A Real Time Leading Economic Indicator based on Text Mining for the Spanish Economy. Fractional Cointegration VAR and Continuous Wavelet Transform Analysis

Carlos Poza

Universidad Francisco de Vitoria, UFV

Manuel Monge*

Universidad Francisco de Vitoria, UFV

Abstract

The main aim of this paper is to build a Real Time Leading Economic Indicator (RT-LEI) that improve Composite Leading Indicators (CLI)'s performance to anticipate GDP trends and turning points for the Spanish economy. The indicator has been constructed by means of a Factor Analysis and is composed of 21 variables concerning motor vehicle activity, financial activity, real estate activity, economic sentiment and industrial sector. The data sources used are Google Trends and Thomson Reuters Eikon-Datastream. This work contributes to the literature, studying the dynamics of GDP, CLI and RT-LEI using Fractional Cointegration VAR (FCVAR model) and Continuous Wavelet Transform (CWT) for its resolution. The results show that the model does not present mean reversion and it is expected the RT-LEI reveals a bear trend in the next two years, alike IMF and Consensus FUNCAS' forecasts. The reasons are mostly associated with escalating global protectionism, uncertainty related to Catalonia and a faster monetary policy normalization.

Keywords: Leading Economic Indicators, Business Cycle, Google Trends, Fractional Cointegration VAR, Wavelet Analysis.

JEL Classification: E32, E37

Corresponding author: Manuel Monge

Universidad Francisco de Vitoria (UFV)

Faculty of Law and Business

E-28223 Madrid

Spain

Email: manuel.monge@ufv.es

* Manuel Monge gratefully acknowledges financial support from an internal Project from the Universidad Francisco de Vitoria (UFV), Madrid.

1. Introduction

One of the interests' key in macroeconomics and business cycles is to identify series that moved in and out of growth and recession before the rest of the economy. Frequently government agencies and research institutions face great challenges to forecast trends and turning points in analysis of countries and international economic outlook. The first challenge in forecasting real activity is related to the data revisions on target and indicator variables which implies that early available information as GDP or early available predictors for GDP may differ from later-available information for a specific point in time. Releasing economic activity indicators with a lag of time presents important implications for evaluating forecasts from different models, since they restrict the ability to accurate assess current conditions.

According to Heinisch and Scheufele (2018), macroeconomic forecasters often ignore the fact that data used for conducting forecasts may be revised from time to time. This is due because forecast typically depend strongly on past observations (available data vintage of a given variable) via parameter estimates and starting levels. Several authors such as Stock and Watson (2003) and Banerjee, Marcellino, and Masten (2005) simulate the real-time situation applying only a rolling or recursive estimation scheme, instead of using data that are available at the time the forecasts are made. Other researchers like Stark and Croushore (2002), Kozicki (2002), and Croushore (2011) criticize this procedure. They advocate the use of real-time data in forecast evaluations.

The publication and the delay of these dataset make nowcasting or the prediction of the present an important source. For this reason, the models should be evaluated with the data available at the point when the forecasts are made. According to Choi and Varian (2012), a real-time daily and weekly index of the volume of queries named Google Trends is often correlated with various economic indicators and may be helpful for short-term economic prediction. In other words, Text Mining, in general, and Google Trends, in particular, may help in predicting the present that is a form of contemporaneous forecasting or nowcasting.

The use of big data, in our case, Google Trends, could be a particular useful for nowcasting and the construction of early estimates, namely for the production of a preliminary estimate for the contemporaneous value of an economic indicator, which has not yet been officially released, like Gross Domestic Product (GDP) which are typically released at least 30-45 days after the end of the reference month or quarter, and later revised. Therefore, in this paper we focus on the use of Google Trends for production of a leading economic indicator that improve existing ones. According to Camacho and Perez-Quiros (2002), the forecasting problem is twofold. First, it is important to identify the group of variables that move in and out recessions before the rest of the economy. And second, we have to find the appropriate filter to extract the signal out of these series.

There are two main purposes for this paper. First, constructing a real time leading economic indicator combining a group of variables from Google Trends (mainly) and Thomson Reuters Eikon to accurate the Spanish GDP behaviour. And secondly, applying a methodology based on wavelets method to analyze and forecast the behavior of that leading economic indicator in time-frequency domain.

The remainder of this paper is structured as follows: Section 2 briefly reviews the literature on these issues. Section 3 presents the data, the model and the methodology applied in the paper. Section 4 presents the main empirical results, while Section 5 concludes the paper.

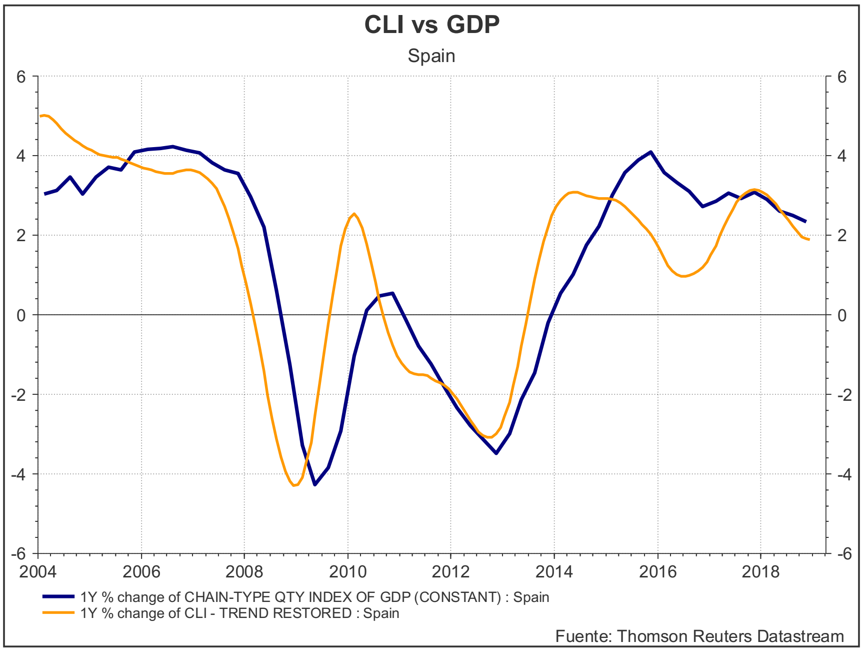
2. Literature review

2.1. Big data and economic indicators: a new approach to measure business cycles

Economic indicators play an important role in policy-making, providing a foundation for discussions and decisions concerning economic measures. However, the timing can be very important at economic policy measures because much of the official data are too delayed to be useful. One example could be GDP figures, that are released quarterly.

In order to identify turning points in business cycle and know ahead (3-6 months) the economic activity trends, many institutions monthly publish composite leading economic indicators (Baumohl, 2009) such as OECD (CLI – Composite Leading Indicator. See Chart 1), The Conference Board (Index of Leading Economic Indicators), The Economic Policy Uncertainty (Economic Policy Uncertainty Index) or The Institute for Supply Management (PMI – Purchasing Managers Index). In Spain, the Ministry of the Economy and Finance releases every month the Synthetic Indicator of Activity.

