International Economics 177 (2024) 100462

Contents lists available at ScienceDirect

International Economics

journal homepage: www.elsevier.com/locate/inteco

Chinese economic behavior in times of covid-19. A new leading economic indicator based on Google trends

Manuel Monge [*], Gloria Claudio-Quiroga, Carlos Poza

Universidad Francisco de Vitoria, Madrid, Spain

ARTICLE INFO

JEL classification:
C00
C22
E32
E37

Keywords:
Leading economic indicators
Business cycle
Google trends
Fractional cointegration
Machine learning

ABSTRACT

Since December 2019 we have been living with a virus called SARS-CoV-2 which has led to health policies being given prevalence over economic ones, causing serious consequences with regard to China's economic growth. For this purpose, we have built a Real Time Leading Economic Indicator based on Google Trends that improves the performance of Composite Leading Indicators (CLIs) to anticipate GDP trends and turning points for the Chinese economy. First, we assess the effectiveness of this new leading indicator relative to China's GDP by analyzing its statistical properties. We use fractional integration techniques to show the high degree of persistence of the new Real Time Leading Economic Indicator (RT-LEI) for China. Second, we observe the same relationship between GDP and RT-LEI in the long term using a Fractional Cointegration VAR (FCVAR) model. Third, we use a multivariate Continuous Wavelet Transform analysis to show which leading indicator best fits GDP and to identify when a structural change occurs. Finally, we forecast, using Artificial Neural Networks and a KNN model based on Machine Learning, our RT-LEI predicting the conclusion of a bearish scenario, after which the recovery begins in mid-2022.

1. Introduction

In the past 300 years there have been 10 pandemics episodes and given their irregularity new pandemics cannot be ruled out in the future (Potter, 2001). SARS-CoV-2 is the latest virus to appear and the cause of the COVID-19 disease which was identified in Wuhan, China, in December 2019 (World Health Organization, 2020; Hui et al., 2020).

Many measures to contain the virus have been applied since the biggest health crisis broke out at the beginning of 2020 (Lee et al., 2020; Chen et al., 2020). The pandemic is causing significant economic consequences: uncertainty, recession, unemployment, etc. For this reason, being able to anticipate the behavior of business cycles can help both economic analysis and decision-making.

As is already known, GDP data are usually released quarterly or yearly, thus the information regarding economic activity, to some extent, lags behind the moment for decision-making. In response to this limitation, some organizations and institutions publish leading economic indicators (LEI) to foresee GDP ups and downs, its trends and turning points.

In China, we can track the economy on a monthly basis through the OECD's Composite Leading Indicator, the Macroeconomic Climate Index from the National Bureau of Statistics, the Leading Economic Index released by The Conference Board or the Economic Policy Uncertainty Index from Economic Policy Uncertainty.[1] All of these works well to an extent, however big data offers new ways to construct high-frequency indicators that can improve the accuracy of GDP anticipation.

* Corresponding author. Universidad Francisco de Vitoria (UFV), Faculty of Law and Business, E-28223, Madrid, Spain.
E-mail address: manuel.monge@ufv.es (M. Monge).

[1] See www.policyuncertainty.com.

https://doi.org/10.1016/j.inteco.2023.100462
Received 15 May 2023; Received in revised form 3 November 2023; Accepted 4 November 2023
Available online 10 November 2023

In this regard, Choi and Varian (2012) suggest that a real-time volume of keyword queries in Google Trends is positively correlated with economic activity indicators, something which might be useful for short-run predictions. In other words, text mining may help in forecasting not only the future, but also the present, understood as a form of concurrent forecasting or nowcasting.

Some recent evidence shows that intentions (web search, for example) are associated with behavior (sales, consumption, or investment). Kim et al. (2016) states that intentions may not be the only foundation of behavior, but they are significant enough to forecast behavior. Anton and Lawrence (2016) and Li et al. (2017), for example, point out that keyword-based online search maintains a statistically significant relationship with the current behavior in the touristic sector.

The main aims of this paper are, firstly, to create a Chinese real time leading economic indicator (RT-LEI) combining Text Mining, Factor Analysis and information from Google Trends and Thomson Reuters Eikon-Datastream to anticipate the Chinese GDP behavior; secondly, to implement fractional integration and cointegration techniques to estimate the degree of persistence and to analyze the behavior of the RT-LEI in the long-term; and finally, to apply an Artificial Neural Networks (ANN) and a K-Nearest Neighbors (KNN) model that are supervised learning models based on Machine Learning to predict the Chinese RT-LEI and GDP in the next 24 months.

The structure of this paper is as follows: Section 2 reviews the literature on these issues. Section 3 describes the data and Section 4 presents the methodology. Section 5 presents the key empirical results. Section 6 concludes the paper.

2. Literature review

Macroeconomic forecasters have some problems: not only do they are based on past observations, but also macroeconomic forecasters do not consider the revision of the data used (Heinisch and Scheufele, 2018). Thus, peaks only will be recognized months late, especially when data are revised significantly. Thanks to big data, we are able to analyze massive data from the Internet in real time to forecast with more up-to-date information. The usefulness of internet search data in forecasting economic indicators. There is enough empirical evidence in scientific literature to confirm the useful of internet search data in forecasting economic indicators.

Some researchers do not use real-time data in forecast evaluations, and they simulate the real-time situation by applying a recursive or rolling estimation scheme (Banerjee et al., 2005). Other authors consider that this is not correct because the forecast evaluations need the use of real-time data (Stark and Croushore, 2002; Kozicki, 2002; Croushore, 2011). For this reason, the data available at the time the forecast are made should be used to evaluate the models.

2.1. Composite leading indicator (CLI)

With the aim of identifying turning points in business cycles and know in advance the trends in economic activity, many institutions publish monthly composite leading economic indicators (Baumohl, 2009). In the case of China, apart from the OECD CLI, there is the Economic Policy Uncertainty Index based on the South China Morning Post and also the Economic Policy Uncertainty and Trade Policy based on mainland newspapers; the Conference Board Leading Economic Index (LEI) for China elaborated by Adams et al. (2010) and based on the study of Guo et al. (2009) which shows it is possible to elaborate an contemporary index of Chinese coincident indicators (CEI) to evaluate changes in economic activity in China; and finally, there is the Macroeconomic Climate Index published by the National Bureau of Statistics.

In this paper, we use the Composite Leading Indicator because our proposal is to build a leading indicator applicable to many countries and the CLI allows us this possibility.

The CLI provides timely and important information on future and current economic situation as well as relevant help in short-term predictions of changes in the economic activity (Saltelli, 2006). It was launched in 1970 (for China in 1992) and is created to bring early signals of turning points in business cycles. It is one of the best leading economic indicators for anticipating GDP behavior, not from a quantitative point of view, but from a qualitative perspective, identifying trends and turning points in the time series. The GDP is used by the CLI uses GDP to identify these trends and turning points in the growth cycle. China is the only country for which the OECD uses the value added of industry at 1995 prices. This system is based on the "growth cycle" where business cycles and turning points are estimated and identified in the deviation-from-trend series (OECD, 2012). This information is highly relevant because it allows a timely analysis of the current and short-term economic activity. By using measures that are very sensitive to upcoming alterations in trading conditions, this system predicts the short-term movements of an economy (Baumohl, 2009).

Despite China not belonging to the OECD, this institution works closely with some of the world's largest economies, considering them key partners. In this regard, the OECD publishes monthly the Chinese CLI which includes seven leading indicators: industrial production of chemical fertilization, production of manufactured crude steel, 5000 industrial enterprises: diffusion index, overseas order level, production of buildings, monetary aggregate: M2, Shanghai Stock Exchange, and turnover and production of motor vehicles (OECD, 2012). In summary, the Chinese CLI is mainly composed of industrial production, construction, and financial indicators, which accordingly provide information concerning Chinese economic fluctuations.

2.2. Forecasting with big data

The limitations of economic indicators can be overcome with the adequate use of big data, which allows us to examine massive data in real time from the Internet to forecast the economic activity with more up-to-date 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; Szármes, 2015).

The creation of automated platforms for monitoring macroeconomic conditions in real time have been made possible by the new methodologies in time-series econometrics. Giannone et al. (2008) combined big data models and filtering techniques to build the first statistical framework known as nowcasting. The term is created from now and forecasting and recently is being used in economics. This term is defined as the prediction of the present, the immediate future and the highly recent past. Big data allows us to anticipate what is going to happen using data as signals (Szármes, 2015). Bok et al. (2018) describe The New York Fed Staff Nowcast which illustrates nowcasting with big data. This nowcasting model elaborates a forecast of the economic series through the obtaining the latent elements that drive movements in the data.

