

Combining the Predictive Ability of Factorial Analysis and Transfer Functions for VAT Revenue Forecasting

César Perez López and Camino González Vasco

Instituto de Estudios Fiscales

Abstract

We propose a two-step methodology combining factorial analysis and a dynamic regression model to produce a valid forecast for VAT revenue. Instead of using final consumption expenditure as the only explanatory variable in a transfer function, we propose a set of indicators covering different areas of the economy (*General, Construction, Labour Market and Service Activity Indicators*). The idea is to enforce parsimony and to avoid multicollinearity with little information loss by performing principal component analysis and regressing not on the full set of indicators but rather on the first two principal components.

We apply the proposed method to quarterly data beginning in 1995 and ending in 2014, providing out of sample estimations for the four quarters of 2015.

Keywords: Principal Components Regression, VAT forecasting, forecast combination, generated regressors.

1. Introduction

Tax analysis and forecasting of revenues are of critical importance to governments in ensuring stability in tax and expenditure policies. To increase timely and effective analysis of the revenue aspects of the fiscal policy, governments have turned toward in-house tax policy units. One of the essential responsibilities of this fiscal policy units is to promote fiscal discipline and address fiscal vulnerabilities through increased transparency, stronger accountability, and a more pronounced medium-term perspective.

To address these goals, many countries have built a Medium-Term Budgetary Framework (MTBF)¹, a system of projections tailored to obtain values of revenues and expenditures in future periods.

These MTBF also serve as meaningful tools that provide a starting point in addressing compliance problems and supporting evasion deters. The main drawback of these prediction systems is that their construction involves the use of techniques related to time series theory, transfer function models and multivariate analysis.

In particular, regarding to Value-Added Tax models, there are several approaches in the literature that either calculate or forecast the VAT revenue.

Most of the models that compute tax revenue of VAT are focused on estimating the VAT base. Prominent examples of such strategy are the models based on the National Accounts approach, used by the U.S. Department of the Treasury and by the International Monetary Fund² for numerous countries, the models based on the Sectoral Approach, and the Input-Output Models approach that is also used for simulation³.

If we focus on VAT revenue forecasting, we often find in the literature methodologies grounded on the GDP based tax forecasting models. As a first step, the models require the construction of data series for tax revenues and their bases for each tax. All these tax bases are assumed to be predetermined and are obtained from macroeconomic variables derived from national accounts and balance of payments aggregates. These historical data series of tax revenues have embedded in them the effects of increases in national income or expenditures, as well as discretionary changes made in the tax system over time. For the VAT revenue model we present in this paper, this changes brought about by discretionary changes are introduced by dummy variables.

The next step for setting up the GDP based forecasting models is to establish an exact relationship between the tax revenue and the economic variables (ie proxy base). In order to do this, it is necessary to determine the correct base for each tax using the *National Accounts*. Subsequently, it is necessary to find out which component of the *National Account* corresponds most closely to the base for a particular tax.

¹ See the IMF working paper "Medium-Term Budgetary Frameworks - Lessons for Austria from International Experience" by Erik J. Lundback

² See e.g. H.H. Zee and J.P. Boding "Aspects of introducing a Value-Added Tax in Sri Lanka" paper prepared for the International Monetary Fund, Fiscal Affairs Department, (August 1992)

³ See "Tax Analysis and Revenue Forecasting-Issues and Techniques" – by Glenn P. Jenkins, Chu-Yan Kuo and Gangadhar P. Shukla, Harvard Institute for International Development, Harvard University, for a detailed description of these models.

In the case of Value-Added Tax, tax revenues are linked with *Total Consumption Expenditure on Goods and Services*⁴. This could be written as a transfer function and a regression analysis is carried out to forecast future revenue collections.

Obviously the predictive ability of this type of models is limited and error margins are large. This is partly because tax revenues are highly sensitive to a wide variety of economic variables and specifically to the economic cycle, and our ability to forecast the path of the economy using only one explanatory variable in a transfer function is restricted.

In order to address this problem, we could explore the possibility of introducing additional variables covering different areas of the economy (*Domestic Demand, Labour Market and Activity Indicators*, among others), but the high degree of linear dependency among this indicators would cause multicollinearity in the model.

Therefore, we propose principal component analysis applied to the entire set of numerical independent variables, to provide orthogonal regressors for the transfer function, ensuring the lack of multicollinearity with little information loss and increasing the forecasting accuracy.

This approach has the advantage of considering the behavioral responses of certain economic sectors (such as *Tourism Industry* or *Construction*) to the introduction of changes in the existing tax laws, and reciprocally, it is able to capture the influence of a decrease in one specific sector on the VAT revenue.

2. Estimation Strategy

Starting from a set of indicators relative to different areas of the economy (*Construction and Services Activity Indicators, Private Consumption variables, Labour Market Indicators and Imports of Goods and Services*) we propose a principal component analysis as a dimension reduction technique for the set of independent variables. The next step is to use the first and second principal components as inputs to the transfer function to estimate the VAT revenue.

The ultimate goal in principal components analysis is to find the minimum number of dimensions that are able to explain the largest variance contained in the initial set of indicators. We intend to simplify the information which gives us the correlation matrix to make it easier to interpret.

⁴ The Institute for Fiscal Studies's public finance forecasting model, which was used to produce forecasts in Britain in each Green Budget up to 2013, was based on the assumption that VAT revenues grew in line with nominal consumer spending. For further information see "Forecasting the PSBR Outside Government: The IFS Perspective" by Christopher Giles and John Hall, *Fiscal Studies* Volume 19, Issue 1 (February, 1998)

Principal component analysis was originated by Pearson (1901) and later developed by Hotelling (1933). The application of principal components is discussed by Rao (1964), Cooley and Lohnes (1971), and Gnanadesikan (1977). Exceptional statistical treatments of principal components are found in Kshirsagar (1972), Morrison (1976), and Mardia, Kent, and Bibby (1979).

Given a data set with p numeric variables, we can compute up to p principal components. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix. The eigenvectors are customarily taken with unit length. The principal components are sorted by descending order of the eigenvalues, which are equal to the variances of the components.

