



Universidade de Vigo

Trabajo Fin de Máster

Modelización de variables de evolución del negocio y de tipo macroeconómico en ABANCA

Juan Anido Rodríguez

Máster en Técnicas Estadísticas

Curso 2014-2015

Propuesta de Trabajo Fin de Máster

Título en galego: Modelización de variables de evolución do negocio e de tipo macroeconómico en ABANCA
Título en español: Modelización de variables de evolución del negocio y de tipo macroeconómico en ABANCA
English title: Modelling of the business outlook and the macroeconomic variables at ABANCA
Modalidad: Modalidad B
Autor/a: Juan Anido Rodríguez, Universidad de Santiago de Compostela
Director/a: Wenceslao González Manteiga, Universidad de Santiago de Compostela; Manuel Febrero Bande, Universidad de Santiago de Compostela
Tutor/a: Belén María Fernández de Castro, ABANCA - Planificación y Estudios
Breve resumen del trabajo: Se desarrollará teoría referente a series de tiempo en una y varias dimensiones con el fin de aplicar los conocimientos mostrados a datos reales referentes al comportamiento del entorno económico español.

Don Wenceslao González Manteiga, Catedrático de la Universidad de Santiago de Compostela, don Manuel Febrero Bande, Catedrático de la Universidad de Santiago de Compostela y doña Belén María Fernández de Castro, Especialista de ABANCA - Planificación y Estudios, informan que el Trabajo Fin de Máster titulado

**Modelización de variables de evolución del negocio y de tipo macroeconómico en
ABANCA**

fue realizado bajo su dirección por don Juan Anido Rodríguez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 7 de Julio de 2015.

Agradecimientos

En primer lugar a mi familia, en particular a mis padres por ser gracias a ellos la persona que soy en estos momentos. También a mis compañeros y amigos pues han sido un gran apoyo en todo este tiempo.

En el aspecto académico a los profesores Wenceslao González y Manuel Febrero como mis directores del proyecto, ya que con sus indicaciones he aprendido nuevos conocimientos que, sin duda, el día de mañana me serán de gran utilidad en el aspecto profesional. Por último, pero no menos importante, agradecerle a la entidad ABANCA que me permitiese introducirme en el área financiera al trabajar con ellos estos nueve meses de duración del proyecto. En especial a Belén Fernández, mi tutora, la cual me ha ayudado a una perfecta adhesión al equipo en el que he tomado parte y me ha aportado el punto de vista económico que en la empresa privada es considerado.

Índice general

Resumen	XI
Prefacio	XIII
I Aspectos metodológicos	1
1. Análisis Univariante de Series de Tiempo	3
1.1. Conceptos previos	3
1.2. Series estacionarias	5
1.3. Series no estacionarias	7
1.4. Estimación, Diagnósis y Predicción de modelos	11
2. Análisis Vectorial de Series de Tiempo	15
2.1. Modelo <i>VAR</i>	15
2.1.1. Estimación y especificación	17
2.2. Modelos <i>VECM</i>	19
2.2.1. Estimación y especificación	21
2.3. Diagnósis y predicción de los modelos	23
3. Regresión con Series Temporales	29
3.1. Modelo dinámico con retardos	29
3.2. <i>Generalized Least Squares</i>	30
3.2.1. Estimación <i>GLS</i> en base a los datos	32

II	Aplicación a datos reales	35
4.	Comentarios previos	37
4.1.	Motivación financiera y bancaria	37
4.2.	Referencias de las bases de datos	39
5.	Aplicación de la metodología	43
5.1.	Creación de los modelos econométricos	43
5.1.1.	Selección y transformación de variables	44
5.1.2.	Modelos en base a la estimación por <i>GLS</i>	45
5.2.	Modelización de series de tiempo	50
5.2.1.	Análisis univariante	50
5.2.2.	Análisis vectorial	54
5.3.	Predicción a través de los modelos econométricos	60
A.	Teorema de la descomposición de Wold	67
B.	Utilización del <i>software</i> estadístico R	69
C.	Implementación con IBM SPSS Modeler v16	87
	Bibliografía	91

Resumen

Resumen en español

El departamento de Planificación y Estudios perteneciente a la entidad bancaria ABANCA está trabajando en el desarrollo de modelos que relacionen variables de evolución del negocio, como puedan ser el saldo de depósitos, el saldo de crédito o la formalización del consumo o de hipotecas, con indicadores macroeconómicos o de entorno, tales como el paro registrado, afiliados a la Seguridad Social, consumo energético, matriculaciones de turismo, demanda nacional o el PIB (Producto Interior Bruto), con el fin de predecir el comportamiento del entorno económico español. El principal objetivo de estas modelizaciones es poder plantear distintos escenarios macroeconómicos en los que evaluar la evolución del entorno económico, tal y como requieren las normativas dictadas por el Banco Central Europeo (BCE). Además, las previsiones obtenidas se utilizarán en la fijación de los objetivos anuales del banco. Para ello están siendo desarrollados modelos de econometría basados en técnicas estadísticas, en particular en teoría de series temporales tanto en una como en varias dimensiones.

English abstract

The Planning and Studies Department belonging to the bank ABANCA is working on model development to relate business outlook variables, such as deposit balances, credit balances or consumption and mortgage lending, with macroeconomic indicators, such as recorded unemployment, Social Security affiliates, energy consumption, total car registrations, domestic demand or GDP (Gross Domestic Product), in order to predict the behaviour of the Spanish economy. The main aim is to use the results from these models to propose different macroeconomic scenarios in which the economic environment will be assessed, as it is required by the regulations issued by the European Central Bank (ECB). To achieve this, econometric models, which are based on statistical techniques, are being developed, in particular on univariate and multivariate time series theory.

Prefacio

Este trabajo se ha llevado a cabo en relación a la demanda por parte del departamento de Planificación y Estudios de la entidad bancaria ABANCA de un desarrollo en profundidad en base a una colección de modelos estadísticos que relacionasen variables de evolución del negocio (variación de saldo de depósitos, saldo de crédito o formalización de hipotecas) con variables macroeconómicas o de entorno. Con esta propuesta es evidente la asociación con modelos econométricos y en particular modelos de serie temporales, por lo que a continuación introducimos ciertas notas sobre lo que la econometría representa por no ser ésta una materia obligada conocimiento para matemáticos y estadísticos.

De entre todas las definiciones que se pueden encontrar en la literatura acerca de lo que se entiende por el término econometría, las que se han recogido en esta introducción aportan las principales ideas de su utilidad en la conexión entre ciencia y economía. La primera a la que hacemos referencia fue aportada en el primer número de la revista *Econometrica* creada por la prestigiosa *Econometric Society* en 1933, en la cual Ragnar Frisch, premio Nobel de Economía en 1961 y pionero en la introducción de la econometría en planes de política económica, comentaba lo siguiente.

”Cada uno de estos tres puntos de vista, la estadística, la teoría económica y las matemáticas, es necesario, pero por sí mismo no suficiente para una comprensión real de las relaciones cuantitativas de la vida económica moderna. Es la unión de los tres aspectos lo que constituye una herramienta de análisis potente. Es la unión lo que constituye la econometría”.

En ella podemos comprobar cómo estadística, matemáticas y economía deben aunarse en un solo ente para poder alcanzar el concepto de econometría, relación que se entenderá con los posteriores comentarios que realicemos acerca de esta rama de la economía. Pero antes citamos a otra importante figura de este campo, Stefan Valavanis, quien a finales de los cincuenta daba la siguiente definición.

”El objetivo de la econometría es expresar las teorías económicas bajo una forma matemática a fin de verificarlas por métodos estadísticos y medir el impacto de una variable sobre otra, así como predecir acontecimientos futuros y dar consejos de política económica ante resultados deseables”.

Podemos comprobar aquí, como las herramientas derivadas de este área no solo explicarán la economía en asociación con matemáticas y estadística, sino que también intentarán predecir el comportamiento de ésta a futuro y de esta forma permitirá tomar diferentes políticas o acciones a los organismos correspondientes. Con tal idea, podemos comentar más claramente cuáles son los objetivos de la econometría, estableciendo tres metas principales.

- El conocimiento de la economía real. Permitiendo estimar las magnitudes económicas situadas en determinado marco espacio-temporal, como por ejemplo en España la última década.
- Simulación de situaciones económicas. Para así, determinar efectos de ciertos escenarios en la economía de estudio, por ejemplo, cuantificar los efectos de una crisis financiera como la acaecida a finales de 2008 y comprobar cómo una entidad bancaria respondería a tal situación.
- Predicción. Cuando hacemos predicciones tratamos de reducir nuestra incertidumbre sobre

el futuro de la economía, pero no es una tarea sencilla, pues las predicciones sólo son satisfactorias cuando no hay cambios drásticos en la economía. Sería muy conveniente predecir también estos cambios bruscos, pero las predicciones tanto con métodos econométricos como con métodos alternativos, por lo general no son muy buenas en esos casos.