Chart 1. Knowing ahead GDP by means of CLI (OECD) in Spain



Notwithstanding, the frequency of releasing of those indicators is not as current as demanded by markets, that try to use others published in a daily basis: interest rates, share prices and so on. In this regard, and thanks to big data, nowadays we have the possibility to analyze massive data in real time from the Internet to forecast the economy with more current information (Antenucci et al., 2014; Bollen et al., 2011; D'Amuri and Marcucci, 2017; Dong et al., 2017; Kozicki, 2002; Pappalardo et al., 2016; Stark and Croushore, 2002; and Szármes, 2015).

Nowcasting has recently become popular in economics. The term "nowcasting" is a contraction of "now" and "forecasting". Standard statistical figures describing the state of an economy (e.g. the GDP) are based on a rigorous method of collecting data, and they are accurate and reliable, but are published with a lag of some weeks or months. Signals about the GDP can be extracted from heterogeneous data sources (e.g., unemployment figures, industrial orders, web search data, etc.) before GDP itself is published. In nowcasting, these data are used to compute sequences of current quarter GDP estimates in relation to the real time flow of data releases. With big data, data is used as signals to anticipate what's going to happen and intervene in time (Szármes, 2015).

Antenucci et al. (2014) analyzed billions of tweets for references to unemployment (hunting for expressions like "axed", "pink slip" or "downsized"). They indexed the findings and compared them to government statistics building a social media signal of job loss that closely tracks initial claims for unemployment insurance. Despite differences in the underlying processes generating unemployment insurance claims and tweets about job loss, the indicator based on social media tracks the official data remarkably well.

Dong et al. (2017) states that thanks to emerging trends in the use of smartphones, online mapping apps and social media along with geo-located data, generate new tools to measure economic dynamics in real time, which directly reflect user's social and economic behavior. To some extent, these new tools overcome limitations on the timeliness and sample size of traditional surveys, and deeply boost empirical research.

These authors measure economic activity in China by means of bottom up view. Firstly, they build indices for gauging employment and consumer trends based on billions of geo-positioning data. Secondly, they advance the estimation of offline store foot traffic via location search data derived from Baidu Maps. And thirdly, they construct consumption indicators to track trends in several service sector industries.

This paper could be included in the research concept of Mobimetrics, which dedicated to quantifying social system dynamics by analyzing massive individual mobility data generated by smartphones, wearable devices, driverless cars and the Internet of Things with machine learning approaches (Dong et al., 2017). In this regard, Varian (2014) demonstrates the possibility of using Google search query indices for short-term economic prediction. In addition, web search data have proven to be helpful in forecasting consumer behavior and financial market activity. Researchers have used Twitter data to create indices to predict economic activity, unemployment and equity market activity. Toole et al. (2015) track employment shocks using mobile phone Call Detail Records in Europe. Pappalardo et al. (2016) propose a data-driven analytical framework to "nowcast" socioeconomic indicators using mobility features extracted from mobile phone data.

Nyman et al. (2014) describe a new approach to economic forecasting, which is based on big data and a new methodological approach termed "Directed Algorithmic Text Analysis". This approach is associated with searching in textual databases. Notwithstanding, this method is based upon a theory of human decision making under radical uncertainty, called "conviction narratives", that is drawn on modern neurobiological theories in which cognition is associated with the functions of observable brain networks and allied neurobiological processes. It is built on the view of emotion as a human resource: excitement about possible future profit versus anxiety about possible future loss.

These authors analyze text archives, such as Reuters news, company emails, internal memoranda or broker reports, to extract time series of relative sentiment shifts that may forecast aspects of the economy. The results show that text analysis with a database of brokers' reports can significantly forecast the Michigan Consumer Index published in the US.

Scott and Varian (2012) state that search engine queries in the "vehicle shopping" category could be good candidates for forecasting automobile sales, while queries such as "file for unemployment" could be useful in forecasting initial claims for unemployment benefits.

Traditional recession prediction literature has relied on findings economic and financial indicators that hold predictive power in forecasting downturns. For example, numerous papers have studied the predictive capabilities of the yield curve spread between the 10-year treasury rate and the three-month treasury rate (Estrella and Trubin, 2006) or the slope, level and curvature of the yield curve, to forecast business cycle (Ang and Piazzesi, 2003).

According to Huang, Rojas and Convery (2018), the importance of understanding sentiment in the context of macroeconomic activity, in order to successfully forecast downturns, has been deeply discussed in literature. Sentiment indicators are crucial to understanding recessions because they capture a sense of animal spirits (Keynes, 1936). Bernanke (2008) states that, from a fundamental level, a sentiment indicator can account for the fact that simply the perception of how the economy is doing influence the actual performance of the economy. Koenig (2002) points out that sentiment indicators have the advantage over traditional leading indicators because the first ones are available in real-time and the second ones with some delay.

Huang, Rojas and Convery (2018) seek to add to the existing recession prediction literature by studying the potential for using not only consumer and business confidence as a measure for sentiment, but also the underlying sentiment of news itself. They examine two components of news: The first is the general positive and negative nature of the articles being published, and the second is the overall concentration of topics in the news. Using keywords, the New York Time and text mining, they build a consistently statistically significant indicator with predictive capabilities. Thus, thanks to big data, a lot of unstructured data may be used as a potential sentiment indicator. For example, Bollen et al. (2011) found that applying sentiment analysis across Twitter feeds could help predict directional stock market movements, Hisano et al. (2013) used topic modelling across news articles to forecast market volatility, D'Amuri and Marcucci (2017) demonstrated that there existed potential for using Google searches in forecasting unemployment in the US, and as aforementioned, Choi and Varian (2012) used Google Trends data to perform short-term forecasts of economic indicators, such as automobile sales, unemployment claims and travel destination planning.

After reviewing many papers regarding the binomial big data – economic activity, we observe three kinds of unstructured data to forecast business cycles: 1) social media (Twitter or Facebook), 2) internet searches (Google Trends) and 3) news sentiment.

As aforementioned in the introduction, the main goal of this paper is to construct a composite leading economic indicator based on Google Trends for Spain, what allows us to deepen in the correlation of key internet searches and economic activity.

As one of the most important LEIs to know ahead GDP is the CLI (OECD), even for Spain, we are going to describe in detail its scope and composition, because the idea is to present a new LEI based on Google Trends statistically correlated with it.

2.2. The Composite Leading Indicator – OECD as a predictor of GDP trend

The OECD system of composite leading indicators was developed in the 1970's to give early signals of turning points of economic activity. This information is of a prime importance for economists, businesses and policy makers to enable timely analysis of the current and short-term economic situation. They are supposed to predict short-term movements in an economy by using measures that are highly sensitive to upcoming changes in business conditions (Baumohl, 2009).

OECD CLIs are constructed to predict cycles in a reference series chosen as a proxy for economic activity. Fluctuations in economic activity are measured as the variation in economic output relative to its long-term potential. The difference between potential and observe output is often referred to as the output gap (OG), and the fluctuation in the output gap as the business cycle. This approach, focusing on turning points (peaks and troughs), results in CLIs that provide qualitative rather than quantitative information on short-term economic movements. The two series (CLI and OG) show strong co-movements, with the turning points of the CLI consistently preceding those of the business cycle; lead time varies, but 6 - 9 months is at what the OECD aims.

The OECD CLIs are compiled and published on a monthly basis for 33 OECD member countries; 6 non-member economies: Brazil, China, India, Indonesia, Russia, South Africa; and another group of countries such as Euro Area, G7, NAFTA...