The principal components meet the following properties (Rao 1964; Kshirsagar 1972):

- The eigenvectors are orthogonal, so the principal components represent jointly perpendicular directions through the space of the original variables.
- The principal component scores are jointly uncorrelated. This property ensures the lack of multicollinearity when we use them as input variables in a regression model.
- The first principal component has the largest variance of any unit-length linear combination of the observed variables. The j th principal component has the largest variance of any unit-length linear combination orthogonal to the first $j-1$ principal components. The last principal component has the smallest variance of any linear combination of the original variables.
- The scores on the first j principal components have the highest possible generalized Variance of any set of unit-length linear combinations of the original variables.
- The first j principal components provide a least squares solution to the model:

$$Y=XB+E$$

Where:

Y is an $n \times p$ matrix of the centered observed variables;

X is the $n \times j$ matrix of scores on the first j principal components;

B is the $j \times p$ matrix of eigenvectors;

E is an $n \times p$ matrix of residuals;

Our goal is to minimize the trace of $E'E$. That means that the first j principal components are the best linear predictors of the original variables among all possible sets of j variables, although any nonsingular linear transformation of the first j principal components would provide an equally good prediction.

In geometric terms, the j -dimensional linear subspace spanned by the first j principal components provides the best possible fit to the data points as measured by the sum of squared perpendicular distances from each data point to the subspace. This is in contrast to the geometric interpretation of least squares regression, which minimizes the sum of squared vertical distances.

3. The Data Set

Our starting point was an extensive collection of time series data comprised of quarterly indicators on a wide range of economic areas valued at current prices (raw data) covering the period from 1995 onwards.

We next selected a subset of indicators, taking into account various attributes: high correlation to the VAT revenue at current prices and quarterly variation rate, speed of publication, operability (easy access), coverage, cyclical sensitivity and frequency.

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	IVGS	IVA
IVGS Large store sales index	1.00000	0.37229 0.0005
IVA VAT revenue	0.37229 0.0005	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	VGEcon	IVA
VGEcon Large firm sales. Consumption goods and services (fixed sample)	1.00000	0.53329 <.0001
IVA VAT revenue	0.53329 <.0001	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	MATRIC	IVA
MATRIC Passenger car registrations	1.00000	0.02092 0.8502
IVA VAT revenue	0.02092 0.8502	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	IMPOR	IVA
IMPOR Good and service imports	1.00000	0.74367 <.0001
IVA VAT revenue	0.74367 <.0001	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	RA	IVA
RA Compensation per employee	1.00000	0.64695 <.0001
IVA VAT revenue	0.64695 <.0001	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	FBCFvivienda	IVA
FBCFvivienda Gross fixed capital formation. Construction.Tangible fixed Assets Dwellings	1.00000	0.40464 0.0001
IVA VAT revenue	0.40464 0.0001	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	CEMEN	IVA
CEMEN Cement Apparent Consumption	1.00000	-0.01332 0.9043
IVA VAT revenue	-0.01332 0.9043	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	CE	IVA
CE Electric power consumption	1.00000	0.78418 <.0001
IVA VAT revenue	0.78418 <.0001	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	CR	IVA
CR Registered contracts	1.00000	0.56870 <.0001
IVA VAT revenue	0.56870 <.0001	1.00000

Pearson Correlation Coefficients, N = 84 Prob > r under H0: Rho=0		
	TURISTAS	IVA
TURISTAS Foreign tourists	1.00000	0.12252 0.2669
IVA VAT revenue	0.12252 0.2669	1.00000

Figure 1: Pearson Correlation Coefficients and associated p-values for VAT tax revenue and the resulting ten partial indicators⁵.

⁵ The correlation coefficients of *Cement Apparent Consumption* and *Passenger Car Registrations* are low compared to the rest of the indicators. We considered this two reference series as useful indicators because of their cycle pattern (Figures 2-4), their relatively short publication lags, and because they belong to economic areas which are sensitive to policy changes.

The next step in the process was to identify the underlying cyclical pattern of the indicators. This goal required the removal of two factors: long term trends and high frequency noise. We decided to remove these factors in a single step using a Fixed length Symmetric Band-Pass Filter (Baxter-King). Figure 2 features the analysis results of the cyclical components of *Cement Apparent Consumption* and *Passenger Car Registrations* obtained by the band-pass filter to validate the inclusion of these partial indicators in the model. The cyclical component of these two time series and the VAT revenue reveal very similar pattern.

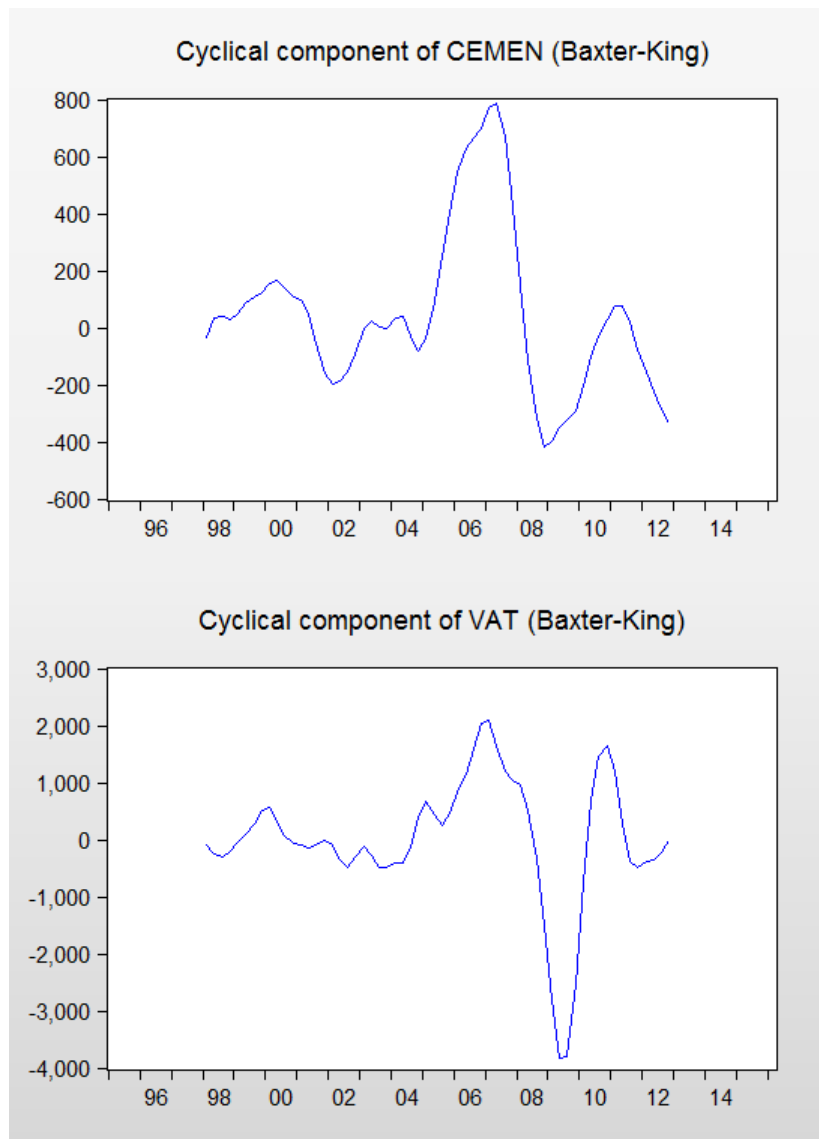


Figure 2: Cyclical pattern of VAT revenue and of Cement Apparent Consumption obtained by a Baxter-King filter.