Antenucci et al. (2014) analyzed billions of tweets for references to unemployment (searching for expressions such as "axed", "pink slip" or "downsized") and elaborated a social media signal of job loss that closely tracks initial claims for unemployment insurance.

Toole et al. (2015) presented algorithms to identify employment shocks at the individual, community, and societal scales from mobile phone data. They show that it is possible to identify the users affected by mass layoffs. Then, they demonstrated that changes in mobility and social interactions, tracked by mobile phone data, predict unemployment rates before official reports.

Dong et al. (2017) carried out the first study that measured China's economic activity by extracting large-scale and fine granular spatial-temporal data. This study belongs to the Mobimetrics research concept that analyses massive individual mobility data generated by smartphones, wearable devices, self-driving cars and the Internet of Things with machine learning approaches to measure the dynamics of the social system in real time. Eagle et al. (2010) discovered that the variety of social interactions of the inhabitants of a municipality is positively relate to a socio-economic indicator of poverty. For this they used a nationwide mobile phone dataset. Papallardo et al. (2016) elaborated a data-driven analytical framework that employs big data to obtain relevant measures of human behavior and estimate indicators for the socio-economic development.

2.3. *Google Trends and forecasting busines cycles*

The scientific literature on big data and the macroeconomic scenario shows us that Google Trends is one of the most important types of unstructured data for the forecasting of business cycles. Choi and Varian (2009b, 2012) affirmed that Google Trends data is able to improve short-term forecasts of economic indicators. For example, internet search indices allow to create forecasts on indicators such as demand for holiday destinations, home sales, retail sales, or car sales and that unemployment, private consumption and house prices. In other words, Google Trends can help to predict the present and is a form of nowcasting.

Schmidt and Vosen (2009) showed that the forecasted results for the two most common private consumption indicators in the US can be improved using Google Trends. Niesert et al. (2020) have used Google search data to predict CPI, consumer confidence and unemployment for the US, the UK, Canada, Germany, and Japan. Monokroussos and Zhao (2020) have constructed a "Google Recession Index" using Google Trends data on internet search popularity.

Other researchers have shown that unemployment forecasting can be improved using Google Trends: McLaren and Shanbhogue (2011) for the UK; D'Amuri and Marcucci (2017) for the US; Fondeur and Karamé (2013) for France; Pescyova (2011) for Slovakia; Vicente et al. (2015) for Spain; Askitas and Zimmermann (2009) for Germany; Suhoy (2009) for Israel; D'D Amuri (2009) for Italy; and Anvik and Gjelstad (2010) for Norway.

The helpfulness of internet search-related data has also been searched in other fields such as the housing market (Beracha and Wintoki, 2013; McLaren and Shanbhogue, 2011), tourism (Bangwayo-Skeete and Skeete, 2015), investor attention (Da et al., 2011) or information demand and supply in the firm and in the market (Vlastakis and Markellos, 2012).

Scott and Varian (2012) showed that the forecast of automobile sales could be improved with search engine queries in the "vehicle shopping" category. The study of Yang et al. (2015) verified that Google and Baidu are very good to significantly decrease forecasting mistakes for numbers of sightseer for a important tourist destination in China.

To the best of our knowledge, this is the first study to build a Real Time Leading Economic Indicator based on Google Trends that improves the performance of Composite Leading Indicators (CLIs) to anticipate GDP trends and turning points for the Chinese economy.

3. Data and variables

3.1. *Data sources*

The data sources utilized in this paper are Google Trends (mainly) and Thomson Reuters Eikon-Datastream, that allow us to track the economy in real time. Regarding Google Trends, we carry out a text mining based on a keyword search related to agriculture, industry, construction and the financial sectors. As far as Thomson Reuters is concerned, we focus on equity and bond markets as well as electricity consumption.

All the data we exploit are released in real time for the Chinese economy, but we construct the indicator on a monthly basis in order to compare the outcomes with the OECD's CLI, which is published every month, and with the Chinese GDP at constant prices and interpolated monthly. The time series start in January 2004 and end in November 2020; hence the number of observations amount to 203.

We use Google Trends for two reasons: first, the literature shows it is useful in forecasting economic indicators while the few who use Baidu do not examine its characteristics or functions in a way that would lead us to consider it more highly (Liu et al., 2012). Besides, several studies state that the data from the two search engines are highly correlated and include basically the same information (Vaughan and Chen, 2015). Second, our purpose is to create a leading indicator that can be applied to any country and Google Trends offers us that possibility, unlike Baidu which is not used in many countries.

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

Variables	Definition	DATA SOURCE
"Electricity Consumption"	Electricity consumption in the agriculture and industrial sectors. Narayan et al. (2011) and Thoma (2004) showed evidence of correlation between electrical energy consumption and the business cycle.	Thomson Reuters Datastream
"Nitrogen"	Keyword linked to chemical fertilization, core in the agricultural sector. Rütting et al. (2018) states: "*Nitrogen is one of the most important inputs in crop production is well documented by the existence of a billion € fertilizer industry. The economic importance of N had already been identified more than 100 years ago*". Moreover, the primary sector accounts for 25% of employment in China (World Health Organization, 2020).	Google Trends
"Machinery" "Steel" "Exports"	Keywords related to industrial sector activity. The steel industry is one of the main pillars of the Chinese economy (ICEX, 2017). Industrial production is often used as a simple leading indicator (The Economist, 2010).	Google Trends
"Crisis" "Bankruptcy"	Keywords focusing on negative expectations and pessimistic economic sentiment (Huang et al., 2018 | Nyman et al., 2014|Musat and Trausan-Matu, 2009 | Pang and Lee, 2008 | Wiebe et al., 2005)	Google Trends
"Buy dwelling" "Mortgage"	Keywords related to the Real Estate sector. Many authors such as Foldvary (1991) and Jaccard (2007) studied the direct relationship between Real Estate activity and the business cycle	Google Trends
"Spread Yield Curve"	Spread yield for 10 year Chinese Bonds vs 1 year Chinese Bonds. (Berganza and Fuertes, 2018 | Bauer and Mertens, 2018 |Estrella and Trubin, 2006|Ang and Piazzesi, 2003)	Thomson Reuters Eikon
"Money"	Keyword linked to the use of money. Central Banks release money supply data to anticipate real sector behavior (ECB, 2020| FED, 2020)	Google Trends
"Shanghai SE"	Shanghai SE is the main stock market index of the Chinese Equity Market. The correlation between the equity market and business cycle has been broadly studied. (Trainer, 2006 | The Economist, 2010)	Thomson Reuters Eikon
"Tires" "Buy car"	Keywords related to private consumption: motor vehicles sector, that are really correlated with economic activity. (Turley, 1976 | Ramey and Vine, 2005)	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 as popular in relation to the maximum value or that there was not enough term data, respectively.
For *Thomson Reuters*: Shanghai SE is measured as an index, the spread yield curve is calculated in percentage points and electricity consumption in millions Kilowatt/Hour.

3.2. Variables

In line with the most recent research about economic indicators and the principle of parsimony, we select 14 variables to build the Real Time Leading Economic Indicator (RT-LEI) for China (see Table 1). The variables are grouped in six latent factors, that are integrated later in a final index. The selection of the variables aims to enhance the CLI capacity to predict GDP fluctuations. This improvement lies in the combination of a higher number of variables, very representative of GDP behavior, along with real-time data, in comparison with the CLI.

4. Methodology

To achieve the research goals, we apply five quantitative methods:

1) A *factor analysis* to create a new Leading Economic Indicator for China.
2) To study the statistical properties of the time series analyzed and to measure the degree of persistence, we use *fractional integration* techniques (see Poza and Monge, 2020; Monge and Gil-Alana, 2020; among others).
3) We use a *Fractional Cointegration VAR* (FCVAR) model to analyze the long-term relationship of the time series (see Johansen and Nielsen, 2010, 2012).
4) To show which leading indicator fits best and to determine with what frequency, we use methodologies based on *wavelet transforms* (see Aguiar-Conraria and Soares, 2014).
5) And, finally, the selected *Machine Learning models* (Artificial Neural Networks and KNN model) give us the opportunity to make more accurate predictions and the possibility of yielding new insights.

4.1. Factor analysis: the multivariate technique to construct the RT-LEI

To aggregate the 14 variables to construct the Chinese Real Time Leading Economic Indicator, we applied a second order factor analysis (principal component) with Promax rotation (non-orthogonal rotation). Each factor analysis is aimed at developing a latent factor by reducing dimensions (from 14 variables to 6 factors, and from 6 factors to the final index), as well as weighting all the variables.