The component series for Spain are (OECD, 2012):

- Manufacturing: Rate of capacity utilization SA (% balance) from European Commission.
- Construction–Employment: future tendency SA (% balance) from European Commission.
- CPI Services less housing (2015=100) inverted from National Institute of Statistics.
- Share prices: IGBM general index (2015=100) from Bank of Spain.
- Passenger car registrations (2015=100) from National Institute of Statistics.
- Consumer Confidence indicator SA (% balance) from European Commission.

In summary, we observe information from industrial, construction and service sectors, as well as equity market, car sales and confidence indicator.

2.3. The Economic Policy Uncertainty (EPU) as a predictor of GDP trend

As we aforementioned, another leading indicator that we could follow to anticipate GDP trend is the Economic Policy Uncertainty Index, published monthly by the Economic Policy Uncertainty for 23 countries, Europe as a whole and the world. The index is composed of three underlying indicators: the first one quantifies the journalistic coverage of policy-related economic uncertainty; the second one reflects the number of federal tax code provisions set to expire in future years; and the third component uses disagreement among economic forecasters as a proxy for uncertainty.

According to the literature (Baker, Bloom and Davis, 2015), it seems that there is a significant dynamic relationship between the economic uncertainty index and the real macroeconomic variables. Specifically, it is observed that an increase in economic uncertainty, measured by this index, anticipates a decrease in economic growth and employment in the following months.

In this paper, we are going to use CLI instead of EPU to compare with our real time leading economic indicator because the correlation between CLI and GDP is higher than EPU and GDP (0.740 vs -0.546), although both are statistically significant.

3. Methodology

3.1. Data

We use data from two kinds of sources: Google Trends (mainly) and Thomson Reuters Eikon-Datastream. As far as Google is concerned, we carry out a text mining based on keywords search related to economic sectors, economic sentiment and consumption. Concerning Thomson Reuters, we exploit data from equity market, bond market and Red Electrica Española (electrical energy).

All the data we use are published in real time for the Spanish economy, but we build the indicator in a monthly basis so as to compare the results with the Composite Leading Indicator, that is released every month by OECD, and with the Spanish GDP at constant prices, that has been interpolated monthly. The time series start in January 2004 and end in December 2018, thus the number of monthly observations amounts to 180.

3.2. Variables included in the Real Time Leading Economic Indicator

To build our Real Time Leading Economic Indicator (RT-LEI), we have used 21 variables that have been grouped in five latent variables, which have been integrated later in a final index. The selection of these 21 variables lies behind the idea of improving CLI's ability of identifying ahead GDP's turning points and trends. This enhancement lays down in a higher number of variables (that represents quite the same parts: industry, construction, consumption, confidence and finance), along with real-time data.

All the variables are included in Table 1, where we can observe not only the definition but also the data source of origin of each one.

Table 1. Set of variables to build the Leading Economic Indicator

Variables	Definition	Data source
"Car registration" "Buy car" "Car insurance"	Keywords related to private consumption, in particular in the motor vehicles sector, that is really correlated with economic activity. (Turley, 1976 │ Ramey and Vine, 2005)	Google Trends
"IBEX 35 - Price Index"	IBEX35 is the main stock market index of Spanish Equity Market, composed by the most liquid companies of the country. The correlation of equity market and business cycle has been broadly studied. (Trainer, 2006 │ The Economist, 2010)	Thomson Reuters Eikon
"Spread Yield Curve"	Spread yield Spanish Bond 10 years vs Spanish Bond 1 year. (Berganza and Fuertes, 2018 │ Bauer and Mertens, 2018 │Estrella and Trubin, 2006) │Ang and Piazzesi, 2003)	
"Home décor" "Buy dwelling" "Buy house" "Mortgage"	Keywords related to Real Estate sector. Many authors such as Foldvary (1991) and Jaccard (2007) studied the direct relationship between Real Estate activity and business cycle	Google Trends
"Boom" "Expansion" "Happiness" "Optimism"	Keywords focused on positive expectations and optimistic economic sentiment (Huang, Rojas and Convery, 2018 │ Nyman et al., 2014 │ Musat and Trausan-Matu, 2009 │ Pang and Lee, 2008 │ Wiebe, Wilson and Cardie, 2005)	Google Trends
"Crisis" "Bankruptcy" "Sadness" "Pessimism"	Keywords focused on negative expectations and pessimistic economic sentiment (Huang, Rojas and Convery, 2018 │ Nyman et al., 2014│Musat and Trausan-Matu, 2009 │ Pang and Lee, 2008 │ Wiebe, Wilson and Cardie, 2005)	Google Trends
"Electricity Consumption"	Electricity consumption in the industrial sector. Narayan et. Al (2011) and Thoma (2004), among others, found evidence between electrical energy and business cycle	Thomson Reuters Datastream
"Petroleum" "Buy Computer" "Machinery"	Keywords related to industrial sector activity. According to the Spanish Bureau of Statistics (2018), oil extraction and refining, production of computers and machinery are ones of the most important variables in industry	Google Trends

Note:
For *Google Trends*: The numbers reflect the search interest in relation to the maximum value in a given region and period (index). A value of 100 indicates the maximum popularity of a term, while 50 and 0 indicate that a term is half popular in relation to the maximum value or that there was not enough term data, respectively.
For *Thomson Reuters*: IBEX35 is measured as an index, spread yield curve is calculated in percentage points and electricity consumption in millions Kilowatt/Hour.

In Table 2, we may observe that all the variables from Thomson Reuters have passed the normality test. However, none of Google Trends' variables present a normal distribution, except the words: car insurance and boom. Additionally, the most popular Google Trends keywords are car registration, buy car and car insurance, followed by buy dwelling and house, expansion, happiness and machinery.

Table 2. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Asymp. Sig. (2-tailed). Kolmogorov-Smirnov Test
Car registration	180	0,00	100,00	45,07	17,66	0,04
Buy car	180	33,00	100,00	57,86	14,20	0,03
Car insurance	180	25,00	100,00	60,25	17,53	**0,20**
IBEX 35 - PRICE INDEX	178	6268,00	15826,00	10209,91	1942,61	**0,07**
Spread Bond 10Y vs 1Y	178	-0,01	3,94	1,89	0,98	**0,49**
Home decor	180	2,00	31,00	7,10	5,63	0,00
Buy dwelling	180	23,00	100,00	43,92	17,40	0,00
Buy house	180	20,00	77,00	35,75	10,92	0,00
Mortgage	180	0,00	9,00	1,68	1,24	0,00
Boom	180	15,00	68,00	36,06	8,72	**0,58**
Expansion	180	28,00	100,00	51,74	12,61	0,05
Happiness	180	21,00	88,00	48,97	15,18	0,02
Optimism	180	9,00	100,00	27,00	13,16	0,02
Crisis	180	12,00	100,00	25,37	12,82	0,00
Bankruptcy	180	7,00	100,00	19,17	11,81	0,00
Sadness	180	3,00	10,00	5,32	0,90	0,00
Pessimism	180	0,00	1,00	0,02	0,15	0,00
Electricity consumption (industry)	179	-9,20	10,60	0,88	3,80	**0,73**
Petroleum	180	7,00	35,00	15,60	5,53	0,00
Buy computer	180	2,00	100,00	18,00	15,45	0,00
Machinery	179	27,00	80,00	40,84	11,16	0,00
N valid	178					

Source: Own elaboration from Google Trends and Thomson Reuters.

