

NUEVAS FUENTES DE DATOS: ANÁLISIS ECONÓMICO Y ESTADÍSTICA OFICIAL I

Javier J. Pérez
Banco de España

Seminario CEMLA
Madrid, 5-7 de junio de 2019

DIRECCIÓN GENERAL DE ECONOMÍA Y ESTADÍSTICA



09:30–11:00 **Nuevas fuentes de datos: análisis económico y estadística oficial I**

Nuevas fuentes de datos para el análisis económico

Aplicaciones de las técnicas de análisis textual

– *Medición de la incertidumbre económica y sobre las políticas*

– *Comunicación cualitativa sobre la situación económica*

Implicaciones de las nuevas fuentes para la estadística oficial

22/05/2019

[Algunas implicaciones de las nuevas fuentes de datos para el análisis económico y la estadística oficial \(499 KB\)](#)

Corinna Ghirelli, Juan Peñalosa, Javier J. Pérez y Alberto Urtasun

- El desarrollo de las nuevas tecnologías y de las redes sociales ha propiciado la aparición de **nuevas fuentes de datos**, comúnmente denominadas como de “big data”
- **Estos nuevos datos presentan una serie de características particulares...**
 - en cuanto a su **volumen y su grado de detalle** (muy superiores a las fuentes tradicionales)
 - su disponibilidad a **altas frecuencias** y su carácter en muchas ocasiones **no estructurado** (esto es, no necesariamente numérico, ni organizado, como, por ejemplo, el proveniente de textos o imágenes)



Increased availability of data...

Microdata: from massive databases to the population (firms; loans; taxes; households; money market transactions;...)

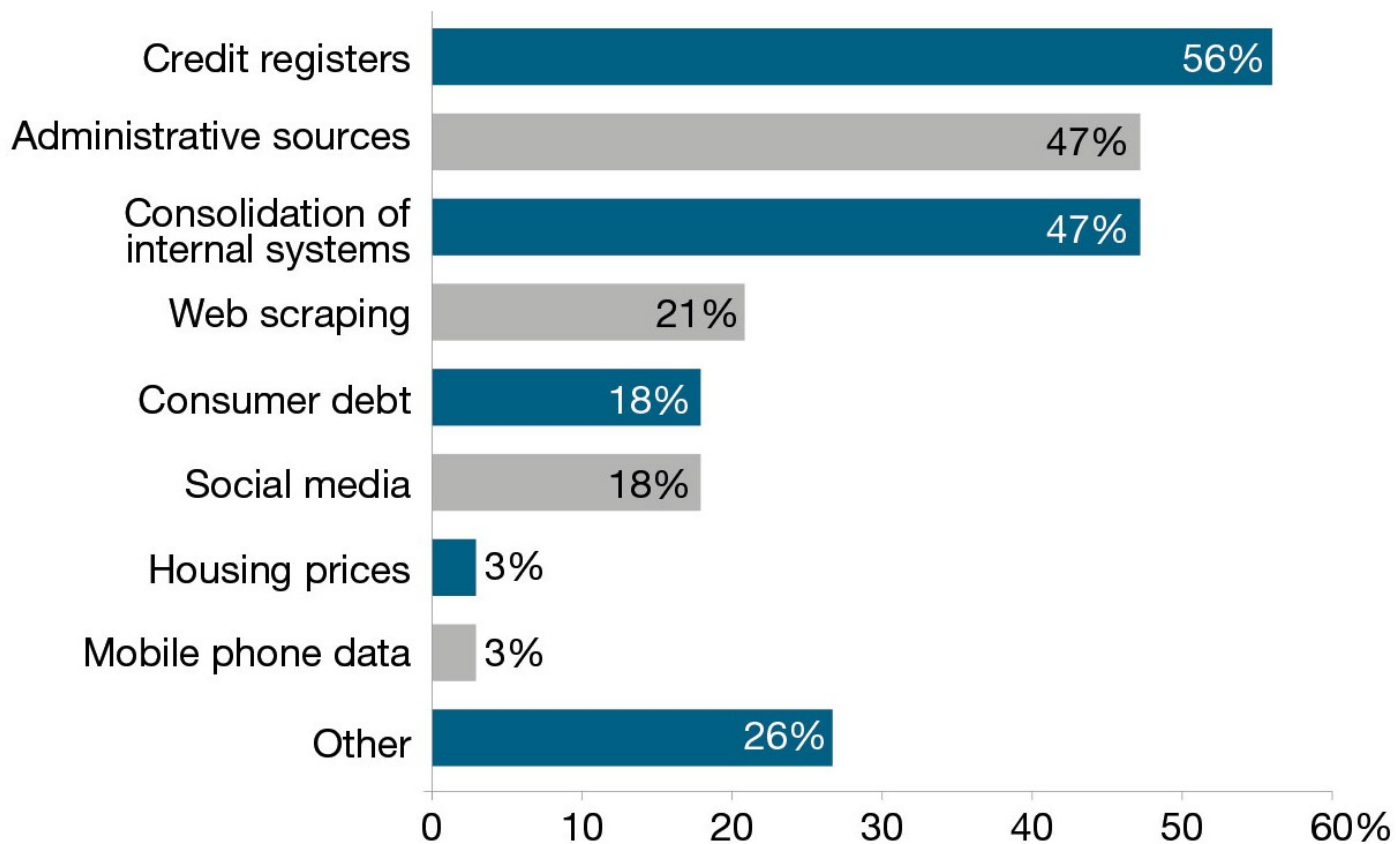
Macrodata: expanding databases – global, more variables

Key: (much) more computing power and storage capacity, better methods



Survey to some 50 central Banks worldwide (Central Banking, 2017): areas in which (big) data projects focused

<https://www.centralbanking.com/central-banks/economics/data/3315546/big-data-in-central-banks-2017-survey>





Another promising avenue: new data sources – UNSTRUCTURED DATA SETS

To deal with new and old phenomena

e.g. agents' behaviour as reflected in internet consumption / search patterns

e.g. mobility through geo-locations

To mitigate informational gaps

e.g. developing countries' data (e.g. satellite-based data)

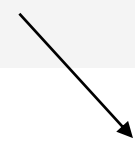
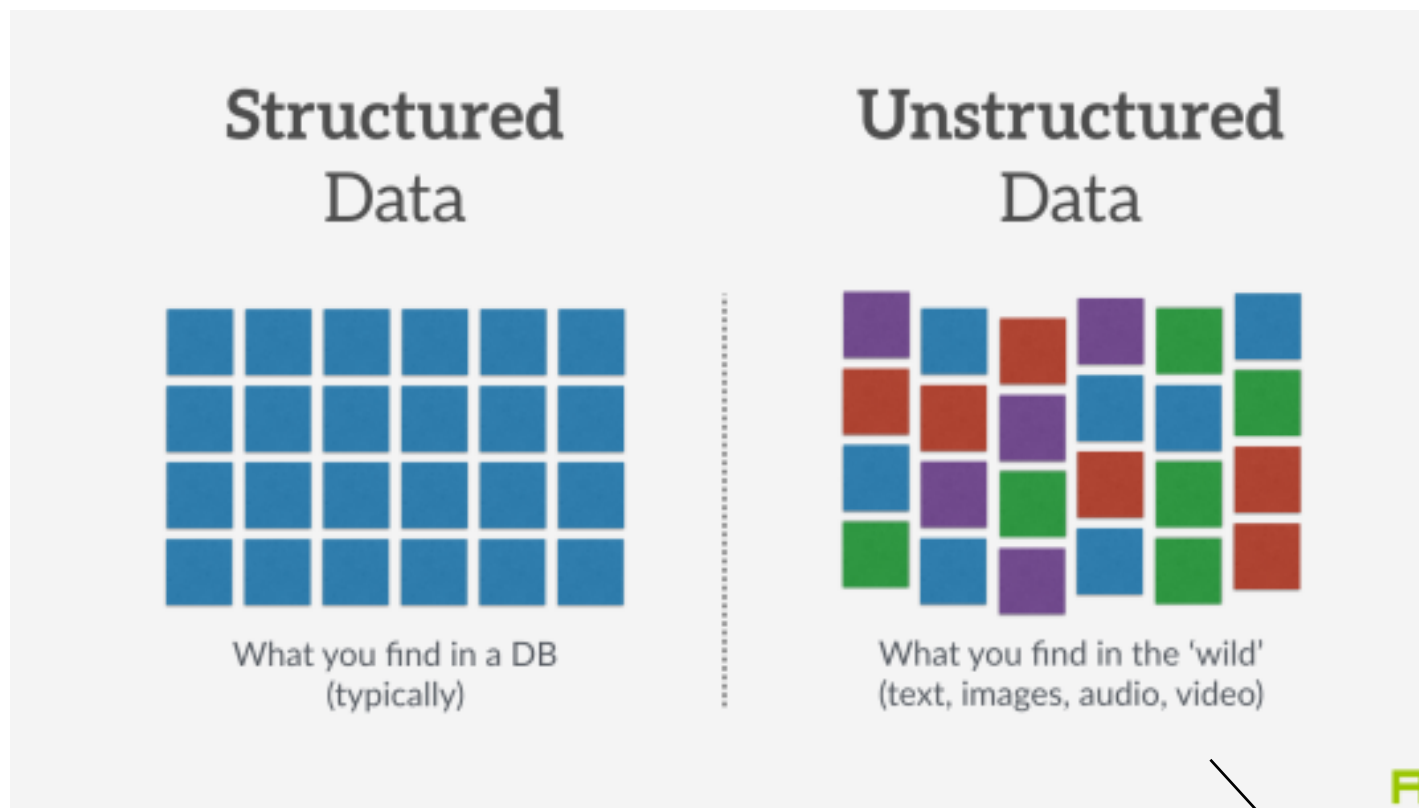
e.g. local data

