplicaciones

**VULNERABILIDADES MACROECONÓMICAS en 2015-2016**  
(cobertura en los medios sobre vulnerabilidades macro en el mundo)



# Algunas aplicaciones

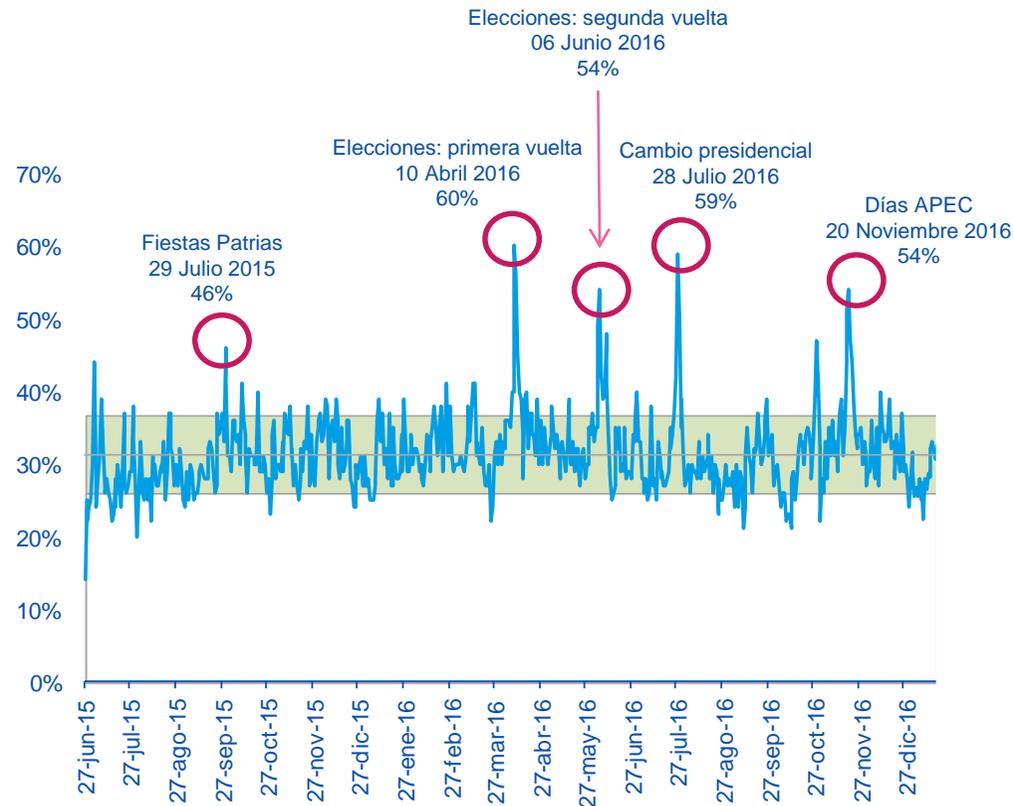
**Dic. 2015: DESACELERACIÓN DE CHINA: preocupaciones y contagios**  
 (cobertura en los medios sobre la desaceleración en China)

- Financiero y comercio
- Impacto en la actividad en Europa
- Impacto en la actividad en Asia emergente
- Exportadores de commodities de África
- Exportadores de commodities de Latam
- Demanda de petróleo