3.3. Multivariate technique to build the RT-LEI

To aggregate the 21 variables for building the Real Time Leading Economic Indicator, we have applied six factor analysis (principal component) with Varimax rotation (orthogonal rotation). Each factor analysis has been focused on elaborating a latent variable or factor by reducing dimensions, that substitutes the main parts of CLI, and weighting all the variables.

In order to apply the factor analysis, we have considered our series of variables (X_1, X_2,..., X_{21}) on our group of observations and we have calculated, from them, a new set of variables F_1, F_2,…, F_p, non-correlated among them, whose variances progressively decrease. Each F_j (where j = 1, ..., p) is a linear combination of the original X_1, X_2, ..., X_{21}, that is:

$$F_j = a_{j1}X_1 + a_{j2}X_2 + \ldots + a_{j21}X_{21} \qquad (1)$$

where a'_j is a vector of constants.

Factor analysis has been broadly utilized to construct indicators because allows researchers to build multivariate indexes by means of linear combinations with non-arbitrary weights (Munda and Nardo, 2005). According to this technique, we can identify the most important variables through the original correlation matrix. Diebold (2000) points out that the principal component methods are rather more sophisticated, not requiring a sharp "in" or "out" decision for each variable, but rather allowing all variables to contribute to an extraction or forecast.

3.4. Empirical methodology

3.4.1. Fractional Cointegrated VAR

Johansen (2008) introduced a method to check for multivariate fractional cointegration denominated Fractionally Cointegrated Vector AutoRegressive (FCVAR) model. Johansen and Nielsen (2010, 2012) expanded this model. It is a step forward of the Cointegrated Vector AutoRegressive model (Johansen, 1996), named also CVAR, which allows for fractional processes of order d that cointegrate to order d–b. To introduce the FCVAR model, we start by refering to the non-fractional CVAR model.

Let $Y_t, t = 1, \ldots, T$ be a p-dimensional I(1) time series. The CVAR model is:

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \varepsilon_t = \alpha\beta' L Y_t + \sum_{i=1}^{k} \Gamma_i \Delta L^i Y_t + \varepsilon_t \qquad (2)$$

To derive the FCVAR model we must replace the difference and lag operators Δ^b and $L_b = 1 - \Delta^b$, respectively. We then obtain:

$$\Delta^{b} Y_t = \alpha\beta' L_b Y_t + \sum_{i=1}^{k} \Gamma_i \Delta^{b} L_b^{i} Y_t + \varepsilon_t \tag{3}$$

which is applied to $Y_t = \Delta^{d-b} X_t$ such that

$$\Delta^{d} X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{i=1}^{k} \Gamma_i \Delta^{b} L_b^{i} Y_t + \varepsilon_t \tag{4}$$

where ε_t is p-dimensional independent and identically distributed, with mean zero and covariance matrix Ω. From the CVAR model we can interpret the parameters. Thus α and β are $p \times r$ matrices, where $0 \leq r \leq p$. The columns of β are the cointegrating relationships in the system, that is to say the long-run equilibria. Γ_i is the parameter that govern the short-run behavior of the variables. The coefficients in α represent the speed of adjustment responses to deviations from the equilibria and the short-run dynamics of the system.

Matlab computer programs for the calculation of estimators and test statistics were provided by Nielsen and Popiel (2018) and it has been employed in several empirical papers (Jones, Nielsen and Popiel, 2014; Baruník and Dvořáková, 2015; Maciel, 2017; Dolatabadi et al., 2016; Gil-Alana and Carcel, 2018; Yaya and Gil-Alana, 2019; Yaya et al., 2019; Tule et al., 2019; etc.).

3.4.2. Wavelet Analysis

The wavelet methodology is used to analyse time series in the time-frequency domain. Following Vacha and Barunik (2012), Aguiar-Conraria and Soares (2011, 2014), Dewandaru et al. (2016), Tiwari et al. (2016), Jammazi et al. (2017), and others that apply Continuous Wavelet Transform (CWT) in finance and economics research, two tools are used in this paper: wavelet coherency and wavelet phase-difference.

There are two reasons for using this methodology: firstly, stationarity is not a requirement to carry out a wavelet analysis and, secondly, it is interesting to study the interaction of both the time and the frequency domains of the time series themselves to find evidence of the potential changes in its pattern.

The wavelet coherency is a two-dimensional diagram that correlates time series and identifies hidden patterns or information in the domain of time and frequency. The $WT_x(a,\tau)$ of a time series $x(t)$, that is obtained by projecting a mother wavelet ψ, is defined as:

$$WT_x(a,\tau) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-\tau}{a}\right) dt, \tag{5}$$

where $WT_x(a,\tau)$ are the wavelet coefficients of $x(t)$; the position of a wavelet in the frequency domain is defined by a, and τ is the position in the time domain. Thus, the wavelet transform provides information concurrently on time and frequency by mapping the original series into a function of τ and a. Morlet wavelet has been chosen like a mother wavelet to carry out our analysis since it is a complex sine wave within a Gaussian envelope, so we will be able to measure the synchronism between time series. (see Aguiar-Conraria and Soares, 2014 for the properties of this wavelet).

To understand the interaction and the integration between the two series we use the wavelet coherence defined as:

$$WCO_{xy} = \frac{SO\left(WT_x(a,\tau) WT_y(a,\tau)^*\right)}{\sqrt{SO(|WT_x(a,\tau)|^2) SO\left(\left|WT_y(a,\tau)\right|^2\right)}}, \tag{6}$$

where *SO* is a smoothing operator in both time and scale. Without the smoothing operator, the wavelet coherency would be always one for all times and scales (see Aguiar-Conraria et al. (2008) for details). Matlab computer programs for the calculation of estimators and test statistics in the CWT are provided in Aguiar-Conraria's website[1].

[1] https://sites.google.com/site/aguiarconraria/joanasoares-wavelets

4. Empirical Results

4.1. Real Time Leading Economic Indicator

After applying the factor analysis using the 21 variables previously defined, we construct 5 dimensions or latent variables that allow us to reduce all the information to build the final indicator. The 5 factors are: car activity, financial activity, real estate activity, economic sentiment and industrial activity. Economic sentiment is calculated as the balance between positive and negative sentiment. Results are displayed across Table 3.

Table 3. Real Time Leading Economic Indicator

Original Variables and Weights		Dimensions or Latent Variables	Final Indicator
Car registration (35,5%) Buy car (33,6%) Car insurance (30,9%)		**CAR ACTIVITY (20%)**	**REAL TIME LEADING ECONOMIC INDICATOR (LEI)**
IBEX35 - Price Index (50%) Spread Yield Curve (50%)		**FINANCIAL ACTIVITY (20%)**	
Home decor (28,3%) Buy dwelling (28%) Buy house (25,5%) Mortgage (18,2%)		**REAL ESTATE ACTIVITY (20%)**	
Boom (28,1%) Expansion (29,3%) Happiness (41,7%) Optimism (0,9%)	POSITIVE SENTIMENT	**ECONOMIC SENTIMENT (20%)**	
Crisis (40%) Bankruptcy (39,8%) Sadness (14,1%) Pessimism (6,2%)	NEGATIVE SENTIMENT		
Electricity Consumption (21,4%) Petroleum (15,6%) Buy computer (31,3%) Machinery (31,7%)		**INDUSTRIAL ACTIVITY (20%)**	

Source: Own elaboration

All the factor analysis applied were consistent according to KMO and Bartlett test (Table 4). In addition, RT-LEI's weight come from component matrices, that gather the importance of each original variable within its dimension.