We consider our set of variables (X_1, X_2, …, X_{14}) in our observations and we calculate, from them, a new group of variables F_1, F_2, …, F_6, non-correlated (or slightly correlated, depending on the type of rotation applied) between themselves, whose variances gradually decline. Each F_j (where $j = 1, …, p$) is a linear combination of the original X_1, X_2, …, X_{14}, that is:

$$F_j = a_{j1}X_1 + a_{j2}X_2 + \ldots + a_{j14}X_{14}$$

where a_j is a vector of constants.

This technique is often used to build indicators because it allows researchers to create composite indexes through linear combinations with non-arbitrary weights (Munda and Nando, 2005; Poza and Monge, 2020).

4.2. Unit roots

Augmented Dicky Fuller (ADF) test, based on Fuller (1976) and Dickey and Fuller (1979), has been used to know the stationarity of the data analyzed in this paper. Also, a non-parametric estimate of spectral density of ut at the zero-frequency based on Phillips (1987) and Phillips and Perron (1988) and deterministic trend estimates based on Kwiatkowski et al. (1992), Elliot et al. (1996) and Ng and Perron (2001) have been used because they have a greater power of estimation. We get the same results.

4.3. ARFIMA (p, d, q) model

To carry out this research, we also employ fractionally integrated methods with the purpose of getting the time series to be stationary. We achieve this objective ($I(0)$) by differentiating the time series with a fractional number.

Following a mathematical notation, a time series $x_t, t = 1, 2, \ldots$ follows an integrated order process d (and denoted as $x_t \approx I(d)$) if:

$$(1 - L)^d x_t = u_t, t = 1, 2, \ldots, \tag{1}$$

where d refers to any real value, L refers to the lag-operator ($Lx_t = x_{t-1}$) and u_t is a covariance stationary process I(0) where the behavior of the spectral density function shows in the weak form a type of time dependence where the function is positive and finite at the zero frequency.

It is said that x_t is ARFIMA (p, d, q) when u_t is ARMA (p, q). So, depending on the value of the parameter d on (1) the reading of the results can be: x_t is anti-persistent if $d < 0$ (see Dittmann et al., 2002); when $d = 0$ in (1) we say that the process is short memory I(0); with a high degree of association over a long time we say that the process is long memory ($d > 0$); $d < 1$ means that the shock is transitory and the series revert to the mean; finally, when $d \geq 1$ we expect that the shocks will be permanent.

We follow the methodology proposed by Sowell (1992) instead of others (see Geweke and Porter-Hudak, 1983; Phillips, 1999, 2007; Sowell, 1992; Robinson, 1994, 1995a,b; etc.) and to select the most appropriate ARFIMA model we use the Akaike information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Akaike, 1979).

4.4. FCVAR

To check a multivariate model from a fractional point of view, Johansen (2008) introduced the Fractionally Cointegrated Vector AutoRegressive (FCVAR). This model was further expanded by Johansen and Nielsen (2010, 2012) being a step ahead of the Cointegrated Vector AutoRegressive (CVAR) model proposed by Johansen (1996), allowing that the series are integrated in order d and cointegrated in order d – b, with b > 0. To understand the FCVAR model, first we introduce the non-fractional CVAR model.

If Y_t, $t = 1, \ldots, T$ is a p-dimensional I(1) time series, the CVAR model can be represented in the following way:

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

Where the difference and the lag operator is represented by Δ^b and $L_b = 1 - \Delta^b$ and are used to derive the FCVAR model, obtaining:

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

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

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

where, ε_t is a mean zero term, Ω is the p-dimensional independent and identical distributed variance-covariance matrix. The α and β matrices have a form $p \times r$ where $0 \leq r \leq p$. β allows us to know the relationship in the long run in terms of cointegration. Γ_i is related to the short-run behavior of the variables and their controls. Finally, α indicates the deviations from the equilibria and the speed in the adjustment.

The FCVAR model is developed in a computer programming language such as Matlab (Nielsen and Popiel, 2018). It has been employed in numerous empirical papers (Baruník and Dvořáková, 2015; Maciel, 2017; Aye et al., 2017; Dolatabadi et al., 2018; Jones, Nielsen and Popiel, 2018; Gil-Alana and Carcel, 2018; Poza and Monge, 2020; etc.).

4.5. Wavelet analysis

Wavelet methodology allows time series to be analyzed in the time-frequency domain. Thus, for this research article we use two tools named wavelet coherency and wavelet phase difference,[2] because the requirement of stationarity is not necessary and studying the interaction in time and frequency domain of the time series reveals evidence of potential changes (structural changes).

Furthermore, the most important information is hidden in the frequency content of the signal. So, as we know, we can define the time series as an aggregation of components operating on different frequencies.

Finally, if we follow the research carried out by Zhou (2008), Podobnik and Stanley (2008), Gu and Zhou (2010) and Jiang and Zhou (2011) we can conclude that misleading results will be found if we apply a typical cross-correlation to study statistical relationships between two multifractal time series.

The wavelet coherency plot represents the correlation of time series and helps us to identify hidden patterns and/or information in the time-frequency domain. The wavelet transform, represented by $WT_x(a,\tau)$, of a time series $x(t)$ obtained by projecting a mother wavelet ψ, is defined as:

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

where the wavelet coefficients of $x(t)$ are represented by $WT_x(a,\tau)$ and provide information on time and frequency by mapping the original time series onto a function of τ and a. Following Aguiar-Conraria and Soares (2014) we choose the Morlet wavelet as the mother wavelet because it is a complex sine wave within a Gaussian envelope, so we will be able to measure the synchronism between time series.

Wavelet coherence helps us understand how one time series interacts with another. We can define this term as:

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

where the smoothing operator in time and scale is represented with the parameter *SO*. This operator is important because without it, the wavelet coherency is always one for all times and scales (Aguiar-Conraria et al., 2008). In Aguiar-Conraria's website[3] we can find the Matlab codes for the CWT resolution.

4.6. Time series forecasting with machine learning models

4.6.1. Forecasting with KNN regression

The KNN regression holds a i-th training instance with a vector of n features $(f_1^i, f_2^i, \ldots, f_n^i)$ that describes the instance and the associated target vector of m attributes $(t_1^i, t_2^i, \ldots, t_n^i)$. At this point, where we do not know the target but we know the features of the new instance $(q_1, q_2, \ldots, q_n)$, the features of this new instance are used to find the k most similar training instances according to the vectors of features and a similarity or distance metric, measured by the Euclidean distance computed as:

$$\sqrt{\sum_{x=1}^{n}\left(f_x^i - q_x\right)^2}$$

Given that KNN learns by analogy, the new instances that are closest to the k training instances are considered the k nearest neighbors (the most similar instances), assuming that the targets of its nearest neighbors are probably similar to its unknown target. Thus, the targets or the k nearest neighbors are aggregated to predict the target of the new instance as:

$$\sum_{i=1}^{k}\frac{t^i}{k}$$

At this point, we have an autoregressive model where the target associated with a training instance is a collection of values of the time series and the features describing the instance are lagged values of the target. So, the use of KNN for time series forecasting is based on the repetitive pattern in the past of the series.

4.6.2. Forecasting with Artificial Neural Networks

A Neural Network is based on a number of hidden and output layers, in each layer we have to consider the number of neurons, the training algorithm parameters and the performance measure.

According to Güler and Übeyli (2005), to find the optimal number of hidden layers there is no general rule, the most popular

[2] Continuous Wavelet Transform (CWT) has been applied in several finance and economics research papers such as Kyrtsou et al. (2009), Crowley and Mayes (2009), Vacha and Barunik (2012), Aguiar-Conraria and Soares (2011, 2014), Dewandaru et al. (2016), Tiwari et al. (2016), Jammazi et al. (2017), among others.

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

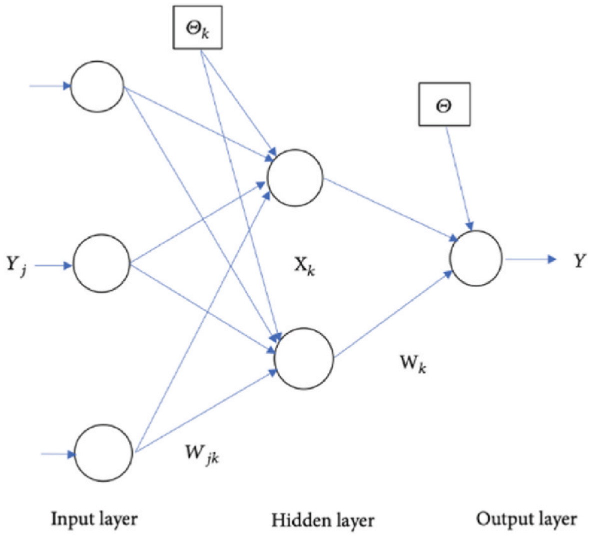

Fig. 1. Artificial Neural Network for time series forecasting.

approach is by trial and error.