Con lo que queda clara la utilidad de esta técnica y se vislumbra como se realizaría la conjunción entre ciencia y teoría económica. Pero para todo esto es necesaria la generación de un modelo estadístico con buenas propiedades que explique de forma correcta el comportamiento de la economía en el escenario considerado. Con este fin, debemos dar, aunque sea de forma esquemática y sin profundizar en gran medida, cuáles serían los pasos a seguir para tal creación, etapas que son clásicas en la elaboración de cualquier modelo estadístico. Éstos quedarán meridianamente claros en la aplicación práctica ulterior.

1. La especificación constituye la primera etapa de la elaboración de un modelo econométrico. En ella se consideran dos elementos: el modelo económico-econométrico y los datos. El primer elemento se especifica al comienzo del proceso mediante la utilización de la teoría económica y da una idea de cómo considerar las técnicas estadísticas aportando la expresión con la cual explicamos el suceso en particular que se está estudiando. El último elemento, los datos, los consideramos al acabar este esquema para poder ampliar posteriormente la idea que aquí se cita.
2. Con la estimación se obtienen los valores numéricos de los coeficientes del modelo econométrico de la primera etapa. Para completar esta etapa se debe disponer de un conjunto de observaciones de todas las variables que aparecen en el modelo especificado, y, por otro lado, es necesario seleccionar el método de estimación apropiado, teniendo en cuenta las implicaciones de esta elección en las propiedades estadísticas de los estimadores de los coeficientes.
3. Por último en la etapa de validación. En esta etapa se evalúa si lo obtenido en la etapa anterior es aceptable, tanto por la teoría económica como desde el punto de vista estadístico. Se analiza, por un lado, si las estimaciones de los parámetros del modelo tienen los signos y magnitudes esperados, es decir, si satisfacen las limitaciones establecidas por la teoría económica. Y desde el punto de vista estadístico, se llevan a cabo contrastes sobre la significación de los parámetros del modelo en los que se utilizan los estadísticos hallados en la etapa de especificación.

Una vez que el modelo econométrico es hallado y se ha comprobado su validez en la situación planteada, se emplea para una etapa adicional de predicción. En ella se intentará dar previsiones sobre el comportamiento a futuro de una cierta variable de estudio para con ellas explicar cierto suceso en la economía y poder tomar de forma adecuada acciones sobre cierto organismo, como se comentaba con anterioridad. Como decíamos, debemos tener en cuenta los datos para la realización de un modelo econométrico, siendo éstos no experimentales, es decir, recogidos mediante la observación del mundo real de una manera pasiva. También es de señalar la diversidad que éstos pueden tener, con lo que aportamos en estos momentos la consideración de tres naturalezas diferentes.

- *Series temporales.* En ellas los datos son observaciones de una variable a lo largo del tiempo, como por ejemplo la cantidad de turistas matriculados o las exportaciones de un cierto estado. Por ello la ordenación de éstos es relevante, siendo errónea la suposición de independencia a lo largo del tiempo, ya que la mayoría de las series económicas se relacionan con su historia reciente. La mayoría de los datos económicos sólo están disponibles para frecuencias bajas, como trimestres o meses, siendo la excepción los datos financieros, donde se dispone de datos para frecuencias más elevadas como días u horas, de forma que los tamaños muestrales podrían ser muy grandes.
- *Datos de corte transversal.* En los datos de corte transversal se dispone de una observación por individuo y se refieren a un punto determinado en el tiempo, los individuos encuestados son personas, hogares, empresas u otros agentes económicos. En éstos, las observaciones deben

ser obtenidas mediante un muestreo aleatorio, lo que implica que las observaciones sean independientes entre sí, y con ello el orden de las observaciones no importa para el análisis econométrico.

- *Datos de panel.* También llamados datos longitudinales combinan elementos de datos de corte transversal y de series temporales, consistiendo en individuos encuestados repetidamente a lo largo del tiempo. En la modelización se adopta generalmente el supuesto de que los individuos son independientes entre sí, pero para un individuo dado, las observaciones a lo largo del tiempo son mutuamente dependientes. Por lo tanto, el orden dentro de un corte transversal de un conjunto de datos panel no importa, pero el orden interno de un dato en particular sí que es relevante.

De entre estos tres tipos de datos que en los modelos econométricos se pueden considerar, los que tendremos en cuenta en el presente trabajo serán las series temporales y sus técnicas estadísticas particulares. Una vez que queda claro lo que la econometría representa, solo queda presentar el esquema en el que se estructura el proyecto para introducirnos en él. El contenido que en las páginas de este trabajo se recoge, será dividido en dos partes bien diferenciadas, comenzando por un desarrollo teórico de métodos asociados a series temporales (Parte I) y continuando con una aplicación a datos económicos reales de la metodología mostrada (Parte II).

En la primera de las partes tomaremos literatura clásica de series de tiempo como pueden ser [Cowpervait, P. *et al.* 2009], [Hamilton, J. D. 1994] o [Peña, D. 2005], entre otros, para exponer las técnicas consideradas en un estudio de forma univariante sobre ciertas variables e introducir los modelos Box-Jenkins o *ARIMA* (Capítulo 1), contenido que podría verse en un curso académico básico sobre series de tiempo, y a continuación se mostrarán análogas técnicas para el caso que se tengan en cuenta relaciones internas entre las correspondientes variables, generando de este modo un estudio multivariante o vectorial, con modelos menos habituales en cursos introductorios como el *VAR* o el *VECM* (Capítulo 2), con referencias esenciales como [Pfaff, B. 2008a] o [Shumway, R. *et al.* 2011]. Por último en este apartado expondremos como realizar modelos estadísticos dinámicos que contengan cierta estructura de correlación en los residuos con la técnica *GLS* (Capítulo 3), donde se emplean referencias como [Cryer, J. *et al.* 2008], [Hamilton, J. D. 1994] o [Kuan, C. 2004].

Una vez que comentemos la metodología que acabamos de citar, pasaremos a aplicarla a ciertas variables, en particular, éstas serán variables de negocio e indicadores macroeconómicos referentes a la economía de España en los últimos diez o quince años. Este estudio tendrá como objetivo hacer previsiones del comportamiento de tal economía en diversos escenarios debido a lo que posteriormente se expondrá en la motivación de tal aplicación (Capítulo 4). Tras esta justificación del trabajo práctico, se emplearán las técnicas introducidas y se obtendrán los resultados que se comentarán al final (Capítulo 5).

Parte I

Aspectos metodológicos

Capítulo 1

Análisis Univariante de Series de Tiempo

Como se ha dicho en la introducción del trabajo, las series de tiempo pueden ser analizadas de un modo univariante, sin tener en cuenta otros procesos, o de forma multivariante, donde sí que se reflejaría una estimación conjunta, siendo esta primera la que desarrollaremos en el presente capítulo. Así, dado un conjunto de K series de tiempo $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_K\}$ seleccionamos la correspondiente al índice j para trabajar con ella, denotada por $\mathbf{y}_j = \{y_{jt}, t \in \mathcal{I}\}$ la identificamos con \mathbf{y} por comodidad en la notación. En base a ella, comenzaremos mostrando algunos conceptos previos que afianzarán la definición de serie temporal y nos darán algunas propiedades básicas que serán empleadas en los próximos párrafos y capítulos (Sección 1.1).

El paso siguiente será introducir ciertos modelos que puedan explicar los procesos que consideremos de forma univariante. Primero suponiendo series estacionarias, donde introduciremos los llamados modelos autorregresivos de medias móviles *ARMA* (Sección 1.2), para a continuación poder plantear la existencia de ciertas alteraciones de tal estacionariedad en nuestros procesos como son la presencia de tendencia, estacionalidad o heterocedasticidad, definiendo en este punto los modelos *ARIMA* (Sección 1.3). En este mismo apartado se explicará cómo poder eliminar éstas y poder llegar al ya citado *ARMA*. Para acabar, mostraremos los contraste que se han de realizar sobre el ajuste correspondiente, así como una predicción en base a dicha estimación (Sección 1.4).

Apuntar que en lo que sigue, y en próximos capítulos, tomaremos un proceso o serie temporal definido como $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$, siendo el intervalo de tiempo \mathcal{I} el asociado a $[1, T]$ con cierto instante final T . En ocasiones explicitaremos tal intervalo, pero no será lo habitual, ya que en ocasiones no será necesario un T específico o simplemente se supondrá un intervalo \mathcal{I} sin mayor pérdida de generalidad. Señalar también que los párrafos que a continuación se desarrollan tomarán ideas de [Hamilton, J. D. 1994], [Peña, D. 2005] y [Shumway, R. *et al.* 2011] para toda la base teórica y [Box, G. *et al.* 2008] para algunas particularidades. No apuntaremos estas referencias cada vez que se empleen, sino que son citadas aquí para no repetirnos demasiado.