Table 4. Factor Analysis Test

	FA1	FA2	FA3	FA4	FA5	FA6
	CAR ACTIVITY	FINANCIAL ACTIVITY	REAL ESTATE ACTIVITY	POSITIVE SENTIMENT	NEGATIVE SENTIMENT	INDUSTRIAL ACTIVITY
KMO	0,663	0,500	0,746	0,466	0,516	0,632
Bartlett test	0,000	0,000	0,000	0,001	0,000	0,000

As expected, according to Ramey and Vine (2005), car registration is the most important variable within the car activity factor. However, equity market is as relevant as spread yield curve to foresee financial cycle (Trainer, 2006). In the real estate factor, buy dwelling or houses and home décor are the most significant variables (Jaccard, 2007). Within economic sentiment, happiness/expansion and crisis/bankruptcy are the main words related to positive or negative expectations, respectively. Finally, machinery is the most useful keyword to measure industrial activity, followed by buy computer variable.

Once RT-LEI has been built, we have calculated correlations among RT-LEI, CLI and real GDP to compare the degree of similarity. Our prior conclusion is that RT-LEI enables to foresee GDP behavior slightly better than OECD CLI, due to the fact that RT-LEI includes more variables and is released in real time. The Spearman correlation coefficient between RT-LEI and real GDP is 0.804 against 0.740 between CLI and real GDP (99% confidence interval, see Table 5).

Table 5. Correlation coefficient RT-LEI, CLI and GDP

Correlations					
			GDP	CLI	RT-LEI
Rho de Spearman	GDP	Correlation coefficient	1.000	**0.740****	**0.804****
		Sig. (bilateral)		0.000	0.000
		N	178	178	178
	CLI	Correlation coefficient	**0.740****	1.000	0.764**
		Sig. (bilateral)	0.000		0.000
		N	178	178	178
	RT-LEI	Correlation coefficient	**0.804****	0.764**	1.000
		Sig. (bilateral)	0.000	0.000	
		N	178	178	178

**. The correlation is significant at the 0.01 level (bilateral).

4.2. Fractional Cointegrating VAR analysis among GDP, CLI and the RT-LEI.

Before to start analyzing the fractional cointegration test, we carried out the univariate analysis by means of univariate tests and the results are displayed in the following table:

Table 6. Univariate analysis: RT-LEI, CLI and GDP

Series	No autocorrelation	Autocorrelation
LEI	**0.86 (0.78, 0.96)**	**0.96 (0.81, 1.15)**
GDP	**1.34 (1.26, 1.45)**	**1.53 (1.31, 1.74)**
CLI	**1.27 (1.18, 1.38)**	**1.40 (1.17, 1.61)**

To do so, we use the Whittle function in the frequency domain (Dahlhaus, 1989) and the results are reported in the table 6.

We see that estimating d for each series, the unit root null is rejected in favor of mean reversion for LEI under no autocorrelation though the unit root cannot be rejected if autocorrelation is permitted. However, for the other two series, GDP and CLI, the values are d are found to be significantly higher than 1, showing lack of mean reversion and permanency of the shocks.

Once we have done an univariate analysis of each variable and before to start analyzing the fractional cointegration test, we are going to define the lag augmentation of the system. Table 1, we have five estimations in the corresponding lag levels for each variable. For each level, WE check the significance through likelihood ratio (LR) test.

In this research we use k=3 as a lag value, following Jones, Nielsen and Popiel (2014). If we see Table 1, results indicate that for all models, the coefficient of highest order lag is significant at 99% confidence level. For this reason, we choose lag = 3.

	k	r	d	b	Log-likelihood	LR	p value	AIC	BIC
Model for RT_LEI, GDP	3	2	1.161	1.161	51.73	3.63	0.458	-65.47	-5.23
	2	2	1.015	1.015	49.92	1.06	0.900	-69.83	-22.28
	1	2	0.997	0.997	49.38	79.17	0.000	-76.77*	-41.89*
	0	2	1.822	1.822	9.80	0.00	0.000	-5.59	16.60
Model for CLI, GDP	3	2	1.543	1.543	754.47	2.38	0.666	-1470.94	1410.70
	2	2	1.602	1.602	753.28	42.37	0.000	-1476.56*	-1429.00*
	1	2	1.832	1.832	732.10	570.18	0.000	-1442.19	1407.32
	0	2	2.000	2.000	447.00	0.00	0.000	-880.01	-857.81
Model for RT_LEI, GDP,	3	3	1.086	1.086	235.71	35.84	0.000	-391.41*	-264.59
CLI	2	3	1.435	1.435	217.79	75.61	0.000	-373.58	-275.29*
	1	3	1.772	1.772	179.98	612.50	0.000	-315.96	-246.21
	0	3	2.000	2.000	-126.27	0.00	0.000	278.53	319.75

*Table 1. Lag Selection Results. Note LR = likelihood ratio; AIC = Akaike information criterion; BIC = Bayesian information criterion. *Indicate the best (that is, minimized) values of the respective information criteria.*

Once we have selected the order lag, we have to determine the number of the cointegrated vectors in the system by the rank, where the hypothesis is $H_0\colon rank = r$ and $H_1\colon rank = p$; $r = 0,1,2,\ldots$ and p is the number is the number of variables in the system. For the alternative ranks, the first nonrejected value is the number of cointegrated vector in the system. According with the results in Table 2 and also, following MacKinnon and Nielsen (2014), we accept $r = 1$ because a single lag is usually sufficient in the fractional model to capture the serial correlation in the residuals.

	Rank	d	b	Log-likelihood	LR statistic	p value
Model for RT_LEI, GDP	0	0.616	0.616	51.544	-3.255	-
	1	0.583	0.583	55.458	-11.082	-
	2	1.015	1.015	49.917	-	-
Model for CLI, GDP	0	1.326	1.326	734.465	37.627	0.000
	1	1.603	1.603	752.353	1.851	0.790
	2	1.602	1.602	753.279	-	-
Model for N_LEI, GDP, CLI	0	1.207	1.207	186.805	61.966	0.000
	1	1.314	1.314	208.707	18.163	0.148
	2	1.432	1.432	216.948	1.680	0.873
	3	1.435	1.435	217.788	-	-

Table 2. LR Tests for Cointegrating Rank. Note. As it is pointed by Nielsen and Popiel (2018), p values for the cointegration rank tests are not provided by the main code. Researcher should use additional code of Jason Rhinelander: https://github.com/jagerman/fracdist/releases. LR = likelihood ratio.