The architecture of the neural network can be generalized as follows:

$$I - (H_1, H_2, H_3, \ldots, H_N) - O$$

where I represents the number of input nodes, H_n the number of neurons in hidden layer n, and O the number of neurons in the output layer.

The training function of an ANN is an optimization process based on a network training function that updates weights and bias values during training (Prasad and Bhagwat, 2002).

According to Fig. 1, Y_j represents the input vector denoted by $Y_j = \{y_1, y_2, y_3, \ldots, y_n\}$; the connection weight vector of the j nodes of the input layer to the k nodes of the hidden layer is represented by $W_{jk}(j = 1,2,3,\ldots,n; k = 1,2,3,\ldots,m)$; the vector of the k neurons in the hidden layer is the $X_k(k = 1, 2, 3, \ldots, m)$ determined by the formula $X_k = f(\sum_{j=1}^{n} W_{jk} Y_j + \Theta_k)$; the connection weights of the k nodes of the hidden layer to the output layer is represented by $W_k(k = 1, 2, 3, \ldots, m)$; the unit output vector for the neural network with one output neuron is Y which is determined by the formula $Y = f(\sum_{k=1}^{n} W_k X_k + \Theta)$. Finally, $\Theta_k(k = 1, 2, 3, \ldots, k)$ is the bias value of the hidden layer nodes and Θ is the bias value of the output layer.

Once the data has been encoded, enriched, cleaned (taking into account the noise and the missing information) normalized to an interval (0,1) in order to prevent saturation of hidden nodes before feeding into the neural networks (Kitapcı et al., 2014), then the network model building process begins.[4]

5. Empirical results

5.1. Real-Time Leading Economic Indicator (RT-LEI)

Through implementing the factor analysis by means of the 14 variables previously outlined, we generate 6 dimensions that condense all the information to create the final indicator. The 6 latent variables are: Chemical-Agriculture Index, Steel Index, Distrust Index, Housing Index, Financial Index and Motor Vehicle Index. Results are displayed in detail across Fig. 2.

RT-LEI = 0.19 Chemical-Agriculture Index +0.20 Steel Index – 0.12 Distrust Index +0.20 Housing Index +0.14 Financial Index +0.14 Motor Vehicle Index.

The two most important dimensions are the "Steel Index" and the "Housing Index", closely followed by the "Chemical-Agriculture Index". Then, the "Financial Index" and the "Motor Vehicle Index" have less weight but complete key information regarding the industry sector and the financial conditions in China. Finally, the "Distrust Index", understood as a sentiment indicator (pessimistic expectations), maintains a negative relationship with economic activity. It is well known that the more pessimism there is in an economy, the less consumption and investment there will be (The Economist, 2010), which results in a decrease in aggregate demand and in the GDP. Consequently, this fall in the economic activity provokes a deterioration in the labor market.

All the factor analysis implemented were consistent according to the KMO and Bartlett test (Table 2). Furthermore, RT-LEI's weights derive from the component matrices (factor loads), that deduce the magnitude of each original variable within its dimension.

[4] For more information see https://cran.r-project.org/web/packages/nnfor/nnfor.pdf.

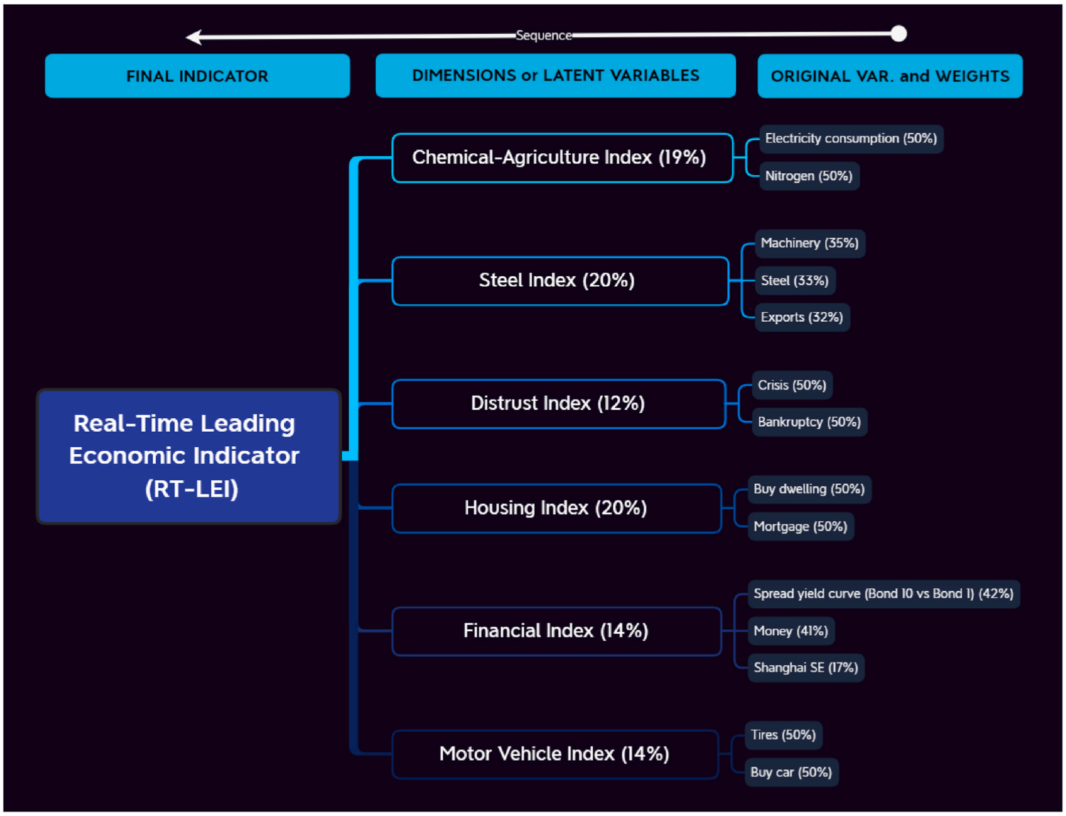

Fig. 2. Real-time leading economic indicator.

Table 2
Factor analysis Test.

Factor Analysis							
	FA1	FA2	FA3	FA4	FA5	FA6	Final FA
KMO	0.500	0.667	0.500	0.500	0.511	0.500	0.687
Bartlett Test	0.000	0.000	0.001	0.000	0.011	N.A.	0.000

Source: Own elaboration

Table 3
Correlation coefficients RT-LEI, CLI and GDP.

	Correlations					
			GDP	GDP (3)	CLI	RT-LEI
Rho de Spearman	GDP	Correlation coefficient	1.000	0.956[a]	0.821[a]	**0.860**[a]
		Sig. (bilateral)	.	0.000	0.000	0.000
		N	201	198	201	189
	GDP (3)	Correlation coefficient	0.956[a]	1.000	0.905[a]	**0.910**[a]
		Sig. (bilateral)	0.000	.	0.000	0.000
		N	198	198	198	186
	CLI	Correlation coefficient	0.821[a]	**0.905**[a]	1.000	0.832[a]
		Sig. (bilateral)	0.000	0.000	.	0.000
		N	201	198	201	189
	RT-LEI	Correlation coefficient	0.860[a]	**0.910**[a]	0.832[a]	1.000
		Sig. (bilateral)	0.000	0.000	0.000	.
		N	189	186	189	190

[a] The correlation is significant at the 0.01 level (bilateral).

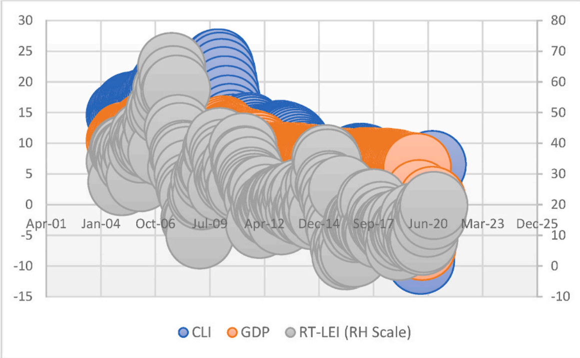

Fig. 3. Bubble chart of economic activity.

Table 4
Unit roots tests.

	ADF			PP		KPSS	
	(i)	(ii)	(iii)	(ii)	(iii)	(ii)	(iii)
CLI	−1.1944	−1.7636	−2.8478	−2.0029	−3.2796	2.5411	0.0544
GDP	−1.356	−2.4495	−4.8514	−1.1639	−3.1269	2.6251	0.1002
RT_LEI	−1.0713	−1.8876	−3.4934	2.2781	−4.4071	2.7213	0.093

(i) Refers to the model with no deterministic components; (ii) with an intercept, and (iii) with a linear time trend. I reflect t-statistic with test critical value at 5%.

