A PROPOSAL OF A SUSPICION OF TAX FRAUD INDICATOR BASED ON GOOGLE TRENDS TO FORESEE SPANISH TAX REVENUES

Manuel Monge

Carlos Poza

Sofía Borgia

Universidad Francisco de Vitoria, Madrid, Spain

Abstract

This article contributes to the relationship between fiscal fraud and tax collection in the Spanish economy, creating a composite suspicion tax fraud indicator (STFI) based on Google Trends searches to study the dynamics and foresee tax revenues evolution in Spain. Also, we expand knowledge in the field of fraud tax indicators, following the UNODC (2020) and OECD (2016) recommendations. To this purpose, we apply factor analysis to create the composite indicator and next we utilize techniques centred on fractional integration (ARFIMA) and fractional cointegration VAR (FCVAR) to assess the STFI behavior against tax collection and GDP. The outcomes indicate that the d is less than 1 in all the time series analyzed. As we can see, the tax collection and the leading indicator have a similar statistical behavior (d=0.49 and d=0.40, respectively), which implies mean reversion. On the other hand, GDP will behave similarly to the other two time series, with d = 0.05, which implies that the shocks will have a temporary effect on the GDP behavior, and these effects will disappear by themselves in the short term and in less time than the other two time series. FCVAR results indicate a short-lived duration of the shock due to the error correction term and their short-run stationary behaviour. In the end, applying wavelet analysis, we determine that the composite suspicion tax fraud indicator maintains a negative association with tax collection, except in 2017 and 2018, when the high economic growth offsets the fiscal fraud.

Keywords: Fiscal Fraud, Composite Indicator, Google Trends, Fractional Integration, FCVAR model, Wavelets.

JEL Classification: C22, E60, E37, H26.

Corresponding author: Prof. Manuel Monge
                            Universidad Francisco de Vitoria
                            Faculty of Law and Business Administration
                            28223 Madrid, Spain
                            Email:

1. INTRODUCTION.

Tax collection is one of the most challenging issues of public finance (OECD, 2006), to the extent many countries experience a high level of shadow economy, fiscal evasion, and avoidance. These problems tend to enormously reduce tax revenues and feed an inefficient public sector.

Tax revenues may be explained by numerous determinants (Castro and Ramírez Camarillo, 2014; Boukbech, et al., 2018). One such is concerned structural reasons over which government has modest influence in the short run. However, there are other components regarding public policy (Brun et al., 2007), which are affected by direct and indirect government decisions.

Within the structural factors, GDP per capita is probably one of the most used to explain tax revenues, according to the literature. Nonetheless, there are others that may affect tax collection intensively (Sen Gupta, 2007), such as output composition, trade and financial openness, public debt, institutional factors (Langford and Ohlenburg, 2015) or fiscal evasion and avoidance (Garg, Goyal and Pal, 2014). In this regard, the World Bank (2017) points out that fiscal fraud negatively influences on tax revenues.

In line with the World Bank (2017), this research paper focuses on the relationship between fiscal fraud and tax collection in the Spanish economy, considering the obvious importance of economic activity. In addition, we expand knowledge in the field of fraud tax indicators, following the OECD (2016) and UNODC (2020) proposals.

The central purpose of this research is to create a composite suspicion of tax fraud indicator (STFI) based on Google Trends searches to foresee tax revenues' evolution in Spain. The contribution of estimating trends may be helpful to make better policy decisions.

The underlying idea of anticipating tax collection trends through a synthetic index has been extrapolated from Choi and Varian (2012), Varian (2014), Szármes (2015). D'Amuri and Marcucci (2017), and Poza and Monge (2020). They use keywords and big data from the Internet (Google Trends, Baidu, Twitter, etc.) to lead to valuable insights in labor market, tourism, or economic performance.

To achieve the objective, we firstly apply a multivariate analysis (factor analysis) to build the indicator; then we implement a descriptive analysis (correlations), and next we use advanced time series techniques such as fractional integration and cointegration methods, along with Wavelets Transform Analysis to assess the degree of association between the STFI and tax collection.

The structure of this research paper is the following: Part 2 concisely reviews the literature on measurement of fraud and tax collection. Part 3 describes the data and the methodology applied in the research paper. Part 4 shows the principal empirical findings. And Part 5 offers the conclusions.

2.   LITERATURE REVIEW.

2.1. Fraud and tax collection

Tax collection may be explained by several factors. According to Sen Gupta (2007), Castro and Ramírez Camarillo (2014) and Boukbech, et al. (2018)), some structural determinants might be highlighted, such as GDP, output composition, trade and financial openness, public debt, institutional factors and fiscal evasion and avoidance. In this regard, the World Bank (2017) affirms that fiscal fraud negatively influences on tax revenues.

Tax fraud is a problem for society;therefore, all countries and international organizations are dedicating more and more resources to the fight against tax evasion and fraud, which can cause GDP losses in a country (Kassa, 2021).

In the same line of research, Putra (2018) advances the idea that non-payment of taxes due to fraud leads to a great loss of resources for the public sector and consequently to a reduction of budgetary resources, both nationally and internationally. Remeur (2019) states that the objective of the fight against fraud is, fundamentally, to avoid the loss of income by the public sector. Fraud has always existed; however, globalization and digitization have caused its growth and expansion. Hence, the fight against it must be approached globally. Nguyen (2020) also affirms that tax non-compliance is a rising concern of public Administrations, due to the negative effects they have on state budgets.

OECD (2017) highlights the importance of controlling the volume of fraud in order to combat it. This OECD work collects information on technological and digital solutions used by 21 countries, most of them developed, to measure and fight against tax fraud. The report emphasizes the effectiveness of the use of technology in the control and fight against fraud that is depleting the public coffers of many countries. The work concludes with the need for collaboration and exchange of information and good practices between the tax authorities of the different countries.

2.2. The measurement of tax fraud

One of the priority objectives of the Agenda for Sustainable Development 2030 is the decrease of Illicit Financial Flows (IFF) (UNODC, 2020). This UNDOC document aims to lay the foundations to create an indicator that measures the scope of the so-called IFF. This indicator has two control bodies: UNODC (United Nations Office on Drugs and Crime) and UNCTAD (United Nations Conference on Trade and Development) that lead

the promotion and improvement of methodologies to quantify IFFs concerning taxes and trade.

IFFs differ among countries, and the absence of a universal framework makes their measurement particularly difficult. For this reason, the two entities, UNODC and UNCTAD, have cooperated to define common statistical definitions and methodologies for the measurement of the several kinds of IFFs.

