FORECASTING SPANISH ECONOMIC ACTIVITY IN TIMES OF COVID-19 BY MEANS OF THE RT-LEI AND MACHINE LEARNING TECHNIQUES

Manuel Monge Moreno

Carlos Poza

Universidad Francisco de Vitoria, Madrid, Spain

Abstract: The main aim of this paper is to analyze and estimate the behavior of the Spanish economic activity in the next 12 months, by means of a Real-Time Leading Economic Indicator (RT-LEI), based on Google Trends, and the real GDP. We apply methodologies based on fractional integration and cointegration to measure the degree of persistence and to examine the long-term relationship. Finally, we carry out a forecast using a Machine Learning model based on an Artificial Neural Network. Our results indicate that the Spanish economy will experience a contraction in 1Q-21 and will require strong measures to reverse the situation and recover the original trend.

Keywords: Leading Economic Indicators, Business Cycle, Google Trends, Fractional Integration, FCVAR model, Machine Learning.

JEL Codes: C22, E32, E37.

Corresponding Author: Prof. Carlos Poza
Universidad Francisco de Vitoria
Faculty of Law and Business Administration
28223 Pozuelo de Alarcón, Madrid, Spain
Email: carlos.poza@ufv.es

1. Introduction.

The impact of COVID-19 in Spain is causing a dramatic health crisis and an unprecedented economic recession. This adverse business cycle has increased the importance of anticipating the economic activity behavior for decision-making. Normally, economic activity is measured by means of GDP growth. Nonetheless, as it is released quarterly, we observe some limitations in terms of making agile decisions. In this sense, it is very common for analysts to use leading economic indicators to anticipate GDP trends and turning points. In Spain, we can track the economy through the CLI (OECD), or ESI (Ministry of Economy).

Following this idea, Poza and Monge (2020) proposed a Real-Time Leading Economic Indicator (RT-LEI) for Spain, inspired by authors such as Choi and Varian (2012) who stated that a real-time index of the volume of queries from Google Trends is frequently correlated with several economic indicators.

The RT-LEI was created using text mining, factor analysis and data from Google Trends and Thomson Reuters monthly. According to Poza and Monge (2020), the RT-LEI works slightly better than CLI to foresee real GDP behavior. It reached the highest correlation versus GDP with a lead of 3 months.

In this paper, our main aims were to apply an improvement in the RT-LEI to anticipate the GDP trends as well as to forecast the economic activity in Spain over the next 12 months.

Concerning the enhancement of the RT-LEI, we implemented the principle of parsimony (Hair et al., 2007) in which we reduced the number of variables used from 21 to 14 to eliminate the unessential variables while maintaining the core information of the index.

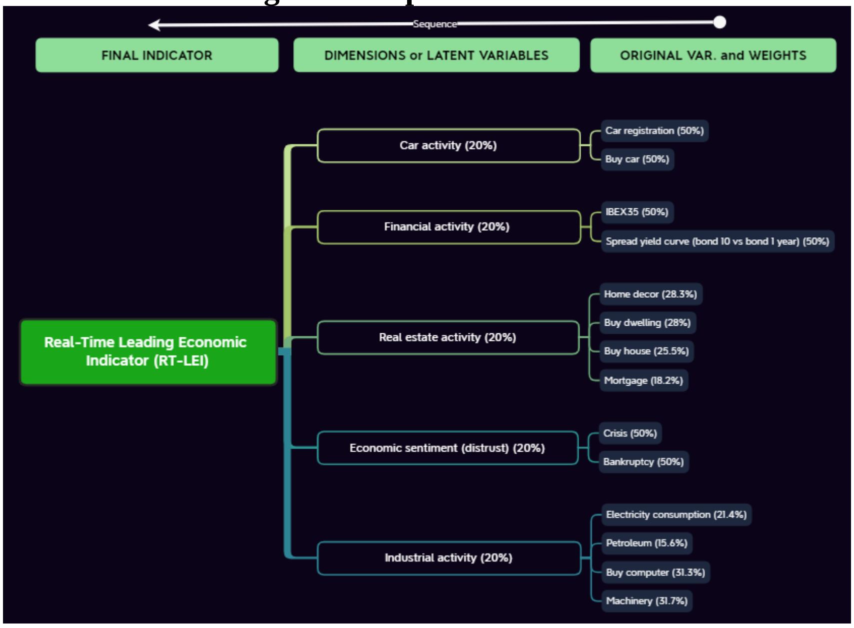
Regarding the forecast, we analyzed the statistical properties of the time-series by applying univariate tests based on long-term memory. Then, we used a Fractional Cointegration VAR (FCVAR) model to examine the relationship RT-LEI–GDP in the long-term. Finally, we applied machine learning techniques based on Artificial Neural Networks (ANN) to show evidence about the consistency of the economic recovery.