Once the RT-LEI has been created, we calculate correlations between RT-LEI, CLI and GDP at constant prices to contrast the degree of association. As in Poza and Monge (2020), our results show that the RT-LEI allows us to estimate GDP behavior slightly better than OECD's CLI. The Spearman correlation coefficient between RT-LEI and real GDP is 0.860 compared with 0.821 between CLI and real GDP. Moreover, the RT-LEI vs GDP correlation improves when a lag of 3 months in GDP is applied: from 0.860 to 0.910 (99% confidence interval, see Table 3), which reinforces the capacity of RT-LEI as a leading indicator.

In this regard, RT-LEI substantially covers GDP trends (Fig. 3). The bubble chart lays out the evolution of CLI, GDP and RT-LEI, and the coincidence seems similar enough to merit a more thorough analysis in the next section by means of cointegration techniques.

5.2. *Unit roots*

Three standard unit roots tests have been calculated to analyze the statistical properties of the CLI, GDP and the constructed Real Time Leading Economic Indicator (RT_LEI).

Table 4 shows the results that we get using the Augmented Dickey-Fuller (ADF) test, the Phillips Perron (PP) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS). The results suggest that since the hypothesis about the existence of a unit root or non-stationary I(1) behavior is not rejected, the shocks in these cases are permanent.

5.3. *Fractional integration*

Using unit root methods in the thee time series we assume that we have to use first differences as we have verified that the data is non-stationary I(1). However, we use fractional alternative, like ARFIMA (p, d, q) models, due to the lower power of the unit root methods[5] to study the persistence of the GDP, CLI and the new Real Time Leading Economic Indicator (RT_LEI) time series.

To get the appropriate AR and MA orders in the model, we consider the Akaike information criterion (AIC; Akaike, 1973) and the Bayesian information criterion (BIC; Akaike, 1979).[6]

For each time series, we show in Table 5 the fractional parameter d and the AR and MA orders obtained using Sowell's (1992) maximum likelihood estimator and taking into consideration p, q $\leq$ 2.

Focusing on the GDP and RT_LEI, we observe that the values of *d* are in the range (0, 1), implying fractional integration and concluding that both time series are mean reverting, where the order of integration for both time series are smaller than 1 implying that shocks have or will have temporary effects and will disappear by themselves in the long run. Additionally, we want to emphasize the

[5] See Diebold and Rudebush (1991), Hassler and Wolters (1994) and Lee and Schmidt (1996).

[6] A point of caution should be adopted here since the AIC and BIC may not necessarily be the best criteria for applications involving fractional models (Hosking, 1981).

Table 5
Results of long memory tests.

Long memory test						
Data analyzed	Sample size (weeks)	Model Selected	d	Std. Error	Interval	I(d)
CLI	189	ARFIMA (1, d, 1)	1.1390903	0.0979081	[0.98, 1.30]	I(1)
GDP	189	ARFIMA (0, d, 2)	0.879694	0,0069677	[0.86, 0.89]	I(d)
RT_LEI	189	ARFIMA (2, d, 1)	0.9229475	0.1721627	[0.91, 0.93]	I(d)

We observe from Table 5 that the estimates of d in the case of GDP and RT_LEI are equal; therefore, we discard the use of the CLI time series since the behavior differs from the GDP.

Table 6
Results of the FCVAR model ($\boldsymbol{d \neq b}$).

	d	Cointegrating equation beta		
		CLI	GDP	RT-LEI
Panel I: CLI, GDP, RT-LEI	d = 1.875 (0.076)	1.000	0.536	−0.833
	b = 1.108 (0.092)			
	$\Delta^d\left(\begin{bmatrix} CLI \\ GDP \\ RT\_LEI \end{bmatrix} - \begin{bmatrix} 14.262 \\ 10.228 \\ 32.787 \end{bmatrix}\right) = L_d \begin{bmatrix} 0.026 \\ -0.036 \\ 0.016 \end{bmatrix} \nu_t + \sum_{i=1}^{2} \widehat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$			
Panel II: CLI, GDP	d = 1.972 (0.070)	1.000	28.761	–
	b = 1.078 (0.070)			
	$\Delta^d\left(\begin{bmatrix} CLI \\ GDP \end{bmatrix} - \begin{bmatrix} 14.274 \\ 10.293 \end{bmatrix}\right) = L_d \begin{bmatrix} -0.019 \\ -0.012 \end{bmatrix} \nu_t + \sum_{i=1}^{2} \widehat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$			
Panel III: RT-LEI, GDP	d = 0.791 (0.222)	–	−2.658	1.000
	b = 0.791 (0.179)			
	$\Delta^d\left(\begin{bmatrix} RT\_LEI \\ GDP \end{bmatrix} - \begin{bmatrix} 32.881 \\ 10.313 \end{bmatrix}\right) = L_d \begin{bmatrix} -0.014 \\ 0.011 \end{bmatrix} \nu_t + \sum_{i=1}^{2} \widehat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$			

high degree of persistence of both time series (GDP and RT-LEI, respectively) that will allow us to perform forecast operations in the long-term without suffering any deviation in our results.

5.4. FCVAR model

To contrast the possible existence of persistence in the long run co-movements of the series, Table 6 summarize the results of the FCVAR model proposed by Johansen and Nielsen (2012).

According to Jones et al. (2014), we use the lag value (k = 3), deterministic components and cointegration rank (r) recommended by them to get our results.

From Panel I in Table 6 we observe that the order of integration of the individual series is about 1.875, while the reduction in the degree of integration in the cointegrating regression is 1.108 implying thus, an order of integration of about 0.767 for the cointegrating relationship. We conclude then that the resulting long run equilibrium time series follows a long memory process, which suggests potential forecasting power at longer horizons (Baillie and Bollerslev, 1994). Also, this value ($d - b$= 0.767) implies that the error correction term follows a nonstationary process, although it is mean-reverting, and the shock duration is long-lived.

From Panel III we observe that the order of integration of RT_LEI and GDP time series is $(d - b) = 0$, because the order of integration of individual series is about 0.791 and the reduction term are equal. This result implies I(0) cointregation errors. So, we cannot reject the hypothesis where the shock duration is short-lived due to the error correction term shows short-run stationary behavior.

5.5. Wavelet analysis

We use a multivariate wavelet analysis based on the time-frequency domain to show which leading indicator (CLI or RT_LEI) best fits GDP. Also, with this methodology we can see when a structural change occurs in the behavior of the leading indicator with respect to our reference time series, GDP.

Fig. 4 represents three different estimations. The first one is the wavelet coherency (a) that represents the interrelations between CLI vs GDP and RT-LEI vs GDP, when they are stronger or not and at which frequencies these points occur. The main regions with statistically significant coherency are in low frequencies (corresponding to cycles between 13 and 64 months with statistically significant coherence) in Panel I and II. The most important ones start around 2007 in all series at which point the coherence indicates how important and strong the relation is between the time series.

The second and the third estimations are the partial phase-difference (b) and the partial gain. The partial phase-difference gives us information about the movement of one time series with respect to the other. Focusing on the periods mentioned before and looking at the 5% significance level, the phase difference is between 0 and $\pi/2$ which means that CLI vs GDP and RT-LEI vs GDP are in phase (positive correlated) with the CLI and RT-LEI leading over GDP and achieving a partial gain (regression coefficient in the regression of