To improve real-time analysis

e.g. credit card transactions (daily) for consumption

e.g. event analysis

“TRADITIONAL” & “NEW” SOURCES OF DATA FOR ECONOMIC ANALYSIS



**Texts
Google**





Target: extract (quantitative) information from texts → analysis

News articles, social media, market intelligence reports, CB speeches...

A document, a collection of documents (a corpus)

How to separate useful from non-useful information in documents?

Pre-processing steps:

tokenize (break up into words/numbers/punctuations); case-fold (to lower case); stopwords removal (“a”, “the”, ...); stemming+lemmatizing; additional steps

Methodologies:

predefined list of words (Boolean, dictionary text mining) versus infer thematic patterns (LSA, LDA)

Nice introductory reference:

Bholat, D., S. Hansen, P. Santos & Ch. Schonhardt-Baley (2015), “Text mining for central banks”, Centre for Central Banking Studies, Bank of England.



Counting the number of words in groups of documents

Why:

The frequency of words reflect the topic or sentiment expressed in texts

TEXT MINING

EXAMPLE I: COMPARING TWO SETS OF PROPOSALS FOR EMU REFORM (1/4)



Reconciling risk sharing with market discipline: A constructive approach to euro area reform

Agnès Bénassy-Quéré, Paris School of Economics and University of Paris 1; Markus Brunnermeier, Princeton University; Henrik Enderlein, Hertie School of Governance and Jacques Delors Institute Berlin; Emmanuel Farhi, Harvard University; Marcel Fratzscher, DIW and Humboldt University Berlin; Clemens Fuest, Ifo Institute and University of Munich; Pierre-Olivier Gourinchas, University of California at Berkeley; Philippe Martin, Sciences Po, Paris and Conseil d'Analyse Économique; Jean Pisani-Ferry, Bruegel, EUI, Hertie School of Governance and Sciences Po; Hélène Rey, London Business School; Isabel Schnabel, University of Bonn and German Council of Economic Experts; Nicolas Véron, Bruegel and Peterson Institute for International Economics; Beatrice Weder di Mauro, INSEAD and University of Mainz; and Jeromin Zettelmeyer, Peterson Institute for International Economics¹

1 Introduction

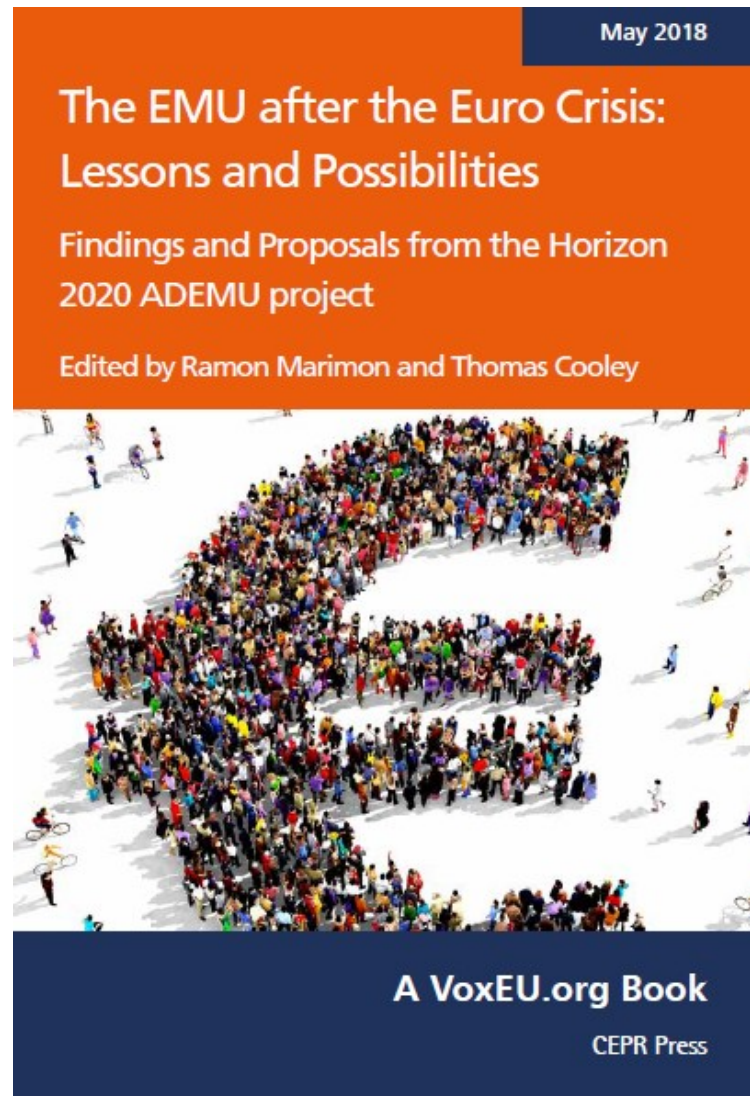
The need to improve the euro area financial architecture to make it less vulnerable to crises and to deliver long-term prosperity to all its members remains as strong as ever. Nevertheless, no meaningful reform has been enacted since the Single Resolution Mechanism was created in July 2014 as part of the banking union initiated in 2012. This reflects a deep disagreement among euro area members on the direction that reforms should take – including between its two largest members, Germany and France. France (along with other members such as Italy) has called for additional stabilisation and risk-sharing mechanisms as well as stronger governance and accountability at the euro area level. In contrast, Germany (along with other members such as the Netherlands) takes the view that the problems of the euro area stem mostly from inadequate domestic policies, that additional euro area stabilisation and risk-sharing instruments could be counterproductive, and that what is really needed is tougher enforcement of fiscal rules and more market discipline.

This Policy Insight was written by a group of independent French and German economists with differing views and political sensitivities but a shared conviction that the current deadlock must be overcome. Reform of the euro area is needed for three reasons: first, to reduce the continued vulnerability of the euro area to financial instability; second, to provide governments with incentives that both encourage prudent macroeconomic policies and deliver growth-enhancing domestic reform; third – and perhaps most importantly – to remove a continuing source of division between euro area members and of resentment of European institutions such as the European Commission and the ECB, which has contributed to the rise of anti-euro populism and which could eventually threaten the European project itself.

We start from the premise that both the French and German positions have a point. To make progress, the major concerns of both sides need to be addressed. The challenge is to do so without ending up with a collection of half-baked compromises. This requires a shift in the euro area's approach to reconcile fiscal prudence with demand policies, and rules with policy discretion. The current approach to fiscal discipline – an attempt to micro-manage domestic policies through complex and

¹ All authors contributed in a purely personal capacity, irrespective of the policy roles they may have or have had in a recent past.