1.1. Conceptos previos

Las definiciones y conceptos mostrados en esta sección se harán en un sentido de una dimensión por simplicidad, ya que la generalización a más dimensiones se realiza de forma natural y no es necesaria, por tanto, su especificación. Comenzamos por comentar la idea de lo que se entiende por un proceso estocástico y como a partir de él se define una serie temporal. Dicho proceso puede ser descrito como el conjunto

$$\mathbf{y} = \{y(s, t), s \in \mathcal{S}, t \in \mathcal{I}\}, \quad (1.1)$$

siendo \mathcal{S} el llamado espacio muestral e \mathcal{I} el correspondiente intervalo de tiempo. Entonces, partiendo de él, se tiene que fijado un $t^* \in \mathcal{I}$ el subconjunto resultante $\{y(s, t^*), s \in \mathcal{S}\}$ es una variable aleatoria sobre el conjunto \mathcal{S} . Y por otro lado, si fijamos un $s^* \in \mathcal{S}$ el resultante subconjunto $\{y(s^*, t), t \in \mathcal{I}\}$ es llamado una realización del proceso original respecto al intervalo de tiempo \mathcal{I} . En particular, lo único que nosotros podemos observar de un proceso estocástico dado por (1.1), no son otra cosa que las propias realizaciones, éstas son conocidas como el proceso de generación de datos (*DGP*). Señalar que en la notación podremos obviar el punto fijado $s^* \in \mathcal{S}$ y denotar las realizaciones simplemente como $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$. Entonces, con lo comentado, renombramos la idea de realización introducida con el concepto de series de tiempo. Con lo que se considerará

$$\mathbf{y} = \{y_t, t \in \mathcal{I}\}, \quad (1.2)$$

como la notación general para las series temporales que ocuparán nuestro estudio en todo el trabajo que sigue. Apuntar que se abusará del lenguaje en múltiples ocasiones ya que se utilizará indistintamente el nombre de serie de tiempo y de proceso, no existirá confusión pues siempre estaremos en las condiciones de la expresión (1.2).

Una vez conocidos los elementos con lo que trabajar, seguimos con la introducción de tres definiciones en referencia a la capital propiedad de estacionariedad de una serie de tiempo. Estos conceptos serán de suma importancia en la teoría que se desarrolle, en particular el primero ya que será una de las condiciones a verificar para un correcto análisis temporal. El segundo, veremos que es un concepto más fuerte que el primero, pero menos utilizado por no ser trivial de calcular en la práctica, y el siguiente será empleado a la hora del desarrollo de un modelo dinámico en la primera parte del tema de regresión.

Definición 1.1. Una serie temporal $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ se dirá que es estacionaria de forma débil si

$$\mathbb{E}(y_t) = \mu < \infty \quad \text{y} \quad \mathbb{E}((y_t - \mu)(y_s - \mu)) = \gamma_{t-s} \quad \text{para todo} \quad t > s \in \mathcal{I}.$$

También es denominado proceso estacionario de segundo orden o en covarianza por basarse en la finitud de sus dos primeros momentos.

Definición 1.2. Una serie temporal $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ se dirá que es estacionaria de forma estricta si el comportamiento en probabilidad de una colección de valores $\mathbf{y} = \{y_t, t \in \mathcal{I}_1\}$ con $\mathcal{I}_1 \subseteq \mathcal{I}$ es idéntica al obtenido con $\mathbf{y} = \{y_{t+h}, t \in \mathcal{I}_1, h \in \mathbb{Z}\}$, esto es,

$$F(\{y_t, t \in \mathcal{I}_1\}) = F(\{y_t, t \in \mathcal{I}_1, h \in \mathbb{Z}\}),$$

donde $F(\cdot)$ es la función distribución conjunta de la serie considerada.

A la luz de estos dos conceptos debemos señalar que, un proceso estacionario de forma estricta con los dos primeros momentos finitos es estacionario de orden dos. Por ello, una serie estacionaria de forma débil podría no serlo de forma estricta. Esto ocurriría en el caso en que la media y la covarianza fueran funciones de momentos de ordenes elevados y no del tiempo. Señalar que en lo que sigue, si se habla de un proceso estacionario, se estará tratando de un proceso estacionario de orden dos. Considerando dos series temporales estacionarias podemos introducir la siguiente asociación entre ambas.

Definición 1.3. Sean dos series temporales $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ y $\mathbf{x} = \{x_t, t \in \mathcal{I}\}$, se dirá que son conjuntamente estacionarias si ambas son estacionarias de orden dos y la función $\mathbb{E}(y_t x_s) = \gamma_{t-s}$ solo depende de la distancia entre los instantes de tiempo considerados.

Sabiendo como se caracteriza la estacionariedad, introducimos el concepto de un proceso de ruido blanco, el cual será empleado en cada modelo de series que mostremos en una o más dimensiones, así como en lo referente al modelo de regresión estimado por el método de *GLS* del Capítulo 3.

Definición 1.4. Sea $\mathbf{u} = \{u_t, t \in \mathcal{I}\}$ un proceso verificando:

$$\mathbb{E}(u_t) = 0, \quad \mathbb{E}(u_t^2) = \sigma_u^2 \quad \text{y} \quad \mathbb{E}(u_t u_s) = 0 \quad \text{para todo} \quad t \neq s \in \mathcal{I}.$$

Esto es, $u_t \sim \mathcal{D}_1(0, \sigma_u^2)$. Para \mathcal{D}_k el conjunto de distribuciones k -dimensionales de cierta media y varianza. Entonces \mathbf{u} es denominado proceso de ruido blanco. Si se tiene que tal proceso verifica la normalidad, hablaremos de ruido blanco gaussiano o normal. Y de ser así, la tercera hipótesis de incorrelación se transformaría de inmediato en una condición de independencia.

Por último comentamos la definición de un proceso invertible en base a ciertos coeficientes y a su convergencia en valor absoluto.

Definición 1.5. Un proceso $\mathbf{x} = \{x_t, t \in \mathcal{I}\}$ que admite una representación del tipo,

$$x_t = c + a_t + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots,$$

verificando que $\sum_{i=1}^{\infty} |\pi_i| < \infty$, se dice que es un proceso invertible.

Como se comprobará cuando se introduzca el concepto del modelo $AR(p)$, un proceso será invertible cuando admita una representación $AR(\infty)$ cuyos coeficientes asociados converjan en valor absoluto. Aunque ésto se comprenderá más adelante.

Mostrados estos conceptos, nos sumergimos en el análisis de series de tiempo propiamente dicho, esto es, mostraremos modelos para su explicación y los analizaremos. Primero de forma estacionaria y luego en base a ciertos efectos que provoquen la violación en el proceso.

1.2. Series estacionarias

Como hemos comentado, comenzamos por las series temporales más simples que se pueden encontrar en la práctica, las que son estacionarias de segundo orden o también llamadas de un forma más sencilla, estables. En esta sección introducimos los tres modelos más sencillos que existen para explicarlas, así como alguna propiedad interesante de los mismos. Los dos primeros, el AR y el MA , se comprenderán posteriormente como casos particulares del tercero que comentaremos, el $ARMA$.

El proceso AR de cierto orden p se interpreta como un modelo donde el valor de la serie en cierto instante t , podría suponerse como el actual, es explicado como una combinación lineal de los p valores anteriores y el error producido en dicho instante t . La siguiente definición muestra formalmente éste en función de su expresión asociada

Definición 1.6. Sea $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ una serie temporal estacionaria de orden dos, se define el proceso autorregresivo de orden p , denotándose por $AR(p)$, como

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t \quad \text{para cada } t \in \mathcal{I}, \quad (1.3)$$

siendo constantes los parámetros c y ϕ_i para todo $i = 1, \dots, p$, y $\mathbf{u} = \{u_t, t \in \mathcal{I}\}$ un proceso de ruido blanco tal que $\mathbb{E}(\mathbf{u}^2) = \sigma_u^2$, esto es, $\mathbf{u} \sim \mathcal{D}(\mathbf{0}, \sigma_u^2)$.

Tomemos ahora el polinomio autorregresivo definido como $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ siendo B el operador retardo verificando que $B^i z_t = z_{t-i}$ para cierto $i \in \mathbb{Z}$, y en base a él reescribimos la expresión (1.3) como $\phi_p(B)y_t = c + u_t$. Supongamos en este instante que $\mathbb{E}(y_t) = 0$, ya que si no lo fuese, bastaría con tomar como proceso de estudio el dado por $y_t - \mathbb{E}(y_t)$, que si tendría media nula. Con esta condición, es trivial comprobar que $c = 0$, por lo que nos centraremos en el proceso $\phi_p(B)y_t = u_t$

Para ver el comportamiento del segundo momento de la serie, primero debemos reflexionar sobre la estabilidad del proceso. Para ello reescribimos $\phi_p(B)y_t = u_t$ como $y_t = \phi_p(B)^{-1}u_t = \psi(B)u_t$ siendo $\psi(B) = \phi_p(B)^{-1}$. Expresión que puede entenderse como una progresión geométrica de razón $1 - \phi_p(B)$,

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j = \sum_{j=0}^{\infty} (1 - \phi_p(B))^j, \quad (1.4)$$

que será convergente, si y solo si, el polinomio $\phi_p(z)$ no posee raíces fuera del círculo unidad complejo, esto es,

$$1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \quad \forall z \in \mathbb{C} \quad \text{tal que} \quad |z| \geq 1.$$

En base a esta condición de estabilidad para el proceso $AR(p)$ y a la suma mostrada en (1.4), aportamos la siguiente expresión para la covarianza del proceso, pudiéndose hallar de forma inmediata,

$$\gamma_{t-s} = \text{Cov}(y_t, y_{t-h}) = \mathbb{E}(y_t y_{t-h}) = \sigma_{\mathbf{u}}^2 \sum_{j=0}^{\infty} \psi_{j+h} \psi_j,$$

para cualquier valor $h \in \mathbb{Z}$, en particular $\gamma_t = \text{Var}(y_t) = \sigma_{\mathbf{u}}^2 \sum_{j=0}^{\infty} \psi_j^2$, y donde $\psi_j = (1 - \phi_p(B))^j$ como antes se apuntaba.