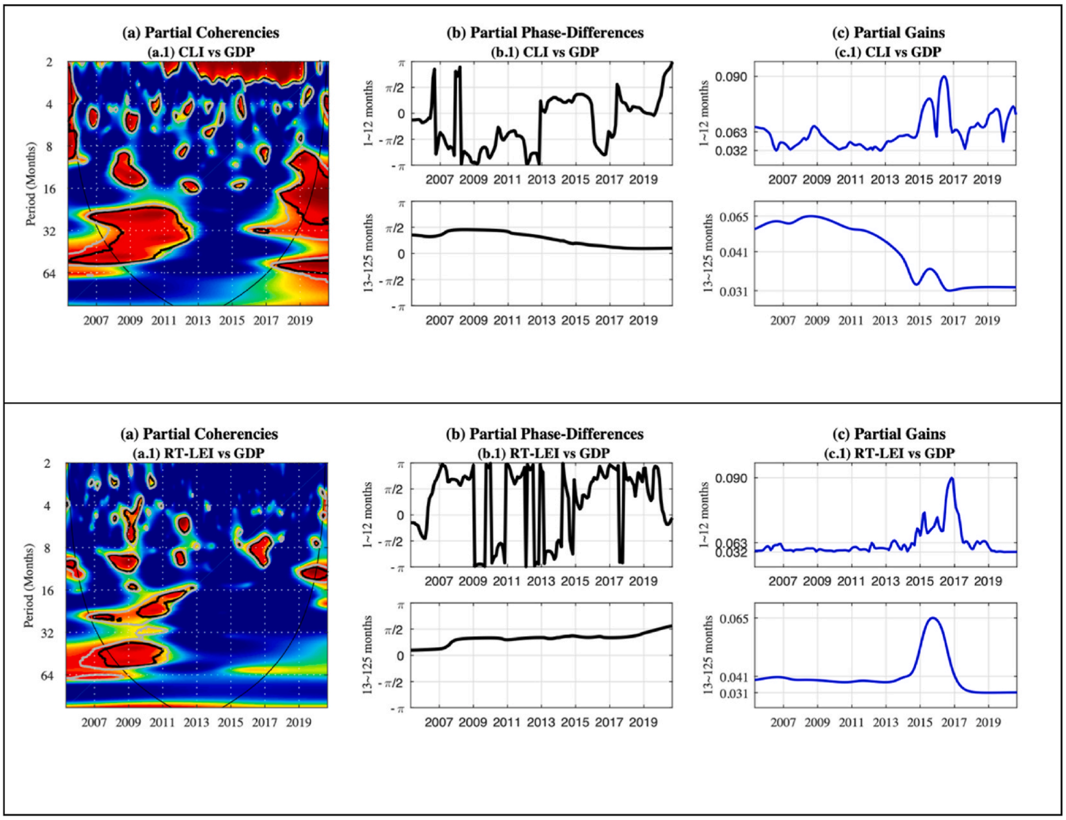

Fig. 4. Wavelet Coherency, Phase-differences and Wavelet gain between leading indicators and GDP.

leading indicators on GDP at each time and frequency) of 0.065.

5.6. *Forecasting for a real time leading indicator using machine learning*

Based on the results obtained up to this point, we use advance computational intelligence techniques based on machine learning to forecast our real time leading economic indicator.

These methodologies present interesting features such as their nonlinearity or the lack of an underlying model (non-parametric model) to obtain the results.

Following Wu et al. (2008), the KNN model is a very popular algorithm used in classification and regression, which tries to find its k most similar examples (nearest neighbors) in accordance with the Euclidean distance, predicting its value as an aggregation of the target values associated with its k most similar examples. To use this model, we have followed the criterion explained in Martínez et al. (2019) and based on Hibon and Evgeniou (2005) to specify the k parameter. If *k* is too small or too large, this may affect the prediction by adding noise or using examples far from the new instance. So, our forecast using the KNN model is not based on a single value of *k* but on a vector of *k* values. The vector of *k* used in our model is equal to [1, 100] and the forecast is the average of the forecasts produced by the hundred models.

Also, we use a Multilayer Perceptron (MLP) neural network for time series prediction, this being one of the most widely implemented neural networks based on the back-propagation rule where the errors are propagated through the network and allow the adaption of the hidden processing elements. In addition, the MLP has massive interconnectivity that means that any element of a given layer feeds all the elements of the next layer and is trained with error correction learning.

To find out which is the most accurate prediction model and following the advice in the literature (Mapuwei et al., 2020), we have used the Root Mean Square Error (RSME). In Table 7 we present the accuracy of the real time leading indicator using the KNN model and the ANN model. The results obtained in both cases are in line with the literature, which states that the most prominent machine learning technique used in time series forecasting is the Artificial Neural Networks, getting a value very close to zero (0.048), indicating that this is the best model with which to predict the real time leading indicator.

To compare the two models, we have plotted the time series in Fig. 5. In both charts, the black line represents the original time series. In the figure on the left, the prediction of the real time leading indicator in the next 24 months is represented in red. In the figure on the right, in blue.

In both estimations we appreciate the impact of SARS-CoV-2 on the Chinese economy and its behavior in the next 24 periods. These

Table 7
Accuracy of the real time leading indicator using the KNN model and the ANN model.

Real Time Leading Economic Indicator for China (RT_LEI)	
	RMSE
KNN model	4.03
Artificial Neural Networks (ANN) model	0.048

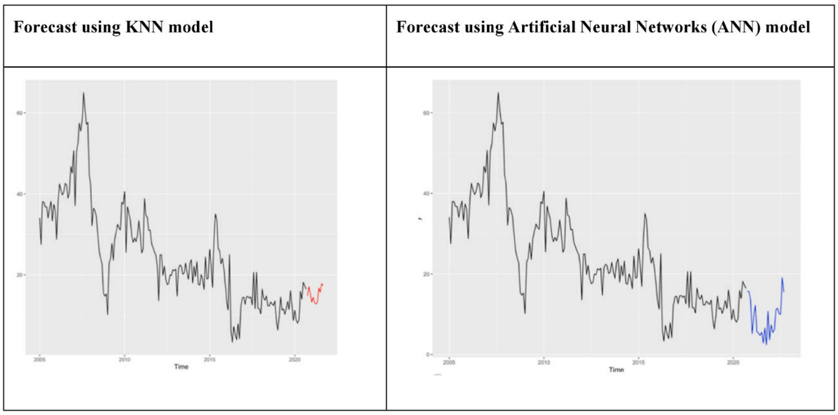

Fig. 5. Forecast machine learning models.

results are not in line with others forecasts such as IMF, in which an annual growth rate of 1.8% is expected in 2020 and 8.2% in 2021 (IMF, 2020). Our predictions indicate that the Chinese economy will begin the recovery end-2021 and will consolidate it in mid-2022, but not before as the IMF stated in the last October WEO. We find two reasons for this discrepancy: 1) the IMF forecast dates from October, when the second and the third waves of COVID had not started in some Chinese international trade partners. However, our model includes data until and including November. And 2) expectations about the vaccine are becoming more moderate, underlining the extent of the challenge to gaining herd immunity. Experts point out it may be achieved at the end of 2021 (Global Times, 2021).

6. Conclusions

The main aim of this article was to create a real time leading economic indicator (RT-LEI) for China, combining Text Mining, Factor Analysis, and data from Google Trends. We also used a Fractional Cointegration Vector AutoRegressive model, Continuous Wavelet Transform and Machine Learning techniques to analyze the dynamics, interconnections and structural changes in the time-series studied, as well as to forecast Chinese economic activity in times of COVID.

Concerning the RT-LEI, we have used 14 variables that have been condensed into six dimensions: Chemical-Agriculture Index, Steel Index, Distrust Index, Housing Index, Financial Index and Motor Vehicle Index. The results show that the proposed RT-LEI can predict GDP trends and turning points slightly better than CLI's OECD. In fact, the highest correlation between the RT-LEI and GDP may be observed with a lead of 3 months, RT-LEI (−3). We could highlight two causes to explain this improvement: RT-LEI can be published in real time while CLI is released on a monthly basis and RT-LEI uses big data and Text Mining, instead of surveys, to extract information from the most representative economic sectors.

The results obtained using fractional integration and the FCVAR Model suggest that the GDP and RT-LEI are mean reverting implying that COVID shocks will have temporary effects and will disappear by themselves in the long term. In addition, the CWT analysis show that RT-LEI fits GDP better than CLI, which is in line with the correlation matrix. And finally, the Machine Learning techniques suggest that the Chinese economy will consolidate its recovery in 2022. After the strong deceleration, its economic activity will show an upward trend in the last quarter 2021, boosted by fiscal stimulus and exports.

Declaration of competing interest

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

References

Adams, B., Bottrelier, P., Ozyildirim, A., Sima-Friedman, J., 2010. On the Selection of Leading Economic Indicators for China. Economics Program Working Papers 10-02. The Conference Board, Economics Program.

Aguiar-Conraria, L., Soares, M.J., 2011. Oil and the macroeconomy: using wavelets to analyze old issues. Empir. Econ. 40, 645–655.

Aguiar-Conraria, L., Soares, M.J., 2014. The continuous wavelet transform: moving beyond uni- and bivariate analysis. J. Econ. Surv. 28, 344–375.

Aguiar-Conraria, L., Azevedo, N., Soares, M.J., 2008. Using wavelets to decompose the time-frequency effects of monetary policy. Phys. Stat. Mech. Appl. 387, 2863–2878.

Akaike, H., 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60 (2), 255–265.

Akaike, H., 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrika 66 (2), 237–242.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. J. Monetary Econ. 50, 745–787.

