



La predicción de la recaudación por el impuesto sobre sociedades: La incidencia de los indicadores adelantados y los indicadores cualitativos en el marco de las series temporales

José Luis Barrón Basterrechea

Cuerpo Superior de Sistemas y Tecnologías de la Información de la Administración del Estado. IEF

Camino González Vasco

Cuerpo Superior de Estadísticos de la Administración del Estado. IEF

Raquel Pajares Rojo

Cuerpo Superior de Técnicos Comerciales y Economistas del Estado. IEF

Este trabajo ha sido seleccionado para su publicación por: don Alberto García Valera, doña Carolina del Campo Azpiazu, doña María Luisa González-Cuéllar Serrano, doña Clara Jiménez Jiménez, don Manuel Lucas Durán y don Jesús Sanmartín Mariñas.

Extracto

Este trabajo de investigación contribuye a la literatura empírica en el desarrollo y la estimación de un modelo de predicción para la recaudación del impuesto sobre sociedades (IS) basándose en el análisis de indicadores macroeconómicos. El estudio que proponemos utiliza un conjunto de predictores que incluye datos de la Central de Balances Trimestral del Banco de España y diversos indicadores opináticos, además de varias variables macroeconómicas relativas a la demanda nacional y al sector exterior, reuniendo un total de 43 variables trimestrales. Para solucionar el problema de la multicolinealidad inherente a estas variables utilizamos una combinación de métodos estadísticos como el análisis de componentes principales y la función de transferencia (modelos autorregresivos con regresores exógenos, ARX). Los resultados del ejercicio de predicción en el intervalo de validación (*backtesting*) muestran la capacidad predictiva del modelo con esta combinación de técnicas e ilustran su utilidad como herramienta de predicción a corto plazo del impuesto.

Palabras clave: modelos ARX; predicción impositiva; impuesto sobre sociedades.

Fecha de entrada: 03-05-2018 / Fecha de aceptación: 10-07-2018 / Fecha de revisión: 26-12-2018

Cómo citar: Barrón Basterrechea, J. L.; González Vasco, C. y Pajares Rojo, R. (2019). La predicción de la recaudación por el impuesto sobre sociedades: la incidencia de los indicadores adelantados y los indicadores cualitativos en el marco de las series temporales. *Revista de Contabilidad y Tributación. CEF*, 431, 91-120.





Forecasting Corporate Income Tax Revenues: The incidence of leading and soft indicators in a time series framework

José Luis Barrón Basterrechea

Camino González Vasco

Raquel Pajares Rojo

Abstract

This paper contributes to the empirical literature on the development and estimation of a forecasting model for Corporate Income Tax (CIT) revenue focusing on the study of macroeconomic indicators. The proposed study uses a set of predictors comprising Central Bank Balance Sheet Data and soft indicators, along with some macroeconomic variables related to domestic demand and the foreign sector, which amounts to a total of 43 initial quarterly variables. To address the problem of multicollinearity, we use a blend of statistical methods such as Principal Components Analysis and ARX (Auto- Regressive with eXogenous input) models. Results of the backtesting exercises show that this combination successfully predicts the evolution of the CIT revenue and illustrate its usefulness as a tool for short-term CIT revenue forecasting.

Keywords: ARX; forecast combination; CIT forecasting.

Citation: Barrón Basterrechea, J. L.; González Vasco, C. y Pajares Rojo, R. (2019). La predicción de la recaudación por el impuesto sobre sociedades: la incidencia de los indicadores adelantados y los indicadores cualitativos en el marco de las series temporales. *Revista de Contabilidad y Tributación. CEF*, 431, 91-120.





Sumario

1. Introducción
 2. Revisión de la literatura
 3. El conjunto de datos
 - 3.1. Criterios para la selección de los indicadores parciales
 - 3.2. Listado de indicadores seleccionados
 4. Estrategia de estimación
 - 4.1. El análisis de componentes principales como técnica de reducción de la dimensión
 - 4.2. Modelos autorregresivos con regresores exógenos (ARX)
 5. Resultados de la estimación
 - 5.1. Determinación de los regresores ortogonales
 - 5.2. Determinación del modelo ARX
 - 5.3. Predicciones fuera de la muestra: ejercicios de *backtesting*
 6. Conclusiones
- Referencias bibliográficas

Nota: Las opiniones expresadas en este documento pertenecen a sus autores y no necesariamente reflejan las opiniones de la Institución para las que trabajan.

1. Introducción

La aplicación de la política fiscal en los países desarrollados parte de los denominados Marcos presupuestarios de medio plazo que contienen los gastos públicos necesarios para lograr los objetivos políticos, sociales y económicos, en un horizonte temporal comprendido normalmente entre los tres y cinco años, así como los ingresos públicos que los financiarían, entre los que destacan por su importancia la recaudación impositiva y la emisión de deuda pública.

Sin embargo, este marco de medio plazo es insuficiente para la ejecución de los presupuestos de gastos y de ingresos aprobados anualmente. Como es bien sabido, el presupuesto de gastos establece límites para cada tipo de gasto público y su ejecución se lleva a cabo con procedimientos presupuestarios legales muy restrictivos mientras que el presupuesto de ingresos es meramente estimativo. Por ello, su ejecución requiere de una previsión precisa de los ingresos públicos a corto plazo de tal modo que el gobierno conozca con suficiente antelación si los ingresos fiscales se ajustan a los presupuestados o si se están desviando. Esto adquiere una especial relevancia cuando se trata de impuestos que representan un alto porcentaje del total de ingresos públicos como sucede con el impuesto sobre la renta de sociedades (IS).

La predicción de la recaudación del IS sobre una base trimestral ha demostrado ser uno de los temas más espinosos dentro de las proyecciones de ingresos públicos debido a algunos rasgos distintivos del impuesto. En primer lugar, las empresas tienen tanto incentivos como margen suficiente para reducir su carga fiscal modificando su comportamiento, por ejemplo, aumentando la proporción de capitales ajenos en su estrategia financiera o alterando los precios de transferencia de las transacciones internas al grupo. En segundo lugar, la conexión entre los beneficios societarios y la base imponible del IS está sujeta a las modificaciones e interpretaciones de las autoridades fiscales. Por último, se trata de un impuesto que se liquida anualmente y la recaudación trimestral consiste en pagos a cuenta realizados por anticipado por las empresas en función de sus expectativas de ganancias anuales o de las ganancias del ejercicio anterior.

La mayor parte de las autoridades fiscales estiman los ingresos impositivos futuros usando metodologías fundamentadas en el excedente neto de explotación de la contabilidad nacional como una forma de aproximar las ganancias corporativas. Ello ha sido así porque la literatura económica y tributaria se ha centrado en la estimación de las elasticidades del IS respecto a alguna variable macroeconómica, generalmente el producto interior bruto (PIB) o el excedente neto de explotación de la contabilidad nacional como una forma de aproximar las ganancias corporativas, de tal modo que las previsiones de ingresos tributarios serían el resultado de aplicar estas elasticidades a las proyecciones del PIB o del excedente neto de explotación. La dependencia de estos modelos econométricos de una única variable exógena para aproximar la base imponible hace que sean incapaces de captar la varia-

bilidad que se produce en la evolución de los distintos sectores económicos y transmitirla a la predicción de recaudación, lo que ha puesto en entredicho su capacidad predictiva.

Para abordar esta debilidad, se han construido nuevos modelos que introducen variables adicionales como regresores en las funciones de transferencia de las estimaciones de la base imponible. Sin embargo, el alto grado de dependencia lineal entre estas variables genera problemas de multicolinealidad en los modelos. Para superarlo, en este trabajo proponemos un enfoque diferente que permite hacer uso de la ganancia de información que se puede obtener utilizando un amplio conjunto de indicadores de diferentes áreas de la economía para predecir los ingresos trimestrales del IS. Este enfoque elimina el problema de la multicolinealidad presente en el conjunto de variables utilizadas mediante un análisis de componentes principales como técnica de reducción de la dimensión. Así, el modelo es capaz de captar las oscilaciones que se producen en diferentes áreas de la economía, que muestran una evolución cíclica claramente diferenciada, y resumirlas en cuatro factores ortogonales (las componentes principales) que servirán de regresores en la función de transferencia final. Esta técnica garantiza la ausencia de multicolinealidad con una mínima pérdida de información a la vez que aumenta la precisión del pronóstico a corto plazo del IS.

Este artículo se estructura como sigue: la *sección 1* contiene la introducción y motiva el trabajo de investigación. La *sección 2* ilustra la revisión de la literatura. La *sección 3* detalla el conjunto de datos. La *sección 4* explica el desarrollo del modelo empleado y describe la técnica de estimación y el marco empírico. La *sección 5* muestra los resultados de la estimación y la evaluación de la predicción en el intervalo de validación. La *sección 6* brinda las principales conclusiones del estudio.

2. Revisión de la literatura

La literatura sobre la predicción de la recaudación del IS se centra principalmente en la estimación de la base imponible del impuesto a partir de variables macroeconómicas y de elasticidades o semi-elasticidades tributarias con respecto a su base, el *output gap* o el PIB. El análisis es más preciso cuando los modelos adoptan un enfoque dinámico que permite, por un lado, diferenciar entre las elasticidades de corto y largo plazo y, por otro, un ajuste asimétrico de la recaudación en las fases expansivas o recesivas del ciclo económico. Wolswijk (2007) aplicó este enfoque dinámico a los Países Bajos y Koester, G. B. *et al.* (2017) lo hicieron para los países de la zona euro. Este es el enfoque más utilizado por los gobiernos.