To download this and other Policy Insights, visit www.cepr.org



The EMU after the Euro Crisis: Lessons and Possibilities

Findings and Proposals from the Horizon 2020 ADEMU project

Edited by Ramon Marimon and Thomas Cooley

A VoxEU.org Book

CEPR Press

TEXT MINING

EXAMPLE I: COMPARING TWO SETS OF PROPOSALS FOR EMU REFORM (2/4)



Report 1 CEPR FR-DE

Report 2 CEPR - ADEMU



TEXT MINING

EXAMPLE I: COMPARING TWO SETS OF PROPOSALS FOR EMU REFORM (3/4)

Report 1 CEPR FR-DE



Report 2 CEPR - ADEMU



Report 1: more focus on fiscal rules + country responsibility – market risks (discipline)

**Report 1
CEPR FR-DE**



**Report 2
CEPR - ADEMU**



Report 2: more focus on macro, monetary-fiscal, “central”



- **Medición de la incertidumbre económica y sobre las políticas económicas**

Economic Policy Uncertainty Index (EPU)

MOTIVATION (I)



- **Uncertainty affects economic activity**
- **How to measure uncertainty**
 - E.g. Implied stock market volatility (Bloom, 2009); expectations dispersion (Bachmann et al., 2013), **EPU index (Baker et al., 2016)**, forecast errors (e.g. Jurado et al., 2015).
- **EPU:**
 - Text-based
 - Replicated for many countries, Spain included.
 - Widely used in empirical applications: e.g. Meinen and Roehle (2017); Fontaine et al. (2017); Caggiano et al. (2017); Colombo (2013).



- **Idea: compute the importance of given topic(s) of interest in the press.**
 1. **Dictionary-based (supervised):**
 - *The researcher chooses a set of keywords representing the topics of interest.*
 - *The frequency of the keywords in the text gives a proxy of the relevance of the topics.*
 - **2. Topic modelling (unsupervised):**
 - *The algorithm extracts the topics in the text based on co-occurrence of words.*
- **EPU index is based on method 1.**

MOTIVATION (II)

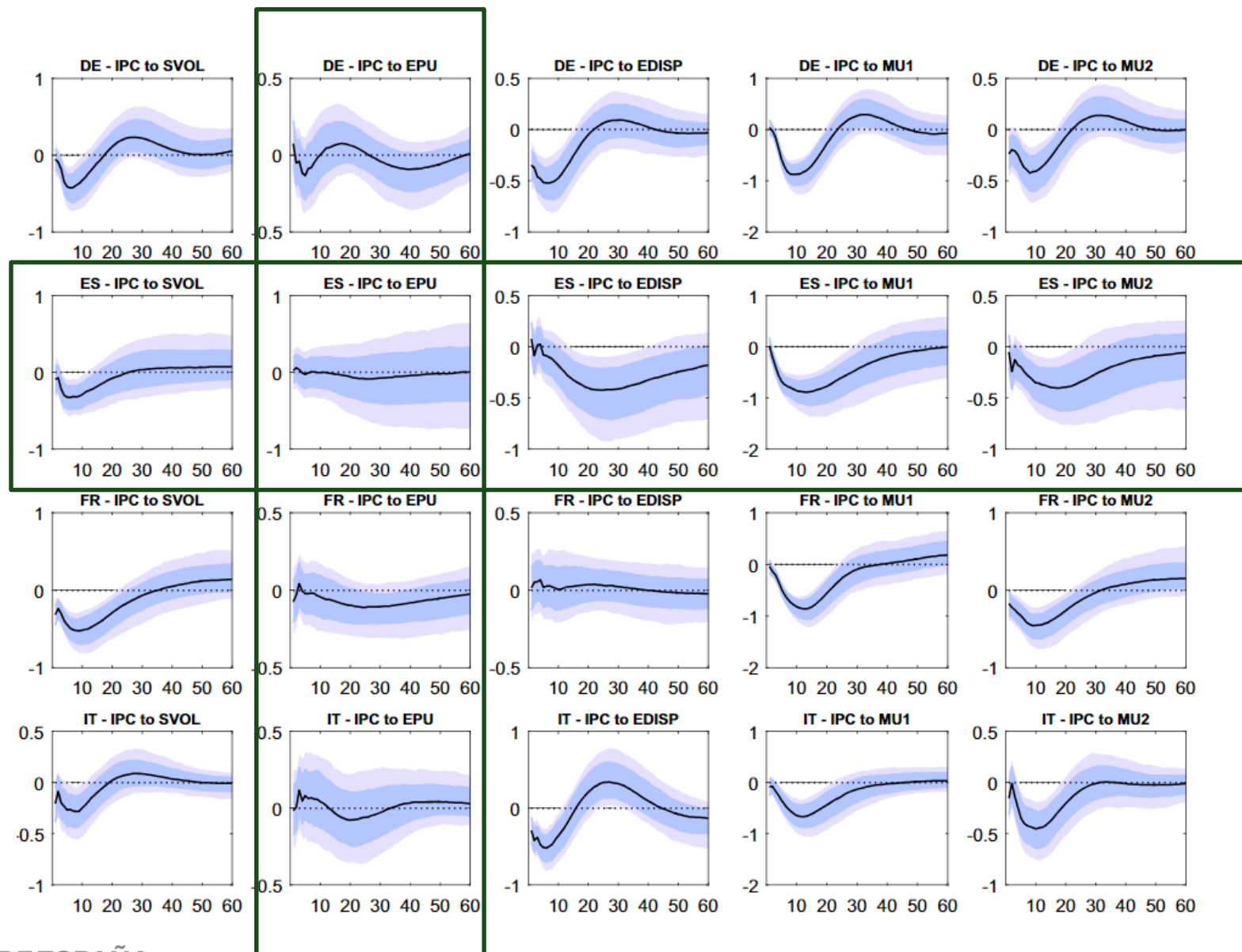


- **Meinen and Roehe (2017):**
 - Impact of different measures of uncertainty on investment: DE, ES, FR, IT
 - *SVOL: stock market volatility (Bloom, 2009)*
 - *EPU: economic policy uncertainty (Baker et al., 2016)*
 - *EDISP: expectation dispersion (Bachmann, 2013)*
 - *MU1: macroeconomic uncertainty (Jurado et al, 2015)*
 - *MU2: macroeconomic uncertainty (Rossi and Sekhposyan, 2015)*
 - EPU-shocks yield zero responses (DE, ES, FR, IT)
 - Negative responses with respect to SVOL, MU1, MU2
- **Gil et al. (2017):**
 - for ES, policy uncertainty yield negative macroeconomic responses

⇒ **Does policy uncertainty affect macroeconomic variables?**

⇒ **Can the construction of EPU be improved?**

MEINEN AND ROEHE (2017): IRF





- **Construct a new EPU index for Spain.**
 - Based on Dow Jones' Factiva database.
 - Follow same methodology of Baker et al (2016), but expand:
 - *Press coverage*
 - *Keywords*
 - *Time span*
- **Compare new vs original EPU (Spain).**
 - Macroeconomic effects of uncertainty based on SVAR.
 - GDP, private consumption and capital goods investment.
- **Findings:**
 - Negative responses in to unexpected shocks in new EPU.
 - Zero effects when using the original EPU.

THE EPU INDEX FOR THE US



- Weighted average of 3 components:
 1. newspaper coverage of policy-related economic uncertainty.
 2. number of federal tax code provisions set to expire in future years
 3. disagreement among economic forecasters (CPI, purchase of goods and services by state, local, federal governments)

FOCUS ON COMPONENT I (which has been replicated for other countries)

- Press coverage: 10 large newspapers (USA Today, the Miami Herald, the Chicago Tribune, the Washington Post, the Los Angeles Times, the Boston Globe, the San Francisco Chronicle, the Dallas Morning News, the New York Times, and the Wall Street Journal)
- Time coverage: From January of 1985
- Keywords:
 - (U) uncertainty, uncertain, uncertainties;
 - (E) economic or economy;
 - (P) congress, deficit, Federal Reserve, legislation, regulation, white house, regulatory, the Fed.

THE BBD EPU INDEX FOR SPAIN



- **Press coverage:**
 - El País, El Mundo
- **Time span:**
 - Monthly, from January 2001 onwards
- **Keywords:**
 - **Uncertainty:** *incierto OR incertidumbre*
 - **Economy:** *económica OR economía*
 - **Policy:** *déficit OR presupuesto OR tarifa OR gastar OR gasta OR gasto OR impuesto OR política OR banco central OR regulación*

OUR NEW EPU INDEX FOR SPAIN (I)



- **Press coverage:**
 - El País, El Mundo, [La Vanguardia](#), [ABC](#), [Expansión](#), [Cinco Das](#), and [El Economista](#).
- **Time span:**
 - Monthly, from January [1997](#) onwards.