If the cyclical profiles are highly correlated⁶, the indicator would provide a signal, not only to approaching turning points, but also to developments over the whole cycle. The cross correlation function between the cyclical component of the partial indicators and the cyclical component of the *VAT revenue*, provides invaluable information on cyclical conformity. The location of the peak of the cross-correlation function is a good indicator of average lead time.

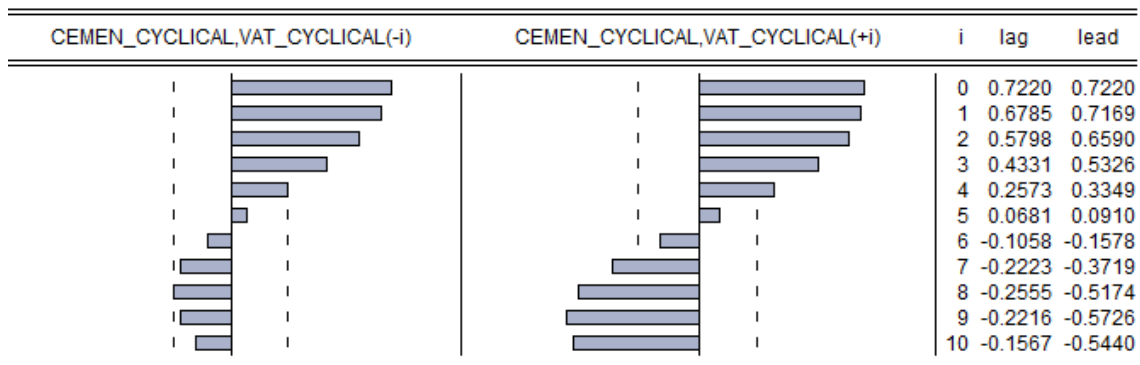


Figure 3: Cross Correlogram of the Cyclical component of *VAT revenue* and of *Cement Apparent Consumption* obtained by a Baxter-King filter.

The cross-correlation analysis of the two cyclical components shows that the cyclical component of *Cement Apparent Consumption* is highly correlated to the cyclical component of *VAT revenue* at lag=0 (Rho=0.7220). A leading relationship (lag=1, Rho=0.6785) could be rationalised on the basis that construction is probably the sector which reacts most quickly to changes in financial conditions.

Figures 4 and 5 feature similar results for the cyclical component of *Passenger Car Registrations*.

⁶ The methodology guideline "OECD System of Composite Leading Indicators" prepared by Gyorgy Gyomai and Emmanuelle Guidetti in April 2012, specifies this approach to select the reference series based on cyclical profiles.

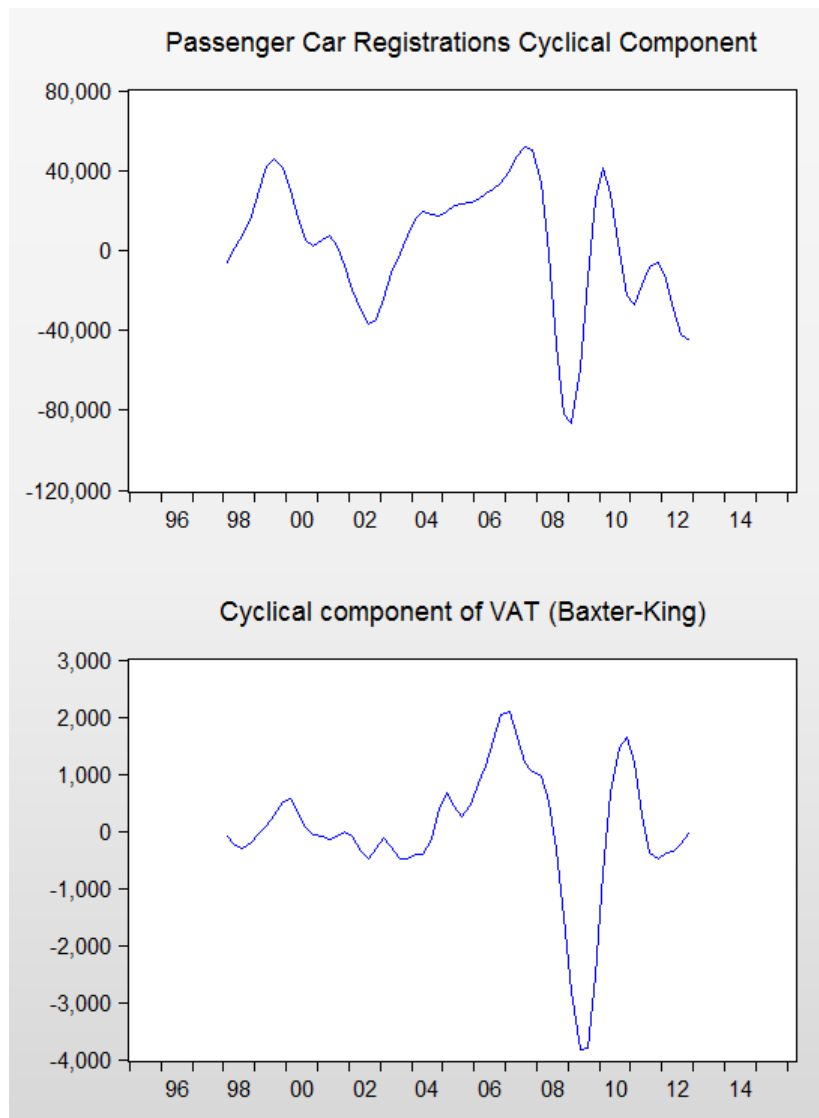


Figure 4: Cyclical pattern of VAT revenue and of Passenger Car Registrations obtained by a Baxter-King filter.