Paralelamente, existe una amplia literatura empírica sobre la calidad de las predicciones de ingresos impositivos. Una revisión de la misma se puede encontrar en Leal *et al.* (2008). La mayor parte de los trabajos se centran en aspectos técnicos como la insesgidez, la racionalidad así como en cuestiones metodológicas (véase, p. ej. Bretschneider *et al.*, 1989; Chatagny, 2015). Existen muy pocos artículos que analicen la influencia política sobre las

predicciones impositivas (véase, por ejemplo, Cassidy *et al.*, 1989; Boylan *et al.*, 2008; Chagny 2015), y algunos de ellos se centran en los ciclos electorales (Boylan *et al.*, 2008).

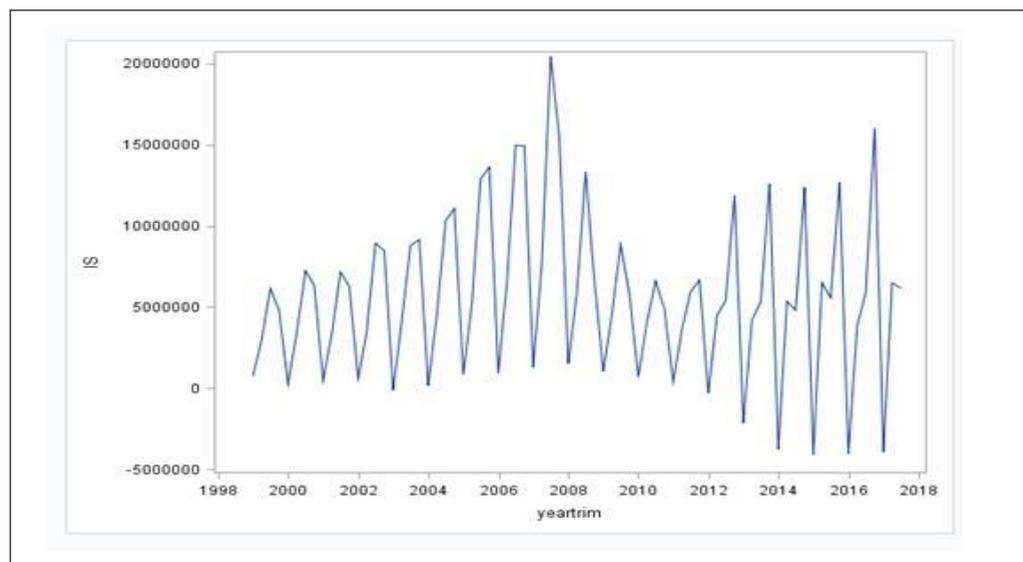
En este trabajo se propone un modelo de estimación y predicción de los ingresos trimestrales de uno de los impuestos más volátiles de nuestro sistema fiscal, el IS, utilizando una combinación de indicadores económicos y técnicas estadísticas para conseguir el objetivo de proporcionar una estimación válida de la recaudación del IS.

3. El conjunto de datos

3.1. Criterios para la selección de los indicadores parciales

La evolución de la recaudación trimestral del IS se puede observar en el siguiente gráfico:

Gráfico 1. Ingresos del IS desde el primer trimestre de 1999 hasta el tercero de 2017. Datos en miles de euros



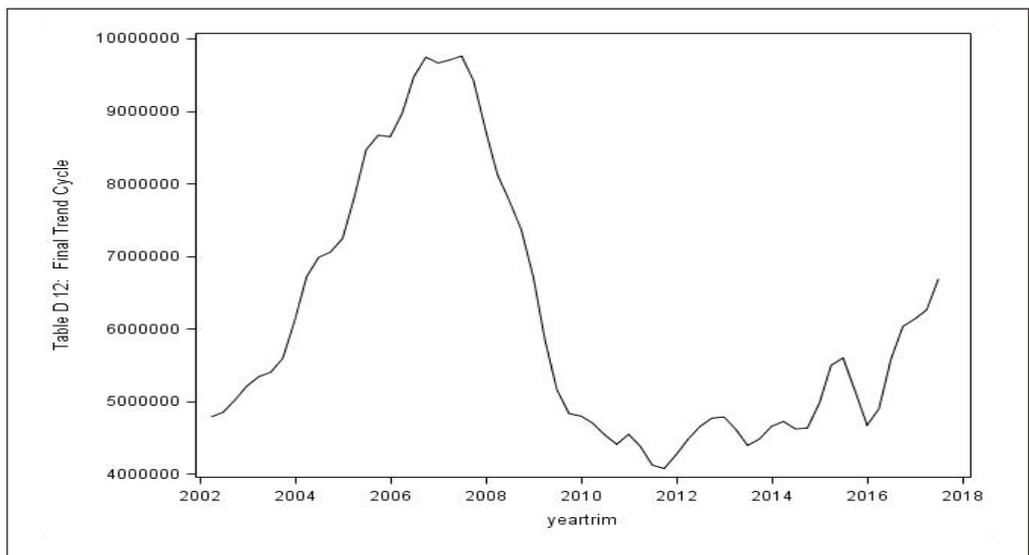
En el gráfico anterior se ve claramente como la recaudación trimestral del IS se ve fuertemente dominada por su componente estacional. Como señalábamos en la introducción de este trabajo, uno de los motivos por el cual las previsiones de los ingresos trimestrales del IS han sido y continúan siendo unas de las más difíciles de abordar son sus especiales características, que hacen que estén sujetas a un mayor grado de incertidumbre y volatilidad que los ingresos que provienen del resto de figuras impositivas, por lo que tradicionalmente muestran los errores de predicción más elevados.

Las razones de esta mayor volatilidad hay que buscarlas en las características del IS:

- a) Las empresas tienen incentivos y un margen de maniobra muy superior al de los sujetos pasivos de otros impuestos, ya que disponen de mecanismos para afectar a sus beneficios gravados y, por tanto, a su base tributaria.
- b) Las empresas también pueden reducir su carga fiscal modificando su estrategia financiera, ya que el coste de la financiación ajena es deducible fiscalmente mientras que el de la financiación propia no.
- c) En una economía abierta como la española, los pagos impositivos pueden reducirse mediante modificaciones en la localización de las actividades gravadas o mediante cambios en los precios de transferencia de las transacciones intragrupo. Estas traslaciones de la carga impositiva son muy difíciles de predecir.

Para aminorar este problema, consideramos la componente de ciclotendencia de la recaudación del impuesto, que nos permite conocer el comportamiento a largo plazo de la recaudación impositiva, ya que disminuye las oscilaciones provocadas por las componentes estacional e irregular de la serie.

Gráfico 2. Componente de ciclotendencia de los ingresos del IS desde el primer trimestre de 2002 hasta el tercero de 2017. Miles de euros. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)



Como se puede observar, durante los trimestres iniciales del periodo muestral hasta el inicio de la crisis financiera internacional (principios de 2008) la componente de ciclotendencia de la recaudación impositiva muestra un fuerte ascenso, en coherencia con la fase expansi-

va del ciclo económico español, alcanzando su valor máximo en el tercer trimestre de 2007. A partir del primer trimestre de 2008, la componente de ciclotendencia del impuesto sufre una dramática caída, con su valor mínimo en el tercer trimestre de 2011. A partir de ese momento, se observan dos máximos locales en la serie de ciclotendencia, correspondientes al primer trimestre de 2013 y al tercero de 2015, pero no es hasta el inicio del año 2016 cuando la componente de ciclotendencia de la serie reanuda de nuevo una trayectoria ascendente.

Por lo que se refiere a los indicadores seleccionados para la estimación del modelo y, teniendo en cuenta la metodología econométrica propuesta, se ha escogido una gama muy amplia con el objetivo de cubrir todo tipo de *shocks* que puedan afectar a la actividad económica y, en consecuencia, a la recaudación del IS. Las condiciones impuestas a los indicadores han sido que deberían estar disponibles a tiempo y elaborarse o agregarse trimestralmente. Asimismo, se han elegido teniendo en cuenta el perfil cíclico de los mismos, con la idea de que puedan «adelantar» o completar las oscilaciones de la recaudación impositiva.

Como ejemplo de los perfiles cíclicos de los indicadores seleccionados, los siguientes gráficos muestran la componente de ciclotendencia de algunos de los indicadores más significativos en el modelo de predicción del IS: a) el Excedente neto de explotación de las sociedades no financieras (de la contabilidad nacional trimestral de España); b) el Índice de cifra de negocios con Base 2010 (Instituto Nacional de Estadística, INE); c) el Índice de producción industrial general, Base 2010 (INE), y d) el Número de sociedades mercantiles constituidas (INE).