OUR NEW EPU INDEX FOR SPAIN (II)



- **Keywords:**

- **Uncertainty:** Inciert* OR incertidumbr* OR inestabl* OR inestabilidad/inestabilidades OR riesgo/riesgos.
- **Economy:** *económic* OR economía.*
- **Policy:** déficit OR presupuest* OR ~~tarifa~~ ~~OR~~ ~~gastar~~ ~~OR~~ ~~gasta~~ ~~OR~~ gasto publico/gastos públicos OR deuda publica/deudas publicas OR (el OR de OR del OR un OR por OR este OR ese OR aquel) w/1 impuesto OR impuestos OR política fiscal/politicas fiscales OR política monetaria/politicas monetarias OR Banco Central Europeo/BCE OR Banco de España/BdE OR Parlamento OR Moncloa OR gobierno central OR Hacienda OR Comisión Europea OR legislación/legislaciones OR reforma/reformas OR norma/normas OR normativ OR regulacion/regulaciones OR reglamento/reglamentos OR ley/leyes.

OUR NEW EPU INDEX FOR SPAIN (III)



Annual Report

2017

1	The euro area's economic recovery was supported by tailwinds and monetary policy	8
1.1	Global economic growth gained momentum	8
1	La recuperación económica de la zona del euro se vio respaldada por factores favorables y por la política monetaria	9
1.1	El crecimiento económico mundial cobró impulso	10

TIME COVERAGE BY NEWSPAPER

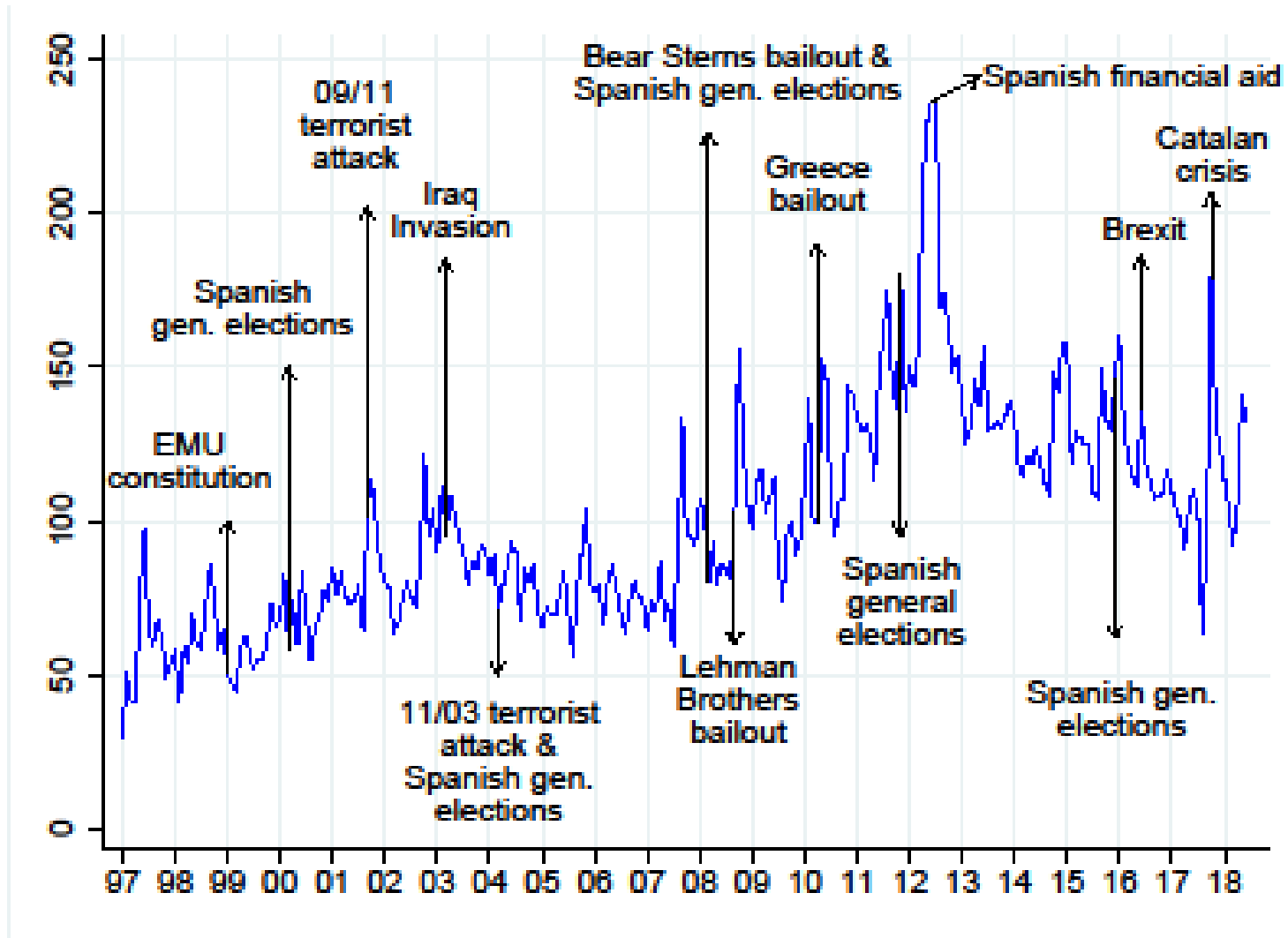


Newspaper	Type	First obs.	Last obs.	# months
El Mundo	Generalist	1997m1	2018m6	258
Expansion	Business	1997m1	2018m6	258
Cinco Dias	Business	1997m1	2018m6	258
La Vanguardia	Generalist	1997m5	2018m6	254
Abc	Generalist	1997m6	2018m6	253
El Pais	Generalist	2001m1	2018m6	210
El Economista	Business	2008m4	2018m6	123

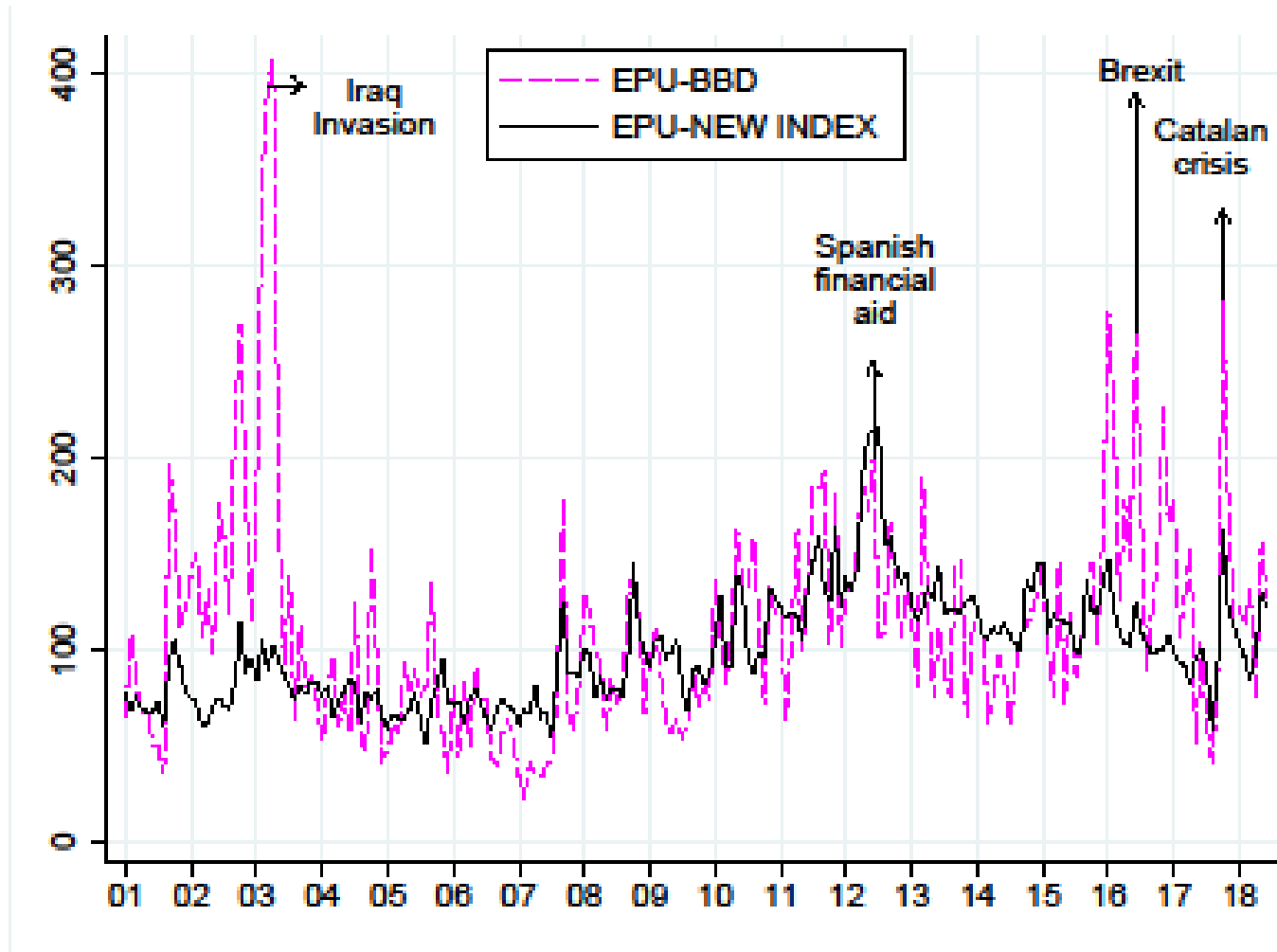
BUILDING THE INDEX

- **Scale raw count by # articles in same newspaper/month.**
- **Standardize monthly series of scaled counts.**
- **Average across the newspapers.**
- **Rescale resulting index to mean 100.**
- **We construct:**
 - Aggregate EPU index;
 - Business/Generalist news EPU index;
 - Newspaper-specific EPU indexes.