UNODC (2020) presents the results derived from the work carried out at the international level to statistically define the IFF and the concepts that allow its measurement. The report carried out has allowed the creation of a conceptual framework, definitions, typologies and statistical methodologies.

The next step will be to test different statistical methodologies that are consistent with the previously given definitions. Although the different countries have data collection mechanisms, it is essential that all strengthen their ability to gather relevant information in order to detect the largest number of illicit financial flows.

In the same line of interest, Schneider (2015) tries to measure the magnitude of the informal economy in the European Union. This study points to tax fraud as a problem that affects all EU countries, given that budget losses due to fraud exceeded one million euros in 2013 (Murphy, 2012). Additionally, it tries to provide solutions in the fight against tax evasion and minimize losses for the public budget.

Other authors, however, study tax fraud from a microeconomic point of view, clustering tax evaders to create a sort of a complex profile.

In this regard, Mehta et al. (2019) present tax evasion by means of massive data analytics, Android Apps, and IT. This approach identifies suspicious dealers utilizing specific parameters and normalizes the process of examination of accounts. They use complex machine learning instruments to forecast suspicious dealers.

Shukla et al. (2018) suggest a method that can help to identify the apparent tax evaders utilizing information from the tax payment and applying big data tools. After pre-processing the available tax data, the authors applied a K-mean clustering to create groups of taxpayers with the same behavior. Next they used decision trees to categorize each cluster into tax payers and patterns are revealed. From these patterns, they train artificial neural network to detect evaders. In this line, Atanasijević et al. (2019) develop some algorithms for identifying the tax evasion risk by means of applying sophisticated big data and machine learning tools in Serbia. This might be useful for fraud prevention.

Houser and Sanders (2018), in the US, use a wide variety of analytic tools to extract relevant data from social media such as Instagram, Facebook, and Twitter. This data mined is mixed with the Internal Revenue Service's (IRS) own proprietary information and studied utilizing pattern identification algorithms, that helps to detect probable tax evaders. Understanding that the IRS is using public data from social media, taxpayers should take into account that their posts might influence their likelihood of being audit.

In the end, OECD (2016) states that the disposal of real-time data will create opportunities for revenue bodies in the future. Instead of catching and exploring transactions that have already occurred, revenue bodies can bear in mind how they might encourage real-time tax assessments. This opportunity may facilitate the creation of new tax indicators.

3. RESEARCH METHODOLOGY

3.1. Data

The data-sources exploited in the research article are Google Trends, the Spanish Ministry of Economy (MINECO) and the National Statistical Office (INE). Regarding Google

Trends, we carry out a search of keywords associated to fiscal fraud, what allows us to track the tax collection in high frequency. These keywords were selected properly from a previous review of thousands of court rulings. As far as MINECO is concerned, we focus on tax revenues. Regarding INE, we extract the GDP data.

All the data we use to build the STFI are published daily by Google Trends for Spain, but we construct the indicator monthly to contrast the results with the tax revenues that are published on a monthly basis. The time series begin in January 2005 and ends in January 2021; hence the quantity of observations reaches 189.

3.2. Variables

We use three variables in the empirical study: 1) a composite indicator concerning tax fraud, 2) a tax collection indicator, and 3) GDP, all of them for Spain.

The composite indicator is called "suspicion of tax fraud indicator" (STFI). It includes the following keywords, extracted from Google Trends: Fiscal evasion, deductible, deductible expenses, subsistence allowance, income tax subsistence allowances, amount of tax / tax liabilities, income tax, invoice, rental income, and Tax Agency (AEAT).

According to Google, the numbers of these variables indicate the search interest compared with the maximum rate in a region and period. A quantity of 100 implies the highest popularity of a word, while 50 suggest that a word is half popular compared with the maximum rate.

All these keywords have been identified previously in the Sentences of the Spanish Supreme Court in the Contentious-Administrative field. For the search and selection of keywords, the Tirant Analytics database was used. It is a tool developed by

Tirant lo Blanch with the collaboration of Universidad Carlos III de Madrid. Tirant Analytics combines artificial intelligence and big data to obtain interesting data, relates them and generates information and interactive graphics in relation to jurisprudence and legislation.

Four jurisprudence searches have been carried out, selecting the following concepts as the object of the process: fraud, Personal Income Tax, Corporation Tax and Value Added Tax, from January 2006 to January 2020. The assumptions that were repeated more frequently in each of the searches carried out have been identified and included in our composite fiscal fraud indicator.

The total number of judgments reviewed was 10,880, having selected those words that referred to the legal-tax assumptions that had a relative weight greater than 50% in relation to other concepts that were repeated to a lesser extent in the judgments.

As far as tax collection is concerned, we have utilized the variable "general government revenue: taxes" in millions of euros, not seasonally adjusted, whose data-source is the Spanish Ministry of Economy.

Regarding the Gross Domestic Product, we have used the GDP at constant prices, released by the Spanish Statistical Office.

3.3. Techniques

3.3.1. Factor analysis to create the composite suspicion of tax fraud indicator (STFI)

The 10 variables have been condensed into the composite suspicion of tax fraud indicator (STFI) through a factor analysis (principal component). Also, the weights have been obtained through the component matrix.

We have studied the set of variables $(X_1, X_2,..., X_{10})$ on our observations and determined, from them, the factor F_1 as a linear combination of $X_1, X_2, ..., X_{10}$, thus:

$$F_j = a_{j1}X_1 + a_{j2}X_2 + ... + a_{j10}X_{10}$$

where a'_j are the variables weights.

This technique is widely applied to create economic indexes because it allows analysts to construct non-arbitrary weights synthetic indicators through linear combinations (Poza and Monge, 2020; Munda and Nardo, 2005).

3.3.2. Unit roots

To analyse the stationarity of our dataset we have employed unit root tests like ADF tests established by Fuller (1976) and Dickey and Fuller (1979). Non-parametric estimates of spectral density of u_t at zero frequency have been used following the metholody proposed by Phillips (1987) and Phillips and Perron (1988). Finally, the unit root tests proposed by Kwiatkowski et al. (1992), Elliot, Rothenberg and Stock (1996) and Ng and Perron (2001) have been used taking into account their deterministic trends. We conclude that all of them have produced the same results.

3.3.3. Fractional integration

The non-stationary nature of numerous financial and economic time series is an important characteristic that can be described by several models. The standard approaches until 1980s was to apply deterministic functions of time where the residuals on the regression model were I(0) stationary. With the research done by Nelson and Plosser (1982) there was a consensus about the non-stationary component of most series was stochastic and the use of unit roots or first differences I (1) was the way to go. On the other hand, to

achieve stationarity I (0), the number of differences does not necessarily have to be an integer value, since it can be any point on the real line and therefore fractional I(d).