Gráfico 3. Componente de ciclotendencia del Excedente neto de explotación de las sociedades no financieras, Base 2010, desde el primer trimestre de 1999 hasta el tercero de 2017. Miles de euros. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

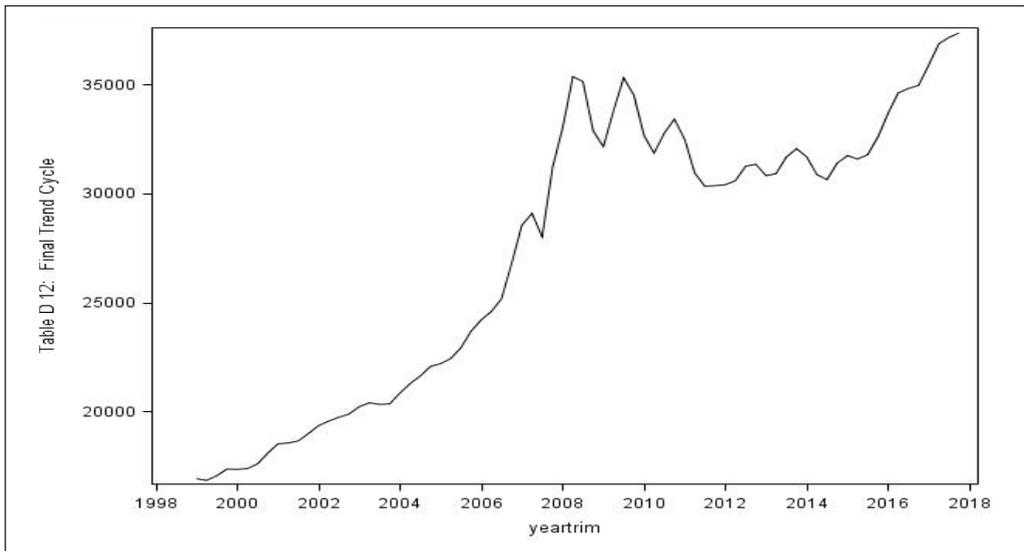


Gráfico 4. Componente de ciclotendencia del Índice de cifra de negocios, Base 2010, desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

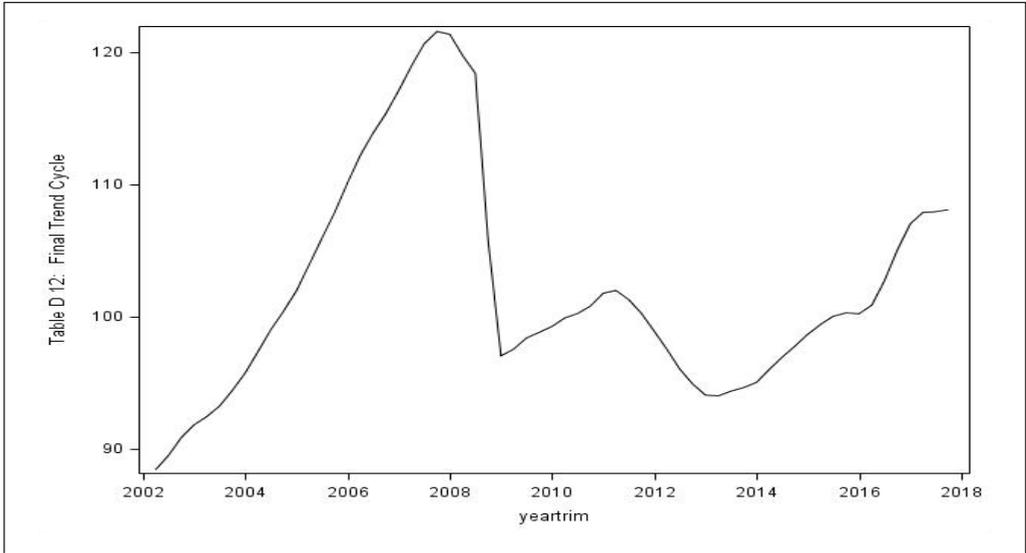


Gráfico 5. Componente de ciclotendencia del Índice de producción industrial general, Base 2010, desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

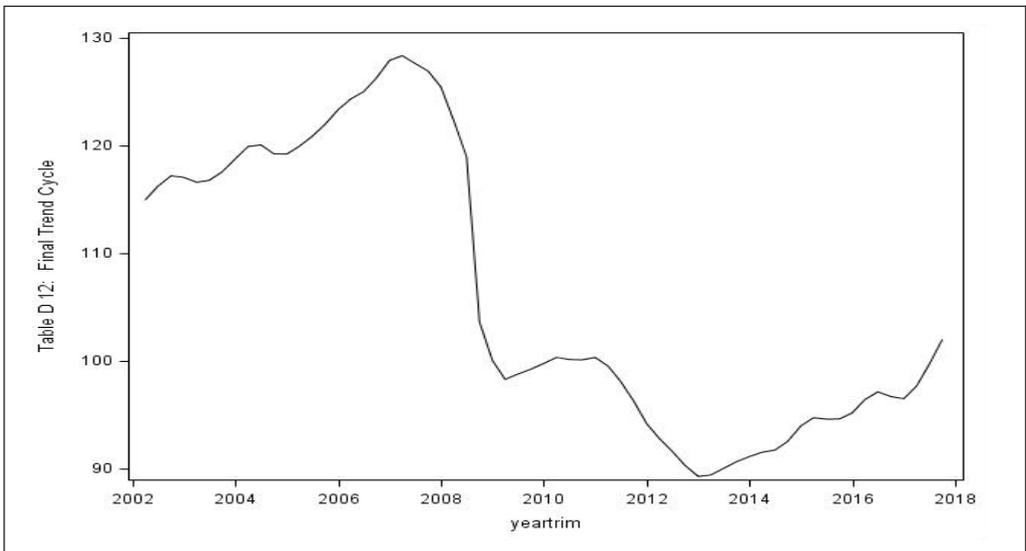
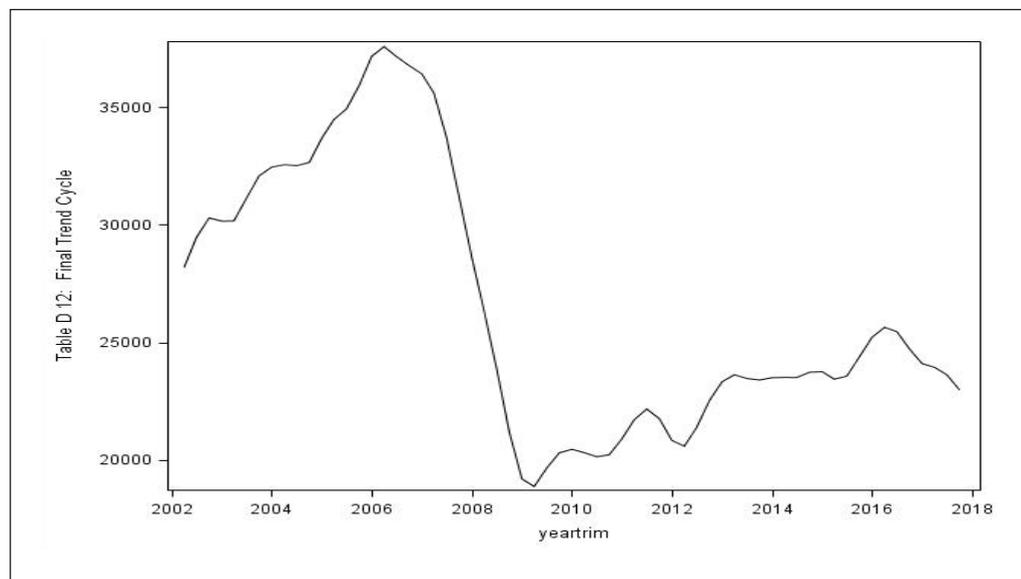


Gráfico 6. Componente de ciclotendencia del Número de sociedades mercantiles constituidas desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)



Como podemos observar en los gráficos anteriores, la señal cíclica de estos indicadores es muy similar, con una trayectoria ascendente en los primeros años de la muestra, hasta el primer trimestre de 2008 cuando se inicia un acusado descenso y con un comienzo de recuperación en los inicios de 2012 o 2013.

Por su parte y a diferencia de los anteriores, los indicadores del mercado laboral muestran perfiles cíclicos heterogéneos dependiendo del sector de actividad económica considerado.

Gráfico 7. Componente de ciclotendencia del Empleo equivalente a tiempo completo. Ocupados. Industria. Miles de puestos de trabajo desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

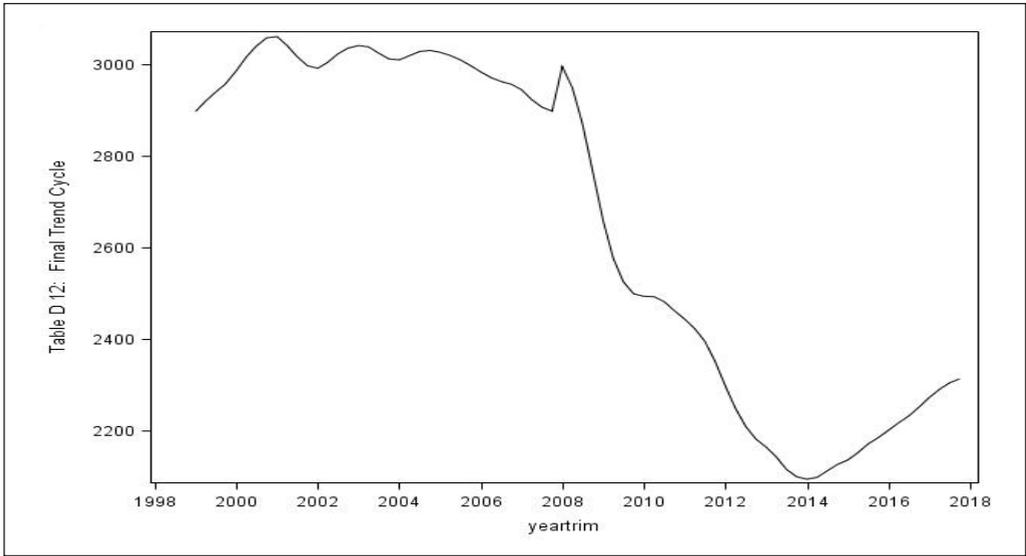


Gráfico 8. Componente ciclotendencia del Empleo equivalente a tiempo completo. Ocupados. Construcción. Miles de puestos de trabajo desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