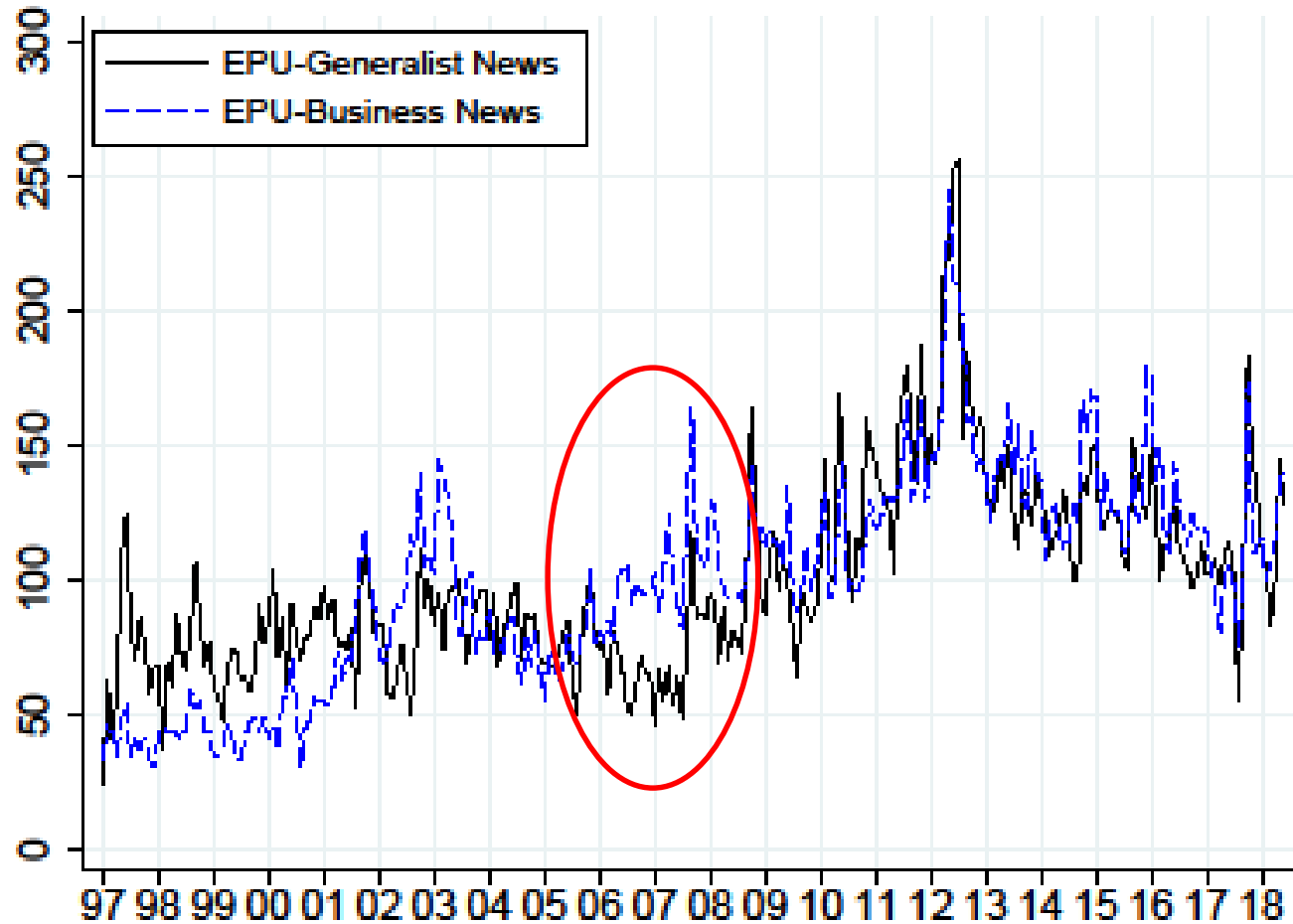
AGGREGATE EPU



NEW EPU VS BBD-EPU



BUSINESS VS GENERALIST NEWS EPU



STRUCTURAL VAR MODEL

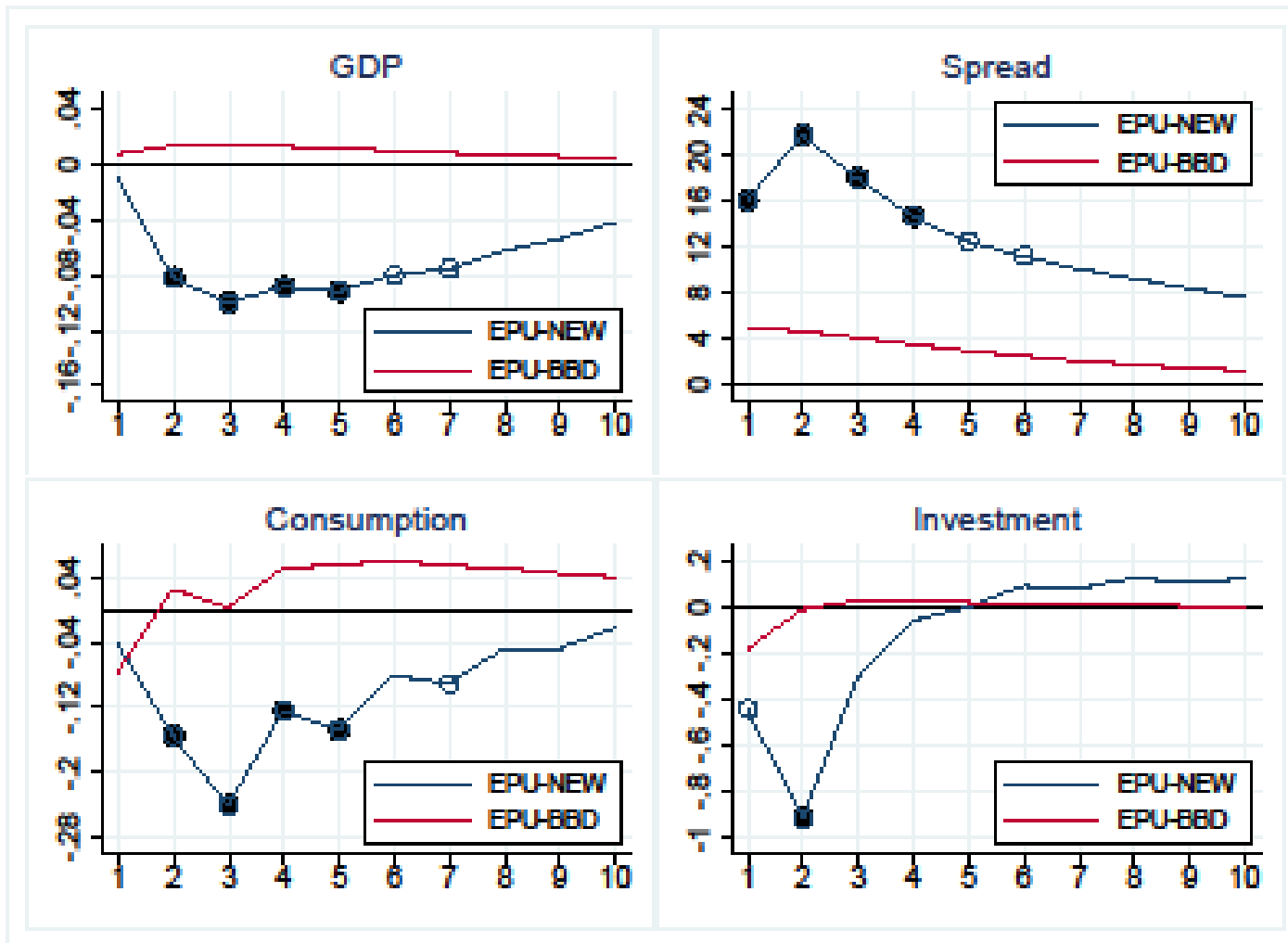


$$X_t = \Pi(L)X_t + \beta GlobalEPU_t + \epsilon_t, \quad \epsilon_t = N(0, \Omega)$$