Este desarrollo muestra que un proceso $AR(p)$ puede expresarse como una suma infinita de sus errores pasados, lo que se asocia directamente con un modelo $MA(\infty)$, como ya se entenderá posteriormente, y demuestra que un proceso autorregresivo siempre es invertible. También decir que a la representación dada por $y_t = \psi(B)u_t$ se la conoce como la descomposición de Wold y los coeficientes ψ_j como coeficientes de Wold, ambos asociados al Teorema de Wold (Apéndice A) que motiva los modelos a desarrollar. Por otro lado notar que, la condición dada en la Definición 1.5 es equivalente a poseer una representación $AR(\infty)$ con cierta condición sobre los coeficientes.

Como caso particular podemos tomar $p = 1$ resultando el proceso $(1 - \phi_1 B)y_t = u_t$, del cual se puede comprobar con sencillez que la condición de estabilidad en base a las raíces del polinomio autorregresivo, en este caso $1 - \phi_1 z \neq 0$ para $z \in \mathbb{C}$ tal que $|z| \geq 1$, implica que $|\phi_1| < 1$. Así mismo, calculando su segundo momento,

$$\gamma_{t,t-h} = \text{Cov}(y_t, y_{t+h}) = \frac{\sigma_{\mathbf{u}}^2 \phi_1^{t-h}}{1 - \phi_1^2},$$

podemos ver que únicamente la condición $|\phi_1| < 1$ asegura que la varianza de la serie sea finita.

Vista alguna característica del proceso AR , prestando atención a la restricción impuesta para su estabilidad, pasamos a definir otro tipo de procesos, los de medias móviles MA . Este tipo de procesos, el MA de orden q , es empleado para explicar el valor de la serie en cierto instante t en función de una combinación lineal tanto del error en t como de los errores q anteriores en el tiempo. La expresión que lo modeliza se define a continuación.

Definición 1.7. Sea la serie de tiempo $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ estacionaria en covarianza, se define el proceso de medias móviles de orden q , denotándose por $MA(q)$, como

$$y_t = c + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q} \quad \text{para cada} \quad t \in \mathcal{I}, \quad (1.5)$$

siendo constantes los parámetros c y θ_j para todo $j = 1, \dots, q$, y $\mathbf{u} = \{u_t, t \in \mathcal{I}\}$ un proceso de ruido blanco tal que $\mathbb{E}(\mathbf{u}^2) = \sigma_{\mathbf{u}}^2$, esto es, $\mathbf{u} \sim \mathcal{D}(\mathbf{0}, \sigma_{\mathbf{u}}^2)$.

Tomemos ahora el polinomio de medias móviles definido como $\theta_p(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ siendo B el ya introducido operador retardo. Con él, la expresión (1.5) de la definición anterior puede reescribirse como $y_t = c + \theta_q(B)u_t$. Supongamos de nuevo que $\mathbb{E}(y_t) = 0$, con lo que es trivial la misma conclusión que antes, $c = 0$.

Debemos señalar que en este tipo de procesos de medias móviles la estacionariedad se verifica sin condición adicional, a diferencia del anterior caso. Este suceso se pone de relevancia con el cálculo del segundo momento de dicho proceso,

$$\gamma_{t-s} = \text{Cov}(y_t, y_{t-h}) = \mathbb{E}(y_t y_{t-h}) = \sigma_{\mathbf{u}}^2 \left(\theta_h + \sum_{j=1}^{q-h} \theta_{h+j} \theta_j \right),$$

siendo $\theta_0 = 1$. Expresión que no es función de ningún instante de tiempo t ni de parámetros que puedan llevar a la no finitud de la misma, por lo que la estacionariedad es alcanzada sin problemas. Pero por el contrario, será invertible, si y solo si, se verifica que,

$$1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0,$$

para todo $z \in \mathbb{C}$ tal que $|z| \leq 1$, esto es, su polinomio característico no contiene ceros en el círculo unidad complejo. Pues si esto no se verificase, al intentar llegar a una representación similar a la de la Definición 1.5, se llegaría a una indeterminación y no se verificaría la invertibilidad. Por otra parte, notemos que el $MA(\infty)$ que citábamos a través del Teorema de Wold en el momento del $AR(p)$ es de inmediato comprendido sin más que llevar el parámetro q de (1.5) hacia infinito.

Definidos los dos procesos anteriores, se introduce un tercero que tiene en cuenta ambas naturalezas que se han explicado por ser unión de ambas, el $ARMA$ de órdenes p y q . Tiene sentido pensar que los resultados en cuanto a sus momentos y sobre su estacionariedad estén ligados a dicha conjunción. La correspondiente interpretación se basa en que el valor en cierto instante t es explicado por los p valores anteriores, el error actual y sus q anteriores errores. A continuación mostramos el concepto formal con su correspondiente expresión.

Definición 1.8. Sea $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ una serie de tiempo estacionaria de orden dos, se define el proceso autorregresivo de medias móviles de órdenes p y q , denotándose por $ARMA(p, q)$, como

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} \quad \text{para cada } t \in \mathcal{I}, \quad (1.6)$$

siendo constantes los parámetros c , ϕ_i para todo $i = 1, \dots, p$ y θ_j para todo $j = 1, \dots, q$, y $\mathbf{u} = \{u_t, t \in \mathcal{I}\}$ un proceso de ruido blanco tal que $\mathbb{E}(\mathbf{u}^2) = \sigma_{\mathbf{u}}^2$, esto es, $\mathbf{u} \sim \mathcal{D}(\mathbf{0}, \sigma_{\mathbf{u}}^2)$.

De forma natural, tomando los dos polinomios introducidos para los dos anteriores procesos, la expresión del $ARMA(p, q)$, puede ser reescrita como $\phi_p(B)y_t = c + \theta_q(B)u_t$ para cada $t \in \mathcal{I}$. Sin calcular sus momentos, realizamos la debida reflexión sobre su estabilidad e invertibilidad, las cual, como no podía ser de otra manera, están ligadas a la parte autorregresiva y de medias móviles, respectivamente, de su expresión. Esto implica que el proceso $ARMA(p, q)$ será estacionario si de nuevo $1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \forall z \in \mathbb{C}$ para $|z| \leq 1$, y será invertible si $1 + \theta_1 z + \cdots + \theta_q z^q \neq 0 \forall z \in \mathbb{C}$ para $|z| \leq 1$. Verificando estas condiciones, este proceso también podría expresarse como una combinación lineal de errores pasados, como ocurría con el AR , es decir,

$$y_t = \frac{c}{1 - \phi_1 B - \cdots - \phi_p B^p} + \frac{1 + \theta_1 + \cdots + \theta_p B^p}{1 - \phi_1 B - \cdots - \phi_p B^p} u_t = \mu + \psi(B)u_t. \quad (1.7)$$

Como se comentó antes, una motivación de estos modelos se refleja en el, ya citado, Teorema de Wold, cuya idea es recogida en el Apéndice A, como consecuencia de tal resultado, los elementos que constituyen $\psi(B)$ son, como ya dijimos, los coeficientes de Wold. Señalar que este tercer proceso es el más importante, ya que a partir de él se realizarán las estimaciones posteriores correspondientes. A continuación mostraremos que en presencia de series no estacionarias, ciertas modificaciones son aplicadas para alcanzar dicho $ARMA$, y a partir de ahí, se estimaría, diagnosticaría y realizarían predicciones de ellos, como veremos en la Sección 1.4.

1.3. Series no estacionarias

Comentados los modelos propios para series estacionarias, en esta sección mostramos las principales causas de la no estacionariedad de una serie temporal, así como los modelos que se definen en base a ellos. Señalar que estos sucesos serán los que producirán que no se verifique la Definición 1.1 y por tanto sea deseable su desaparición. Podemos suponer la dependencia del primero de los momentos del proceso con respecto al tiempo, originando una componente de tendencia o de estacionalidad si esa dependencia tiene lugar según el periodo de tiempo en el que nos encontremos. O

se puede estar en presencia de una dependencia entre el segundo momento estadístico, referente a la variabilidad de nuestros datos, y el propio tiempo, produciendo el fenómeno de heterocedasticidad. Existen muchos más efectos que provocan la no estacionariedad de una serie, pero se indagará en estos tres por ser los más habituales en la práctica.

Señalar que una vez que se trate la no estacionariedad de la serie y se alcance una serie estable, podremos modelizarla mediante un *ARMA* de ciertos órdenes. Y en este punto se pasará a las etapas de estimación, diagnosis y predicción, como comentábamos. Con todo, mostramos a continuación los diferentes efectos y procedimientos para alcanzar una serie explicada por un *ARMA*.