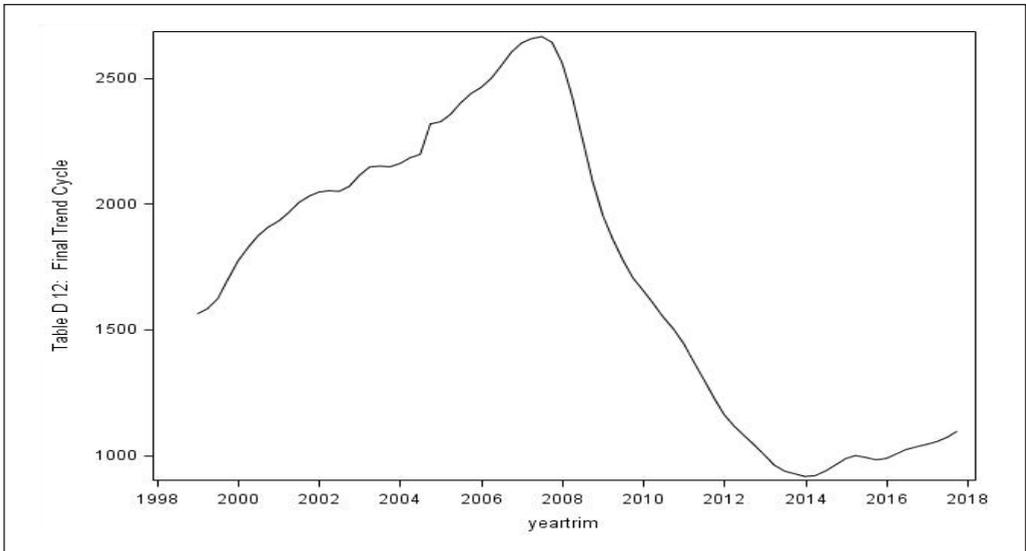
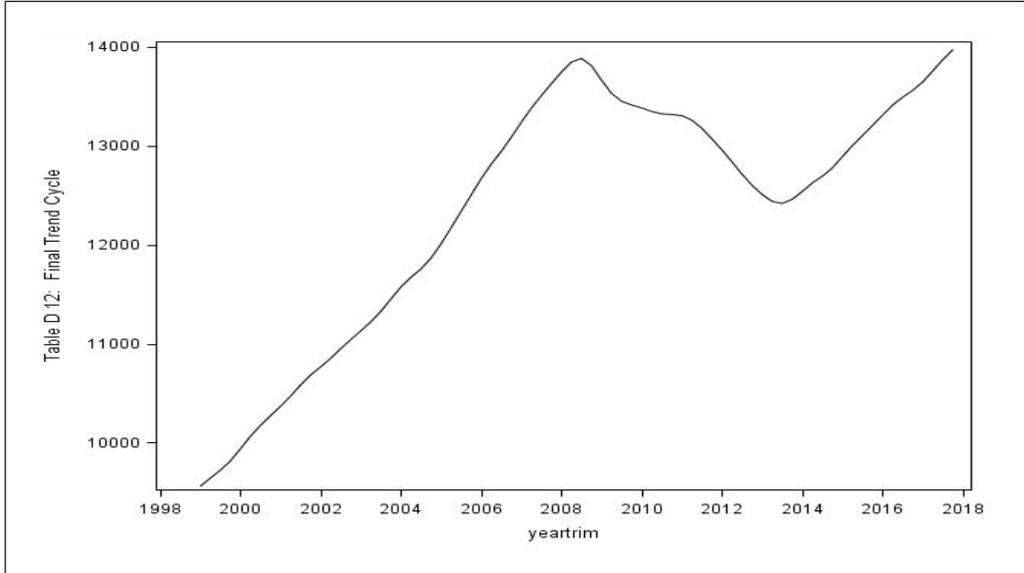


Gráfico 9. Componente ciclotendencia del Empleo equivalente a tiempo completo. Ocupados. Servicios. Miles de puestos de trabajo desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)



Como podemos observar en los tres gráficos anteriores, la señal cíclica de estos indicadores es heterogénea y depende del sector económico considerado. En todos los sectores se ha acusado el descenso en el primer trimestre de 2008, sin embargo mientras que en el sector industrial y el de la construcción la recuperación en el empleo se inicia lentamente a partir de 2014, en el sector servicios la recuperación se inicia con anterioridad y posteriormente se alcanzan los niveles anteriores a la crisis.

La señal cíclica de los indicadores de opinión que elabora la Unión Europea para España es similar entre los indicadores considerados. En todos ellos se revela la mala percepción de la situación durante la crisis pero se observan rápidos síntomas de recuperación a partir de 2009. Ejemplos de la componente de ciclotendencia de este tipo de indicadores de opinión se muestran en los gráficos 10, 11 y 12.

Gráfico 10. Componente de ciclotendencia de Confianza en la industria manufacturera en España desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

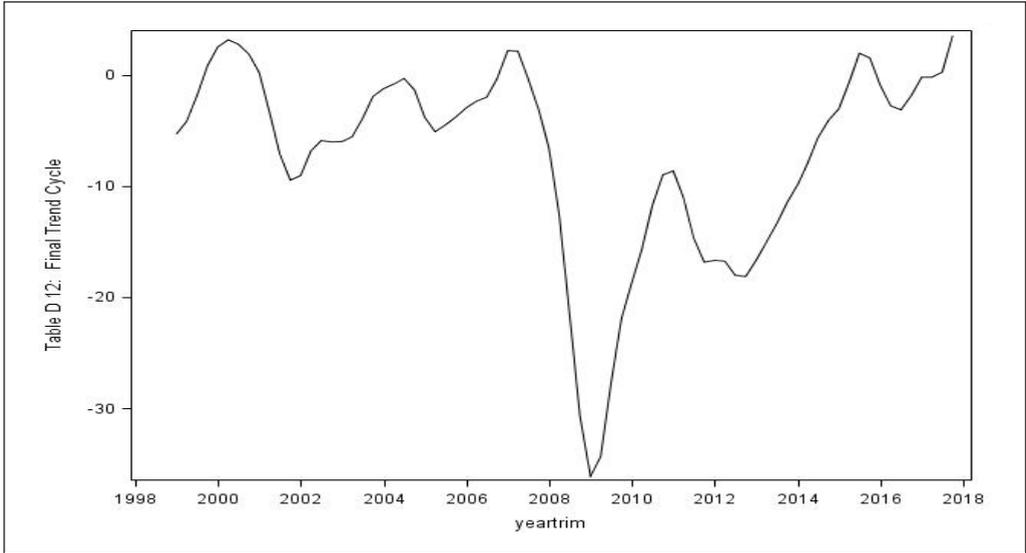


Gráfico 11. Componente de ciclotendencia de Confianza en el comercio minorista en España desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)

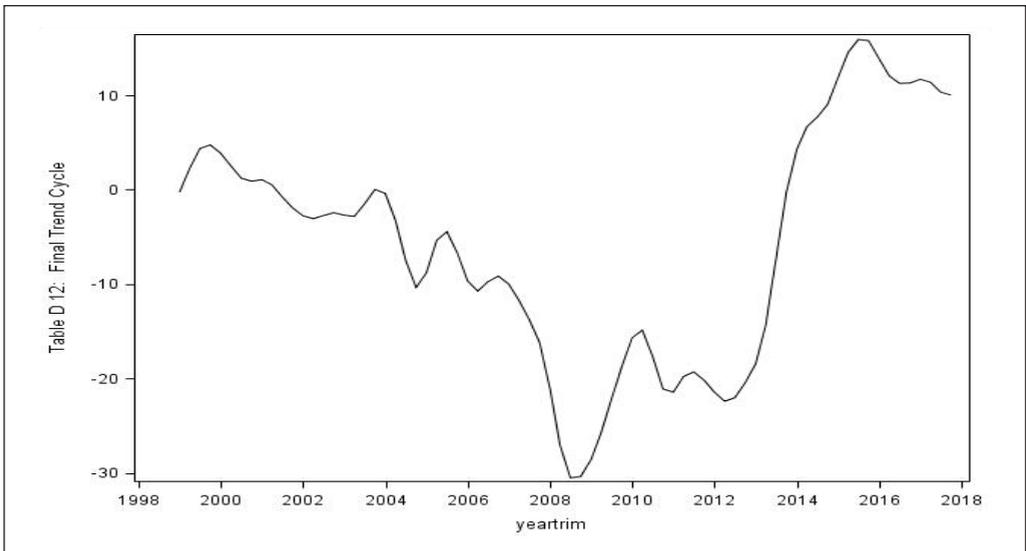
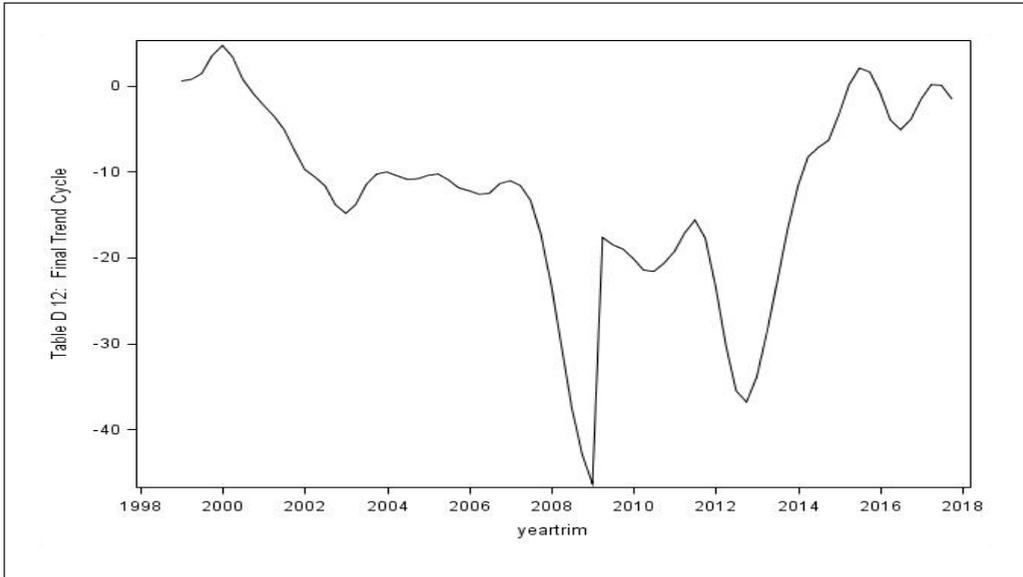


Gráfico 12. **Componente de ciclotendencia de Confianza del consumidor en España desde el primer trimestre de 1999 hasta el tercero de 2017. Método de desestacionalización X12 ARIMA (U.S. Census Bureau)**



3.2. Listado de indicadores seleccionados

Basándonos en las propiedades de los indicadores descritas en el epígrafe anterior, se han seleccionado un total de 43 indicadores parciales que se clasifican en cinco grandes bloques:

Tabla 1. **Indicadores que aproximan los beneficios empresariales**

Índice de cifra de negocios. General.
Índice de cifra de negocios. Industrias extractivas y manufactureras.
Índice de cifra de negocios. Suministro de energía eléctrica y agua, saneamiento y gestión de residuos.
Índice de cifra de negocios. Comercio.
Índice de cifra de negocios. Servicios no financieros de mercado.
Excedente de explotación neto. Sociedades no financieras.