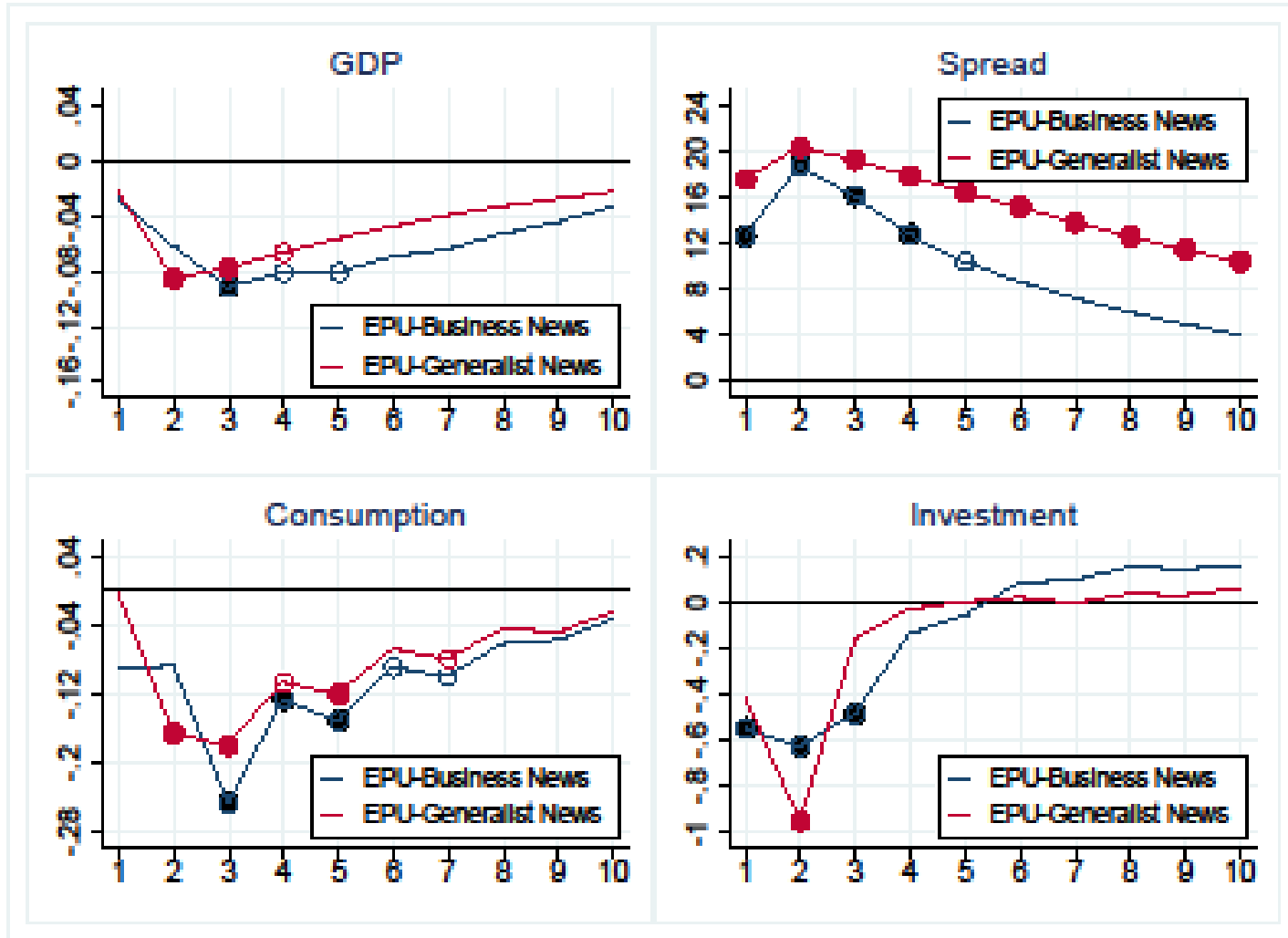
$$X_t = \begin{pmatrix} EPU \\ spread \\ real\ GDP\ (*) \\ inflation\ rate \end{pmatrix}$$

- (*)Alternatively: private consumption, capital goods investment.
- OLS estimation.
- Optimal lag length (SIC).
- IRF to shocks in EPU => Cholesky decomposition.
- Robustness: (i) variable ordering; (ii) lag length.