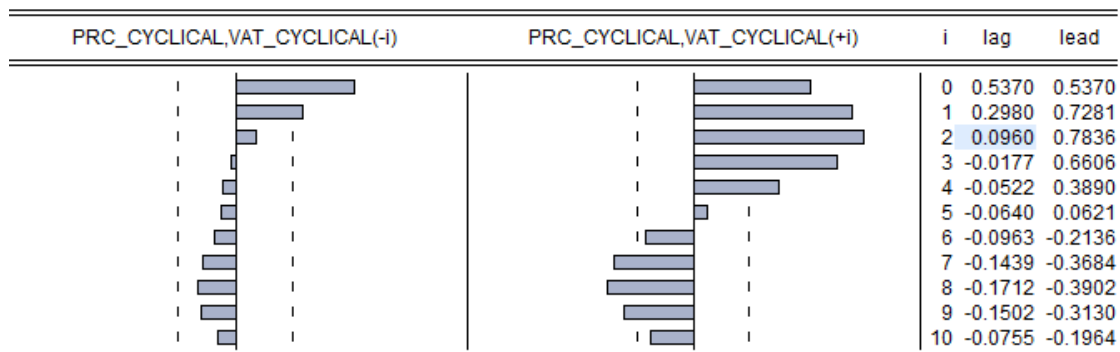


Figure 5: Cross Correlogram of the Cyclical component of VAT revenue and of Passenger Car Registrations obtained by a Baxter-King filter.

The cross-correlation analysis of the two cyclical components shows that the cyclical component of *Passenger Car Registrations* is highly correlated to the cyclical component of *VAT revenue* at lag=0 ($\text{Rho}=0.5370$) but the maximum occurs at lag=2 ($\text{Rho}=0.7836$). That result reveals a lagging relationship between these two variables.

A similar cyclical pattern analysis of the rest of the candidate reference series reveals, for the selected partial indicators, that *Fixed Capital Formation in Construction*, *Goods and Services Imports*, and *Foreign Tourists Arrivals* display movements that precede those of the *VAT revenue* (the average lead is two quarters). *Large Store Sales Index*, *Large Firm Sales (Consumption)*, *Cement Apparent Consumption*, *Electric Power Consumption* and *Registered Contracts* are more significant in providing contemporaneous information. *Passenger Car Registrations* and *Compensation per Employee* perform as lagging indicators.

Note that whereas the correlation value of the peak provides a measure of how well the cyclical profiles of the indicators match, the size of the correlations cannot be the only indicators used for component selection.

Higher correlations in quarterly variation rate maintain a similar structure and correspond to general indicators (*Electric Power Consumption*), consumption indicators (*Large Firm Sales in Consumption Goods and Services*), construction indicators (*Gross Fixed Capital Formation in Construction*) and services indicators (*Foreign Tourists Arrivals*).

We fitted the model using a training data set from $t=1$ (first quarter of 1995) to $t=T$ (last quarter of 2014) and then we tested its performance computing one-step ahead forecasts on a test data set (first, second and third quarter of 2015, corresponding to the latest update released by the AEAT in its “*Tax revenue monthly report*”).

We also provide forecasts for the last quarter of 2015 and the first quarter of 2016. The latest predictions are obtained by extending the partial indicators using seasonal ARIMA models.

The following sections provide a more detailed description of the various steps highlighted above.

4. Estimation Results

4.1 Finding the Two Orthogonal Regressors

As indicated before, the purpose of the principal component analysis is to compute two variables that best summarize all ten partial indicators.

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Commuality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 10 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	5.86515592	3.54002011	0.5865	0.5865
2	2.32513580	1.18389974	0.2325	0.8190
3	1.14123607	0.87975303	0.1141	0.9332
4	0.26148304	0.06464110	0.0261	0.9593
5	0.19684193	0.09649771	0.0197	0.9790
6	0.10034422	0.04954868	0.0100	0.9890
7	0.05079554	0.01803972	0.0051	0.9941
8	0.03275582	0.01654268	0.0033	0.9974
9	0.01621314	0.00617463	0.0016	0.9990
10	0.01003852		0.0010	1.0000

Table 1: Eigenvalues of the Correlation Matrix

		Correlation Matrix									
		IVGS	VGEcon	MATRIC	IMPOR	RA	FBCFvivienda	CEMEN	CE	TURISTAS	CR
IVGS	Large store sales index	1.0000	0.8719	0.4460	0.5673	0.5662	0.8870	0.4756	0.6464	0.0024	0.5031
VGEcon	Large firm sales.Consumption	0.8719	1.0000	0.2777	0.8583	0.8425	0.8511	0.2569	0.8465	0.3493	0.7593
MATRIC	Passenger car registrations	0.4460	0.2777	1.0000	0.0015	-.1344	0.5334	0.8664	-.0134	0.0433	0.3072
IMPOR	Good and service imports	0.5673	0.8583	0.0015	1.0000	0.9477	0.6269	-.0301	0.9341	0.4650	0.8098
RA	Compensation per employee	0.5662	0.8425	-.1344	0.9477	1.0000	0.6099	-.1164	0.9371	0.4089	0.7114
FBCFvivienda	Gross fixed capital formation.Construction.Dwellings	0.8870	0.8511	0.5334	0.6269	0.6099	1.0000	0.6431	0.6742	0.2143	0.5882
CEMEN	Cement Apparent Consumption	0.4756	0.2569	0.8664	-.0301	-.1164	0.6431	1.0000	0.0361	0.0884	0.2543
CE	Electric power consumption	0.6464	0.8465	-.0134	0.9341	0.9371	0.6742	0.0361	1.0000	0.3543	0.7377
TURISTAS	Foreign tourists	0.0024	0.3493	0.0433	0.4650	0.4089	0.2143	0.0884	0.3543	1.0000	0.6883
CR	Registered contracts	0.5031	0.7593	0.3072	0.8098	0.7114	0.5882	0.2543	0.7377	0.6883	1.0000

Table 2: Correlation Matrix of the partial indicators.

Table 1 and Table 2 display the results of the principal component analysis. We compute principal components from the correlation matrix (Table 2). The set of partial indicators show a high correlation between the variables, validating the relevance of prior principal component analysis to avoid problems of multicollinearity.