The results in Table 3 shows us the following interesting comments.

	d	Cointegrating equation beta		
		RT_LEI	GDP	CLI
Panel I: RT_LEI, GDP	0.583 (0.038)	1.000	-0.583	-
	$\Delta^d\left(\begin{bmatrix} RT\_LEI \\ GDP \end{bmatrix} - \begin{bmatrix} 8.456 \\ 3.230 \end{bmatrix}\right) = L_d \begin{bmatrix} -0.252 \\ -0.028 \end{bmatrix} \nu_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$			
Panel II: CLI, GDP	1.603 (0.062)	-	-1.537	1.000
	$\Delta^d\left(\begin{bmatrix} CLI \\ GDP \end{bmatrix} - \begin{bmatrix} 4.918 \\ 3.223 \end{bmatrix}\right) = L_d \begin{bmatrix} -0.000 \\ 0.016 \end{bmatrix} \nu_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$			
Panel III: RT_LEI, GDP, CLI	1.314 (0.051)	1.000	1.845	-6.152
	$\Delta^d\left(\begin{bmatrix} N\_LEI \\ GDP \\ CLI \end{bmatrix} - \begin{bmatrix} 93.387 \\ 3.201 \\ 4.918 \end{bmatrix}\right) = L_d \begin{bmatrix} -2.225 \\ -0.004 \\ -0.000 \end{bmatrix} \nu_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$			

Table 3. Fractionally Cointegrating Vector Autorregression: Estimation Results

In the first case, the relationship between Real Time Leading Economic Indicator (RT_LEI) and GDP is negative. The value for beta in the cointegrating equation is -0.583. The same scenario occurs in the case of CLI and GDP, where the relationship also is negative and the beta term is -1.537.

Analyzing the relationship between RT_LEI, GDP and CLI in Panel III we find a negative relationship between CLI with RT_LEI and GDP. The value for beta in the cointegrating equation is -6.152. This result indicates that when CLI increases, it would provide a decrease in the other indices. This fact has sense because, CLI has less correlation with respect to the N_LEI.

Also, we have calculated the value of the fractional differencing parameter, which is is 0.583 (0.038), meaning that the cointegrating system presents non-stationarity. This results give us feed about the behavior in the long-run of the time-series analyzed and we can conclude that is mean reverting, assuming that, faced with a shock in the economy, it indicates that in the future GDP and the RT_LEI returns to the previous trend. Also, we can conclude that the combination of both series is not very persistent, indicating that the trend is not well defined. Therefore, we can detect areas with lower probability success.

In the second and third case, we have an opposite result. For the second case, the cointegrating structure is 1.603 (0.062) and for the third case is 1.314 (0.051). These means that the cointegrating system presents non-stationarity and are not mean reverting in both cases, where the cointegrated time series have a different behavior, not following the trend before the structural change or shock.

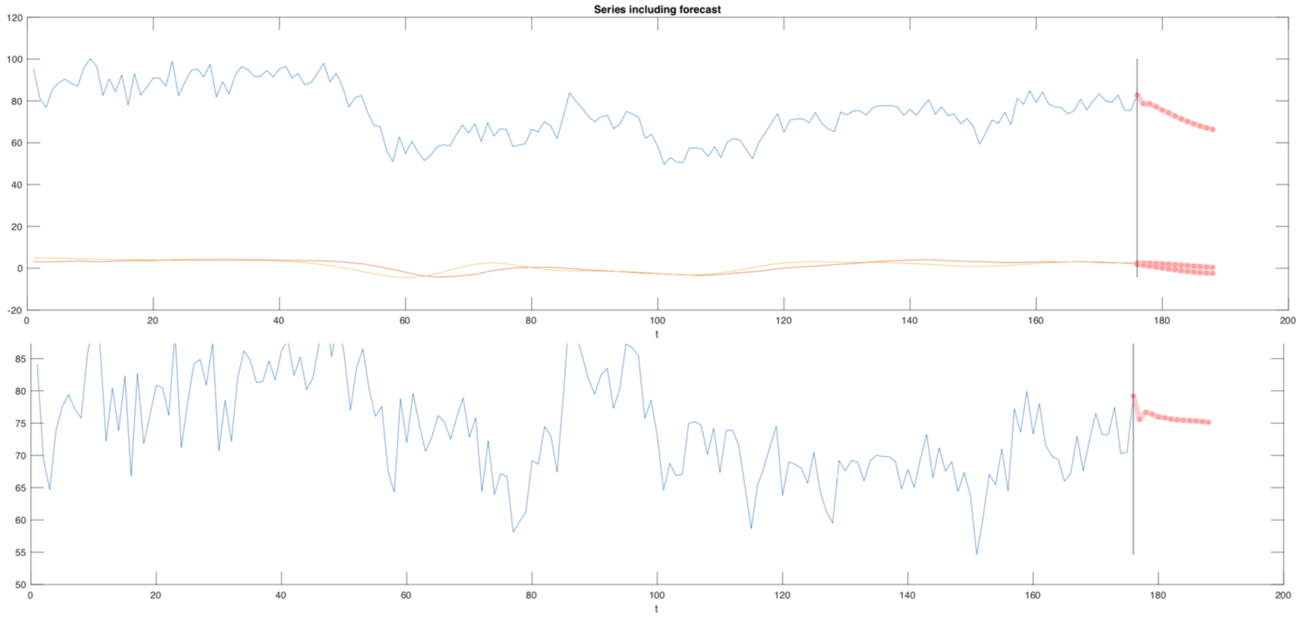


Figure 1. Fractionally Cointegrating Vector Autorregression: Forecast Results for Spain of final model 12 steps ahead.

Finally, in Figure 1 we show the forecasting of the trend of the three time series analyzed (blue = RT_LEI; orange = GDP; yellow = CLI; red = forecast). We also show that the forecasted trend of the three cointegrated time series, where obviously is in line with the forecasting done by several international and national organism in the U.S., Europe and Spain, clearly reflecting the fall of the GDP in Spain.

IMF (2018) expects the Spanish economy starts to decelerate and show a bear trend in the next two years. According to this institution, there are several downside risks clouding the medium-term outlook. Externally, they comprise sudden changes in investors' global risk appetite, escalating global protectionism, and weakening conditions in emerging economies. Domestically, they include pressure to reverse reforms, continued procyclical fiscal policy, and prolonged uncertainty related to Catalonia. In addition, a faster monetary policy normalization could increase interest rates and moderate private consumption and investment (FUNCAS, 2019). These could hurt the economy particularly in an environment of high public debt and structural unemployment as well as sluggish productivity growth, which is set to slow Spain's income convergence.

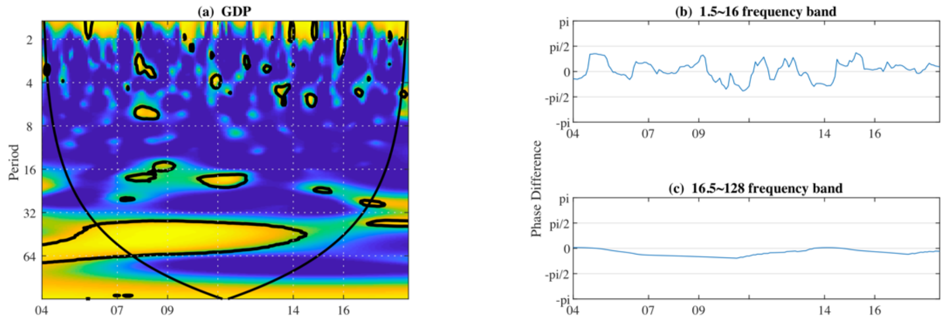
For these reasons, IMF (2018) encourages the authorities to persevere with policies and reforms aimed at further enhancing economic resilience, reducing public debt, improving productivity, reducing inequality and increasing employment, especially raising long-term and youth employment.

4.3. Wavelet analysis between GDP and the RT-LEI

After the cointegrating analysis between RT_LEI, GDP and CLI, we are going to analyze in time-frequency domain the dynamic correlation between the RT_LEI and GDP and CLI and GDP to achieve a robust analysis.