In the next sections, we first describe the key empirical results and then present the conclusions.

2. Empirical Results.

Using Poza and Monge’s (2020) methodology, the 14 original variables were saturated into 5 factors: car activity, financial activity, real estate activity, economic confidence, and industrial activity. The 5 dimensions went into the final indicator (Figure 1).

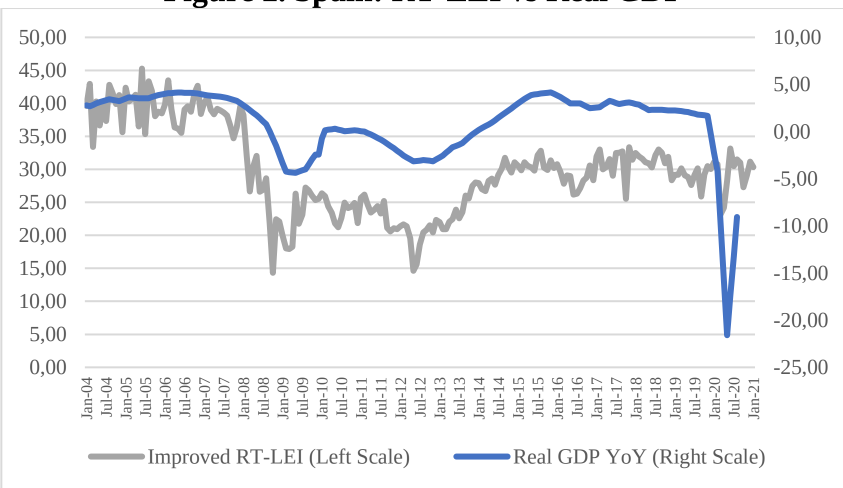
Figure 1. Improved RT-LEI



Source: Own elaboration

The correlation RT-LEI–GDP is 0.800, but it increases at 0.826 if we apply a lead of 3 months to the RT-LEI (99% level of confidence). In Poza and Monge (2020), the values were 0.804 and 0.797, respectively. The evidence shows that lean data management allows us to improve them slightly (Figure 2).

Figure 2. Spain: RT-LEI vs Real GDP



Source: Own elaboration from Google Trends and Thomson Reuters

Using this dataset, we conducted a standard unit root test (Dickey and Fuller, 1979; Elliot et al., 1996; Phillips and Perron, 1988 and Kwiatkowski et al., 1992) on the Spanish GDP and RT-LEI. The results suggest that both time-series are non-stationary I(1).

According to Diebold and Rudebusch (1991), Hassler and Wolters (1994) and Lee and Schmidt (1996), the unit root methods have very low power under fractional alternatives. For this reason, we use the ARFIMA (p, d, q) model where the mathematical notation is:

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

where the time-series $x_t, \ \ t = 1, 2, \dots$ follows an integrated order process d (and denoted as $x_t \approx I(d)$) and where d refers to any real value, L refers to the lag-operator ($Lx_t = x_{t-1}$) and u_t refers to I(0). The Akaike information criterion (Akaike, 1973) and Bayesian information criterion (Akaike, 1979) were used to select the appropriate AR and MA orders in the models.

The d parameter has been estimated considering all combinations of AR and MA terms ($p, q \ \leq 2$) for each time-series and their confidence bands at 95%. The results have been displayed in Table 1.