Observations that are far apart in time but highly correlated is a feature of the long memory and it is able to be captured by the I(d) models. The form of this models is:

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

where x_t represents a time series, d represents any real value, L is the lag-operator ($L x_t = x_{t-1}$) and the covariance stationary process is represented by u_t where the spectral density function, that is positive and finite at the zero frequency, displays a type of time dependence in the weak form. For this reason, we can state if u_t is ARMA (p, q), x_t is ARFIMA (p, d, q).

Adenstedt (1974), Granger and Joyeux (1980), Granger (1980; 1981) and Hosking (1981) were the authors that introduced the idea of fractional integration.

From equation (1), the polynomial $(1 - L)^d$ is expressed in terms of binomial expansion where for all real d, x_t depends not only on a finite number of past observations but also on the whole of its history. So, a higher value of d implies a higher level of association between the observations of the series.

Given the parameterization in (1), one can differentiate between various cases depending on the value of the parameter d, and several specifications based on (1) can be noted. Following Hosking (1981), the next table 1 summarizes the different values of d and the corresponding consequences for the mean (or trend), variance, and duration of the shock.

Table 1. Interpretation of the results of d for the ARFIMA models.

d	Mean (or trend) and variance	Shock duration
$d = 0$	Short-run mean-reversion Finite variance	Short-lived
$0 < d < 0.5$	Long-run mean reversion Finite variance	Long-lived
$0.5 \leq d < 1$	Long-run mean reversion Infinite variance	Long-lived
$d = 1$	No mean-reversion Infinite variance	Infinite
$d > 1$	No mean-reversion Infinite variance	Infinite; effect increase as time moves forward

Following the Akaike information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Akaike, 1979) we choose the most appropriate ARFIMA model. Also, we present our results following the likelihood process suggested by Sowell (1992) in contrast other several procedures (see Geweke and Porter-Hudak, 1983; Phillips, 1999; 2007; Robinson, 1994;1995a,b; etc.)

3.3.4. FCVAR model

The Fractionally Cointegrated Vector AutoRegressive (FCVAR) model is a generalization of Johansen's (1996) Cointegrated Vector AutoRegressive (CVAR) model to allow for fractional processes of order d that co-integrate to order $d - b$ with $b > 0$. The FCVAR model was introduced by Johansen (2008), developed by Johansen and

Nielsen (2010; 2012). The advantage of this model is the ability to use stationary and non-stationary time series.

To introduce the FCVAR model, we present first the non-fractional CVAR model. Being Y_t, $t = 1, \dots, T$ a p-dimensional I(1) time series, the CVAR model is:

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

where Δ^b and $L_b = 1 - \Delta^b$ represent the difference and the lag operator. These are utilized to derive the FCVAR model. We then find:

$$\Delta^b Y_t = \alpha\beta' L_b Y_t + \sum_{i=1}^{k} \Gamma_i \, \Delta L_b^i Y_t + \varepsilon_t, \qquad (2)$$

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

$$\Delta^d X_t = \alpha\beta' L_b \Delta^{d-b} X_t + \sum_{i=1}^{k} \Gamma_i \, \Delta^b L_b^i Y_t + \varepsilon_t, \qquad (3)$$

where, ε_t is p-dimensional independent and identically distributed with mean zero, and variance-covariance matrix Ω. The parameters α and β are $p \times r$ matrices, where $0 \leq r \leq p$. In matrix β the columns are the cointegrating relationships and $\beta' X_t$ are the stationary combination, i.e., the long-run equilibrium, which is integrated to order d, and the short-run parts from the long-term equilibrium are integrated to order $d - b$. The coefficients in α correspond to the velocity of adjustment unto equilibrium. Hence, $\alpha\beta'$ is the adjustment long-term and Γ_i represents the short-run behaviour of the variables.

Compared with the CVAR model, there are two other parameters in the FCVAR model. The order of fractional integration of the observable time series is represented by the parameter d. The degree of fractional cointegration, that is, the reduction in fractional integration order of $\beta' X_t$ compared to X_t itself, is represented by the parameter b.

The relevant ranges for b are $(0,\frac{1}{2})$, in which case the equilibrium errors are fractional of order greater than $1/2$ and hence non-stationary although mean-reverting, and $(\frac{1}{2},1]$, in which case the equilibrium errors are fractional of the order less than $1/2$ and are stationary (Dolatabadi et al., 2015). Mention that for $d = b = 1$, the FCVAR models is reduced to the CVAR model, which is therefore nested in the FCVAR model as a special case.

As an intermediate step toward the final model, we consider a version of model (2) with d = b as an assumption of no persistence in the cointegration vectors and a constant mean term for the cointegration relations. That is to say:

$$\Delta^d X_t = \alpha(\beta' L_d X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L_d{}^i X_t + \varepsilon_t \qquad (5)$$

The simple model considered is the following:

$$\Delta^d(X_t - \mu) = L_d \alpha \beta'(X_t - \mu) + \sum_{i=1}^{k} \Gamma_i \Delta^d L_d{}^i X_t + \varepsilon_t \qquad (6)$$

where μ represents the level parameter that shifts each of the series by a constant to avoid the bias related to the starting values in the sample (Johansen and Nielsen, 2016). $\beta'\mu = -\rho'$ represents the mean stationary cointegrating relations.

The asymptotic analysis in Johansen and Nielsen (2012) shows that the maximum likelihood estimators of $(d, \alpha, \Gamma, ..., \Gamma_2)$ are asymptotically normal, while the maximum likelihood estimator of (β, ρ) is asymptotically mixed normal when $d_0 < 1/2$ and asymptotically normal when $d_0 > 1/2$. Several empirical papers like Baruník and Dvořáková, 2015; Maciel, 2017; Aye et al., 2017; Dolatabadi et al., 2018; Jones, Nielsen and Popiel, 2014; Gil-Alana and Carcel, 2020; Poza and Monge, 2020; Monge and Gil-Alana, 2021a,b; among others have been used FCVAR models.

The MATLAB codes for the FCVAR model estimation are provided in Nielsen and Popiel (2018).

3.3.5. Continuous Wavelet Transform (CWT)

Continuous Wavelet Transform (CWT) has been applied to several research papers in finance and economics in order to understand the behaviour of time series in the time-frequency domain (see Aguiar-Conraria and Soares, 2011and 2014; Vacha and Barunik, 2012; Dewandaru, Masih and Masih, 2016; Tiwari, Mutascu and Albulescu, 2016; Jammazi et al., 2017; Poza and Monge, 2020; Monge and Gil-Alana, 2021a and b; among others).