Excedente de explotación neto. Instituciones financieras.

Central de Balances Trimestral. Resultado bruto de explotación.

Central de Balances Trimestral. Resultado ordinario neto.

El primer conjunto de indicadores incluye aquellas variables que mejor aproximan la base imponible del impuesto. Dentro de este grupo se encuentran el excedente de explotación neto de la contabilidad nacional (sociedades no financieras e instituciones financieras), el índice de cifra de negocios y el resultado bruto de explotación de la Central de Balances Trimestral.

Tabla 2. **Indicadores que aproximan la actividad empresarial (producción, ventas, utilización de la capacidad productiva, formación bruta de capital)**

Formación bruta de capital fijo. Maquinaria, bienes de equipo y cultivados. Precios corrientes.

Formación bruta de capital fijo. Equipos de transporte. Precios corrientes.

Formación bruta de capital fijo. Otra maquinaria y equipo. Precios corrientes.

Utilización de la capacidad productiva. General (%).

Utilización de la capacidad productiva. Sector Industrial (%).

Utilización de la capacidad productiva. Bienes de inversión (%).

Utilización de la capacidad productiva. Bienes de consumo (%).

Utilización de la capacidad productiva. Bienes intermedios (%).

Central de Balances Trimestral. Producción total.

Ventas interiores en grandes empresas.

Ventas totales en grandes empresas.

Índice de Producción Industrial. Índice general. Base 2010.

Índice de Producción Industrial. Bienes de equipo. Base 2010.

Índice de Producción Industrial. Bienes intermedios. Base 2010.

Índice de Producción Industrial. Energía. Base 2010.

El segundo conjunto de indicadores contiene información relativa al nivel de actividad de la empresa. Entre los indicadores de este bloque se incluyen aquellos que tradicionalmente se han utilizado para capturar la evolución de la actividad económica empresarial tales como los índices de producción, la utilización de la capacidad productiva e indicadores de inversión como la formación bruta de capital. También se incluyen en este bloque las ventas interiores y ventas totales en grandes empresas.

Tabla 3. **Indicadores del mercado laboral**

Remuneración bruta media. Grandes empresas.
Empleo equivalente a tiempo completo. Ocupados. Total sector industrial.
Empleo equivalente a tiempo completo. Ocupados. Total sector construcción.
Empleo equivalente a tiempo completo. Ocupados. Total sector servicios.

La distribución del empleo por sectores a partir de los datos de la contabilidad nacional y la remuneración bruta media de las grandes empresas a partir de los datos de la estadística de *Ventas, Empleo y Salarios en las Grandes Empresas* de la Agencia Tributaria forman el tercer bloque de indicadores del mercado laboral español. Indicadores tales como el número de afiliaciones a la Seguridad Social, el número de contratos registrados o el número de ocupados según la Encuesta de Población Activa (INE) han sido descartados en un análisis de sensibilidad como información redundante y sustituidos por los indicadores mencionados.

Tabla 4. **Sociedades mercantiles**

Número de sociedades mercantiles constituidas.
Número de sociedades mercantiles que incrementan capital.
Número de sociedades mercantiles que reducen capital.
Número de sociedades mercantiles disueltas.

Los indicadores en este bloque miden la demografía de las sociedades mercantiles, ofreciendo información mensual de las sociedades mercantiles creadas, disueltas y aquellas en las que se han realizado modificaciones de capital. La fuente de información es el Registro Mercantil Central, que recoge toda la información provincial sobre la inscripción de sociedades y empresarios, así como los actos mercantiles que determina la ley.

Tabla 5. **Indicadores opináticos o cualitativos**

Indicador de clima económico. España.
Indicador de confianza en la industria manufacturera. España.
Indicador de confianza del comercio minorista. España.
Indicador de confianza del consumidor. España.
Indicador compuesto adelantado de España.
Índice PMI. Industria manufacturera.
Índice PMI. Servicios.

El quinto bloque consiste en indicadores cualitativos relacionados con España, donde el indicador de sentimiento económico juega un papel fundamental por su pronta disponibilidad. Los indicadores adelantados se basan en el enfoque de la componente cíclica, donde esta componente de la serie temporal capta las oscilaciones debidas al ciclo económico y se identifican los puntos de giro a partir de las desviaciones respecto a la componente tendencial.

Finalmente hemos considerado algunos ratios financieros relativos a empresas como la rentabilidad del activo neto de la Central de Balances Trimestral o la diferencia entre rentabilidad y coste financiero de la misma fuente. De la misma forma, hemos incluido el indicador de importaciones de grandes empresas de la operación estadística Ventas, empleo y salarios en las grandes empresas. Teniendo en cuenta el peso de la actividad turística en el PIB, hemos considerado el indicador de llegadas de turistas extranjeros.

La combinación seleccionada de los 43 indicadores trimestrales ha sido la que mejores resultados ha conseguido de cara a la predicción del impuesto y la que conlleva una interpretación económica mejor. El conjunto final de indicadores seleccionado demostró ser robusto frente a ampliaciones del modelo en varias direcciones. Como análisis de sensibilidad del modelo, se han utilizado versiones desagregadas de las variables trimestrales ya incluidas en el modelo, así como indicadores alternativos en los bloques considerados.

Por ejemplo, se incluyeron en el modelo el número de afiliaciones a la Seguridad Social, el número de contratos registrados y el número de ocupados proporcionados por la EPA en lugar de los indicadores de ocupados a tiempo completo de la contabilidad nacional en el bloque 3 relativo al mercado de trabajo para ampliar el conjunto de variables consideradas. Sin embargo, los resultados fueron que los componentes principales resultantes no fueron mejores predictores en el modelo ARX que los del modelo anterior, y la ampliación del bloque del mercado laboral no resultó en una ganancia en precisión en el modelo final.

Se obtuvieron resultados similares con algunas variables del sector exterior, como las exportaciones de bienes y servicios o variables financieras como el crédito a las empresas.

4. Estrategia de estimación

Nuestro punto de partida ha sido el conjunto de indicadores trimestrales descritos en la sección anterior. El primer paso en la metodología consiste en realizar un análisis de componentes principales como técnica de reducción de la dimensión para el conjunto de 43 variables independientes. El siguiente paso consiste en utilizar las cuatro primeras componentes principales como variables de entrada al modelo ARX para estimar la recaudación del IS. A continuación se describen las técnicas utilizadas.

4.1. El análisis de componentes principales como técnica de reducción de la dimensión

El principal objetivo del análisis de componentes principales es encontrar el mínimo número de dimensiones que son capaces de explicar el mayor porcentaje de varianza contenida en el conjunto inicial de indicadores. Así, se trata de simplificar la información que proporciona la matriz de correlaciones del conjunto inicial de indicadores para hacerla más fácil de interpretar.

El análisis de componentes principales fue inicialmente publicado por Pearson (1901) y más tarde desarrollado por Hotelling (1933). La aplicación de las componentes principales se debate en Rao (1964), Cooley y Lohnes (1971) y Gnanadesikan *et al.* (1977). Se pueden encontrar tratamientos estadísticos excepcionales de los componentes principales en Kshirsagar (1972), Morrison (1976), y Mardia, Kent y Bibby (1979).

Dado un conjunto de datos con p variables numéricas podemos calcular hasta p componentes principales. Cada componente principal es una combinación lineal de variables originales con coeficiente de ponderación que coinciden con los autovalores de la matriz de correlaciones o de covarianzas.

Las componentes principales tienen las siguientes propiedades (Rao 1964; Kshirsagar 1972):

- Los autovectores son ortogonales, por lo que los componentes principales representan de forma conjunta direcciones perpendiculares a través del espacio de las variables originales.
- Los resultados del análisis de componentes principales (factores ortogonales) están incorrelados. Esta propiedad garantiza la ausencia de multicolinealidad cuando utilizamos esos factores como variables de entrada en un modelo de regresión.

- El primer componente principal tiene la mayor varianza de cualquier combinación lineal cuyos pesos estén normalizados de las variables observadas. El j -ésimo componente principal tiene la mayor varianza de cualquier combinación lineal normalizada ortogonal a los primeros componentes principales $j-1$.
- El último componente principal tiene la varianza más pequeña de cualquier combinación lineal de las variables originales.
- Las puntuaciones en los primeros componentes j principales tienen la mayor varianza generalizada posible de cualquier conjunto de combinaciones lineales de longitud uno de las variables originales.
- Los primeros j componentes principales proporcionan una solución de mínimos cuadrados para el modelo:

$$Y = XB + E$$

Donde:

- Y es la $n \times p$ matriz de las variables observadas centradas.
- X es la $n \times j$ matriz de puntuaciones de las primeras j componentes principales.
- B es la $j \times p$ matriz de autovectores.
- E es la $n \times p$ matriz de residuos.