Table 1: Results of the long memory tests

Data Analyzed	Model Selected	d	Std. Error	Interval	$I(d)$
		Level values			
Real GDP	ARFIMA (0, d, 2)	0.9186614	0.0651306	[0.81, 1.03]	I(1)
RT-LEI	ARFIMA (1, d, 2)	1.0353103	0.1639207	[0.76, 1.30]	I(1)
		Logged values			
Log_GDP	ARFIMA (0, d, 2)	0.809281	0.0610574	[0.71, 0.91]	I(d)
Log_RT-LEI	ARFIMA (0, d, 0)	0.681281	0.0501896	[0.60, 0.76]	I(d)

The estimate of d for the Spanish GDP (real and in logarithms) and our leading indicator in logs are lower than 1 ($d < 1$), observing in the three cases that there is a high degree of persistence and finding evidence of mean reversion. For the case of RT-LEI and real GDP we cannot reject the hypothesis of I(1). Confidence intervals and the results show great uncertainty due to COVID-19, although the RT-LEI and real GDP results are telling us that the Spanish economic growth will require an extraordinary boost to reverse the situation and recover the original trend. It is important to highlight that the RT-LEI includes the data from 4Q-20 and Jan-21, while the latest GDP figures in the dates from are 3Q-20.

Next, we followed Johansen and Nielsen (2012) and their FCVAR model as notated in the next equation:

$$\Delta^d X_t = \ \alpha\beta' L_b \Delta^{d-b} X_t + \sum_{i=1}^{k} \Gamma_i \, \Delta^b L_b^i Y_t + \varepsilon_t \tag{2}$$

Where ε_t is a term with mean zero and variance-covariance matrix Ω that is p-dimensional independent and identically distributed; α and β are $p \times r$ matrices where $0 \leq r \leq p$. The relationship in the long-term equilibria in terms of cointegration in the system is due to the matrix β. Controlling the short-term behavior of the variables is due to parameter Γ_i. Finally, the deviations from the equilibria and their speed in the adjustment is due to parameter α.

The results of the FCVAR model have been summarized in Table 2.

Table 2: Results of the FCVAR model

	$d = b$	Cointegrating equation beta	
		RT-LEI	GDP
Panel I: RT-LEI and GDP (YoY)	0.893 (0.072)	1.000	-3.886
	$\Delta^d \left(\begin{bmatrix} RTLEI \\ GDP\ (YoY) \end{bmatrix} - \begin{bmatrix} 38.958 \\ 3.087 \end{bmatrix} \right) = L_d \begin{bmatrix} 0.030 \\ 0.022 \end{bmatrix} v_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$		
Panel II: RT-LEI and Real GDP	0.830 (0.062)	1.000	1.511
	$\Delta^d \left(\begin{bmatrix} RTLEI \\ GDP \end{bmatrix} - \begin{bmatrix} 38.961 \\ 90.736 \end{bmatrix} \right) = L_d \begin{bmatrix} -0.040 \\ 0.003 \end{bmatrix} v_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$		
Panel III: Log_RT-LEI and Log_Real GDP	0.837 (0.080)	1.000	4.111
	$\Delta^d \left(\begin{bmatrix} Log\_RTLEI \\ Log\_GDP \end{bmatrix} - \begin{bmatrix} 1.581 \\ 1.958 \end{bmatrix} \right) = L_d \begin{bmatrix} -0.046 \\ 0.001 \end{bmatrix} v_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$		

We get the estimate of common order of integration, considering level values and logged values for the GDP and our leading indicator in two different panels (I and II). The values of the fractional differencing parameter are lower that 1 ($d < 1$) for the two panels. Following Tkacz (2001), we conclude that the stochastic trend is of a fractionally nature and possess stationarity with long memory and the shock duration is long-lived. Also, we can conclude that the combination of both time series is very persistent, indicating that the trend is well defined.

Considering the high degree of persistence of both time-series and their degree of cointegration, we can use advanced computational intelligence techniques based on machine learning to forecast the RT-LEI created.

For this purpose, we have used the Multilayer Perceptron (MLP) neural network to predict the time-series. The motivation is that the underlying model (non-parametric model) is required to get the results. It also presents interesting features such as its nonlinearity. The MLP Neural Network method is based on the back-propagation rule where the errors are propagated throughout the network and allow for the adaptation of the hidden processing elements. It has a massive level of interconnectivity that means that any element of a given layer feeds all the elements of the next layer. It is trained using error correction learning (Güler and Ubeyli, 2005; Martínez et al., 2019; Mapuwei et al., 2020).

The GDP and RT-LEI accuracy using the ANN model measured by Root Mean Square Error (RSME) is very close to zero (0.02 and 0.04, respectively), what allows us to forecast the Spanish economic activity (Figures 3 and 4). We expect a GDP contraction of 6.9% YoY in 1Q-21. The economic growth will then recover gradually and with a high uncertainty. The RT-LEI estimations show a similar behavior, but the economic activity does not consolidate the recovery unless policymakers apply a stimulus such as European funds. These measures should be mainly focused on the health system to finance herd immunity. Therefore, we do not expect a real recovery until 2Q-22.