Tendencia

Comencemos por el suceso de relación entre media y paso del tiempo originando un componente de tendencia. Ésta puede darse en dos posibles sentidos, de forma determinista, supongamos el proceso dado por el modelo

$$y_t = \beta_t + x_t \quad \text{para todo } t \in \mathcal{I},$$

siendo $\beta_t = \beta_1 + \beta_2 t$ para cada t , con β_1 y β_2 constantes, y $\mathbf{x} = \{x_t, t \in \mathcal{I}\}$ un proceso estacionario de forma débil. El proceso \mathbf{y} poseerá una componente de tendencia determinista aportada por los elementos β_t , lo que implica que no es estacionario. Se dirá que es un proceso no estacionario con tendencia determinista. Y un segundo sentido es una tendencia de forma estocástica, supongamos el proceso anterior $\beta = \{\beta_t\}_t$ como un paseo aleatorio, entonces el proceso \mathbf{y} posee tendencia estocástica, lo que implica que no es estacionario. Se dirá que es un proceso no estacionario con tendencia estocástica.

Para el análisis de una serie que presente tales sucesos se definirá un nuevo tipo de modelos, los *ARIMA*, siendo su idea principal la de tomar diferencias en el proceso correspondiente. En el primer caso considerado, tomando diferencias regulares de primer orden se obtiene $y_t - y_{t-1} = \beta_2 + x_t - x_{t-1}$ para cada t , que ya no posee tendencia y, por tanto, es un proceso estacionario. De igual forma, en el segundo caso obtenemos $y_t - y_{t-1} = \beta_t - \beta_{t-1} + x_t - x_{t-1}$, que, por la definición de paseo aleatorio que se puede ver en cualquier manual de procesos estocásticos como por ejemplo [Bhattacharya, R. *et al.* 2009] donde asegura que $\beta_t - \beta_{t-h} \sim \mathcal{N}(0, h)$ para $h \in \mathbb{Z}$, ya no posee tendencia y por tanto se alcanza la estacionariedad.

De una forma general se introduce el siguiente concepto referente a la asociación entre la estacionariedad de un proceso y la cantidad de diferencias regulares necesarias para alcanzarla.

Definición 1.9. Dado un proceso $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ no estacionario con tendencia, diremos que es integrado de orden $d > 0$, denotándose por $\mathbf{y} \sim I(d)$, si para todo $t \in \mathcal{I}$ se verifica que $(1 - B)^{d-1}y_t$ no es estacionario y $(1 - B)^d y_t$ si lo es.

En base a esta definición se introduce el citado concepto del modelo *ARIMA*, el cual es un proceso que después de aplicarle d diferencias regulares se convierte en un *ARMA*. Esto es, se origina a través de procesos $\mathbf{y} \sim I(d)$.

Definición 1.10. Dado $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ una serie de tiempo no estacionaria con tendencia, se define el proceso integrado autorregresivo de medias móviles de ordenes p , d y q , denotándose por *ARIMA*(p, d, q), como

$$\phi_p(B)(1 - B)^d y_t = c + \theta_q(B)u_t \quad \text{para cada } t \in \mathcal{I}, \quad (1.8)$$

siendo c una constante, y $\phi_p(B)$ y $\theta_q(B)$ los polinomios en base a operadores retardos asociados a los modelos ya introducidos.

Por tanto, en presencia de una serie de tiempo en estas circunstancias, tendremos que realizar cierto número de diferencias regulares sobre ella y así obtendremos la estabilidad deseada. Usualmente es suficiente con $d < 3$, lo que sería un polinomio de grado 2 en una tendencia determinística que representaría un crecimiento cuadrático

Estacionalidad

Siguiendo con la suposición de dependencia entre media y tiempo, tomamos ahora ésta con diferentes periodos, estando en presencia de estacionalidad. Para analizar este suceso se realiza un proceso análogo al mostrado en el apartado anterior, definiendo un nuevo tipo de modelos, los *ARIMA* estacionales. Pero primero supongamos

$$y_t = s_t + x_t, \quad \text{para todo } t \in \mathcal{I},$$

siendo $\mathbf{x} = \{x_t, t \in \mathcal{I}\}$ un proceso estacionario de forma débil y $\mathbf{s} = \{s_t, t \in \mathcal{I}\}$ un proceso no estacionario que puede ser de dos naturalezas diferentes, determinista si verifica $s_t = s_{t-s}$ para cierto periodo s , o estocástica si $s_t = s_{t-s} + w_t$ para $\mathbf{w} = \{w_t, t \in \mathcal{I}\}$ un cierto paseo aleatorio. Vemos que éstas son análogos a las componentes de tendencia ya vistas. Por tal razón trabajamos de forma similar, tomando diferencias, pero en este caso estacionales de periodo s en la serie. De esta manera se puede comprobar en detalle que la serie obtenida como $y_t - y_{t-s}$ es estacionaria de segundo orden independiente de la naturaleza del proceso \mathbf{s} .

De una forma más general se introduce el siguiente concepto para entender las diferencias que se toman en este tipo de series temporales.

Definición 1.11. Dado un proceso $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ estacional de periodo s sin tendencia, diremos que es integrado de orden $D > 0$, denotándose por $\mathbf{y} \sim I_s(D)$, si para todo $t \in \mathcal{I}$ se verifica que $(1 - B^s)^{D-1}y_t$ no es estacionario y $(1 - B^s)^D y_t$ si lo es.

A partir de esta definición introducimos otro nuevo tipo de modelo, los citados *ARIMA* estacionales, los cuales son procesos que después de aplicarles D diferencias estacionales se convierten en un *ARMA*. Esto es, se originan a través de procesos $\mathbf{y} \sim I_s(D)$. Señalar que podría existir más de una estacionalidad como se va a reflejar en el siguiente concepto.

Definición 1.12. Dado un proceso $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ una serie temporal con varias estacionalidades de periodos s_1, \dots, s_J sin tendencia, se define el proceso *ARIMA* de periodos s_1, \dots, s_J , denotándose por $ARIMA(P_1, D_1, Q_1)_{s_1} \times \dots \times (P_J, D_J, Q_J)_{s_J}$, como

$$\Phi_{P_1}(B^{s_1}) \dots \Phi_{P_J}(B^{s_J})(1 - B^{s_1})^{D_1} \dots (1 - B^{s_J})^{D_J} y_t = c + \Theta_{Q_1}(B^{s_1}) \dots \Theta_{Q_J}(B^{s_J}) u_t \quad (1.9)$$

para cada $t \in \mathcal{I}$, siendo c una constante, y $\Phi_{P_j}(B^{s_j})$ y $\Theta_{Q_j}(B^{s_j})$ los polinomios en base a operadores retardo asociados a modelos ya introducidos para $j = 1, \dots, J$.

El cual es similar al modelo *ARIMA* con tendencia, pero en este caso las diferencias tomadas no son regulares sino que son estacionales con ciertos periodos s_j . Si ahora tenemos en cuenta la Definición 1.10 y 1.12, podemos introducir el llamado *ARIMA* multiplicativo, el cual plantea la existencia simultánea de tendencia y estacionalidades.

Definición 1.13. Dado un proceso $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ una serie de tiempo con tendencia y estacionalidades de periodos s_1, \dots, s_J , se define el proceso *ARIMA* multiplicativo, denotándose por $ARIMA(p, d, q) \times (P_1, D_1, Q_1)_{s_1} \times \dots \times (P_J, D_J, Q_J)_{s_J}$, como

$$\begin{aligned} \phi_p(B) \Phi_{P_1}(B^{s_1}) \dots \Phi_{P_J}(B^{s_J})(1 - B)^d (1 - B^{s_1})^{D_1} \dots (1 - B^{s_J})^{D_J} y_t = \\ c + \theta_q(B) \Theta_{Q_1}(B^{s_1}) \dots \Theta_{Q_J}(B^{s_J}) u_t \end{aligned} \quad (1.10)$$

siendo c una constante, y $\phi_p(B)$, $\theta_q(B)$, $\Phi_{P_j}(B^{s_j})$ y $\Theta_{Q_j}(B^{s_j})$ los polinomios en base a operadores retardo asociados a los modelos ya introducidos.

Siendo esta definición de *ARIMA* una de las más generales y de la cual se puede llegar a gran parte de los modelos de esta naturaleza como las antes introducidas. Éstos son conocidos como modelos de Box-Jenkins, por ser introducidos por esta pareja de estadísticos. Una vez que hemos mostrado como definir los procesos que en media no son estacionarios y sus transformaciones a realizar para llegar a un modelo *ARMA* estacionario, mostramos un ejemplo que resume estas dos etapas y las aclara gráficamente.

Ejemplo 1.14. Con este sencillo ejemplo mostraremos de forma esquemática como actuar cuando los efectos que acabamos de desarrollar, y por los que la estacionariedad falla, están presentes. Para ello consideramos la serie de tiempo $\mathbf{y} = \{y_t, t = 1, \dots, T\}$ para $T = 200$ definida como adición de un proceso $AR(1)$, un movimiento browniano y una cierta secuencia periódica. El gráfico secuencial correspondiente puede verse en la Figura 1.14 (a). Notar que éste no parece indicar la presencia de un cambio en la variabilidad de los datos a lo largo del tiempo, con lo que solo nos centraremos en lo referente a los cambios de la media introducidos.