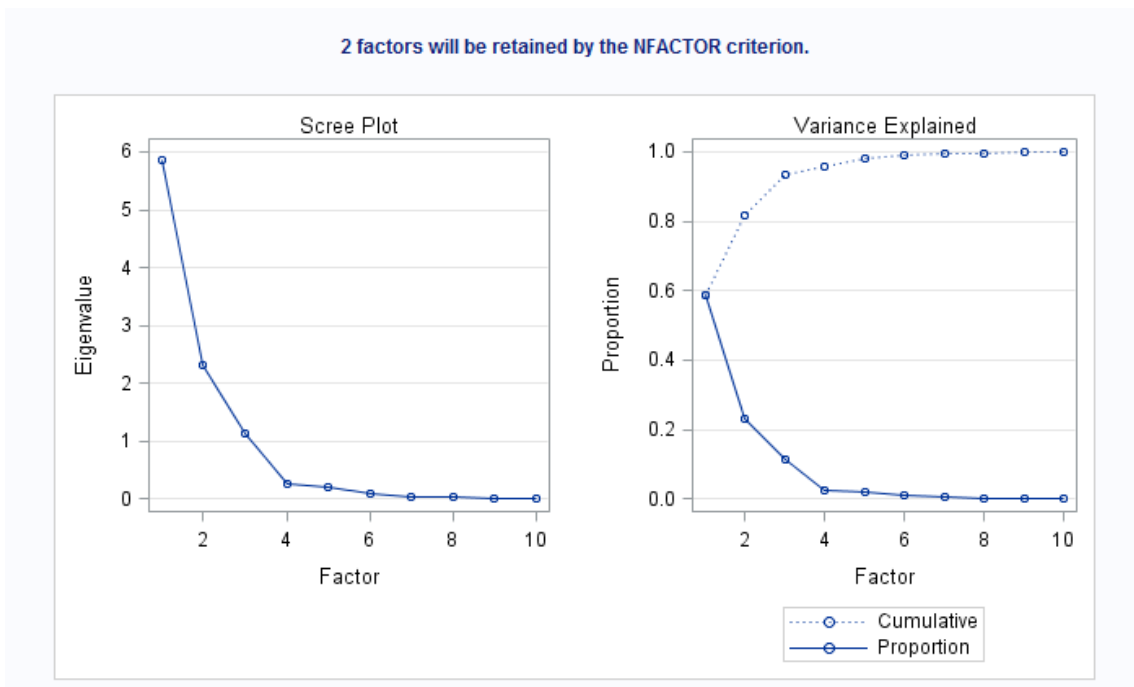


Figure 6: Scree Plot and Variance Explained Plot.

The Scree Plot on the left in Figure 6 shows that the eigenvalue of the first component is approximately 5,9 and the eigenvalue of the second component is largely decreased to 2,35. The Variance Explained Plot on the right in Figure 6 shows that the first two principal components account for nearly 82% of the total standardized variance, which indicates that two components provide a good summary of the data.

Factor Pattern			
		ivafactor1	ivafactor2
IVGS	IVGS	0.80693	0.34099
VGEcon	VGEcon	0.96450	-0.00306
MATRIC	MATRIC	0.31570	0.87375
IMPOR	IMPOR	0.90216	-0.36754
RA	RA	0.86602	-0.44716
FBCFvivienda	FBCFvivienda	0.86793	0.39339
CEMEN	CEMEN	0.33424	0.88981
CE	CE	0.90023	-0.31011
TURISTAS	TURISTAS	0.46228	-0.24458
CR	CR	0.85502	-0.08868

Table 3: Factor Pattern of the two principal components.

The factor pattern (Table 3) shows that the first component (labeled " Ivafactor1") has large positive loadings for all ten variables. The second component is basically a contrast of *Large Store Sales Index* (0,341), *Passenger Car Registrations* (0,873), and the two construction indicators (*Gross Fixed Capital Formation in Construction* and *Cement Apparent Consumption*) against the rest.

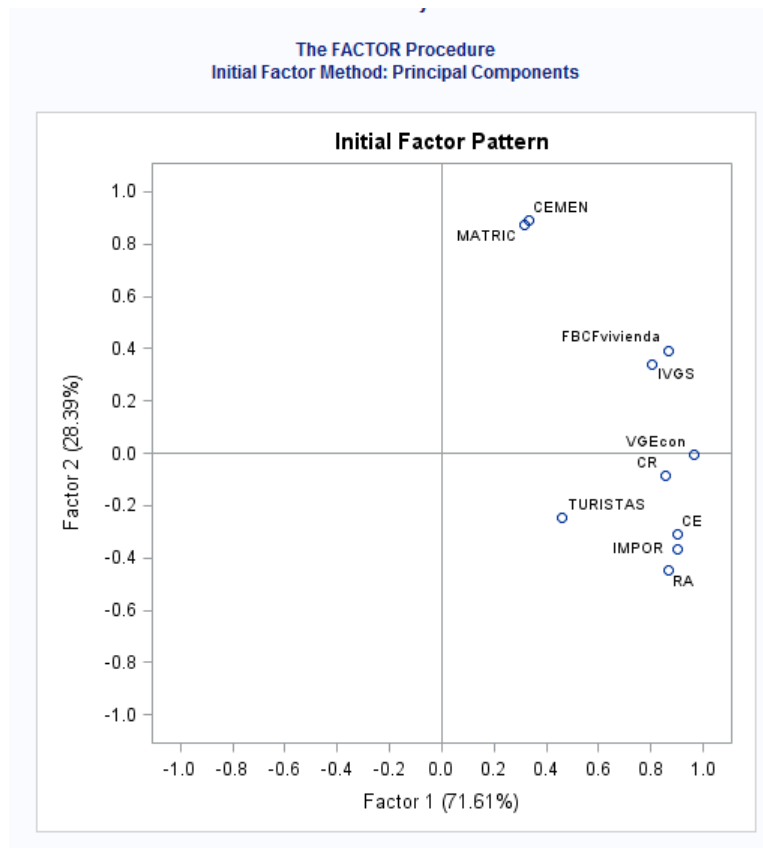


Figure 7: Unrotated Factor Pattern Plot of the two Principal Components.

The unrotated factor pattern (Figure 7) reveals three clusters of variables, with the variables *Cement Apparent Consumption* and *Passenger Car Registrations* at the positive end of Factor2 axis, and *Compensation per Employee*, *Goods and Services Imports* and *Electric Power Consumption* at the negative side. The rest of the variables remain between these two clusters.

The results of the Varimax rotation are shown in Table 4 and Figure 8.

Standardized Scoring Coefficients			
		ivafactor1	ivafactor2
IVGS	IVGS	0.06698	0.18960
VGEcon	VGEcon	0.15109	0.06494
MATRIC	MATRIC	-0.10186	0.36570
IMPOR	IMPOR	0.20440	-0.08286
RA	RA	0.21254	-0.11669
FBCFvivienda	FBCFvivienda	0.06743	0.21442
CEMEN	CEMEN	-0.10175	0.37330
CE	CE	0.19417	-0.06037
TURISTAS	TURISTAS	0.11447	-0.06460
CR	CR	0.14881	0.02372