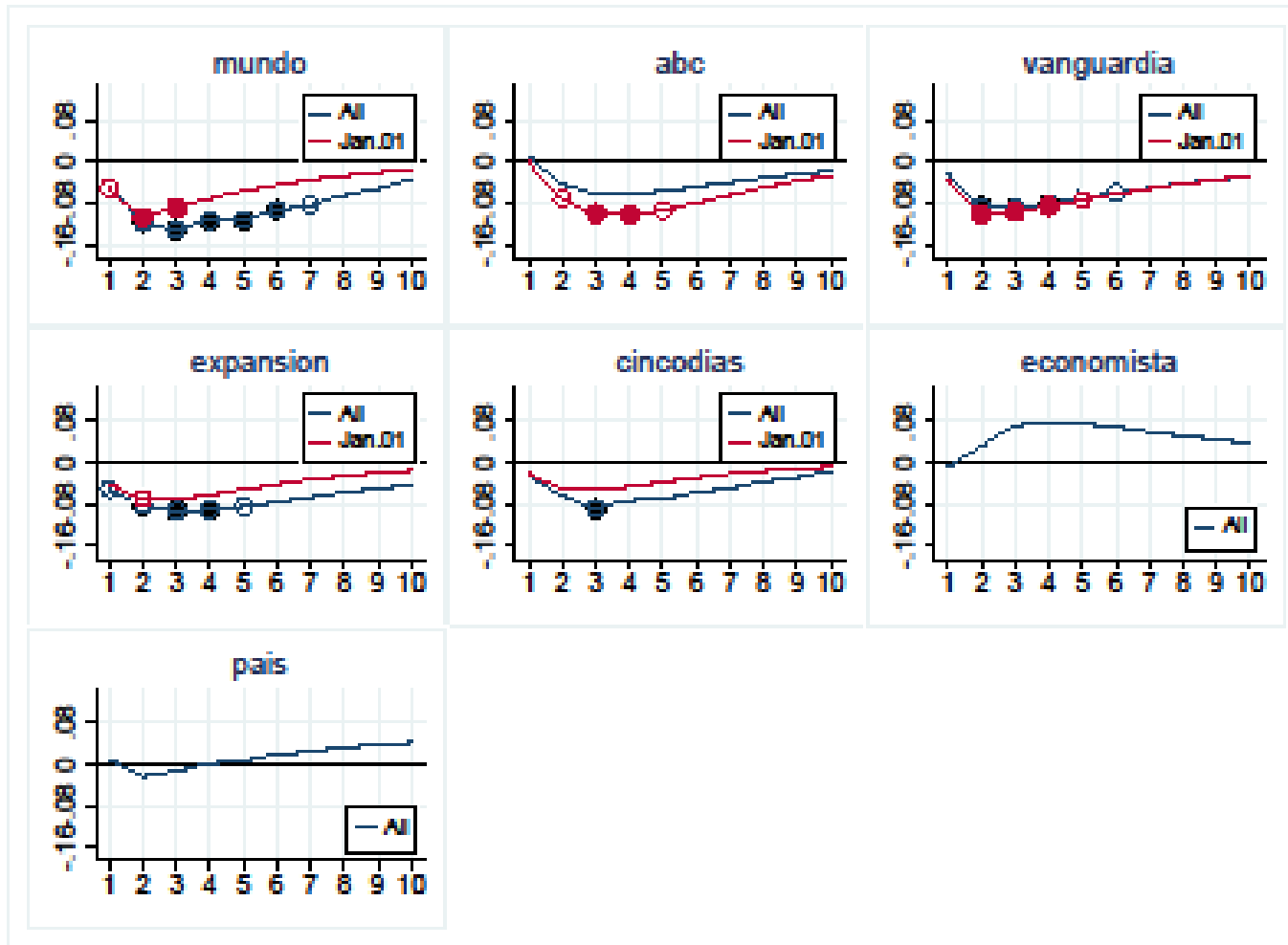
IRF: NEW EPU VS BBD-EPU



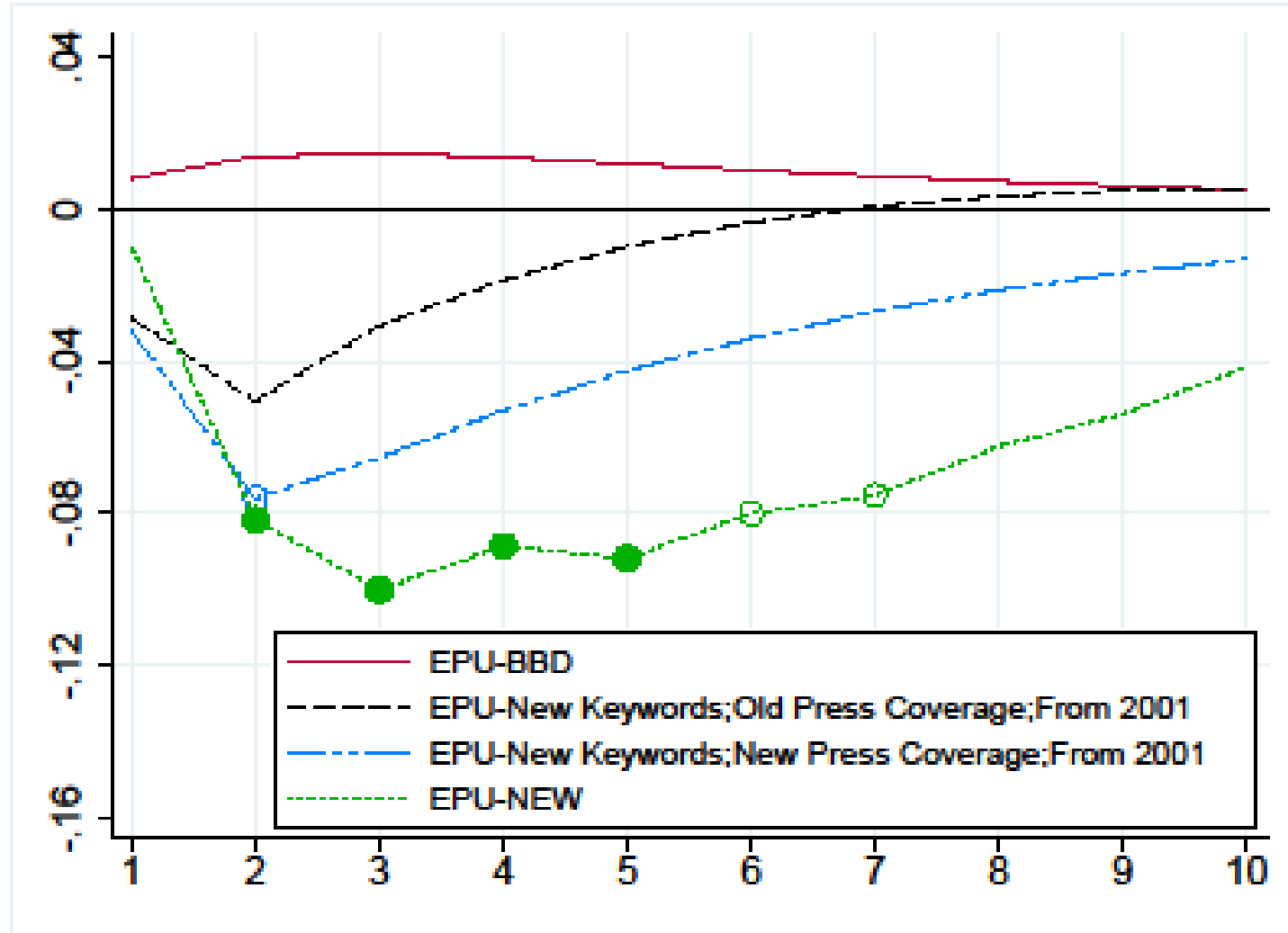
IRF: BUSINESS VS GENERALIST EPU



IRF: NEWSPAPER-SPECIFIC EPU



RELATIVE CONTRIBUTION OF KEYWORDS, PRESS, AND TIME COVERAGE: GDP RESPONSES





- **New EPU shows more consistent chronology of relevant events.**
- **Macroeconomic effects of EPU:**
 - Significant dynamic relationship between new EPU and macroeconomic variables.
 - Not significant effect with BBD-EPU.
- **Heterogeneous effects by press/time coverage and keywords:**
 - Widening the press coverage and enriching the set of keywords is key.
 - Enlarging the time coverage improves the precision of the estimates



- **Comunicación cualitativa sobre la situación económica: Informe Trimestral del Banco de España desde 1T 1999 hasta la actualidad**

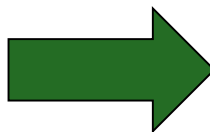
1. Rasgos básicos

NOTA PREVIA

Como consecuencia del comienzo de la Unión Económica y Monetaria (UEM), el pasado 1 de enero, el Banco de España está revisando sus publicaciones con el fin de adaptarlas al entorno de una moneda única y una política monetaria común. Sin embargo, tanto este Boletín como el Boletín estadístico del mes de enero, de próxima aparición, se encuentran en fase de transición, ya que, si bien los nuevos marcos analíticos y conceptuales están bien definidos, el ámbito temporal de estas publicaciones sigue estando referido, básicamente, a la etapa previa al inicio de la UEM (datos hasta diciembre de 1998). A pesar de que estas ediciones presentan ya numerosas novedades (por ejemplo, se generaliza la difusión de la información expresada en euros), en los próximos meses se irán incorporando los cambios pendientes hasta la total adaptación a la nueva situación.

* * *

En un entorno internacional incierto, debido a la persistencia de factores desestabilizadores en algunos mercados financieros, la economía española ha finalizado 1998 y ha iniciado su camino dentro de la Tercera Fase de la Unión Económica y Monetaria (UEM) en unas condiciones favorables. El ritmo de crecimiento real es elevado y la suave tendencia a la desaceleración que se ha apreciado en la segunda mitad del año se debe a la caída que han experimentado las exportaciones, como consecuencia del menor dinamismo que se registra en algunos países industriales (véanse cuadro 1 y gráfico 1). En el cuarto trimestre, la demanda nacional mantuvo un alto ritmo de expansión y el crecimiento del empleo fue notable, conservando un buen tono los niveles de confianza de los consumidores, lo mismo que ocurre en los restantes países de la UEM (véase gráfico 2). Solo la actividad industrial ha dado signos de debilidad, y así lo han reflejado las opiniones de los empresarios, por lo que se ha atenuado el crecimiento de la inversión en equipo; la fortaleza de la inversión en capital fijo, sin embargo, siguió siendo considerable, a lo que ha contribuido decididamente el sector de la construcción. La continuidad de la fase expansiva, el descenso registrado en los tipos de interés y la persistencia en el control del gasto han permitido que la economía española reduzca el déficit de sus cuentas públicas más allá de lo que se había previsto, mientras que el incremento interanual de los precios de consumo al final del año resultó históricamente bajo, siendo de medio punto porcentual el diferencial con los países del área del euro, en términos de los índices armonizados (véase gráfico 3). La econo-



INFORME TRIMESTRAL DE LA ECONOMÍA ESPAÑOLA 1 PRINCIPALES MENSAJES

En el tramo final de 2018 ha proseguido la expansión de la actividad mundial, si bien han ido haciéndose patentes algunas muestras de desaceleración, observándose además un comportamiento por áreas geográficas más heterogéneo que al comienzo del año. Esta evolución estaría reflejando la materialización parcial de algunos de los focos de incertidumbre que vienen acechando a la economía mundial desde el inicio del año, en particular las tensiones comerciales y un cierto endurecimiento de las condiciones financieras globales. Ello se ha traducido en varios episodios de volatilidad en los mercados financieros, que se han saldado con retrocesos de las cotizaciones bursátiles y repertes de los diferenciales corporativos.

A futuro, no cabe descartar la posibilidad de un recrudecimiento de la incidencia de estas dos fuentes de incertidumbre, a las que cabe añadir otras adicionales, en el ámbito europeo, como el devenir de la política presupuestaria en Italia o la incertidumbre en torno a la forma en que se producirá finalmente la salida del Reino Unido de la Unión Europea. Además, más recientemente, han resurgido las preocupaciones acerca del curso de la economía china, donde han aparecido señales crecientes de debilidad de la actividad, a pesar de los estímulos proporcionados por las políticas de demanda. En un contexto en el que el endeudamiento de los agentes ha aumentado muy rápidamente en los últimos años, existe el temor de que una hipotética profundización de la desaceleración en esta economía pueda afectar negativamente a la estabilidad de su sistema financiero. En este contexto, una incógnita es la capacidad de las autoridades para revertir la desaceleración del producto a través de la aplicación de una nueva ronda de estímulos monetarios y fiscales.