Our forecasts are in line with IMF (2020) in terms of economic activity trends (Figures 3 and 4). Nonetheless, our estimations are more pessimistic because of two reasons: Firstly, we used Big Data to forecast the economic activity and the institution did not. Secondly, we exploited more recent data, therefore we have considered the second and the third COVID waves. The institution has not. Furthermore, the latest data (Dec-20 and Jan-21) shows a relapse in economic activity to some extent. This is related to the severe vaccine delays.

Figure 3. Forecasting using RT-LEI

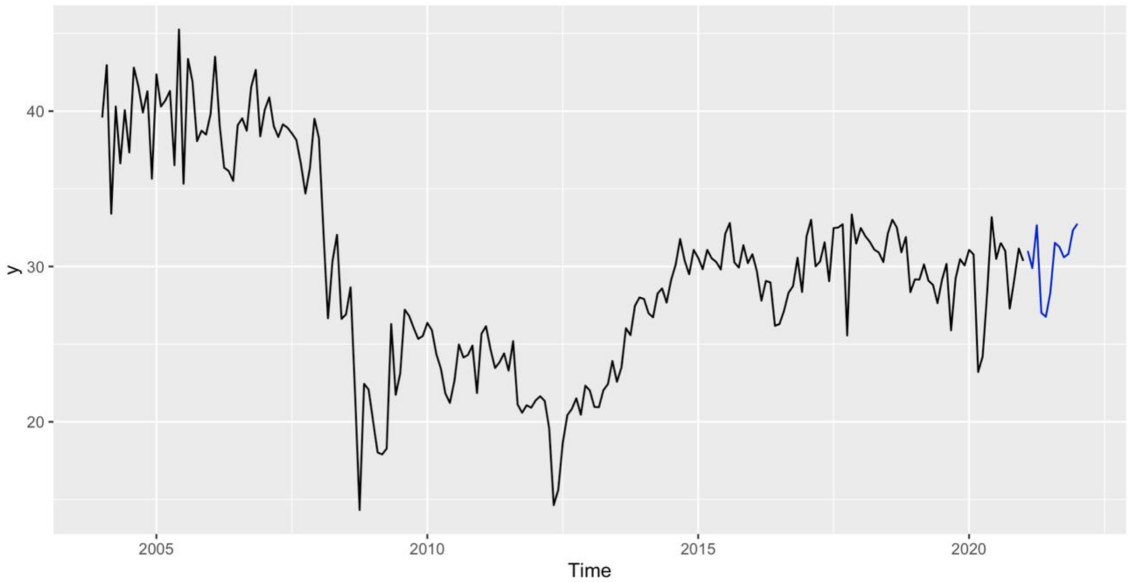
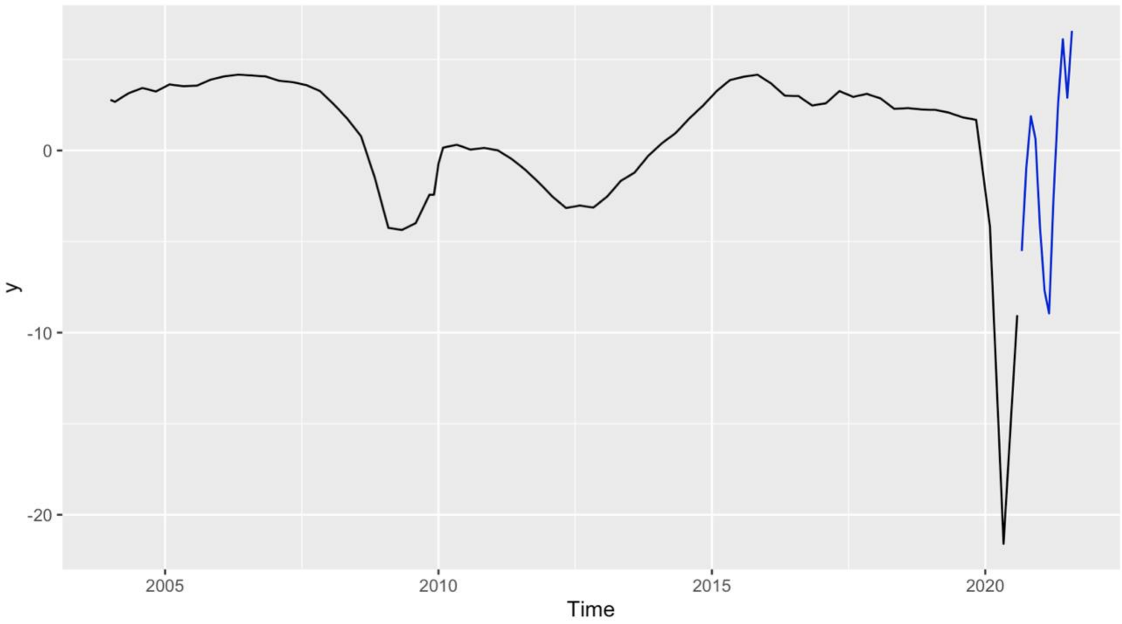