Stationary is not required to carry out a wavelet analysis. Also, this methodology allows us to find evidence of changes in patterns between the time series analysed.

Two tools are used with this methodology to get the results: wavelet coherency and wavelet phase-difference.

First, we apply wavelet coherency that shows the correlation in time and frequency between both time series in a dimensional diagram into a function of τ and a, identifying hidden patterns or information.

The $WT_x(a, \tau)$ 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 $WT_x(a, \tau)$ refers to wavelet coefficients of $x(t)$; a refers to the position of a wavelet in the frequency domain, τ refers to the position in the time domain. The mother wavelet (ψ) that has been chosen for this research paper has been Morlet because we are going to be able to measure the synchronism between time series due to its characteristics

of complex sine wave within a Gaussian envelope (see Aguiar-Conraria and Soares, 2014).

Second, we use the wavelet coherence to interpret and understand the integration of the time series obtained in the previous step. So, we define the wavelet coherence as:

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

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, Azevedo and Soares, 2008 for details). Aguiar-Conraria's website provides the CWT's MATLAB code for the calculations of estimators and test statistics.[1].

4.   EMPIRICAL RESULTS.

4.1. Suspicion of tax fraud indicator (STFI)

Through using the factor analysis by means of the 10 variables earlier outlined, we generate the final indicator (see Figure 1).

The technique implemented was consistent in accordance with KMO and Bartlett test (0,801 and p-value<0.05, respectively). Additionally, STFI's weights derive from the component matrix (factor loads), that deduce the magnitude of each original variable within the composite indicator.

---

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

**Figure 1. Suspicion of Tax Fraud Indicator (STFI)**
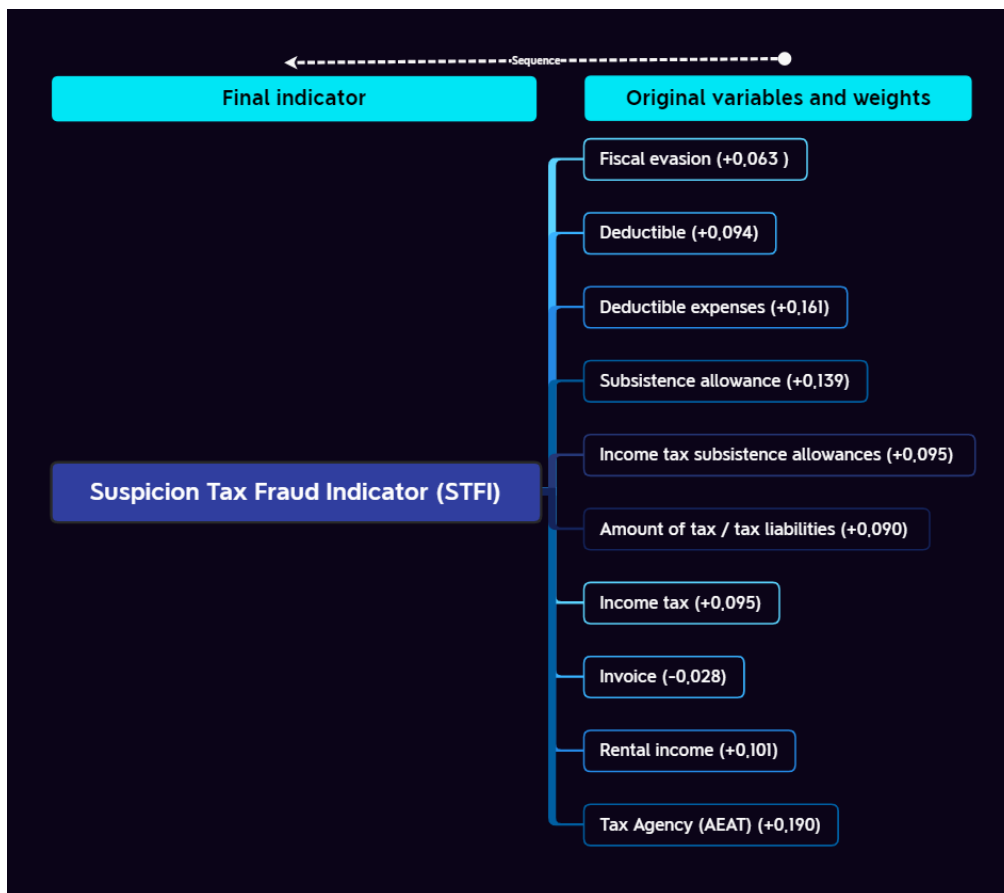
Source: Own elaboration from Google Trends

In terms of equation:

STFI = 0,063 fiscal evasion + 0,094 deductible + 0,161 deductible expenses + 0,139 subsistence allowance + 0,095 income tax subsistence allowances + 0,090 amount of tax / tax liabilities + 0,095 income tax − 0,028 invoice + 0,101 rental income + 0,190 Tax Agency (AEAT).

The three most important variables to measure the suspicion of tax fraud in Spain are "tax agency (AEAT)", "deductible expenses" and "subsistence allowance", according to the weights. Positive sign means the more the level is (the use of the keyword is high), the more the fiscal fraud will be; therefore, the less the tax revenue will probably be. And

vice versa if the sign is negative. Graph 1 shows the inverse association between STFI and tax collection.