En este mismo gráfico observamos la presencia de una tendencia, que sabemos que es estocástica al ser generada a través de la simulación del movimiento browniano. También puede verse la presencia de esta componente con la ACF (Figura 1.14 (b)) en base a sus altos valores de retardos muy lejanos. Por tal razón, debemos diferenciar de forma regular el proceso de estudio y obtener así

$$\mathbf{y}^1 = \{y_t^1\}_t = (1 - B)\mathbf{y}.$$

El correspondiente gráfico secuencial se muestra en la Figura 1.14 (c), donde se observa la desaparición de la tendencia. Si ahora volvemos a representar la ACF (Figura 1.14 (d)) se observa la eliminación de toda la significación que antes se hallaba y se comprueba la significación de retardos cercanos a un periodo de 10 instantes de tiempo, lo que sugiere la estacionalidad en periodo $s = 10$ de la serie temporal. Por ello, diferenciamos ésta estacionalmente con periodo 10, obteniendo entonces

$$\mathbf{y}^2 = \{y_t^2\}_t = (1 - B^{10})\mathbf{y}^1 = (1 - B)(1 - B^{10})\mathbf{y}.$$

Originado el gráfico secuencial de la Figura 1.14 (e), donde se puede vislumbrar la estacionariedad de la serie, aunque de forma más notoria este suceso se concluye con las representaciones de la ACF y $PACF$ (Figura 1.14 (f) y (g), respectivamente). En ellas se ve la desaparición de la significación para retardos próximos a 10, quedando únicamente éstos significativos. Por lo tanto se comprueba que la nueva serie \mathbf{y}^2 es estacionaria, que era el objetivo a buscar. Una vez que llegamos a este punto, obtención de un proceso estacionario de orden dos, estamos en condiciones de proceder al siguiente paso, estimación de la serie. Señalar que para esta nueva etapa partimos de $d = 1$ y $D = 1$.

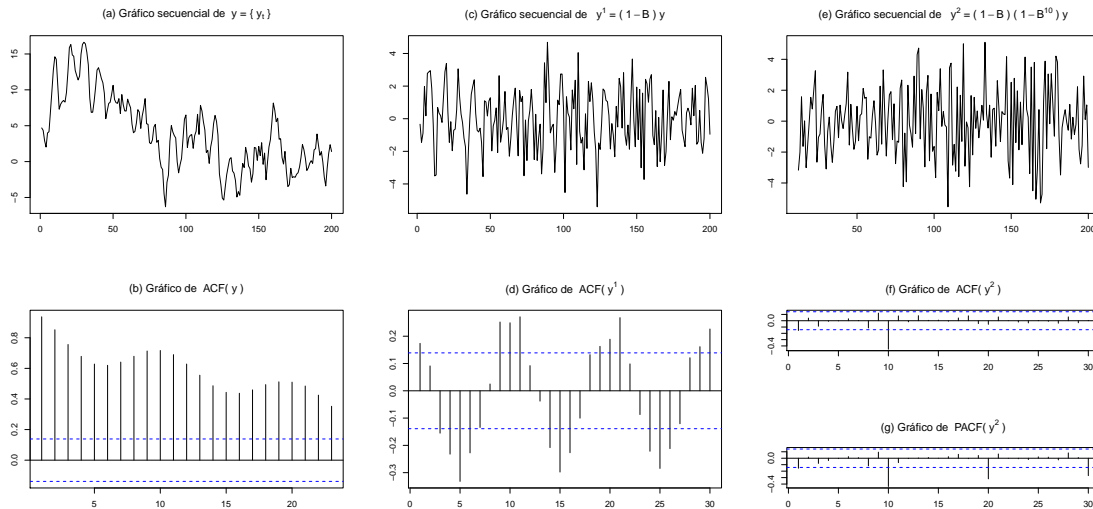


Figura 1.1: (a) Gráfico secuencial de la serie temporal \mathbf{y} definida en el Ejemplo 1.14. (b) Representación de la ACF correspondiente a la serie citada. (c) Gráfico secuencial de la serie dada por $\mathbf{y}^1 = (1 - B)\mathbf{y}$. (d) Representación de la ACF correspondiente a la serie \mathbf{y}^1 . (e) Gráfico secuencial de la serie dada por $\mathbf{y}^2 = (1 - B)(1 - B^s)\mathbf{y}$ para $s = 10$. (f) Representación de la ACF correspondiente a la serie \mathbf{y}^2 . (g) Representación de la $PACF$ correspondiente a la serie \mathbf{y}^2 .

Heterocedasticidad

Por último, supongamos ahora la asociación con el tiempo de la variabilidad de los datos, presencia pues de heterocedasticidad. En la literatura existen muchas transformaciones para eliminar este suceso, pero será la función de Box-Cox la que mostraremos en este apartado. Comenzamos tomando una serie temporal $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ tal que $\text{Var}(y_t) = \sigma_t^2$, es decir, verificando que su varianza varía con respecto al tiempo, con lo que nuestra serie no sería estacionaria. Para solventar tal problema, G. Box y D. Cox plantean la siguiente función continua sobre los datos,

$$y_t^{BC} = \begin{cases} \log y_t & \text{si } \lambda = 0, \\ \frac{y_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0. \end{cases} \quad (1.11)$$

Señalar que esta transformación solo es válida para series temporales cuyos valores sean positivos para todo instante de tiempo, debido a la presencia de la función log. Pero este aspecto se solventa con la selección de una constante fija $\zeta \in \mathbb{R}$ que verifique $y_t + \zeta > 0$ para todo $t \in \mathcal{I}$. Comentar que esta función es continua debido a que $\lim_{\lambda \rightarrow 0} (y_t^\lambda - 1)/\lambda = \log y_t$, lo cual será una propiedad a tener en cuenta.

La transformación óptima viene determinada por la relación en regresión entre media y varianza de los datos, a partir de la cual se obtiene el parámetro λ óptimo, esto es, debido a que cuanto menor relación lineal exista entre éstas, menos dependerá la varianza del tiempo. Por tal razón se seleccionará, entre un cierto rango de valores, el λ asociado al modelo de regresión con menor significación para la media. Una vez encontrado el valor de λ óptimo para emplear, se tomará como nueva serie $\mathbf{y} = \mathbf{y}^{BC}$. Y se trabajará a partir de aquí con una serie homocedástica.

1.4. Estimación, Diagnosis y Predicción de modelos

Para el siguiente desarrollo podría suponerse una serie generada por un *ARIMA* multiplicativo y así mostrar el caso más general de todos, pero por simplicidad supondremos que el *DGP* sigue un modelo *ARMA*(p, q) estacionario en covarianza con órdenes p y q conocidos. En referencias como [Lütkepohl, H. *et al.* 2004] o [Cryer, J. *et al.* 2008] se parte del modelo citado, pero se emplean otras técnicas de las que se aportarán en este trabajo, ya que en éste nos centraremos en las clásicas vistas en cursos de series temporales univariantes. Una vez hechas estas consideraciones, comenzamos por indicar como se estimaría un proceso *ARMA* con sus respectivos órdenes.

Estimación

Lo primero a realizar en esta fase es una estimación del proceso en base a ciertos parámetros p y q generales, para después en base a ella, seleccionar, de cierta forma, que ya se indicará, los órdenes más adecuados para el ajuste de nuestros datos. Para dicho objetivo existen dos posibles vías, la primera es mediante mínimos cuadrados aplicados a los residuos generados del correspondiente ajuste, dando de este modo los siguientes estimadores,

$$(\hat{c}, \hat{\phi}, \hat{\theta}) = \arg \min_{c, \phi, \theta} \sum_{t \in \mathcal{I}} \hat{u}_t^2,$$

siendo $\hat{u}_t = \hat{\phi}_p(B)y_t - \hat{c} - \hat{\theta}_q(B)\hat{u}_{t-1}$ los residuos del ajuste. Lo que genera el problema de estimar \hat{u}_k para $k = 1, \dots, p$ por estar dados en función de los valores no observados y_0, \dots, y_{1-p} . Con esta idea en mente se origina el método de mínimos cuadrados condicionados, estableciendo los estimadores

$$(\hat{c}, \hat{\phi}, \hat{\theta}) = \arg \min_{c, \phi, \theta} \sum_{t \in \mathcal{I} \setminus \{1, \dots, p\}} \hat{u}_t^2,$$

con $\hat{u}_j = 0$ para $j = p, \dots, p - q + 1$, que ya no dependería de valores no observados. El segundo posible camino es tomado la estimación generada a través del método de máxima verosimilitud,

donde se toma,

$$\left(\hat{c}, \hat{\phi}, \hat{\theta}, \hat{\sigma}_u^2\right) = \arg \max_{\tilde{c}, \tilde{\phi}, \tilde{\theta}, \tilde{\sigma}_u^2} L_{y_1, \dots, y_t} \left(\tilde{c}, \tilde{\phi}, \tilde{\theta}, \tilde{\sigma}_u^2\right),$$

con $\tilde{\sigma}_u^2$ el estadístico de la varianza de los residuos generados y siendo $L(\cdot)$ la función de densidad conjunta de un vector aleatorio $\tilde{\mathbf{y}}$ procedente de un proceso *ARMA* con parámetros \tilde{c} , $\tilde{\phi}$, $\tilde{\theta}$ y $\tilde{\sigma}_u^2$. Esta última opción bajo adecuadas condiciones genera estimadores asintóticamente óptimos para tamaños T grandes, esto es, insesgados, eficientes y con distribución normal, siendo la última condición necesaria para la construcción de intervalos de predicción. A mayores, el estimador para σ_u^2 es consistente. Señalar que estos dos posibles caminos a tomar en la estimación de los parámetros del modelo serán equivalentes si las innovaciones o errores presentes en él son gaussianos, reflexión clásica en cursos introductorios de estas dos técnicas.