El objetivo del análisis es minimizar la traza de $E'E$. Esto significa que las primeras j -ésimas componentes principales son los mejores predictores lineales de las variables originales entre todos los posibles conjuntos de j -variables, aunque cualquier transformación lineal no singular de los primeros componentes principales proporcionaría una predicción igualmente buena.

4.2. Modelos autorregresivos con regresores exógenos (ARX)

Supóngase que la serie temporal y_n de dimensión 1 consiste en p variables de salida

$$s_n = (s_n(1), \dots, s_n(p))^T$$

y q -variables de entrada:

$$r_n = (r_n(1), \dots, r_n(q))^T$$

de forma que

$$1 = p + q \text{ y } y_n = (s_n^T, r_n^T)^T$$

El modelo ARX con entrada y salida está dado por:

$$s_n = \sum_{j=1}^m a_j s_{n-j} + \sum_{j=1}^m b_j r_{n-j} + u_n$$

Donde a_j y b_j son matrices ($p \times p$) y ($p \times q$), y u_n es un ruido blanco p -dimensional con matriz de covarianzas $W_{r,m}$.

Nótese que un modelo ARX es una parte del modelo AR para una serie temporal 1-dimensional:

$$y_n = \sum_{j=1}^m A_j y_{n-j} + v_n$$

Con la relación

$$A_j = \begin{bmatrix} a_j & b_j \\ * & * \end{bmatrix}, v_n = \begin{bmatrix} u_n \\ * \end{bmatrix}, W_m = \begin{bmatrix} W_{r,m} & * \\ * & * \end{bmatrix}$$

El símbolo * indica que esta parte de la matriz no se usa en el modelo ARX. Esto significa que los parámetros del modelo ARX se obtienen como parte de un modelo AR multivariante para la serie temporal y_n .

5. Resultados de la estimación

5.1. Determinación de los regresores ortogonales

Como indicamos en el apartado anterior, el propósito de este análisis de componentes principales es calcular cuatro variables que mejor resuman los 43 indicadores parciales iniciales.

Tabla 6. Autovalores de la matriz de correlaciones

Eigenvalues of the Correlation Matrix: Total = 43 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	15.6770899	7.3451104	0.3646	0.3646



Eigenvalues of the Correlation Matrix: Total = 43 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
▶				
2	8.3319795	1.1592345	0.1938	0.5584
3	7.1727451	3.8259975	0.1668	0.7252
4	3.3467475	1.1631585	0.0778	0.8030
5	2.1835890	0.9791472	0.0508	0.8538
6	1.2044418	0.2511264	0.0280	0.8818

Fuente: Prior Communality Estimates: ONE

Los resultados de los autovalores del análisis de componentes principales se muestran en la tabla 6. En la primera columna se observan los autovalores que corresponden a cada componente principal. Cuanto mayor es el autovalor, mayor es la capacidad de «resumen» de la información que ostenta la componente principal asociada. En la última columna de la parte derecha de la tabla observamos el porcentaje de varianza explicada por el conjunto de componentes principales a medida que vamos agregando componentes.

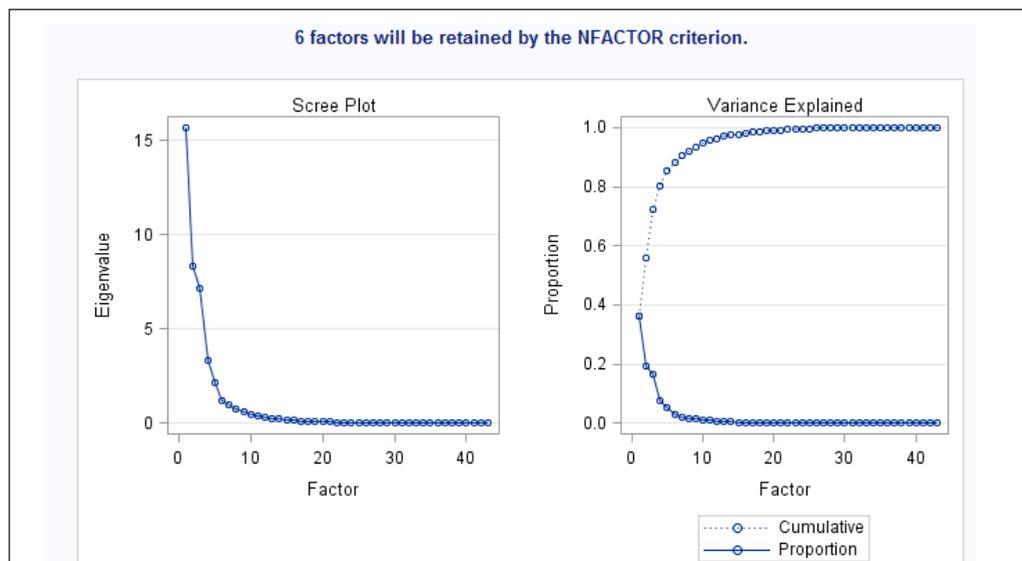
Así, si quisiéramos resumir la información proporcionada por el conjunto de 43 indicadores parciales mediante una sola componente principal (la primera componente) solo seríamos capaces de explicar el 36,46 % de la variabilidad del conjunto de indicadores. Si escogiéramos dos componentes principales para resumir el conjunto inicial de indicadores, podríamos explicar el 55,84 % de la variación del conjunto de indicadores parciales. Con tres componentes principales explicaríamos el 72,52 % de la variabilidad inicial y con las cuatro primeras componentes principales somos capaces de explicar el 80,30 % de la variación del conjunto de 43 indicadores, lo que constituye un buen resumen de la información.

Hemos calculado las componentes principales a partir de la matriz de correlaciones. El conjunto inicial de 43 indicadores muestra una gran correlación entre las variables validando la relevancia del análisis de componentes principales como primer paso en nuestra metodología para evitar multicolinealidad.

A pesar del hecho de que si siguiéramos la regla de Kaiser deberíamos seleccionar 6 componentes principales (son aquellas componentes cuyos autovalores asociados son superiores a 1), las dos últimas componentes principales no son regresores significativos en el modelo ARX posterior para predecir la recaudación del impuesto.

Teniendo en cuenta las cuatro primeras componentes principales que resultan regresores significativos en el modelo ARX, somos capaces de explicar más del 80 % de la varianza total del conjunto inicial de indicadores como muestra el gráfico de varianza total explicada (gráfico 13)

Gráfico 13. Gráfico de autovalores y de varianza inicial explicada



El gráfico de autovalores asociados a las primeras componentes principales (*Scree Plot*) (gráfico 13 izquierda) muestra, como ya hemos visto en la tabla, que el autovalor de la primera componente principal supera el valor 15 y los asociados a la segunda y tercera componente decrecen hasta 8,3 y 7,2. El gráfico de varianza inicial explicada (gráfico 13 derecha) indica que las primeras cuatro componentes principales resumen casi el 80 % de la variabilidad del conjunto inicial de indicadores, lo que valida que las cuatro primeras componentes principales constituyen un buen resumen del conjunto inicial de variables.

Los cuatro factores ortogonales obtenidos del análisis de componentes principales serán las variables de entrada exógenas en el modelo ARX.

5.2. Determinación del modelo ARX

La segunda parte de la metodología utiliza las cuatro componentes principales como variables de entrada (regresores) del modelo de función de transferencia.

Tabla 7. Estimación por mínimos cuadrados condicionales para la predicción de los ingresos tributarios del IS

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	Variable	Shift
AR1,1	-0.23990	0.13947	-1.72	0.0914	1	IS	0
AR1,2	-0.17253	0.14644	-1.18	0.2441	2	IS	0
NUM1	-1017231.9	699375.0	-1.45	0.1518	0	isfactor1	0
NUM2	-538530.0	200782.7	-2.68	0.0098	0	isfactor2	0
NUM3	1998445.7	333779.7	5.99	<.0001	0	isfactor3	0
NUM4	-534010.1	280707.5	-1.90	0.0627	0	isfactor4	0
NUM5	-5792249.5	950940.8	-6.09	<.0001	0	is2008q1	0

En esta tabla la primera columna nos indica los estimadores de los parámetros. La cuarta columna, Approx PR > |t|, nos indica el p-valor del contraste de significación individual de cada uno de los parámetros del modelo. Los parámetros más importantes en este modelo son las cuatro componentes principales que nos servirán de «guía» en la predicción, y que, a su vez, resumen la información proporcionada por los 43 indicadores parciales. Son las filas 3.º a la 6.º en la tabla. La última fila de la tabla representa una variable de tipo escalón, es decir, que vale 0 hasta el primer trimestre de 2008 y luego vale 1. Se introduce en el modelo para captar el cambio de nivel de la serie de recaudación a partir de la crisis financiera internacional.

Las dos primeras filas de la tabla 7 corresponden a dos parámetros autorregresivos del modelo, es decir, dependencias de la variable de recaudación en un trimestre t con respecto a los dos trimestres anteriores $t-1$, $t-2$. Los estimadores de ambos parámetros están alejados del valor 1 y resultan significativos al 75 %.

El modelo ARX utiliza las cuatro primeras componentes principales obtenidas en el subpartado anterior $ISFACTOR1_t, \dots, ISFACTOR4_t$ como regresores exógenos en el modelo.

De acuerdo con la tabla 7 todos los parámetros pueden considerarse estadísticamente significativos al 75 %. Nótese que el parámetro autorregresivo de segundo orden podría eliminarse de la ecuación dado que su p-valor en el contraste de significación individual es mayor que 0,20. Sin embargo, dado que el ajuste del modelo es muy bueno incluido este parámetro, se ha decidido trabajar con un nivel de significación del 75 %.