Graph 1. STFI vs Tax Collection in Spain
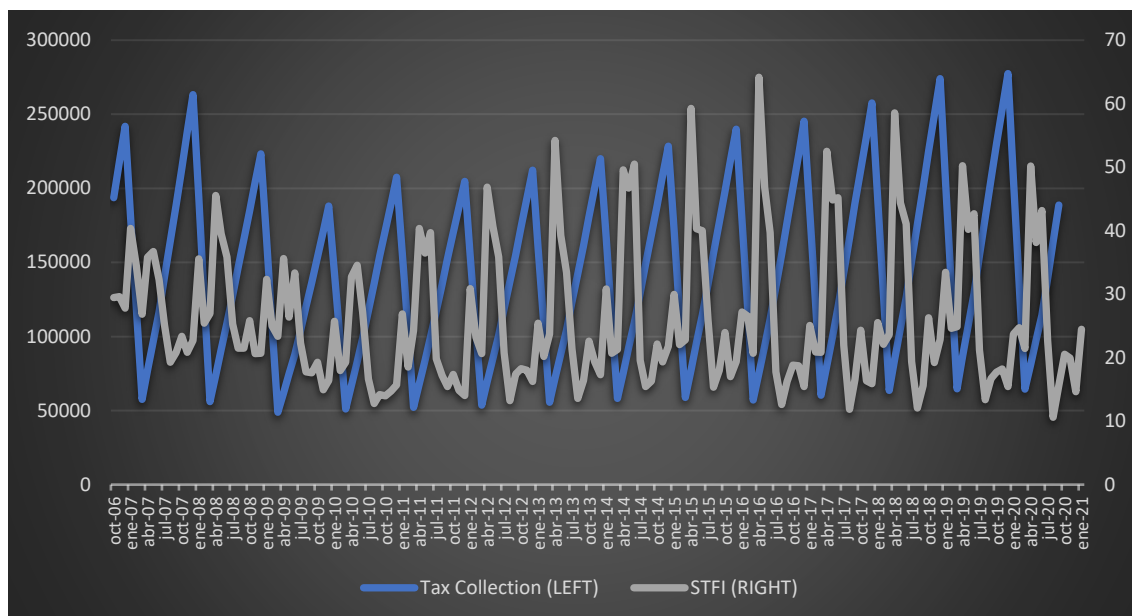
Source: Own elaboration from Google Trends and MINECO

Once STFI has been created, we calculate correlations among STFI, Tax Collection and real GDP to contrast the degree of association. The Pearson correlation coefficient of STFI vs Tax Collection is -0.500, and STFI vs Tax Collection (t-1) -0.506 (99% confidence interval, see Table 2), which implies a slight improvement of leading tax revenues. As expected, GDP vs Tax Collection presents a positive and statistically significant correlation.

Table 2. Correlation coefficients STFI, Tax Collection and GDP

			STFI	TAX COLLECTION	TAX COLLECTION (t − 1)	GDP
Correlations						
Pearson	STFI	Correlation coefficient	1	-,500**	-,506**	-,031
		Sig. (bilateral)		,000	,000	,685
		N	172	168	169	170
		Correlation coefficient	-,500**	1	,811**	,287**

TAX COLLECTION	Sig. (bilateral)	,000		,000	,000
	N	168	189	188	189
TAX COLLECTION (t – 1)	Correlation coefficient	-,506**	,811**	1	,287**
	Sig. (bilateral)	,000	,000		,000
	N	169	188	189	189
GDP	Correlation coefficient	-,031	,287**	,287**	1
	Sig. (bilateral)	,685	,000	,000	
	N	170	189	189	191

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

These moderate correlations seem interesting enough to deeply analyze the relationship between the STFI and Tax Collection through fractional integration and cointegration procedures in the next section.

4.2. Unit roots tests

To assess the statistical properties of the original series and its differences to achieve robust outcomes, we start the analysis by performing the Augmented Dickey-Fuller test (ADF), Phillips and Perrron (PP, 1988) and Kwiatkowski et al. (KPSS, 1992). Table 3 displays the results, which suggest that the tax collection and the new indicator of fiscal suspicion are stationary I(0). Finally, we have to perform first differences on GDP time series to get stationary I(0) behavior.

Table 3. Unit roots tests.

	ADF			PP		KPSS	
	(i)	(ii)	(iii)	(ii)	(iii)	(ii)	(iii)
Tax collection	-2.8489*	-9.946*	-10.1916*	-5.2847*	-5.3494*	0.206*	0.06*
Indicator of Fiscal Suspicion	-2.2994*	-7.2222*	-7.2011*	-7.3723*	-7.3479*	0.0573*	0.0506*
GDP	-0.1395	-3.0596*	-3.2391	-2.1095	-2.1918	0.884	0.479

(i) Refers to the model with no deterministic components; (ii) with an intercept, and (iii) with a linear time trend. * Denotes a statistical significance at the 5% level.

4.3. Results of long memory tests

According to the methodology explained before, and the low power of the unit root methods under fractional alternatives[2], we use ARFIMA (p, d, q) models to study the persistence of the tax collection, the new indicator of fiscal suspicion and the GDP. AIC criterion (Akaike, 1973) and BIC (Akaike, 1979) criterion were selected to select AR and MA orders in the models[3]

The estimates of d following the maximum likelihood estimator proposed by Sowell (1992) are in table 4. We have considered for each time series various ARFIMA (p, d, q) specifications with all combinations of $p, q \leq 2$.

Table 4. Results of long memory tests

Long memory test						
Data analyzed	Sample size (weeks)	Model Selected	d	Std. Error	Interval	I(d)
Tax collection	168	ARFIMA (0, d, 0)	0.493679	0.0086885	[0.48, 0.51]	I(d)
Indicator of Fiscal Suspicion	168	ARFIMA (0, d, 0)	0.4001668	0.0782432	[0.27, 0.53]	I(d)
GDP	168	ARFIMA (1, d, 2)	0.057581	0.283567	[-0.41, 0.52]	I(d), I(0)

We see from Table 4 that the estimates of d in the case of Tax Collection, Indicator of Fiscal Suspicion and GDP are equal, where the values of d are in the range (0, 1), implying fractional integration. The findings show that the d is less than 1 in all the time

---

[2] See Diebold and Rudebusch (1991), Hassler and Wolters (1994) and Lee and Schmidt (1996).
[3] According to Hosking (1981) and Beran et al. (1998), the AIC and BIC criterion could not be the best criteria under the assumptions of fractional models.

series analyzed. As we can see, the tax collection and the leading indicator have a similar statistical behavior (d=0.49 and d=0.40, respectively), which implies mean reversion. On the other hand, GDP will behave similarly to the other two time series, with d = 0.05, which implies that the shocks will have a temporary effect on the GDP behavior, and these effects will disappear by themselves in the short-term and in less time than the other two time series.