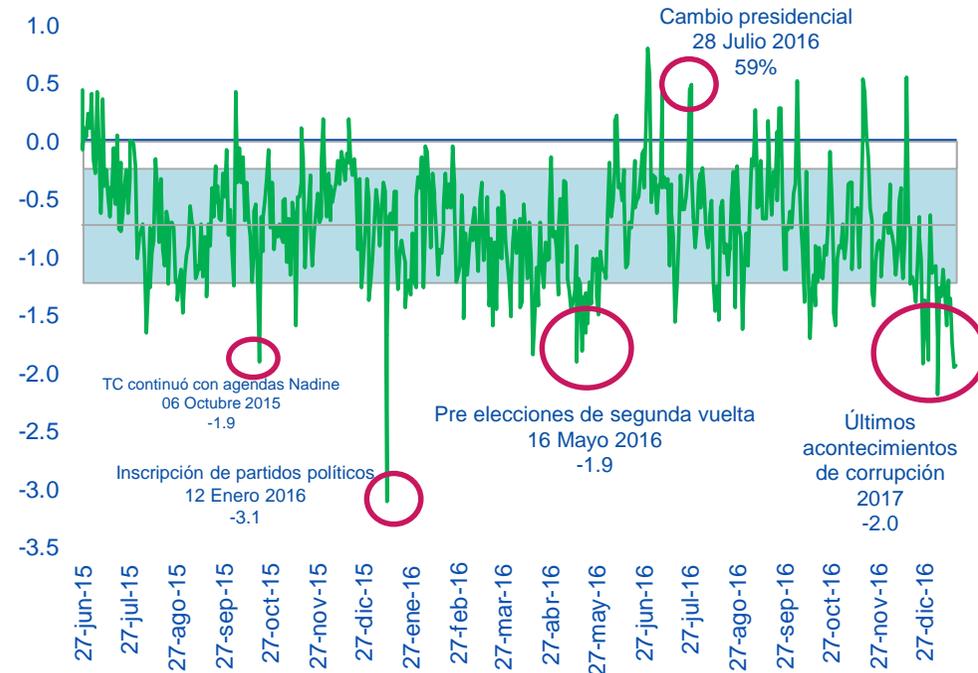


# Tensiones políticas

**PERÚ: ÍNDICE DE INTENSIDAD TEMAS POLÍTICOS**  
(cobertura de temas políticos en medios de comunicación)



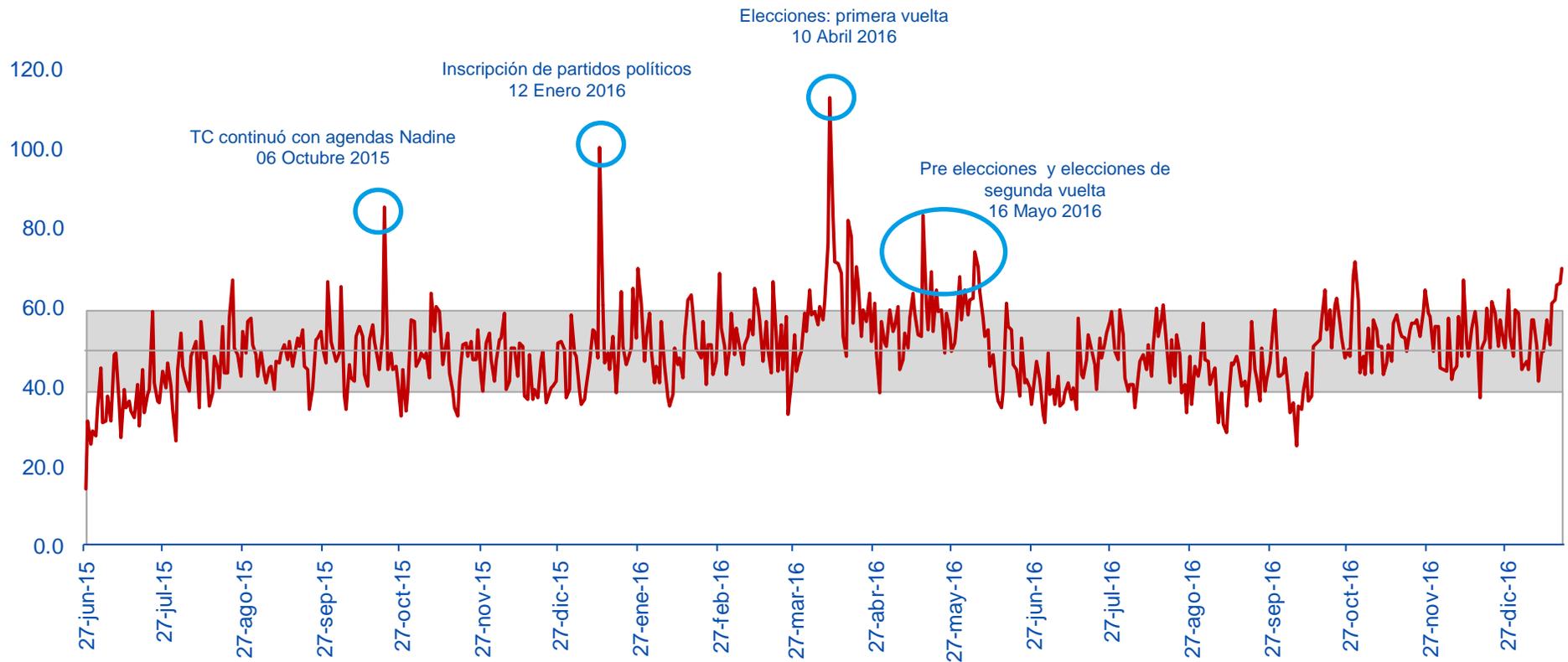
**PERÚ: TONO DE TEMAS POLÍTICOS**  
(<0 connotación negativa, >0 connotación positiva)



Fuente: BBVA Research y GDELT  
Elaboración: BBVA Research

# Tensiones políticas

## PERÚ: ÍNDICE DE TENSIONES POLÍTICAS



# Conclusiones

- **El uso de Big Data nos ofrece oportunidades para la investigación**, ya que, de alguna manera, nos permite cuantificar lo que solo se conoce de manera cualitativa.
- **Nos brinda información que usualmente no manejamos.** Nos permite mejorar nuestros modelos al incorporar el comportamiento de las personas (búsquedas y noticias). No solo para mejorar estimaciones, si no también proyecciones.
- **Además, es información en tiempo real**, lo que nos permite adelantarnos a las publicaciones de datos oficiales y, como hemos visto en el caso de intención de compra y consumo, es información relevante para el análisis de la economía peruana.
- **Estas aplicaciones son el comienzo de futuras investigaciones en el campo de Big Data.**

# El uso del Big Data en el análisis de variables económicas

Seminario de Investigación  
Banco Central de Reserva del Perú

Yalina Crispin  
Vanessa Belapatiño

25 de enero, 2017