Table 4: Standardized Factor Scoring Coefficients

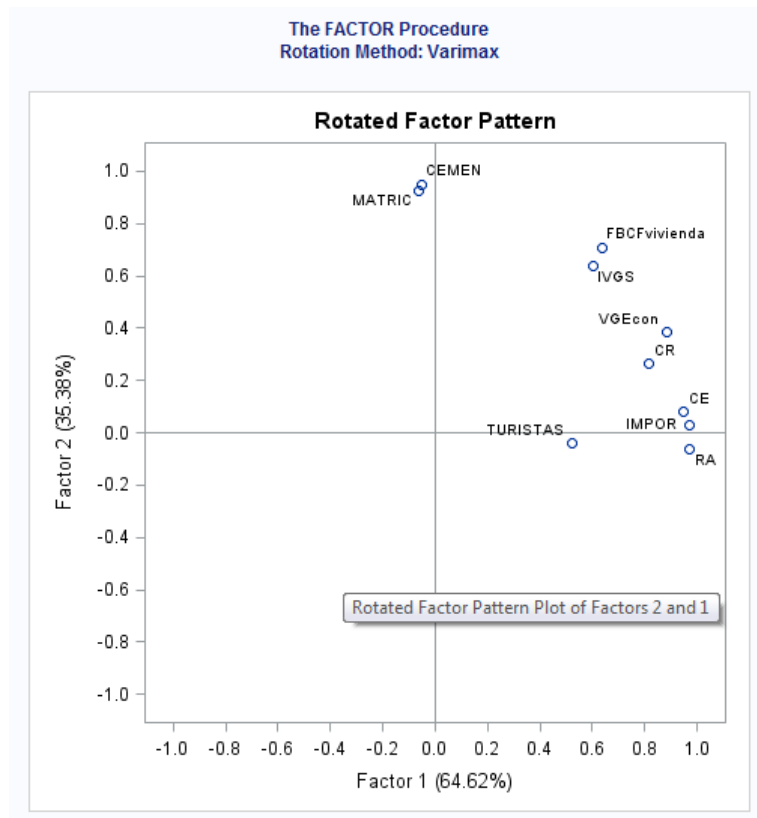


Figure 8: Graphical Plot of the VARIMAX Rotated Factor Loadings.

The graphical plot of the Varimax-rotated factor loadings clearly features that *Cement Apparent Consumption* and *Passenger Car Registrations* cluster together on the Factor 2 axis, while *Compensation Per Employee*, *Goods and Services Imports*, *Electric Power Consumption*, *Foreign Tourists Arrivals*, *Registered Contracts*, *Large Firm Sales (Consumption)* cluster together on the Factor 1 axis. The standardized scoring coefficients of *Gross Fixed Capital Formation (Construction)* and *Large Store Sales Index* are larger in factor 2 than in factor 1.

4.2 Determining the Transfer Function.

The second part of the methodology makes use of these factors as input variables for a transfer function:

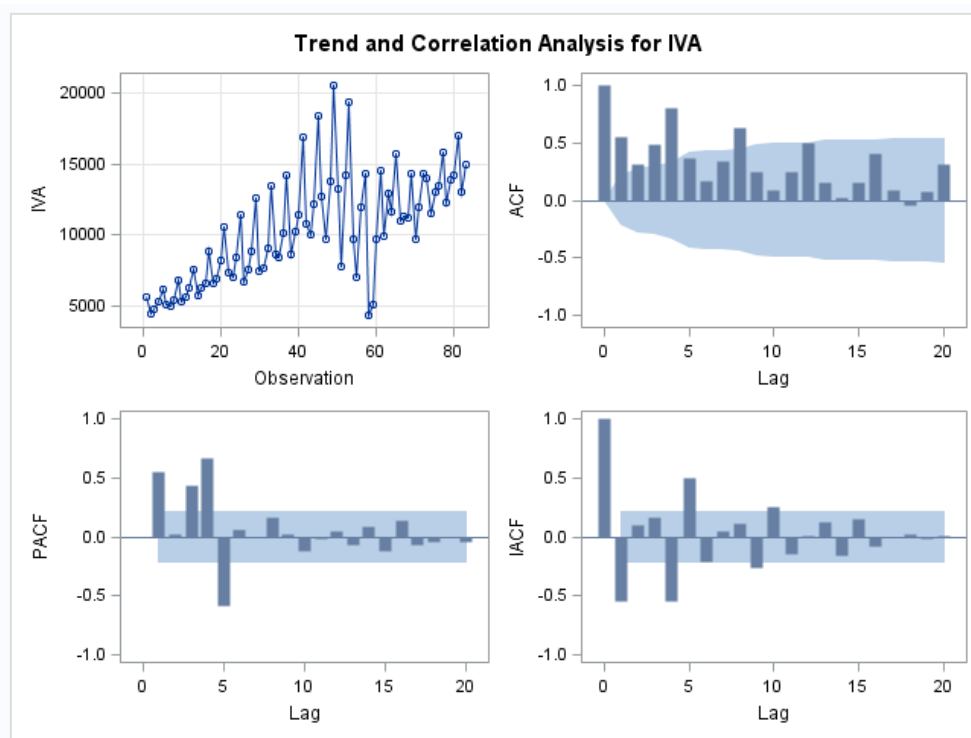


Figure 9: Correlation analysis panel for VAT revenue. Sample Autocorrelation Function plot (ACF), Partial Autocorrelation Function plot (PACF) and Sample Inverse Autocorrelation Function plot (IACF) of VAT revenue.

We introduce in the model two level shifts corresponding to the second quarter of 2010 and the third quarter of 2012 (VAT reform). The parameter estimates table and goodness-of-fit statistics for this model are shown in the conditional Least Squares Estimation table (Table 5).

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	Variable	Shift
MU	9698.5	732.69700	13.24	<.0001	0	IVA	0
MA1,1	0.22207	0.11330	1.96	0.0537	1	IVA	0
AR1,1	0.94349	0.04880	19.33	<.0001	4	IVA	0
NUM1	1783.9	403.69503	4.42	<.0001	0	ivafactor1	0
NUM2	2058.5	275.18582	7.48	<.0001	0	ivafactor2	0
NUM3	4732.6	499.06158	9.48	<.0001	0	Is2010q2	0
NUM4	2255.8	467.58735	4.82	<.0001	0	Is2012q3	0

Table 5: Table of parameter estimates. Method: Conditional Least Squares.

As shown in Table 5, all parameters are statistically significant, although the moving average parameter MA1,1 is close to the 5% significance level.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.35	4	0.3608	0.035	-0.175	0.073	0.014	-0.097	0.051
12	10.93	10	0.3632	0.196	-0.016	-0.004	-0.098	0.071	-0.126
18	13.64	16	0.6252	-0.025	0.092	-0.033	0.004	-0.023	-0.122
24	18.14	22	0.6975	-0.060	0.073	0.066	0.150	-0.056	0.018

Table 6: Check for White Noise Residuals.

The autocorrelations checks on the residuals (Table 6) features there is no autocorrelation of residuals at any lag. Test statistics fail to reject the no-autocorrelation hypothesis at a high level of significance ($p = 0.3608$ for the first six lags). This result seems fairly robust to changes in the number of lags.