Antenucci, D., Cafarella, M., Levenstein, M., Re, Che, Shapiro, M.D., 2014. Using Social Media to Measure Labor Market Flows. National Bureau of Economic Research.

Anton, C.E., Lawrence, C., 2016. The relationship between place attachment, the theory of planned behavior and residents' response to place change. J. Environ. Psychol. 47, 145–154.

Anvik, C., Gjelstad, K., 2010. Just Google it. Forecasting Norwegian Unemployment Figures with Web Queries. Center for Research in Economics and Management. Working Paper N. 11.

Askitas, N., Zimmermann, K.F., 2009. Google econometrics and unemployment forecasting. Appl. Econ. Q. 55 (2), 107–120.

Aye, G.C., Carcel, H., Gil-Alana, L.A., Gupta, R., 2017. Does gold act as a hedge against inflation in the UK? Evidence from a fractional cointegration approach over 1257 to 2016. Resour. Pol. 54, 53–57.

Baillie, R.T., Bollerslev, T., 1994. Cointegration, fractional cointegration, and exchange rate dynamics. J. Finance 49 (2), 737–745.

Banerjee, A., Marcellino, M., Masten, I., 2005. Leading indicators for euro-area inflation and GDP growth. Oxf. Bull. Econ. Stat. 67, 785–813.

Bangwayo-Skeete, P.F., Skeete, R.W., 2015. Can google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. Tourism Manag. 46, 454–464.

Baruník, J., Dvořáková, S., 2015. An empirical model of fractionally cointegrated daily high and low stock market prices. Econ. Modell. 45, 193–206.

Bauer, M.D., Mertens, T.M., 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.

Beracha, E., Wintoki, M.B., 2013. Forecasting residential real estate price changes from online search activity. J. R. Estate Res. 35 (3), 283–312.

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

Bok, B., Caratelli, D., Giannone, D., Sbordone, A.M., Tambalotti, A., 2018. Macroeconomic nowcasting and forecasting with big data. Annual Rev. Econ. 10, 615–643.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. J. Comput. Sci. 2, 1–8.

Chen, S., Yang, J., Yang, W., Wang, C., Bärnighausen, T., 2020. COVID-19 control in China during mass population movements at New Year. Lancet 395 (10226), 764–766.

Choi, H., Varian, H., 2009b. Predicting initial claims for unemployment insurance using google trends. Technical Report.

Choi, H., Varian, H., 2012. Predicting the present with google trends. Econ. Rec. 88, 2–9.

Croushore, D., 2011. Frontiers of real-time data analysis. J. Econ. Lit. 49, 72–100.

Crowley, P.M., Mayes, D.G., 2009. How fused is the Euro area core? An evaluation of growth cycle co-movement and synchronization using wavelet analysis. OECD J.: J. Bus. Cycle Meas. Anal. 2008 (1), 63–95.

D Amuri, F., 2009. Predicting Unemployment in Short Samples with Internet Job Search Query Data. University Library of Munich, Germany. MPRA Paper 18403.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. J. Finance 66 (5), 1461–1499.

Dewandaru, G., Masih, R., Masih, A.M.M., 2016. Contagion and interdependence across Asia-Pacific equity markets: an analysis based on multi-horizon discrete and continuous wavelet transformations. Int. Rev. Econ. Finance 43, 363–377.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74 (366a), 427–431.

Diebold, F.X., Rudebush, G.D., 1991. On the power of Dickey-Fuller tests against fractional alternatives. Econ. Lett. 35, 155–160.

Dittmann, I., Granger, C., W, J., 2002. Properties of nonlinear transformations of fractionally integrated processes. J. Econom. 110, 113–133.

Dolatabadi, S., Narayan, P.K., Nielsen, M, Ø., Xu, K., 2018. Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. J. Futures Mark. 38 (2), 219–242.

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

D'Amuri, F., Marcucci, J., 2017. The predictive power of google searches in forecasting unemployment. Int. J. Forecast. 33 (4), 801–816.

Eagle, N., Macy, M., Claxton, R., 2010. Network diversity and economic development. Science 328 (5981), 1029–1031.

Elliot, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.

ECB, 2020. Monetary Aggregates. ECB Data Portal. European Central Bank. Available in: https://data.ecb.europa.eu/data/datasets/BSI/BSI.M.U2.Y.V.M30.X.I.U2.2300.Z01.A.

Estrella, A., Trubin, M.R., 2006. The yield curve as a leading indicator: some practical issues. Curr. Issues in Econ. Finance 12.

FED, 2020. Money Stock Measures. Board of Governors of the Federal Reserve System. Available in: https://www.federalreserve.gov/releases/h6/current/default.htm.

Foldvary, F.E., 1991. Real estate and business cycles: henry george's theory of the trade cycle. In: Lafayette College Henry George Conference. June 1991.

Fondeur, Y., Karamé, F., 2013. Can google data help predict French youth unemployment? Econ. Modell. 30 (C), 117–125.

Fuller, W.A., 1976. Introduction to Statistical Time Series. JohnWiley. FullerIntroduction to Statistical Time Series1976, New York.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Anal. 4 (4), 221–238.

Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: the real-time informational content of macroeconomic data. J. Monetary Econ. 55, 665–676.

Gil-Alana, L.A., Carcel, H., 2018. A fractional cointegration var analysis of exchange rate dynamics. N. Am. J. Econ. Finance V51 (C).

Global Times, 2021. Herd Immunity in China Could Be Realized at End of 2021: Expert. Global Times. January 2021.

Gu, G.F., Zhou, W.X., 2010. Detrending moving average algorithm for multifractals. Phys. Rev. 82 (1), 011136.

Güler, İ., übeyli, E.D., 2005. An expert system for detection of electrocardiographic changes in patients with partial epilepsy using wavelet-based neural networks. Expet Syst. 22 (2), 62–71.

Guo, F., Ozyildirim, A., Zarnowitz, V., 2009. On the measurement and analysis of aggregate economic activity for China: the coincident economic indicators approach. China Econ. J. 2 (2), 159–186.

Hassler, U., Wolters, J., 1994. On the power of unit root tests against fractional alternatives. Econ. Lett. 45 (1), 1–5.

Heinisch, K., Scheufele, R., 2018. Should forecasters use real-time data to evaluate leading indicator models for GDP prediction? German evidence. Ger. Econ. Rev. 20, 170–200.

Hibon, M., Evgeniou, T., 2005. To combine or not to combine: selecting among forecasts and their combinations. Int. J. Forecast. 21 (1), 15–24.

Hosking, J.R.M., 1981. Modelling persistence in hydrological time series with using fractional differentiation. Water Resour. Res. 20, 1898–1908.

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

Hui, D.S., Azhar, E.I., Madani, T.A., Ntoumi, F., Kock, R., Dar, O., Ippolito, G., Mchugh, T.D., Memish, Z.A., Drosten, C., Zumla, A., Petersen, E., 2020. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. Int. J. Infect. Dis. 91, 264–266.

ICEX, 2017. Informe económico y comercial sobre China. In: Oficina Económica y Comercial en Pekín. Secretaría de Estado de Comercio.

IMF, 2020. World Economic Outlook. International Monetary Fund, USA. October 2020.