En el área del euro, se ha prolongado en el segundo semestre del año la fase de crecimiento más moderado ya observada desde su inicio. En parte, estos desarrollos se deberían al encadenamiento de distintos factores de naturaleza transitoria. El último de ellos habría sido el impacto sobre la industria del automóvil de la introducción de una nueva norma acerca de las emisiones contaminantes de los vehículos. No obstante, parecen acumularse cada vez mayores indicios de la existencia de un componente más persistente, ligado al debilitamiento del contexto exterior del área, que explicaría esta etapa de crecimiento más contenido. En esta dirección apunta la pérdida de dinamismo de las exportaciones dirigidas hacia el resto del mundo. Estos desarrollos han tenido su reflejo en las últimas proyecciones del Eurosistema, que han revisado a la baja en una décima el crecimiento esperado del PIB en 2018 y 2019. El Consejo de Gobierno del BCE decidió en diciembre, al objeto de asegurar que la inflación converja a su objetivo de medio plazo, que, después del cese de las compras netas de activos a fin de año, reinvertirá, durante un tiempo prolongado tras el inicio de la subida de tipos de interés, los activos de la cartera adquiridos en el marco del programa que vayan amortizándose.

La economía española habría crecido, en el cuarto trimestre del año, de acuerdo con la información disponible hasta el momento, a una tasa intertrimestral del 0,6%, al igual que en los dos trimestres precedentes (véanse gráfico 1 y cuadro 1). La desaceleración de la actividad en España a lo largo de 2018 habría sido notablemente menos intensa que en el conjunto del área del euro, lo que, en parte, podría deberse a la inclusión en los Presupuestos Generales del Estado (PGE) para 2018, que entraron en vigor en el verano, de algunas medidas que han impulsado las rentas de los hogares. Además, la introducción



Indicador de positividad del lenguaje basado en el texto de los informes trimestrales del Banco de España (en particular, de la sección de rasgos básicos)

La base es un diccionario con palabras y valores asociados, por ejemplo:

dinamismo: +1.00	retroceso: -0.50
robusto: +1.00	desequilibrio: -0.50
fortaleza: +1.00	turbulencias: -0.75
crecimiento: +0.75	recesión: -1.00
favorable: +0.50	colapso: -1.50

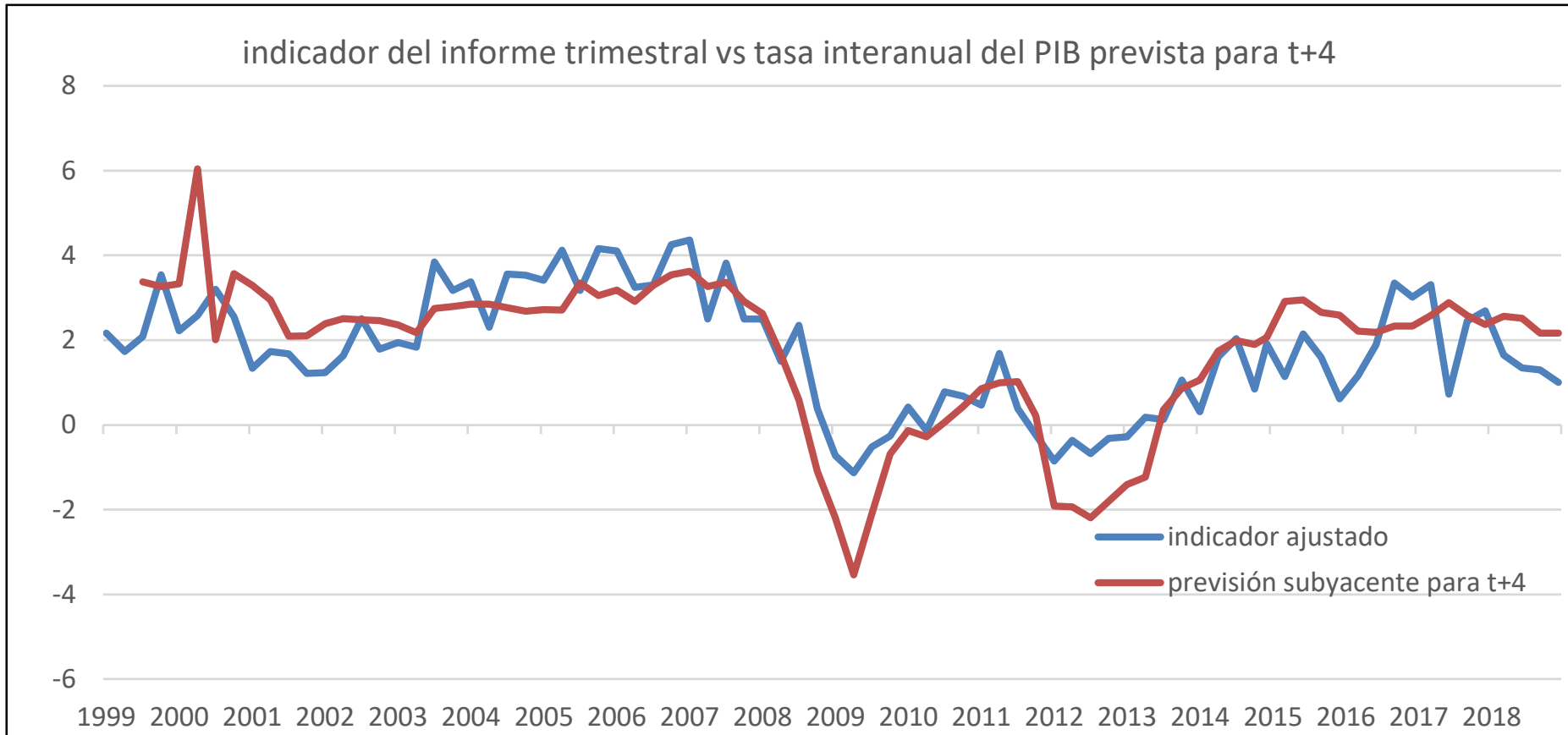
(la lista completa contiene 90 términos y expresiones, más sus variantes sintácticas)

Aplicando ese diccionario sobre los textos, se construye un indicador trimestral, y se compara con la previsión de crecimiento trimestral del PIB (nowcast) que se publicó en el informe correspondiente, y con la previsión a un año vista (interanual de t+4) que estaba implícita en las previsiones que sirvieron de base para escribir el informe

Referencias: Tones, Sinclair and Stekler (2018), ...



El indicador resultante, que cuantifica la positividad del lenguaje con que se escribió cada informe trimestral, se aproxima bastante a las previsiones subyacentes:



Además, las diferencias entre las variables pueden explicarse en términos de riesgos de la previsión (crisis exterior en 2001, dudas sobre la recuperación desde 2015)



Estos resultados nos indican que, incluso si no se publicaran las cifras de las previsiones, el lector podría deducir una señal adecuada sobre ellas a partir del lenguaje utilizado en los informes

IMPLICACIONES DE LAS NUEVAS FUENTES PARA LA ESTADÍSTICA OFICIAL



IMPLICACIONES DE LAS NUEVAS FUENTES PARA LA ESTADÍSTICA OFICIAL



- Las nuevas fuentes también abren **posibilidades** hasta ahora sin explotar para la elaboración de estadísticas, aunque se viene trabajando con información granular desde hace mucho tiempo
- Un efecto positivo para la estadística oficial del auge de estas nuevas fuentes de datos procede del desarrollo del **instrumental técnico** para manejar esa ingente información
- **En el ámbito de las estadísticas no financieras** ha habido posiblemente más iniciativas para explotar las nuevas fuentes de información
 - medición de precios (mediante técnicas de web scraping)
 - algunas partidas del comercio exterior (estimando, por ejemplo, los movimientos de turistas mediante el seguimiento de las redes de telefonía móvil)
 - En algunos países con más dificultades para establecer una infraestructura estadística sólida, estas nuevas fuentes están comenzando a utilizarse incluso para realizar estimaciones de algunos agregados de la Contabilidad Nacional [véase Hammer et al. (2017)].
- **“Competencia”**: finalmente, debe reseñarse que la abundancia de información disponible supone también un reto para la estadística oficial, pues permite que, a partir de esos datos, cualquier usuario sea capaz de generar su propia medición de algún fenómeno económico y difundirla