Figure 2 displays the wavelet coherency and the phase difference for the monthly data of the cited time-series, showing evidence of varying dependence between both time series across different frequencies and over time.

i) GDP vs RT_LEI



ii) GDP vs CLI

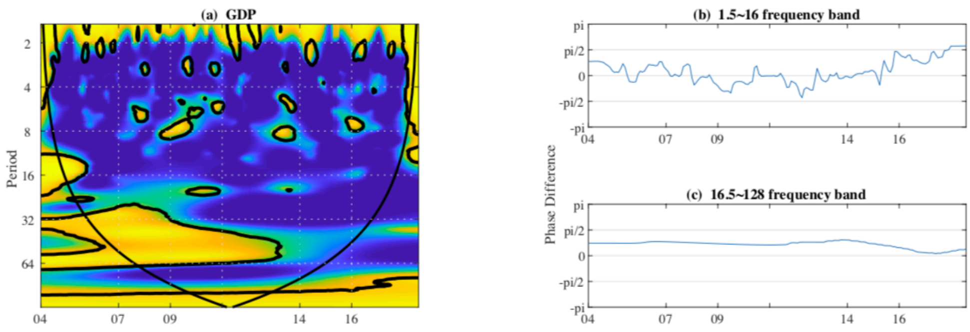


Figure 2. Wavelet coherency and phase difference between time series. Left: Wavelet coherency. The contour designates the 5% significance level. Coherency ranges from blue (low coherency) to yellow (high coherency). Right: Phase difference prices at 1.5-16 months (top) and 16.5-128 months (bottom) frequency bands. The cone of influence is shown with a thick line, which is the region subject to border distortions

Analyzing the wavelet coherence between GDP and RT_LEI (i) and GDP and CLI, we appreciate that the time series were strongly related at lower frequencies (long-term), increasing the dependence between them.

The level of dependence starts early 2006, reaching high levels of dependence centered at lower frequencies (from 36 to 64 months) in the year 2015 in the case of figure (i). After 2015 dependence for the long run dissipated. In the case of figure (ii) it occurs the same but reach the levels of dependence at frequencies that are located between 28 to 64 months in the year 2013. After 2013 dependence for the long run dissipated.

If we analyze the phase difference during the period of dependence, in the case of figure (i) it is between 0 and $-\pi/2$. We can conclude that the correlation of the series is negative, and they move together, suggesting that GDP data is lagged with respect to RT_LEI. In other words, an increase in GDP is led by the increase in RT_LEI. This insinuates that the RT_LEI better explains the advanced behavior of the GDP.

In the case of GDP and CLI the phase difference during the period of dependence is between 0 and $\pi/2$, concluding that the correlation is positive and the results suggest that GDP leads the behavior of the CLI.

5. Conclusions

After reviewing the literature concerning business cycle and big data, we have constructed a real time leading economic indicator (RT-LEI) for Spain based on Thomson Reuters Eikon-Datastream and mainly Google Trends. We use Fractional Cointegration Vector AutoRegressive model (FCVAR model) and Continuous Wavelet Transform (CWT) to analyze the dynamics, interconnections and structural changes in the time-series analyzed.

As far as RT-LEI is concerned, we have used 21 variables that have been saturated in five factors: car activity, financial activity, real estate activity, economic sentiment and industrial activity. The indicator has been built by means of a factor analysis. The results show that our RT-LEI is able to anticipate GDP trends, ups and downs and turning points slightly better than CLI from OECD, according by the correlation matrix. Two reasons may explain the improvement: RT-LEI might be released in real time while CLI is published monthly and RT-LEI uses keywords that capture agent's behavior with the same accuracy (or even more) as CLI, in which some variables come from surveys. Thus, we compare massive data with samples.

The results obtained using CWT suggest that the new RT_LEI is reflecting and advancing us the responsiveness of the GDP. Using the FCVAR model we can confirm the negative relationship between these time series. Also, we conclude that the cointegrating system presents non-stationarity and the behavior in the long-run is mean reverting.

The forecasting made by the FCVAR model shows a bear trend of each of the time series (top part of the figure) and of the cointegrated series (bottom part of the figure). These results are similar to IMF and FUNCAS' forecasting: it is expected that Spanish economy begins to slow down in the next two years. According to these institutions, there are several downside risks that explain this bear trend. Externally, we expect an escalating global protectionism and weakening conditions in emerging economies. Nationally, they highlight a possible reverse reforms and more uncertainty related to Catalonia. Furthermore, a faster tapering in monetary policy could put pressure on interest rates and moderate private consumption and investment.

6. References

Aguiar-Conraria, L. and Soares, M. J. (2011). Oil and the macroeconomy: using wavelets to analyze old issues. Empirical Economics, 40, 645-655.

Aguiar-Conraria, L. and Soares, M. J. (2014). The continuous wavelet transform: Moving beyond uni- and bivariate analysis. Journal of Economic Surveys 28, 344–375.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics, 50,745-787.

Antenucci, D.; Cafarella, M.; Levenstein, M.; Ré, Che and Shapiro, M.D. (2014). Using Social Media to Measure Labor Market Flows. Working Paper No. 20010*, National Bureau of Economic Research*, March 2014, URL: http://www.nber.org/papers/w20010.pdf

Baker, S.R.; Bloom, N. and Davis, S.J. (2015). Measuring Economic Policy Uncertainty. CEP Discussion Paper No 1379. October 2012.

Banerjee, A., Marcellino, M., and Masten, I. (2005). Leading indicators for euro-area inflation and GDP growth. Oxford Bulletin of Economics and Statistics, 67, 785-813.

Baruník, J., Dvořáková, S. (2015). An empirical model of fractionally cointegrated daily high and low stock market prices. Economic Modelling, 45, 193-206.

Bauer, M.D., y T.M. Mertens (2018). Economic forecasts with the yield curve, FRBSF Economic Letter, 2018-07.

Baumohl, B. (2009). The Secrets of Economic Indicators. Hidden Clues to Future Economic Trends and Investment Opportunities. Wharton School Publishing.

Berganza, J.C. and Fuertes, A. (2018). El aplanamiento de la curva de rendimientos en Estados Unidos. Boletín Económico 1/2018, Banco de España.

Bernanke, B.S. (2008). Stabilizing the financial markets and the economy, Speech at the Economic Club of New York.

Bollen, J.; Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market, Journal of Computational Science, 2, 1-8.

Camacho, M. and Pérez-Quiros, G. (2002). This is what the leading indicators lead. Journal of Applied Econometrics, 17, 61-80.

Choi, H.; Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88, 2-9.

Croushore, D. (2011). Frontiers of real-time data analysis. Journal of economic literature, 49, 72-100.

D'Amuri, F. and Marcucci, J. (2017). The Predictive Power of Google Searches in forecasting US unemployment, International Journal of Forecasting, 33, 801-816.

Dewandaru, G., Masih, R., and Masih, A. M. M. (2016). Contagion and interdependence across Asia-Pacific equity markets: An analysis based on multi-horizon discrete and continuous wavelet transformations. International Review of Economics & Finance, 43, 363-377.

Diebold, F.X. (2000). Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting. University of Pennsylvania and NBER.

Dolatabadi, S., Nielsen, M. Ø., and Xu, K. (2016). A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets. Journal of Empirical Finance, 38, 623-639.

Dong, L.; Chen, S.; Cheng, Y.; Wu, Z.; Li, C. and Wu, H. (2017). Measuring economic activity in China with mobile big data. EPJ Data Science 6, 29.