The probability of white noise is clearly high (Figure 10).

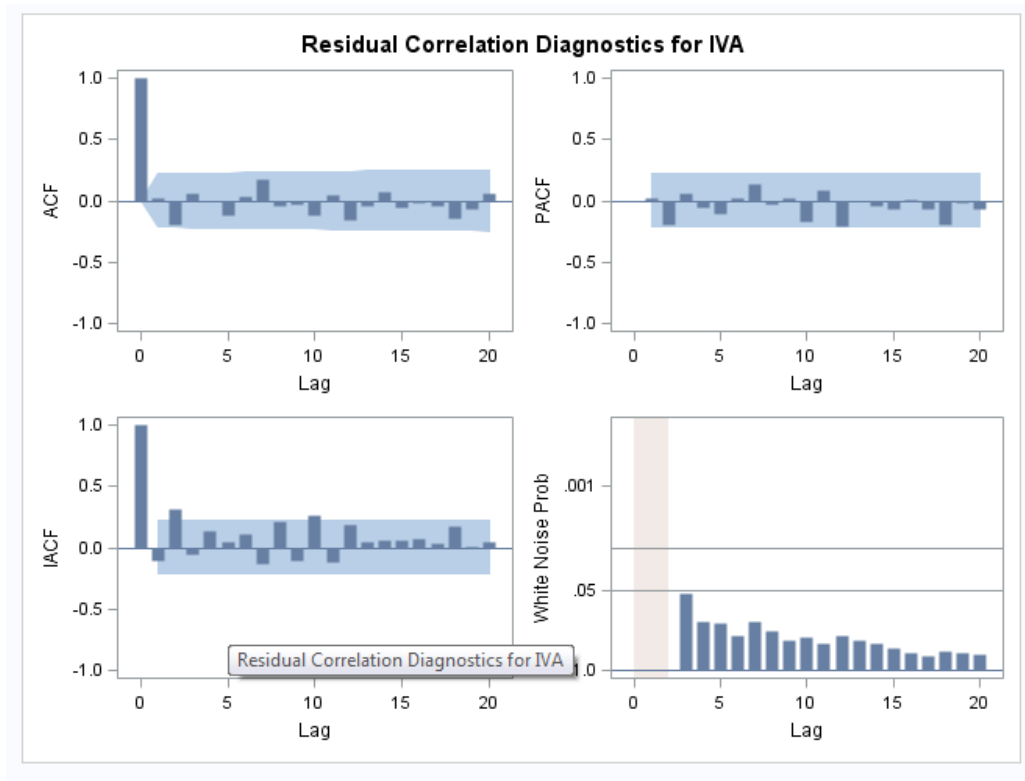


Figure 10: Correlation analysis panel for residuals. ACF, PACF and IACF plots of the residuals. White Noise Probability plot.

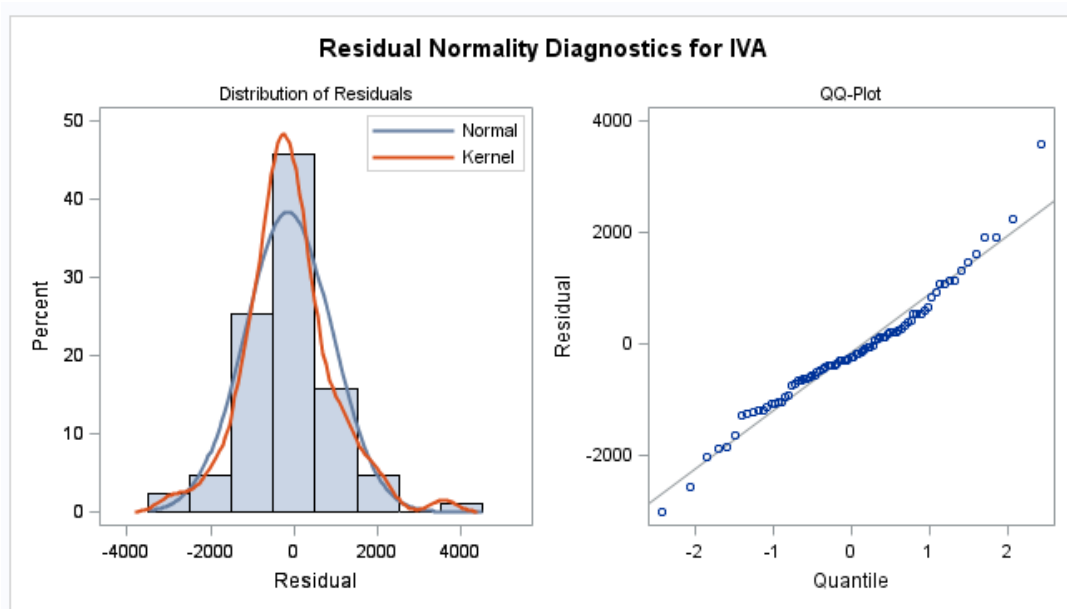


Figure 11: Residual Normality Diagnostics.

As showed in Figure 11 residuals of the model follow a *Normal* distribution.

4.3 Out of Sample Forecasts

1. For the first quarter of 2015, the observed *VAT revenue* at current prices in Millions Euros was: 16997,655. Table 7 shows the predicted values for *VAT revenue* by the model.

Forecasts for variable IVA				
Obs	Forecast	Std Error	95% Confidence Limits	
81	16804.4492	1114.6663	14619.7434	18989.1550

Table 7: Forecast and Confidence Limits of *VAT revenue*. Out of Sample estimations. First quarter of 2015. Million Euros.

2. For the second quarter of 2015, the observed VAT revenue at current prices in Millions Euros was: 13032,929. Table 8 shows the predicted values for *VAT revenue* by the model.

Forecasts for variable IVA				
Obs	Forecast	Std Error	95% Confidence Limits	
82	13349.7536	1106.5157	11181.0226	15518.4846

Table 8: Forecast and Confidence Limits of *VAT revenue*. Out of Sample estimations. Second quarter of 2015. Million Euros.

3. For the third quarter of 2015, the observed VAT revenue at current prices in Millions Euros was: 14976,823. Table 9 shows the predicted values for *VAT revenue* by the model.

Forecasts for variable IVA				
Obs	Forecast	Std Error	95% Confidence Limits	
83	15000.9714	1098.8438	12847.2771	17154.6657

Table 9: Forecast and Confidence Limits of *VAT revenue*. Out of Sample estimations. Third quarter of 2015. Million Euros.