El modelo ARX puede escribirse como sigue:

$$(1 - B) y_t^4 = \frac{1}{(1 + 0.23990B + 0.17253B^2)} u_t - \frac{1}{(1 - B)} 5792249.51 I_{2008q1} - 1017231.9 ISFACTOR1_t - 538530 ISFACTOR2_t + 1998445.7 ISFACTOR3_t - 534010.1 ISFACTOR4_t$$

Donde:

- y_t son los ingresos tributarios por el IS.
- B es el operador retardo.
- u_t es la innovación.
- $ISFACTOR1_t, \dots, ISFACTOR4_t$ son las cuatro componentes principales.
- I_{2008q1} es el indicador dicotómico correspondiente al primer trimestre de 2008.
 - $I_{2008q1} = 1$ si $t \geq 2008q1$
 - $I_{2008q1} = 0$ si $t < 2008q1$

La diagnosis del modelo se completa contrastando si los residuos del mismo (aquella parte de la variación de la serie de recaudación tributaria que no somos capaces de explicar con el modelo) son ruido blanco y siguen una distribución normal.

El contraste de autocorrelación para los residuos del modelo se muestra en la tabla 8. La hipótesis nula es que no existe autocorrelación en los residuos hasta los retardos 6, 12, 18 y 24. La conclusión es que los residuos carecen de autocorrelación y que este resultado es independiente del número de retardos.

Tabla 8. Contraste de autocorrelación para los residuos

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.89	4	0.4217	-0.007	0.024	-0.029	0.140	-0.192	-0.026
12	8.82	10	0.5493	0.063	-0.074	-0.166	0.100	-0.066	-0.127
18	16.20	16	0.4392	-0.025	-0.023	-0.043	-0.149	-0.025	0.122

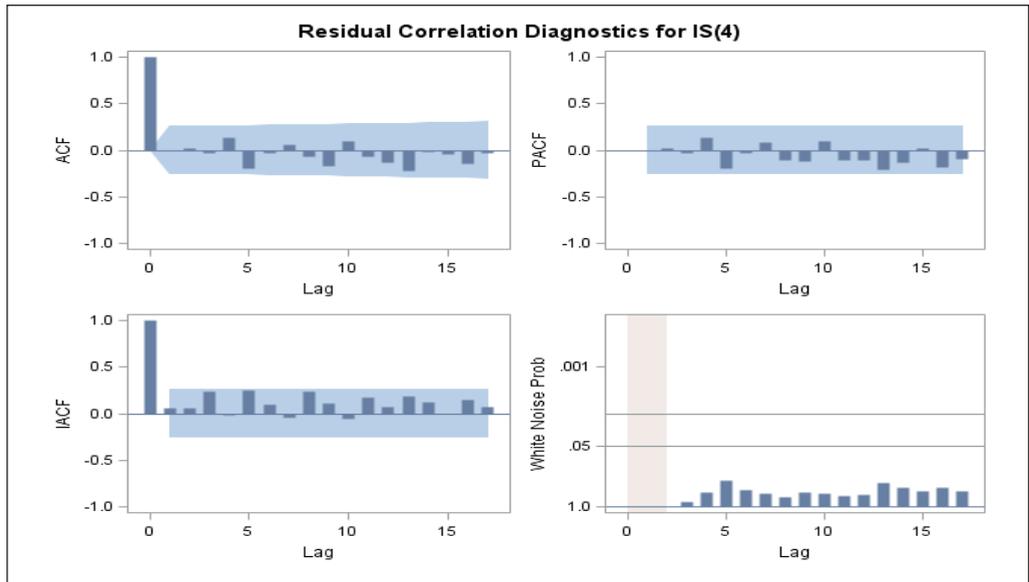


Autocorrelation Check of Residuals										
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations						
▶										
24	21.41	22	0.4953	0.146	-0.043	0.168	0.058	-0.031	-0.018	

Como se observa en la tabla 8 los residuos del modelo ARX no presentan autocorrelación en ninguno de los retardos especificados. El contraste falla a la hora de rechazar la hipótesis nula de no autocorrelación con altos niveles de significación, como se indica en la tercera columna de la tabla $pr > ChiSq$. ($p = 0.4217$ para los seis primeros retardos). Este resultado se muestra robusto ante cambios en el número de retardos.

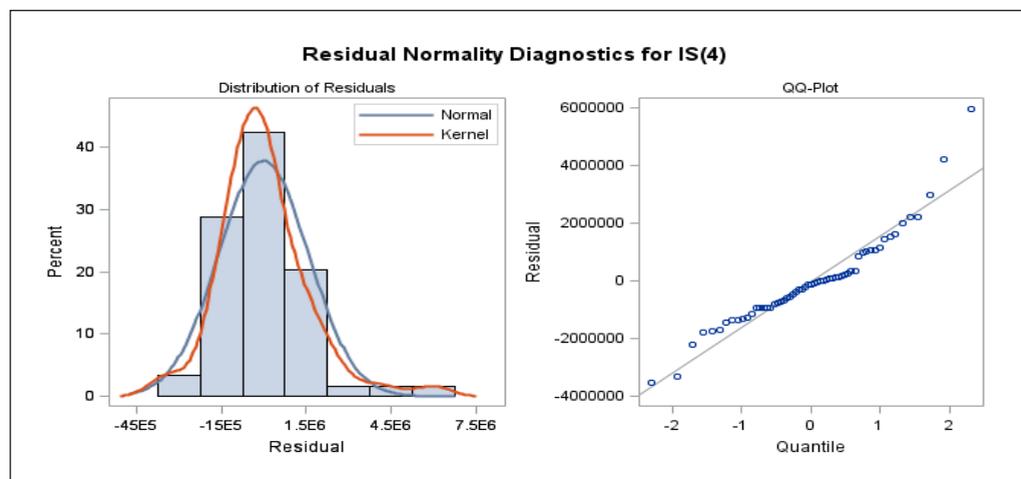
A continuación se muestra un gráfico con la estructura de autocorrelación de los residuos:

Gráfico 14. Panel de correlaciones para los residuos. La función de autocorrelación simple muestral (FASM) se muestra en la esquina superior derecha. En la esquina superior izquierda se muestra la función de autocorrelación parcial muestral (FAPM). El gráfico inferior derecho muestra la probabilidad de ruido blanco de los residuos en función del retardo



Las funciones de autocorrelación simple y parcial muestrales de los residuos se encuentran dentro de los intervalos de confianza al 95 % del cero en todos los retardos del residuo, por lo que podemos suponer que se anulan. La probabilidad de ruido blanco es muy alta.

Gráfico 15. Diagnóstico de normalidad residual



Si se observa el gráfico de cuantiles (QQ-plot) para comparar los cuantiles de la distribución del residuo con los cuantiles de la distribución normal teórica, se puede concluir que los valores se aproximan razonablemente a la diagonal. De forma análoga comparamos el histograma de los residuos del modelo (en rojo) con el de una distribución normal (en azul) obteniendo un parecido razonable.

El gráfico de la derecha indica que los residuos del modelo siguen pautas de normalidad.

5.3. Predicciones fuera de la muestra: ejercicios de *backtesting*

En esta subsección presentamos los resultados de los ejercicios de *backtesting*. Con el objeto de evaluar la capacidad predictiva del modelo, hemos dividido la muestra en dos partes:

- **Muestra de entrenamiento:** sirve para la estimación de los parámetros del modelo. Abarca desde el primer trimestre de 1995 hasta el último trimestre de 2010).
- **Muestra de validación:** desde el primer trimestre de 2011 hasta el tercer trimestre de 2017 (fecha de cierre de datos observados). Durante este periodo calculamos la predicción recursiva de un trimestre hacia delante con el modelo y observamos la desviación de las predicciones y los intervalos de confianza al 95 % con respecto a los valores de recaudación del impuesto.

La predicción recursiva un paso hacia adelante consiste en un método iterativo en el que solo calculamos la predicción del impuesto para el siguiente trimestre y, una vez calculada, la consideramos como un dato observado, la «añadimos» a la serie observada, y volvemos a calcular la predicción del modelo para el trimestre siguiente.

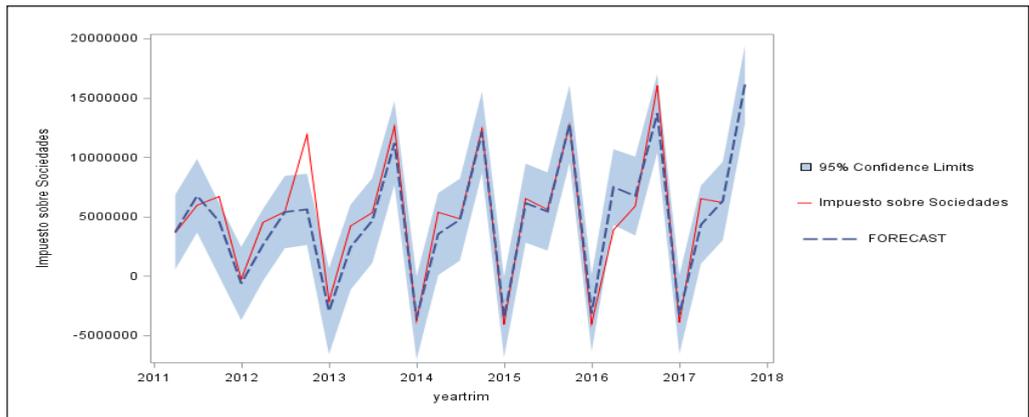
Inicialmente calculamos la predicción de un trimestre hacia adelante con respecto a nuestro último dato de recaudación oficial a la finalización de este trabajo (tercer trimestre de 2017). Así, la predicción para el último trimestre de 2017 según el modelo es:

Tabla 9. Estimación de la recaudación correspondiente al cuarto trimestre de 2017

Forecasts for variable IS				
Obs	Forecast	Std. Error	95% Confidence Limits	
76	16089459.0	1671018	12814234.1	19364593.8

A continuación presentamos los resultados de la predicción recursiva en el intervalo de validación del modelo:

Gráfico 16. Evaluación de la predicción de la recaudación tributaria por IS en la muestra de validación. Ejercicio de *backtesting* durante el periodo comprendido entre el primer trimestre de 2011 y el cuarto de 2017. Predicción recursiva un paso hacia adelante



El gráfico 16 muestra la capacidad predictiva del modelo, durante el intervalo temporal que abarca desde el primer trimestre de 2011 hasta el tercero de 2017, comparando los valores reales de recaudación publicados por la Agencia Tributaria (en rojo) con los predichos por el modelo (en azul discontinua). Los límites de confianza al 95 % se representan en el área sombreada en azul. Como se puede observar, si bien existen desviaciones trimestrales en los puntos de giro del impuesto hasta finales de 2012, a partir de 2013 la predicción del modelo se ajusta de forma razonable a los verdaderos valores que intenta predecir.