4.4.   Results of FCVAR model

To contrast the possible existence of persistence in the long-run co-movements of the series, we use FCVAR model proposed by Johansen and Nielsen (2012). Table 5 recaps the outcomes of the FCVAR model:

Table 5. Results of the FCVAR model ($d \neq b$)

	d	Cointegrating equation beta		
		Tax collection	Indicator of Fiscal Suspicion	GDP
Panel I: Tax Collection, Indicator of Fiscal Suspicion and GDP	d = 0.501 (0.044) b = 0.501 (0.012)	1.000	-3860.691	-2875.266

$$\Delta^d\left(\begin{bmatrix} \text{Tax collection} \\ \text{Indicator Fiscal Suspicion} \\ \text{GDP} \end{bmatrix} - \begin{bmatrix} 148917.532 \\ 28.012 \\ 100.380 \end{bmatrix}\right)$$

$$= L_d \begin{bmatrix} -0.409 \\ 0.001 \\ -0.000 \end{bmatrix} v_t + \sum_{i=1}^{2} \hat{\Gamma}_i \Delta^d L_d^i (X_t - \mu) + \varepsilon_t$$

Following the Jones, Nielsen and Popiel (2014) suggestions regarding the lag value (k = 3). Also, we consider deterministic components and cointegration rank (r) to get our results. In Panel I we indicate the estimate of common order of integration for the Tax Collection, Indicator of Fiscal Suspicion and GDP. The value of the fractional

differencing parameter is lower than 1 ($d < 1$) for the panel. So, the order of integration in the cointegrating regression is precisely of the same magnitude, 0.501, implying that the order of integration $(d - b) = 0$, which, in turn, implies I(0) cointegration errors. So, we cannot refuse the hypothesis where the shock duration is short-lived due to the error correction term shows short-run stationary behaviour.

4.5.   Results of Wavelet Continuous Transform

To identify the dependence along time and to identify when structural changes occur between composite suspicion of tax fraud indicator (STFI) and tax collection time series we have calculated the wavelet coherency and phase difference. The results are displayed in figure 2.

Figure 2: Wavelet coherency and phase difference results
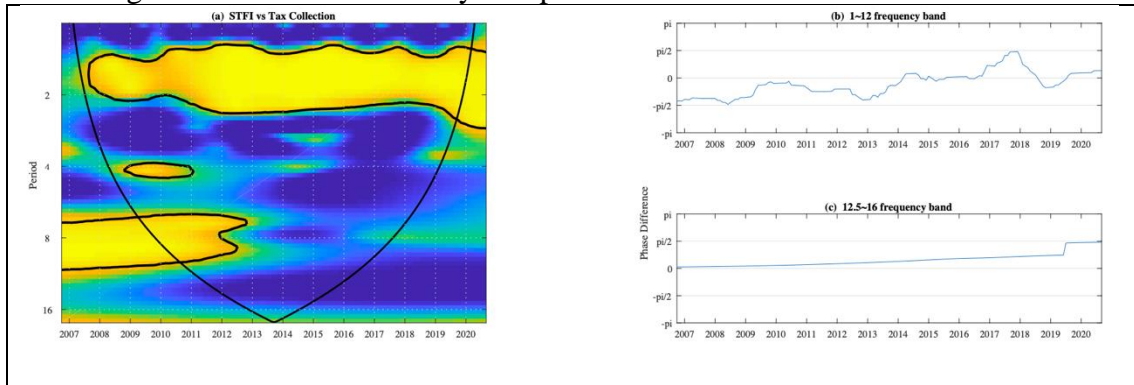
The contour designates the 5% significance level. Coherency ranges from blue (low coherency) to yellow (high coherency). (a) Wavelet coherency. (b)-(c)Phase difference.

Panel (a) represents the wavelet coherency that explains the interrelations between both time series, showing the strength of the relationship and in which frequency occurs. From the scale 1 (a single month) up to scale 18 (approximately a year and a half), the frequencies are displayed on the vertical axis. On the hand, all the sample period is represented in the horizontal axis.

To identify the high frequency and the high coherence with statistical significance of local correlations between time-frequency domain we have used Monte Carlo simulations, where the regions with significance values at 5% are represented by the black contour. The relation of both time series, one in relation to the other, is understandable using the frequency bands represented in panel (b) and (c), that are 1.5-12 frequency band and 12.5-16 frequency band, respectively.

The principal regions with statistically significant coherency are in low frequency (short term) that correspond to cycles between 1 to 2 months. If we focus on the phase-difference, we observe that this region stays between 0 and $-\pi/2$, which means that tax collection is negative correlated with STFI, except for the period between the last quarter of 2017 and mid-2018.

The first result is consistent with the economic theory, the more the fiscal fraud is, the less the tax collection will be (Sen Gupta, 2007; World Bank, 2017). However, we can observe an interesting outcome in 2017 and 2018. In these years, the correlation between STFI and tax collection is positive, which may seem illogical, nonetheless the high economic growth experienced by the Spanish economy in those years offsets the evolution of fiscal fraud. We must consider that tax revenues are explained by numerous factors (Castro and Ramírez Camarillo, 2014; Boukbech, et al., 2018).

According to AEAT (2017), tax revenues in 2017 increased by 4.1%, reaching 193,951 million euros. This variation did not have its main origin in regulatory changes, but in the evolution of tax bases and the result of the annual income tax and corporation tax returns with which the year 2016 was settled. The tax bases of the main taxes grew by 5.6% in 2017 thanks to job creation and higher profits. For its part, spending increased, and with special intensity in the first part of the year, due to the price increases observed at the end of 2016 and the beginning of 2017.

Additionally, AEAT (2018) states that, in 2018, tax revenues reached the figure of 208,685 million euros, which represented a growth of 7.6%. The increase compared to 2017 is affected by the change in the VAT filing dates of taxpayers under the Immediate Supply of Information System[4], a change that occurred in July 2017 and which meant a displacement of income from 2017 to 2018. If this element is corrected, revenues would have grown by 5.4%. 2018 was a year of deceleration; however, domestic demand slowed less than real GDP and compensation of wage earners grew more than in 2017, thus the aggregate base of the main taxes grew by 6.1%, one point more than in 2017.

5.   CONCLUSIONS.

Article 31.1 of the Spanish Constitution determines that "we will all contribute to the support of public expenditures...". Therefore, it is the responsibility of the Tax Agency to employ the tax system in this sense to comply with the constitutional principle by virtue of which everyone must contribute to the maintenance of public expenditures in accordance with their economic capacity. The main objective of the Tax Agency is to promote voluntary compliance by citizens with tax obligations to defray public expenses. Thus, it forms two routes of decision-making: firstly, the provision of information and assistance services to the taxpayer to diminish the indirect costs related to compliance with tax obligations and, secondly, the detection and regularization of tax breaches through actions control and fight against fraud.

In recent years, the generalization of services to assist citizens in complying with their tax obligations and the permanent improvement of the quality of these services, has

---

[4] It is a new VAT processing system, developed by the Spanish Tax Agency, for the electronic management of VAT registration books, which means an almost immediate supply of billing records.

allowed the Spanish Tax Agency to be considered a reference Administration throughout the world in providing services to taxpayers.