4.4 One Step Forward Forecast

We calculate the forecast for the fourth quarter of 2015:

Forecasts for variable IVA				
Obs	Forecast	Std Error	95% Confidence Limits	
84	15260.0001	1091.1343	13121.4161	17398.5841

Table 10: One-step-Forward Forecast and Confidence Limits of VAT revenue. Fourth quarter of 2015. Million Euros.

4.5 Forecasts from Extended Partial Indicators

We also provide a forecast for the first quarter of 2016 obtained by extending the partial indicators using seasonal ARIMA models.

Forecasts for variable IVA				
Obs	Forecast	Std Error	95% Confidence Limits	
85	17719.2223	1085.6321	15591.4225	19847.0222

Table 11: Two-step-Forward Forecast and Confidence Limits of VAT revenue. First quarter of 2016. Million Euros.

The last updated values of VAT revenue corresponds to November 2015. Thus, the quarterly one-step-forward prediction refers to the last quarter of 2015 and it is calculated from observed data of all partial indicators.

The forecast for the first quarter of 2016 (two-steps forward) is calculated from Seasonal Arima estimates of all partial indicators and accumulates greater bias.

5. Conclusions

As mentioned in the introduction, the final aim of this paper is to propose a methodology that successfully combines principal components analysis and transfer function theory to forecast VAT revenue. This approach offers advantages to Value-Added Tax forecasting models based on the National Accounts approach, and specifically, to those using *Total Consumption Expenditure on Goods and Services* as the only input explanatory variable. The pre-selection of the reference series and the dimension reduction technique enables to incorporate in advance changes in specific fields of the economy that may affect tax revenues.

The analysis of which partial indicators contain useful leading or lagging information about the dependent variable and the filtering process aimed to identify underlying cyclical pattern of the candidate component series was not simple. The approach taken does not in any sense attempt to construct an optimal set of partial indicators, but has the more limited aim of assessing which indicators contain information that is useful for VAT revenue short-term forecasting. Among other criteria, we selected those variables that exhibited a cyclical profile highly correlated to the cyclical pattern of VAT revenue. *Fixed Capital Formation in Construction, Goods and Services Imports*, and *Foreign Tourists Arrivals* were found to add significant leading information to the model. The usefulness of the rest of indicators arises

from the contemporaneous relationships between the variables, and their inclusion in the model found some support in less sophisticated methods such as correlation analysis. Consumption indicators selected have also the advantage of having shorter publications lags than the National Accounts.

The output factors obtained from the dimension reduction technique were highly significant as explanatory variables in the transfer function. Thus, their influence has been crucial for achieving such high predictive power of the model.

Model predictions are consistent with the extension of the expansionary phase of Spanish economy activity expected from the Central Bank of Spain in the final stretch of 2015 and throughout 2016, and its projections for the *Private Consumption*, expected to grow at slightly over 3% during 2015 and 2016⁷. Along the same lines, the European Commission *Country Report Spain 2016*⁸ expects growth to remain robust in 2016 and 2017, while *Private Consumption* and *Investment* are set to remain the main drivers of growth in these two coming years.

References

- [1]. Artola, C., and Galán, E. (2012). Tracking the future on the Web: construction of leading indicators using internet searches. Banco de España. Documentos Ocasionales N.º 1203.
- [2]. Bandholz, Harm (2005) New Composite Leading Indicators for Hungary and Poland. Ifo Working Paper, No. 3. Available at <http://www.econstor.eu/handle/10419/73832>.
- [3]. Central Bank of Spain (2015) Quarterly Report on the Spanish Economy. Economic Bulletin, December 2015. Available at <http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/BoletinEconomico/descargar/15/Dic/Files/be1512-coye.pdf>
- [4]. European Commission. (2016) Country Report Spain 2016. Including an In-Depth Review on the prevention and correction of macroeconomic imbalances (February 26, 2016). Commission staff working document. Available at http://ec.europa.eu/europe2020/pdf/csr2016/cr2016_spain_en.pdf
- [5]. European Commission (2015), VAT Rates Applied in the Member States of the European Union. Situation at 1st September 2015– Taxud.c.1(2015) – EN, Available at http://ec.europa.eu/taxation_customs/resources/documents/taxation/vat/how_vat_works/rates/vat_rates_en.pdf.

⁷ See Quarterly Report on the Spanish Economy, Economic Bulletin, December 2015 . Banco de España. Projections cut-off date: 17 December 2015.

⁸ See Country Report Spain 2016. Including an In-Depth Review on the prevention and correction of macroeconomic imbalances. European Commission (February 26, 2016).

- [6]. Glenn P. J., Chu-Yan K. and Gangadhar P. S. (2000) Tax Analysis and Revenue Forecasting-Issues and Techniques. Harvard Institute for International Development, Harvard University.
- [7]. Giles, C. and Hall, J. (1998), Forecasting the PSBR Outside Government: The IFS Perspective, Fiscal Studies Volume 19, Issue 1.
- [8]. Gyomai, G. and E. Guidetti (2012) OECD system of composite leading indicators. Available at <http://www.oecd.org/std/leading-indicators/41629509.pdf>
- [9]. Lundback, E.J. (2008). Medium-Term Budgetary Frameworks - Lessons for Austria from International Experience. IMF working paper WP/08/163, Available at <https://www.imf.org/external/pubs/ft/wp/2008/wp08163.pdf>.
- [10]. Labeaga, J. M. and A. López (1994), Estimation of the welfare effects of indirect tax changes on Spanish households: an analysis of the 1992 VAT reform, Investigaciones Económicas, Vol. XVIII(2), May, pp.289-311.
- [11]. Michael Keen (2013) The Anatomy of the VAT. IMF Working Paper. Fiscal Affairs Department. WP/13/111. Available at <https://www.imf.org/external/pubs/ft/wp/2013/wp13111.pdf>.
- [12]. Pérez, C. (2006) Econometría de las series temporales. Prentice Hall.
- [13]. Pérez, C. (2007) Econometría básica. Prentice Hall.
- [14]. Pérez, C. (2008) Econometría avanzada. Técnicas y herramientas.. Prentice Hall.
- [15]. Pérez, C. (2010) El Sistema Estadístico SAS. Garceta Grupo Editorial
- [16]. Pérez, C. (2013) Análisis multivariante de datos. Garceta Grupo Editorial
- [17]. Țitan, E., Boboc, C., Ghita, S., Todose, D (2011) Econometric Analysis of the Correlations between the Social Security Budget and the Main Macroeconomic Aggregates in Romania. Theoretical and Applied Economics Volume XVIII (2011), No. 2 (555), pp. 117-126.
- [18]. Zee, H.H. and Boding, J.P (1992) Aspects of introducing a Value-Added Tax in Sri Lanka. Paper prepared for the International Monetary Fund, Fiscal Affairs Department.