Una vez sabemos como estimar el modelo debemos hallar los órdenes p y q para nuestros datos. Para ello empleamos las conocidas funciones de autocorrelaciones (*ACF*) y autocorrelaciones parciales (*PACF*), para así establecer ciertos valores de dichos parámetros. Y de entre todos los posibles pares (p, q) que podamos tomar generadores de un ajuste cuya diagnosis sea favorable (en un sentido que se precisará más adelante), se seleccionará el que minimice ciertos criterios de información. En el sentido univariante, éstos están dados por la expresión

$$IC(p, q) = \delta + T^{-1}g(p + q)$$

donde $\delta = \log(\hat{\sigma}_u^2)$ para $\hat{\sigma}_u^2$ la estimación de la varianza de los residuos que antes hemos introducido. Y será en base a la función denotada por g sobre la cual se definen diferentes criterios, entre otros, introducimos los siguientes,

$$\begin{aligned} AIC(p, q) &= \delta + T^{-1}2(p + q), \\ AICC(p, q) &= \delta + T^{-1} \frac{(p + q)(T - 1) + 2}{T - p - q - 2}, \\ BIC(p, q) &= \delta + T^{-1} \log T(p + q), \\ HQ(p, q) &= \delta + T^{-1} \log \log T(p + q). \end{aligned}$$

Siendo éstos los de Akaike, su corrección, el de Bayes y el de Hamilton-Quinn. Con lo que se seleccionarán los parámetros p y q que minimicen estos criterios, quedándonos con el correspondiente modelo *ARMA*(p, q) ajustado.

Diagnosis

Una vez estimado el modelo para ciertos órdenes, debemos dar la diagnosis que ha de pasar el ajuste y que ya en el apartado anterior hacíamos referencia. En este sentido lo que se intentará comprobar es, si los residuos producidos por la susodicha estimación se pueden entender como un proceso de ruido blanco (Definición 1.4), esto es, que sean estacionarios de media cero e incorrelados. También se intentará ver si son gaussianos para todo retardo en aras de la construcción de intervalos de predicción, podría demostrarse que para retardos altos se verifica aproximadamente la gaussianidad. Señalar que los respectivos contrastes solo los citaremos sin ahondar en sus expresiones, debido a que son bien conocidos. Donde sí que se profundizará más es cuando hablemos del caso vectorial por ser un tema más novedoso.

Para contrastar la incorrelación de los residuos se echará mano del test de Ljung-Box para ciertos retardos. Para ver si son estacionarios, bastará con visualizar la serie residual de forma gráfica, sus *ACF* y *PACF*, y para comprobar la media nula se empleará la prueba t de Student. Hasta aquí las hipótesis necesarias para la buena definición del ajuste, ahora se plantea la normalidad de forma gráfica con el qq -plot y de forma analítica en base al test presentado por shapiro-wilks y al expuesto por Jarque-Bera, ambos con la hipótesis nula de la gaussianidad

Predicción

Una vez realizado todo el proceso de saber como es el modelo que ajustamos a nuestra serie temporal y comprobado que éste es válido, podemos realizar predicciones a cierto horizonte h en el tiempo. Para ello retomamos la expresión del $ARMA(p, q)$ dado en (1.6) ajustada según se ha mostrado hasta ahora, pero para un instante $T + h$. Entonces para cada $t \in \mathcal{I}$,

$$y_{T+h} = \hat{\phi}_1 \tilde{y}_{T+h-1} + \cdots + \hat{\phi}_p \tilde{y}_{T+h-p} + \hat{u}_{T+h} + \hat{\theta}_1 \hat{u}_{T+h-1} + \cdots + \hat{\theta}_q \hat{u}_{T+h-q},$$

con $\tilde{y}_t = y_t$ y $u_t = 0$ para $t \leq T$. Que tomando la expresión presentada en (1.7), y verificando la invertibilidad del proceso, podemos escribir como $y_{T+h} = \hat{\mu} + \hat{\psi}(B)\hat{u}_{T+h}$. Si ahora se verifica que \hat{u}_t es ruido blanco gaussiano, podemos asegurar que la predicción a horizonte h se distribuye tal que

$$y_{T+h} | \mathcal{F}_t \sim \mathcal{N} \left(y_{T+h|t}, \hat{\sigma}^2(h)(1 + \hat{\psi}_1^2 + \cdots + \hat{\psi}_{h-1}^2) \right),$$

donde los $\hat{\psi}_i$ para $i = 1, \dots, h-1$ son los coeficientes de Wold asociados al proceso $ARMA(p, q)$ estimado, como comentábamos anteriormente, y $\hat{\sigma}(h)$ es la desviación típica correspondiente al proceso a horizonte h . Por tanto se tiene que el intervalo de confianza a nivel $1 - \alpha$ viene dado por

$$\left(y_{T+h} | \mathcal{F}_t \pm \mathcal{Z}_{1-\alpha/2} \hat{\sigma}(h) \sqrt{1 + \hat{\psi}_1^2 + \cdots + \hat{\psi}_{h-1}^2} \right), \quad (1.12)$$

siendo, como es usual, α el nivel de significación del contraste y $\mathcal{Z}_{1-\alpha/2}$ el cuantil de nivel de significación α para la normal en un contraste bilateral.

Capítulo 2

Análisis Vectorial de Series de Tiempo

Hasta este punto hemos visto los modelos *ARIMA* empleados para explicar las series de tiempo de forma univariante, en este capítulo nos introduciremos en una visión vectorial o multivariante sobre las series temporales. Es decir, tomaremos el conjunto \mathbf{y} de K series temporales, $\mathbf{y} = \{\mathbf{y}_t\}_t = \{(y_{1t}, \dots, y_{Kt})\}_t$ que mencionábamos al introducir el primer capítulo, pero en vez de estudiar el j -ésimo proceso, lo que haremos es analizar las K variables conjuntamente. La motivación a esta siguiente etapa del estudio es debido a la más que probable existencia de relaciones internas entre algunas series de tiempo que se consideren en el estudio, suceso que no se ha tenido en cuenta hasta ahora. Este fenómeno será denominado cointegración y tendrá como base el concepto de integración dado en el anterior capítulo.

A lo largo de los próximos párrafos introduciremos los dos modelos vectoriales más importantes, los procesos autorregresivos vectoriales, *VAR* (Sección 2.1), y los modelos de correlación de error vectorial, *VECM* (Sección 2.2). De ellos daremos su expresión y comentaremos como poder estimarlos, para luego mostrar algunos contrastes para ambos procesos y que se deben realizar para comprobar la validación de cada modelo (Sección 2.3). Siguiendo así un esquema análogo al desarrollado en el Capítulo 1.

Señalar que a lo largo del actual capítulo, como en el anterior, emplearemos el intervalo de tiempo denotado por \mathcal{I} , que si no se especifica será el correspondiente al $[1, T]$. Otro apunte que debemos realizar antes de entrar en materia, es que los modelos que se muestren y sus consiguientes pasos, podrían presentar elementos que reflejasen variables exógenas, esto es, variables que se toman fijas en el modelo por ser conocidas, pero por simplicidad las omitiremos. Esta consideración será muy útil a la hora de la predicción en la parte práctica, ya que si de ciertas variables se sabe el comportamiento a un horizonte h , ello será utilizado para mejorar la precisión a un cierto horizonte de otras variables endógenas.

Para este capítulo tomaremos como referencia básica el [Lütkepohl, H. *et al.* 2004] y en particular lo desarrollado en su tercer capítulo, *Vector Autoregressive and Vector Error Correction Models*. También echaremos mano de [Pfaff, B. 2008a] para comprender cómo estos modelos deben aplicarse y cómo facilitar tal aplicación con un adecuado desarrollo teórico.

2.1. Modelo *VAR*

Comenzamos por los procesos de carácter autorregresivo multivariante que, como hemos dicho, son una visión multidimensional de los introducidos *ARIMA*, en particular son una generalización

de los procesos AR como ya se entenderá, de hecho, veremos como muchos resultados son análogos a los hallados en el anterior capítulo, simplemente transformando un valor real en un vector o una matriz. Primero definiremos el concepto de VAR con un cierto orden p , para luego especificar y estimar el modelo resultante comprobando su validez.