However, the undoubtedly positive assessment that citizen information and assistance services deserves is not accompanied by a similar assessment when it comes to the other basic line of action of the Tax Agency: tax control and the fight against tax fraud. Consequently, the Tax Agency, in compliance with the purposes established in its creation law, must rethink its strategy for the prevention and control of tax fraud.

As seen in the literature review, fraud has a clear impact on the income of the public sector and, therefore, on the quality of the public services provided by the state. At the same time, it generates situations of tax injustice among those taxpayers who comply with their obligations and are forced into a situation of unfair competition against those who do not meet their tax obligations.

To face the necessity of identifying fraud situations, several authors have developed some tax indicators and techniques based on big data analytics, which are helping authorities to create a complex profile of evaders. In this regard, we have built a composite suspicion of tax fraud indicator to foresee tax collection in Spain. A statistical analysis has been carried out to know the degree of persistence and the long-run relationship between time series utilizing fractional integration and FCVAR approaches. The first step in this research paper has been to perform ADF (Dickey and Fuller, 1979), PP (Phillips and Perron, 1988), and KPSS (Kwiatkowski et al., 1992) unit root tests. The results suggest that the tax collection and the new indicator of fiscal suspicion are stationary I(0). We have to perform first differences on GDP time series to get stationary I(0) behavior.

As we mentioned previously, time series depend on all observations from history. So, fractional integration methods are relevant in this research paper because allow us to

know the degree of dependence of the time series and also if there is a shock on fiscal fraud and tax collection on Spanish GDP, how would it be, temporary or permanent.

The outcomes that we achieve applying the ARFIMA (p, d, q) model indicate that the d is less than 1 in all the time series analyzed. As we can see, the tax collection and the leading indicator have a similar statistical behavior (d=0.49 and d=0.40, respectively), which implies mean reversion. On the other hand, GDP will behave similarly to the other two time series, with d = 0.05, which implies that the shocks will have a temporary effect on the GDP behavior, and these effects will disappear by themselves in the short run and in less time than the other two time series.

Using FCVAR for the multivariate assumption, we have achieved results very similar to the previous one, where the parameter d that represents the order of integration of individual series is 0.501, being the short-run part from the long-term equilibrium integrated to order $(d - b) = 0$. Therefore, the hypothesis that the duration of the shock is short-lived we cannot be rejected.

We use Continuous Wavelet Transform in order to know the relationship in time-frequency between STFI and tax collection and to know how these relationships have evolved over time. We observe that the tax collection is negatively correlated with STFI, being consistent with the economic theory where the more the fiscal fraud is, the less the tax collection will be (World Bank, 2017; Sen Gupta, 2007). However, we can observe an interesting outcome in the last quarter of 2017 and mid-2018 where they indicate a positive correlation because the high economic growth offsets the fiscal fraud (AEAT 2017 and 2018).

This contribution is in line with the idea of searching for tools that facilitate Administration the measurement and prevention of tax fraud, recommended by OECD (2016).

The STFI that we present can be of great help in this regard since it allows us to track trends in the intention of fraud and, based on them, design specific control policies. This indicator puts technology and data analysis at the service of the fight against fraud and, therefore, at the service of the whole of society.

A technological tool, such as the one proposed, does not constitute a single solution to the challenge of tax evasion, but, if it is applied effectively, it can be a complement and an important advance in this field.

In any case, the design of instruments to solve this problem, which is international in nature and which the UN Agenda for Sustainable Development 2030 has indicated as one of its priorities, must be accompanied by a series of measures that guarantee the success. All solutions must be accompanied by adequate legislative measures that guarantee greater control and compliance by all citizens. Likewise, the effectiveness of any measure will depend largely on the collaboration of taxpayers and international cooperation.

References.

Adenstedt, R.K. (1974). On Large-Sample Estimation for the Mean of a Stationary Random Sequence. *The Annals of Statistics,* 2(6), pp.1095-1107.

AEAT (2017). *Annual Report on Tax Collection year 2017.* Tax Studies and Statistics Service.

AEAT (2018). *Annual Report on Tax Collection year 2018*, Tax Studies and Statistics Service.

Aguiar-Conraria, L., Azevedo, N. and Soares, M. J. (2008). Using wavelets to decompose the time-frequency effects of monetary policy. *Physica A: Statistical Mechanics and its Applications*, 387, pp. 2863-2878.

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

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

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

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

Atanasijević, J.,Jakovetić, D., Krejić, N., Krklec-Jerinkić, N. and Marković, D. (2019). Using Big Data Analytics to Improve Efficiency of Tax Collection in the Tax Administration of the Republic of Serbia. *Ekonomika Preduzeca*, 67, pp. 115-130.

Aye, G.C., Carcel, H., Gil-Alana, L.A. and Gupta, R. (2017). Does gold act as a hedge against inflation in the UK? Evidence from a fractional cointegration approach over 1257 to 2016. *Resources Policy*, 54, pp.53-57.

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

Beran, J., Bhansali, R. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, *Biometrika,* 85, 4, pp.921-934.

Boukbech, R., Bousselhamia, A. and Ezzahid, E. (2018). *Determinants of tax revenues: Evidence from a sample of Lower Middle-Income countries.* MPRA Paper No. 90268. MPRA.

Castro, G.A. and Ramírez Camarillo, D.B. (2014). Determinants of tax revenue in OECD countries over the period 2001–2011. *Contaduría y Administración*, 59(3), pp. 35-59. https://doi.org/10.1016/S0186-1042(14)71265-3.

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88, pp.2–9.

D'Amuri, F. and Marcucci, J. (2017). The predictive power of google searches in forecasting US unemployment. *International Journal of Forecasting*, 33, pp.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 and Finance*, 43, pp.363-377.

Dickey, D.A. and Fuller, W.A. (1979). Distributions of the estimators for autoregressive time series with a unit root, *Journal of American Statistical Association*, 74 (366), pp.427-481.

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

Dittmann, I. and Granger, C.W.J. (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, 110, pp.113-133.

Dolatabadi, S., Nielsen, M. Ø., & Xu, K. (2015). A fractionally cointegrated VAR analysis of price discovery in commodity futures markets. Journal of Futures Markets, 35(4), 339-356.

Dolatabadi, S., Narayan, P.K., Nielsen, M.Ø. and Xu, K. (2018). Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. *Journal of Futures Markets*, 38(2), pp.219-242.

Elliot, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root, *Econometrica*, 64, pp.813-836.

Fuller, W.A. (1976). *Introduction to Statistical Time Series*, New York: John Wiley.