Jaccard, I., 2007. House Prices, Real Estate Returns, and the Business Cycle. Swiss Finance Institute Research Paper, pp. 6–37.
Jammazi, R., Ferrer, R., Jareño, F., Shahzad, S.J.H., 2017. Time-varying causality between crude oil and stock markets: what can we learn from a multiscale perspective? Int. Rev. Econ. Finance 49, 453–483.
Jiang, Z.Q., Zhou, W.X., 2011. Multifractal detrending moving-average cross-correlation analysis. Phys. Rev. 84 (1), 016106.
Johansen, S., 1996. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, New York, NY.
Johansen, S., 2008. A representation theory for a class of vector autoregressive models for fractional processes. Econom. Theor. 24, 651–676.
Johansen, S., Nielsen, M.Ø., 2010. Likelihood inference for a nonstationary fractional autoregressive model. J. Econom. 158, 51–66.
Johansen, S., Nielsen, M.O., 2012. Likelihood inference for a fractionally cointegrated vector autoregressive model. Econometrica 80, 2667–2732.
Jones, M.E., Nielsen, M.Ø., Popiel, M.K., 2014. A fractionally cointegrated VAR analysis of economic voting and political support. Can. J. Econ. 47, 1078–1130.
Kim, E., Lee, J.H., Sung, Y., Choi, S.M., 2016. Predicting selfie-posting behavior on social networking sites: an extension of theory of planned behavior. Comput. Hum. Behav. 62, 116–123.
Kitapcı, O., Özekicioğlu, H., Kaynar, O., Taştan, S., 2014. The effect of economic policies applied in Turkey to the sale of automobiles: multiple regression and neural network analysis. Procedia-Social and Behavioral Sci. 148, 653–661.
Kozicki, S., 2002. Comments on 'Forecasting with a real-time data set for macroeconomists. J. Macroecon. 24, 541–557.
Kwiatkowski, D., Phillips, P.C., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. J. Econom. 54 (1–3), 159–178.
Kyrtsou, C., Malliaris, A., Serletis, A., 2009. Energy sector pricing: on the role of neglected nonlinearity. Energy Econ. 31 (3), 492–502.
Lee, V.J., Chiew, C.J., Khong, W.X., 2020. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. J. Trav. Med. 27 (3), taaa039.
Li, X., Pan, B., Law, R., Huang, X., 2017. Forecasting tourism demand with composite search index. Tourism Manag. 59, 57–66.
Liu, Y., Lv, B., Peng, G., Yuan, Q., 2012. A preprocessing method of Internet search data for prediction improvement: application to Chinese stock market. Knowledge Discovery and Data Mining (KDD) 1–7 article 3.
Maciel, L., 2017. Technical analysis based on high and low stock prices forecasts: evidence for Brazil using a fractionally cointegrated VAR model. Empir. Econ. V58 (4), 1513–1540.
Mapuwei, T.W., Bodhlyera, O., Mwambi, H., 2020. Univariate time series analysis of short-term forecasting horizons using artificial neural networks: the case of public ambulance emergency preparedness. J. Appl. Math. 2020, 1–11.
Martínez, F., Frías, M.P., Charte, F., Rivera, A.J., 2019. Time series forecasting with KNN in R: the tsfknn package. The R J. 11 (2), 229–242.
McLaren, N., Shanbhogue, R., 2011. Using internet search data as economic indicators. Bank Engl. Q. Bull. 51 (2), 134–140.
Monge, M., Gil-Alana, L.A., 2020. Water prices: persistence, mean reversion and trends. Water Pol. 22 (6), 1200–1216.
Monokroussos, G., Zhao, Y., 2020. Nowcasting in real time using popularity priors. Int. J. Forecast. 36 (3), 1173–1180.
Munda, G., Nando, M., 2005. Constructing Consistent Composite Indicators: the Issue of Weights. European Commission Directorate-General Joint Research Centre.
Musat, C.C., Trausan-Matu, S., 2009. In: Cruz, Díaz (Ed.), Sentiment Analysis on Economic Texts in Troyano, vol. 9. WOMSA, pp. 71–78.
Narayan, P.K., Narayan, S., Smyth, R., 2011. Energy consumption at business cycle horizons: the case of the United States. Energy Econ. 33, 161–167.
Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 519–1554.
Nielsen, M.Ø., Popiel, M.K., 2018. A Matlab Program and User's Guide for the Fractionally Cointegrated VAR Model (Queen's Economics Department Working Paper No 1330). Canada, Ontario. K7L 3N6.
Niesert, R.F., Oorschot, J.A., Veldhuisen, C.P., Brons, K., Lange, R.-J., 2020. Can Google search data help predict macroeconomic series? Int. J. Forecast. 36 (3), 1163–1172.
Nyman, R., Ormerod, P., Smith, R., Tuckett, D., 2014. Big data and economic forecasting: a top-down approach using directed algorithmic text analysis. In: Workshop on Big Data for Forecasting and Statistics. ECB, Frankfurt, pp. 7–8.
OECD, 2012. OECD System of Composite Leading Indicators. OECD. March 2012.
Pang, B., 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., Giannotti, F., 2016. An anaytical framework to nowcast well-being using mobile phone data. Int. J. Data Sci. Analytics 2, 75–92.
Pescyova, M., 2011. Vygooglime si budúcnos. In: Institute for Financial Policy at the Ministry of Finance of the Slovak Republic, Commentary No, 29, December.
Phillips, P.C., 1987. Time series regression with a unit root. Econometrica: J. Econom. Soc. 277–301.
Phillips, P.C.B., 1999. Discrete Fourier Transforms of Fractional Processes. Department of Economics, University of Auckland.
Phillips, P.C.B., 2007. Unit root log periodogram regression. J. Econom. 138 (1), 104–124.
Phillips, P.C., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.
Podobnik, B., Stanley, H.E., 2008. Detrended cross-correlation analysis: a newmethod for analyzing two nonstationary time series. Phys. Rev. Lett. 100 (8), 084102.
Potter, C.W., 2001. A history of influenza. J. Appl. Microbiol. 91 (4), 572–579.
Poza, C., Monge, M., 2020. A real time leading economic indicator based on text mining for the Spanish economy. Fractional cointegration VAR and Continuous Wavelet Transform analysis. Int. Econ. 163, 163–175.
Prasad, J.Y., Bhagwat, S.S., 2002. Simple neural network models for prediction of physical properties of organic compounds. Chem. Eng. Technol. 25, 1041–1046.
Ramey, V., Vine, D., 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.
Robinson, P.M., 1994. Efficient tests of nonstationary hypotheses. J. Am. Stat. Assoc. 89, 1420–1437.
Robinson, P.M., 1995a. Gaussian semi-parametric estimation of long-range dependence. Ann. Stat. 23, 1630–1661.
Robinson, P.M., 1995b. Log periodogram regression of time series with long range dependence. Ann. Stat. 23 (3), 1048–1072.
Rütting, T., Aronsson, H., Delin, S., 2018. Efficient use of nitrogen in agriculture. Nutrient Cycl. Agroecosyst. 110, 1–5. https://doi.org/10.1007/s10705-017-9900-8.
Saltelli, A., 2006. Composite indicators between analysis and advocacy. Soc. Indicat. Res. 81 (1), 65–77.
Schmidt, T., Vosen, S., 2009. Forecasting private consumption: survey-based indicators vs. Google trends. J. Forecasting 30 (6), 565–578.
Scott, S.L., Varian, H.R., 2012. Bayesian Variable Selection for Nowcasting Economic Time Series. National Bureau of Economic Research.
Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J. Econom. 53 (1–3), 165–188.
Stark, T., Croushore, D., 2002. Forecasting with a real-time data set for macroeconomists. J. Macroecon. 24, 507–531.
Suhoy, T., 2009. Query Indices and a 2008 Downturn: Israeli Data. Bank of Israel. Technical Report.
Szármes, P., 2015. Creating new types of business and economic indicators using big data technologies. Sci. J. Bus. Manag. 18–24.
The Economist, 2010. Guide to Economic Indicators. Making Sense of Economics, sixth ed. Bloomberg Press, USA.
Thoma, M., 2004. Electrical energy usage over the business cycle. Energy Econ. 26, 463–485, 2004.
Tiwari, A.K., Mutascu, M.I., Albulescu, C.T., 2016. Continuous wavelet transform and rolling correlation of European stock markets. Int. Rev. Econ. Finance 42, 237–256.
Toole, J.L., Colak, S., Sturt, B., Alexander, L.P., Evsukoff, A., Gonzales, M.C., 2015. The path most traveled: travel demand estimation using big data resources. Transport. Res. C Emerg. Technol. 58 (B), 162–177.
Trainer, E.M., 2006. Using Economic Indicators to Improve Investment Analysis, third ed. Wiley, United States.
Turley, J., 1976. Automobiles Sales in Perspective. Federal Reserve Bank of Sr, Louis, USA. June 1976.
Vacha, L., Barunik, J., 2012. Co-movement of energy commodities revisited: evidence from wavelet coherence analysis. Energy Econ. 34, 241–247.
Vaughan, L., Chen, Y., 2015. Data mining from web search queries: a comparison of google trends and Baidu index. J. Association for Inf. Sci. Technol. 66 (1), 13–22.
Vicente, M.R., López-Menéndez, A.J., Pérez, R., 2015. Forecasting unemployment with internet search data: does it help to improve predictions when job destruction is skyrocketing? Forecasting and Social Change 132–139.
Vlastakis, N., Markellos, R., 2012. Information demand and stock market volatility. J. Bank. Finance 36 (6), 1808–1821.

Wiebe, J., Wilson, T., Cardie, C., 2005. Annotating expressions of opinions and emotions in language. Comput. Humanit. 39, 165–210.
World Health Organization, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline#event-0.
Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., Zhou, Z.H., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14 (1), 1–37.
Yang, M., Sommez, M.M., Li, Q., Duan, Y., 2015. The power of triple contexts on customer-based brand performance—a comparative study of Baidu and Google from Chinese netizens' perspective. Int. Bus. Rev. 11–22.
Zhou, W.X., 2008. Multifractal detrended cross-correlation analysis for two nonstationary signals. Phys. Rev. 77 (6), 066211.