La siguiente tabla ilustra el rendimiento del modelo en el ejercicio de *backtesting* con los límites de confianza.

Tabla 10. Evaluación de la predicción de los ingresos tributarios por IS en el intervalo de validación entre el primer trimestre de 2011 y el cuarto de 2017. Predicción recursiva un paso hacia adelante

Predicción recursiva impuesto sobre sociedades (en miles de euros)						
Obs	yeartrim	FORECAST	REAL	L95	U95	standar_error_of_forecast
1	2011Q2	3677762.1	3608553	527152.1	6828372.1	1607484
2	2011Q3	6735306.6	5935536	3643510.7	9827102.6	1577476
3	2011Q4	4578017.1	6669956	1528796.4	7627237.7	1555753
4	2012Q1	-657803.5	-311228	-3742687.3	2427080.4	1573949
5	2012Q2	2600060.0	4498005	-434978.7	5635098.7	1548518
6	2012Q3	5370500.5	5391138	2321409.7	8419591.3	1555687
7	2012Q4	5584261.2	11857250	2583182.7	8585339.6	1531191
8	2013Q1	-2999917.5	-2156548	-6630132.7	630297.6	1852185
9	2013Q2	2411917.1	4190800	-1172041.0	5995875.2	1828584
10	2013Q3	4656806.7	5322248	1090417.4	8223196.0	1819620
11	2013Q4	11183054.6	12588694	7660623.0	14705486.2	1797192
12	2014Q1	-3584536.8	-3791664	-7085183.8	-83889.7	1786077
13	2014Q2	3509719.2	5342079	54990.5	6964447.9	1762649
14	2014Q3	4725213.7	4794293	1273488.7	8176938.7	1761117
15	2014Q4	121011897.8	12368351	8693532.9	15510262.7	1738994
16	2015Q1	-3518605.4	-4110293	-6886003.8	-151206.9	1718092
17	2015Q2	6116759.6	6496762	2785125.1	9448394.1	1699845
18	2015Q3	5407518.9	5568815	2112956.8	8702081.0	1680930
19	2015Q4	12779695.4	12693588	9522447.4	16036943.4	1661892
20	2016Q1	-3169953.7	-4054064	-6390899.6	50992.2	1643370
21	2016Q2	7465957.6	3849722	4270306.1	10661609.1	1630464
22	2016Q3	6706368.7	5871295	3386339.1	10026398.2	1693924
23	2016Q4	13677503.9	16011373	10384801.1	16970206.8	1679981
24	2017Q1	-3277260.0	-3936065	-6594592.6	40072.7	1692548
25	2017Q2	4301316.8	6487877	1012610.0	7590023.5	1677942
26	2017Q3	6303363.2	6152987	2996517.9	9610208.5	1687197
27	2017Q4	16089459.0	-	12814324.1	19364593.8	1671018

6. Conclusiones

Como se ha mencionado en la introducción de este trabajo, el objetivo final de esta investigación es proponer una metodología que combine con éxito indicadores que abordan la actividad empresarial, los beneficios, el mercado laboral, las sociedades mercantiles y los indicadores cualitativos y opináticos para generar un modelo de previsión a corto plazo para los ingresos tributarios por el IS.

Debido a la estructura del impuesto, el IS es uno de los impuestos más complicados de predecir en la mayor parte de los países y la precisión en la predicción de este impuesto es particularmente baja comparada con la de otros impuestos como el IVA o el IRPF.

Desde nuestra experiencia podemos afirmar que, en lo relativo a la predicción del IS, la precisión depende fuertemente del tiempo transcurrido desde el último dato publicado y de la selección de los indicadores introducidos en el modelo.

A diferencia de los métodos de predicción que se basan únicamente en el excedente de explotación (bruto o neto) de la contabilidad nacional como variable que «guía» la predicción, hemos podido aprovechar las ventajas derivadas del uso de un conjunto muy amplio de indicadores. Para evitar los problemas de multicolinealidad que surgen cuando se combinan variables de naturaleza similar hemos utilizado como paso inicial en la metodología el análisis de componentes principales como método de reducción de la dimensión.

Los cuatro componentes obtenidos de este análisis han servido como variables de entrada en la función de transferencia (modelo ARX) que, junto con una intervención relativa al cambio en la recaudación del impuesto durante la crisis del 2008 (cambio de nivel), han resultado muy significativos para explicar y predecir el comportamiento de la recaudación del impuesto. Los resultados del ejercicio de *backtesting* durante el periodo comprendido entre el primer trimestre de 2011 y el tercero de 2017 ilustran la utilidad de este modelo como un modelo a corto plazo para predecir los ingresos tributarios por el IS.

Referencias bibliográficas

- Auerbach, A. J. (1999). On the performance and use of government revenue forecasts. *National Tax Journal*, 767-782.
- Bayer, O. (2015). Relevance of Input Data Time Series for Tax Revenue Forecasting. *Procedia Economics and Finance*, 25, 518-529.
- Bischoff, I. y Gohout, W. (2010). The political economy of tax projections. *International Tax and Public Finance*, 17(2), 133-150.
- Boylan, J. E.; Syntetos, A. A. y Karakostas, G. C. (2008). Classification for forecasting and stock control: a case study. *Journal of the operational research society*, 59(4), 473-481.
- Bretschneider, S. I.; Gorr, W. L.; Grizzle, G. y Klay, E. (1989). Political and organizational influences on the accuracy of forecasting state government revenues. *International Journal of Forecasting*, 5(3), 307-319.

- Buettner, T. y Kauder, B. (2010). Revenue forecasting practices: differences across countries and consequences for forecasting performance. *Fiscal Studies*, 31(3), 313-340.
- Cassidy, G.; Kamlet, M. S. y Nagin, D. S. (1989). An empirical examination of bias in revenue forecasts by state governments. *International Journal of Forecasting*, 5(3), 321-331.
- Chatagny, F. (2015). Incentive effects of fiscal rules on the finance minister's behavior: Evidence from revenue projections in Swiss Cantons. *European Journal of Political Economy*, 39, 184-200.
- Cooley, W. W. y Lohnes, P. R. (1971). *Multivariate data analysis*. J. Wiley.
- Danninger, M. S. (2005). *Revenue Forecasts as Performance Targets (EPub)* (No. 5-14). International Monetary Fund.
- Feenberg, D. R.; Gentry, W. M.; Gilroy, D. y Rosen, H. S. (1988). *Testing the rationality of state revenue forecasts*.
- Gnanadesikan, R.; Kettenring, J. R. y Landwehr, J. M. (1977). Interpreting and assessing the results of cluster analyses. *Bulletin of the International Statistical Institute*, 47(2), 451-463.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 417.
- Jochimsen, B. y Lehmann, R. (2017). On the political economy of national tax revenue forecasts: evidence from OECD countries. *Public Choice*, 170(3-4), 211-230.
- Keene, M. y Thomson, P. (2007). *An analysis of tax revenue forecast errors* (No. 07/02). New Zealand Treasury.
- Koester, G. B. y Priesmeier, C. (2017). Revenue elasticities in euro area countries. An analysis of long-run and short-run dynamics. *ECB working paper No. 1989*.
- Kshirsagar, A. (1972). Likelihood ratio tests; optimality of principal components. *Multivariate analysis*, Marcel Dekker Inc. New-York.
- Kyobe, A. y Danninger, M. S. (2005). *Revenue forecasting: how is it done? Results from a survey of low-income countries* (No. 5-24). International Monetary Fund.
- Leal, T.; Pérez, J. J.; Tujula, M. y Vidal, J. P. (2008). Fiscal forecasting: lessons from the literature and challenges. *Fiscal Studies*, 29(3), 347-386.
- Mardia, K. V.; Kent, J. T. y Bibby, J. M. (1979). *Multivariate analysis, Probability and Mathematical Statistics: A Series of Monographs and Textbooks*. Academic Press, Harcourt Brace Jovanovich Publishers.
- Mikesell, J. L. y Ross, J. M. (2014). State revenue forecasts and political acceptance: The value of consensus forecasting in the budget process. *Public Administration Review*, 74(2), 188-203.
- Morrison, D. G. (1976). Measurement problems in cluster analysis. *Management science*, 13(12), B-775.
- Paleologou, S. M. (2005). Political manoeuvres as sources of measurement errors in forecasts. *Journal of Forecasting*, 24(5), 311-324.
- Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2), 559.
- Pike, T. y Savage, D. (1998). Forecasting the public finances in the Treasury. *Fiscal Studies*, 19(1), 49-62.
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329-358.
- Wolswijk, G. (2007). Short-and long-run tax elasticities: the case of the Netherlands. *ECB working paper No. 763*.