Figure 4. Forecasting using GDP (YoY)



3. Conclusions.

The RT-LEI proposed in this paper improves the first version released by Poza and Monge (2020), thanks to the principle of parsimony. The correlation GDP–RT-LEI (t = -3) reaches 0.826. The cointegration of GDP versus the improved RT-LEI enhances the fit when compared to the original version.

These are the main empirical conclusions:

- We conducted a standard unit root test on the Spanish GDP and the RT-LEI. The results suggest that both time-series are non-stationary I (1).
- The ARFIMA model shows that the GDP and RT-LEI presents fractional I(d) behavior, but confidence intervals and the results show great uncertainty due to COVID-19 that tell us that the Spanish economic growth could need strong measures to recover the pre-COVID trend, financed by European funds.
- After applying the FCVAR, we conclude that the resulting long-run equilibrium time-series follows a long-memory process, where the shock duration is long lived. Also, the results suggest powerful potential forecasting in the long-term.
- Due to the high degree of persistence and cointegration, we have used Machine Learning techniques (ANN) to forecast the GDP and RT-LEI. The estimations show that the Spanish economy will suffer from a new contraction in 1Q-21 and the recovery will be slow and with uncertainty. In fact, we do not expect a real recovery until 2Q-22 when herd immunity takes place. Our predictions are in line with those of institutions in terms of trend, but they are more pessimistic because of the third COVID wave and the vaccine delays.

References

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.

Baillie, R.T., and Bollerslev, T. (1994). Cointegration, fractional cointegration, and exchange rate dynamics. The Journal of Finance, 49(2), 737-745.

Choi, H. and Varian, H. (2012). Predicting the present with google trends. Economic Record, 88, 2–9.

Dickey, D., and Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time-Series with a Unit Root. Journal of the American Statistical Association, 74, 427–431.

Diebold, F.X., and Rudebush, G.D. (1991). On the power of Dickey-Fuller tests against fractional alternatives. Economics Letters, 35, 155-160.

Elliot, G., Rothenberg, T.J., and Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit Root. Econometrica, 64, 813–836.

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

Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. (2007). Análisis multivariante. 5ª edición. Pearson Prentice Hall.

Hassler, U., and Wolters, J. (1994). On the power of unit root tests against fractional alternatives. Economics Letters, 45(1), 1-5.

IMF (2020). World Economic Outlook. October 2020. International Monetary Fund. USA.

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

Kwiatkowski, D., Phillips, P.C., Schmidt, P., and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure are We that Economic Time-Series Have a Unit Root? Journal of Econometrics, 54, 159–178.

Lee, D., and Schmidt, P. (1996). On the power of the KPSS test of stationarity against fractionally-integrated alternatives. Journal of Econometrics, 73(1), 285-302.

Martínez, F., Frías, M.P., Charte, F., and Rivera, A.J. (2019). Time-Series Forecasting with KNN in R: the tsfknn Package. The R Journal, 11(2), 229-242.

Mapuwei, T.W., Bodhlyera, O., and Mwambi, H. (2020). Univariate Time-Series Analysis of Short-Term Forecasting Horizons Using ANN: The Case of Public Ambulance Emergency Preparedness. Journal of Applied Mathematics.

Phillips, P., and Perron, P. (1988). Testing for a Unit Root in Time-Series Regression. Biometrika, 75, 335–346.

Poza, C. and 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. International Economics, 163 (163-175).

Tkacz, G. (2001). Estimating the fractional order of integration of interest rates using a wavelet OLS estimator. Studies in Nonlinear Dynamics & Econometrics, 5(1).