Estrella, A. and Trubin, M.R. (2006). The Yield Curve as a Leading Indicator: Some Practical Issues. Current Issues in Economics and Finance, 12.

Foldvary, F.E. (1991). Real Estate and Business Cycles: Henry George's Theory of the Trade Cycle. Lafayette College Henry George Conference. June 1991.

FUNCAS (2019). Panel de previsiones de la economía española. Marzo de 2019. Madrid, España.

Gil-Alana, L.A., and Carcel, H. (2018). A fractional cointegration var analysis of exchange rate dynamics. The North American Journal of Economics and Finance.

Heinisch, K., and Scheufele, R. (2018). Should Forecasters Use Real-Time Data to Evaluate Leading Indicator Models for GDP Prediction? German Evidence. German Economic Review, 20, 170-200.

Hisano, R.; Sornette, D.; Mizuno, T.; Ohnishi, T.: Watanabe, T. (2013). High Quality Topic Extraction from Business News Explains Abnormal Financial Market Volatility. PLoS ONE 8(6): e64846. https://doi.org/10.1371/journal.pone.0064846

Huang, M.Y.; Rojas, R.R. and Convery, P.D. (2018). News Sentiment as Leading Indicators for Recessions. Department of Economics. University of California, USA.

International Monetary Fund (2018). 2018 Article IV Consultation Spain. IMF Country Report No. 18/330. November 2018. Washington DC, USA.

Jaccard, I. (2007). House Prices, Real Estate Returns, and the Business Cycle. Swiss Finance Institute Research Paper, 06-37.

Jammazi, R., Ferrer, R., Jareño, F., and Shahzad, S. J. H. (2017). Time-varying causality between crude oil and stock markets: What can we learn from a multiscale perspective?. International Review of Economics & Finance, 49, 453-483.

Johansen, S. (1996). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, 2nd ed., Oxford University Press, Oxford.

Johansen, S. (2008). A Representation Theory for a Class of Vector Autoregression Models for Fractional Processes. Econometric Theory, 24, 651-676.

Johansen, S., and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. Journal of Econometrics, 158, 51–66.

Johansen, S., and M. O. Nielsen (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model, Econometrica, 80, 2667–2732.

Jones, M.; Nielsen, M.O. and Popiel, M.K. (2014). A fractionally cointegrated VAR analysis of economic voting and political support. Canadian Journal of Economics 47, 1078–1130.

Keynes, J.M. (1936). The General Theory of Employment, Interest, and Money. The Quarterly Journal of Economics, 51, 209-223.

Koenig, E. (2002). Using the purchasing managers' index to assess the economy's strength and the likely direction of monetary policy. Federal Reserve Bank of Dallas and Economic and Financial Policy Review, 1, 1-14.

Kozicki, S. (2002). Comments on 'Forecasting with a real-time data set for macroeconomists'. Journal of Macroeconomics, 24, 541-557.

Maciel, L. S. (2017). Technical analysis based on high and low stock prices forecasts: Evidence for Brazil using a fractionally cointegrated VAR model. Empirical Economics, 1-18.

MacKinnon, J. G. and M. Ø. Nielsen (2014). Numerical distribution functions of fractional unit root and cointegration tests. Journal of Applied Econometrics 29, 161–171.

Munda, G. and Nando, M. (2005). Constructing Consistent Composite Indicators: the Issue of Weights. European Commission Directorate-General Joint Research Centre.

Musat, C.C. and Trausan-Matu, S. (2009). Sentiment Analysis on Economic Texts in Troyano, Cruz and Díaz (Eds.): WOMSA'09, 71-78.

Narayan, P.K.; Narayan, S. and Smyth, R. (2011). Energy consumption at business cycle horizons: The case of the United States. Energy Economics, 33, 161-167.

Nielsen, M.A. and Popiel, M.K. (2018). A Matlab Program And User's Guide For The Fractionally Cointegrated Var Model, Working Paper 1330, Economics Department, Queen's University.

Nyman, R.; Ormerod, P.; Smith, R. and Tuckett, D. (2014). Big Data and Economic Forecasting: A Top-Down Approach Using Directed Algorithmic Text Analysis. Workshop on Big Data for Forecasting and statistics, ECB. Frankfurt 7-8 April.

OECD (2012). OECD System of Composite Leading Indicators. OECD, March 2012.

Pang, B. and Lee L. (2008). Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, 2, 1-135.

Pappalardo, L.; Vanhoof, M.; Gabrielli, L.; Smoreda, Z.; Pedreschi, D. and Giannotti, F. (2016). An anaytical framework to nowcast well-being using mobile phone data". International Journal of Data Science and Analytics, 2, 75-92.

Ramey, V. and D. Vine (2005). Tracking the source of the decline in GDP volatility: An analysis of the Automobile Industry. Finance and Economics discussion Series, No. 2005-14, Federal Reserve Board, Washington.

Scott, S.L. and Varian, H.R. (2012). Bayesian Variable Selection for Nowcasting Economic Time Series. National Bureau of Economic Research.

Stark, T., and Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. Journal of Macroeconomics, 24, 507-531.

Stock, J. H., and W Watson, M. (2003). Forecasting output and inflation: The role of asset prices. Journal of Economic Literature, 41, 788-829.

Szármes, P. (2015). Creating new types of business and economic indicators using big data technologies. Science Journal of Business and Management, 3, 18-24.

The Economist (2010). Guide to Economic Indicators. Making Sense of Economics. 6th Edition. Bloomberg Press. USA.

Tiwari, A. K., Mutascu, M. I., and Albulescu, C. T. (2016). Continuous wavelet transform and rolling correlation of European stock markets. International Review of Economics & Finance, 42, 237-256.

Thoma, M. (2004). Electrical energy usage over the business cycle. Energy Economics 26 (2004) 463–485.

Toole, J.L.; Lin, Y.R.; Muehlegger, E.; Shoag, D.; Gonzalez, M.C. and Lazer, D. (2015). Tracking employment shocks using mobile phone data. Journal of The Royal Society Interface, 12, 20150185.

Trainer, E.M. (2006). Using Economic Indicators to Improve Investment Analysis. Third edition, Wiley. United States.

Tule, M. K., Salisu, A. A. and Ebuh, G. U. (2019). A test for inflation persistence in Nigeria using fractional integration & fractional cointegration techniques. Economic Modelling.

Turley, J. (1976). Automobiles Sales in Perspective. Federal Reserve Bank of Sr. Louis. June 1976. USA.

Vacha, L., and Barunik, J. 2012. Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. Energy Economics, 34, 241-247.

Varian, H.R. (2014). Big Data: new tricks for econometrics. Journal of Economic Perspective, 28, 3-28.

Wiebe, J.; Wilson, T. and Cardie, C. (2005). Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation, 39, 165-210.

Yaya and Gil-Alana 2019. Yaya, O.S., Ogbonna, E.A. and Olubusoye, O.E. (2019). "How Persistent and Dynamic Inter-Dependent are pricing Bitcoin to other Cryptocurrencies Before and After 2017/18 Crash?" Physica A, Statistical Mechanics and Applications, https://doi.org/10.1016/j.physa.2019.121732.

Yaya, O.S. and Gil-Alana, L.A. (2019). High and Low Intraday Commodity Prices: A Fractional Integration and Cointegration approach. MPRA paper. So, authors need to understand cointegration well.