Garg, S., Goyal, A. and Pal, R. (2014). *Why Tax Effort Falls Short of Capacity in Indian States: A Stochastic Frontier Approach*. Indira Gandhi Institute of Development Research (IGIDR), WP-2014-032.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4), pp.221-238.

Gil-Alana, L.A. and Carcel, H. (2020). A fractional cointegration var analysis of exchange rate dynamics. *The North American Journal of Economics and Finance*, 51.

Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* 14(2), pp.227–238.

Granger, C.W.J. (1981). Some properties of Time Series data and their use in Econometric Model Specification. *Journal of Econometrics*, 16, pp.121-131.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, pp.15–29.

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

Hosking, J.R.M. (1981). Modelling persistence in hydrological time series with using fractional differentiation. *Water Resources Research*, 20, pp.1898-1908.

Houser, K. and Sanders, D. (2018). The Use of Big Data Analytics by the IRS: What Tax Practitioners Need to Know. *Journal of Taxation*, 128(2).

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 and Finance*, 49, pp.453-483.

Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models*. New York, NY: Oxford University Press.

Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, 24, pp.651-676.

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

Johansen, S. and Nielsen, M.Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, 80, pp.2667-2732.

Johansen, S. and Nielsen, M.Ø. (2014). *The role of initial values in nonstationary fractional time series models*. QED working paper 1300, Queen's University.

Jones, M.E., Nielsen, M.Ø. and Popiel, M.K. (2014.). A fractionally cointegrated VAR analysis of economic voting and political support. *Canadian Journal of Economics*, 47, pp.1078-1130.

Kassa, E.T. (2021). Factors influencing taxpayers to engage in tax evasion: evidence from Woldia City administration micro, small, and large enterprise taxpayers. *Journal of Innovation and Entrepreneurship,* 10, p.8.

Kwiatkowski, D., Phillips, P.C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3), pp.159-178.

Langford, B. and Ohlenburg, T. (2015). *Tax revenue potential and effort: An empirical investigation*. International Growth Centre.

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

Maciel, L. (2017). Technical analysis based on high and low stock prices forecasts: Evidence for Brazil using a fractionally cointegrated VAR model. *Empirical Economics*, 58(4), pp.1513-1540.

Mehta, P., Mathews, J., Kumar, S., Suryamukhi, K., Babu, Ch.S.,Visweswara Rao, S.V.K., Shivapujimath, V. and Bisht, D. (2019). Big Data Analytics for Tax Administration. In: Kő, A., Francesconi, E., Anderst-Kotsis, G., Tjoa, A. and Khalil, I. (eds.), *Electronic Government and the Information Systems Perspective. EGOVIS 2019. Lecture Notes in Computer Science*, vol 11709. Springer. https://doi.org/10.1007/978-3-030-27523-5_4

Monge, M. and Gil-Alana, L.A. (2021a). Lithium industry and the US crude oil prices. A Fractional Cointegration VAR and a Continuous Wavelet Transform analysis. *Resources Policy*, 72, p. 102040.

Monge, M. and Gil-Alana, L. A. (2021b). Spatial crude oil production divergence and crude oil price behaviour in the United States. *Energy*, 232, p. 121034.

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

Murphy, R. (2012). "*Closing the European tax gap", A Report for Group of the Progressive Alliance of Socialists & Democrats in the European Parliament*. Norfolk: Tax Research LLP.

Nelson, C.R. and Plosser, C.I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics,* 10(2), pp.139-162.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69, pp.519-1554

Nguyen, L., Nguyen, A., Le, H.D., Le, A.H. and Vu Truong, T.T.V. (2020). The Factors Affecting Corporate Income Tax Non-Compliance: A Case Study in Vietnam, *The Journal of Asian Finance, Economics and Business*, 7(8), pp. 103-115.

Nielsen, M.Ø., and Morin, L. (2014). *FCVAR model: a MATLAB software package for estimation and testing in the fractionally cointegrated VAR model.* QED Working Paper 1273, Queens University.

Nielsen, M.Ø. and 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. Ontario, Canada, K7L 3N6.

OECD (2006). *Report on Identity Fraud: Tax Evasion and Money Laundering Vulnerabilities. Centre for Tax Policy and Administration*. Organisation for Economic Cooperation and Development.

OECD (2016). *Technologies for Better Tax Administration. A Practical Guide for Revenue Bodies*. Organisation for Economic Cooperation and Development.

OECD (2017). *Technology Tools to Tackle Tax Evasion and Tax Fraud* Organisation for Economic Cooperation and Development

Phillips, P.C. (1987). Time series regression with a unit root. *Econometrica*, 55(2), pp.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. *Journal of Econometrics*, 138(1), pp.104-124.

Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75, pp.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, pp. 163-175.

Putra, P.D., Syah, D.H. and Sriwdari, T. (2018). Tax avoidance: Evidence as proof of agency theory and tax planning. *International Journal of Research and Review*,. 5 (9), pp.2454-2223.

Remeur, C. and Dobreva, A. (2019). *The fight against tax fraud*,. European Parliamentary Research Service. .

Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association*, 89, pp.1420-1437.

Robinson, P.M. (1995a). Gaussian semi-parametric estimation of long range dependence. *Annals of Statistics,* 23, pp.1630-1661.

Robinson, P.M. (1995b). Log periodogram regression of time series with long range dependence. *Annals of Statistics*, 23(3), pp.1048-1072.

Schneider, F., et al. (2015). Shadow economy and tax evasion in the EU. *Journal of Money Laundering Control*, 18(10), pp.34-51.

Sen Gupta, A. (2007). *Determinants of Tax Revenue Efforts in Developing Countries*. WP/07/184. International Monetary Fund. USA.

Shukla, Y., Sidhu, N., Jain, A., Patil, T.B. and Sawant-Patil, S.T. (2018). Big Data Analytics Based Approach to Tax Evasion Detection. *International Journal of Engineering Research in Computer Science and Engineering*,5(3).

Sipos, A. (2015). Determining Factors of Tax-morale with Special Emphasis on the Tax Revenues of Local Self-governments. *Procedia Economics and Finance*, 30, pp.758-767. https://doi.org/10.1016/S2212-5671(15)01325-8.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1-3), pp.165-188.

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

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 and Finance*, 42, pp.237-256.

UNODC (2020). *Conceptual Framework for the statistical measurement of Illicit Financial Flows*. United Nations Office on Drugs and Crime.

Varian, H.R. (2014). Big Data: new tricks for econometrics. *Journal of Economic Perspectives,*28, pp.3–28.

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

World Bank (2017). *Tax Revenue Mobilization. Lessons from World Bank Group Support for Tax Reform*. World Bank Group.