Definición 2.1. Sea $\mathbf{y}_t = (y_{1t}, \dots, y_{Kt})$ la representación de K variables endógenas en el instante $t \in \mathcal{I}$. Se define el modelo autorregresivo vectorial de orden p , denotándose por $VAR(p)$, como

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{C} \mathbf{D}_t + \mathbf{u}_t, \quad (2.1)$$

siendo $\mathbf{A}_i \in \mathcal{M}_{K \times K}$ para $i = 1, \dots, p$ y $\mathbf{C} \in \mathcal{M}_{K \times M}$ los coeficientes asociados a los regresores deterministas dados por $\mathbf{D}_t \in \mathbb{R}^M$, entre los que se encuentran la tendencia o la estacionalidad de la serie. Y el vector $\mathbf{u}_t \in \mathbb{R}^K$ un proceso de ruido blanco K -dimensional con $\mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\Sigma}_u$, esto es, $\mathbf{u}_t \sim \mathcal{D}_K(\mathbf{0}, \boldsymbol{\Sigma}_u)$.

Como comentábamos, este modelo es muy similar al dado en la ecuación (1.3) del modelo $AR(p)$, simplemente debemos generalizarlo a una concepción vectorial. Por tanto de igual modo que en la primera sección, podemos reescribir la expresión (2.1) en función del polinomio matricial definido como $\mathbf{A}_p(B) = \mathbf{I}_K - \mathbf{A}_1 B - \dots - \mathbf{A}_p B^p$, obteniendo de esta forma

$$\mathbf{A}_p(B) \mathbf{y}_t = \mathbf{C} \mathbf{D}_t + \mathbf{u}_t.$$

Y es en base a este polinomio sobre el cual, de nuevo, podremos reflexionar sobre la estabilidad de este proceso. Este modelo será estacionario, si y solo si, todas las raíces del polinomio $\mathbf{A}_p(z)$ no se encuentran en el círculo unidad complejo, es decir,

$$\det(\mathbf{A}_p(z)) = \det(\mathbf{I}_K - \mathbf{A}_1 z - \dots - \mathbf{A}_p z^p) \neq 0 \quad \forall z \in \mathbb{C}^K \quad \text{tal que} \quad |z| \leq 1. \quad (2.2)$$

Si existiese algún z^* raíz de $\mathbf{A}_p(z)$ tal que no verificase la expresión (2.2) anterior, podríamos afirmar la existencia de alguna variable integrada en el sentido de la Definición 1.9 en nuestra consideración inicial y si el modulo de tal raíz fuese uno nos introduciríamos en el problema de la raíz unitaria, del cual trataremos más adelante y que ahora dejamos a un lado.

Observación 2.2. El proceso introducido $VAR(p)$ siempre puede ser visto de una forma más sencilla como un $VAR(1)$ sin más que aumentar la dimensión de los vectores considerados. Tomemos el modelo dado por la expresión

$$\boldsymbol{\zeta}_t = \mathbf{A} \boldsymbol{\zeta}_{t-1} + \mathbf{v}_t \quad (2.3)$$

donde $\boldsymbol{\zeta}_t = (\mathbf{y}_t, \dots, \mathbf{y}_{t-p+1})' \in \mathbb{R}^{Kp}$, $\mathbf{v}_t = (\mathbf{u}_t, 0, \dots, 0) \in \mathbb{R}^{Kp}$ y

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{pmatrix} \in \mathcal{M}_{Kp \times Kp}.$$

Y en base a esta visión de un modelo más simple también se puede reflexionar sobre la estacionariedad del proceso. Ya que si suponemos que $\lambda_{\mathbf{A}}^i$ para $i = 1, \dots, Kp$ son los autovalores de la matriz \mathbf{A} dada en la anterior expresión, verificándose que $|\lambda_{\mathbf{A}}^i| < 1$ para todo i , entonces el modelo definido en (2.3) es estacionario y por tanto lo es también el asociado original $VAR(p)$. En la práctica este sentido será un sencillo camino para comprobar la estacionariedad de un proceso VAR .

Observación 2.3. Por último, comentar que, de forma análoga a la ya introducida, el proceso $VAR(p)$ puede ser reescrito en función de una suma infinita de sus errores suponiendo su estacionariedad, como ya pasaba en el caso de una dimensión, de nuevo nos basamos para ello en el Teorema de la Descomposición de Wold (Apéndice A.1), resultando

$$\mathbf{y}_t = \phi_0 \mathbf{u}_t + \phi_1 \mathbf{u}_{t-1} + \phi_2 \mathbf{u}_{t-2} + \dots \quad (2.4)$$

donde $\phi_0 = \mathbf{I}_K$ y $\phi_i = \sum_{j=1}^i \phi_{i-j} \mathbf{A}_j$ para $i = 1, 2, \dots$ verificando $\mathbf{A}_j = 0$ si $j > p$.

2.1.1. Estimación y especificación

Una vez presentado el modelo VAR , análogamente a lo ya desarrollado para una dimensión, supondremos un valor p genérico y mostraremos cómo se estimaría dicho modelo. Cuando sepamos cómo ajustar el correspondiente modelo de entre un rango de parámetros p que generan un ajuste favorable en cuanto a la diagnosis multivariante se refiere (la especificaremos más tarde), se seleccionará el orden p en base a ciertos criterios que ya comentaremos.

Para la etapa de estimación supongamos el proceso $VAR(p)$ dado por la expresión (2.1), sin considerar la componente determinista $\mathbf{C}\mathbf{D}_t$ por simplicidad, con lo que el modelo resulta ser

$$\mathbf{A}_p(B)\mathbf{y}_t = \mathbf{u}_t. \quad (2.5)$$

Éste se estimará por medio de mínimos cuadrados ordinarios (OLS), ya que este ajuste separadamente resulta ser de la misma eficiencia que si procediésemos a través de mínimos cuadrados generalizados, GLS (método que se verá en el Capítulo 3). Consideramos entonces la forma matricial de la expresión (2.1)

$$\mathbf{Y} = \mathbf{A}\mathbf{Z} + \mathbf{U}, \quad (2.6)$$

donde $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T) \in \mathcal{M}_{K \times T}$, $\mathbf{A} = (\mathbf{A}_1 : \dots : \mathbf{A}_p) \in \mathcal{M}_{K \times Kp}$, $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_T) \in \mathcal{M}_{K \times T}$ y $\mathbf{Z} = (\mathbf{Z}_0 : \dots : \mathbf{Z}_{T-1}) \in \mathcal{M}_{Kp \times T}$ con $\mathbf{Z}_{t-1} = (\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p})' \in \mathbb{R}^{Kp}$. Planteado el modelo general, el estimador empleando OLS de la matriz \mathbf{A} viene dado por

$$\hat{\mathbf{A}}_{OLS} = (\hat{\mathbf{A}}_1 : \dots : \hat{\mathbf{A}}_p) = \mathbf{Y}\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1} \sim \mathcal{N}_K(\mathbf{A}, T^{-1}\Sigma_{\hat{\mathbf{A}}_{OLS}}), \quad (2.7)$$

donde la matriz de covarianzas resultante verifica $\Sigma_{\hat{\mathbf{A}}_{OLS}} \xrightarrow{p} (\mathbf{Z}\mathbf{Z}'/T)^{-1}\Sigma_{\mathbf{u}}$, con lo que, si tomamos el límite en la expresión anterior (2.7), se tiene $\hat{\mathbf{A}}_{OLS} \sim \mathcal{N}_K(\mathbf{A}, (\mathbf{Z}\mathbf{Z}')^{-1} \otimes \Sigma_{\mathbf{u}})$. Siendo \otimes el producto tensorial de matrices.

Observación 2.4. Reflexiones sobre la relación entre el estimador y la integrabilidad de las variables en el sentido de la Definición 1.9.

- Si cada elemento \mathbf{y}_t de nuestro modelo verifica que todas las variables que lo componen son $I(0)$, esto es, $y_{kt} \sim I(0)$ para todo $k = 1, \dots, K$, y se tiene también la pertinente gaussianidad, entonces, como en estadística clásica, el estimador alcanzado mediante OLS es idéntico al de máxima verosimilitud, *i.e.*, $\hat{\mathbf{A}}_{OLS} = \hat{\mathbf{A}}_{ML}$.
- Con la premisa anterior, la matriz $\Sigma_{\hat{\mathbf{A}}_{OLS}}$ es no singular si las todas las variables son $I(0)$.
- Por el contrario, la matriz $\Sigma_{\hat{\mathbf{A}}_{OLS}}$ es singular si existe alguna variable integrada, es decir, $\exists k = 1, \dots, K$ tal que $y_{kt} \sim I(1)$. A pesar de ello, el resultado (2.7) se mantiene cierto, aunque algunos coeficientes de dicha estimación convergen con mayor velocidad que $T^{1/2}$. Lo que implica que los t -, χ^2 - o F - test podrían no ser válidos. Este suceso es conocido como el problema de la raíz unidad o unitaria, veremos en el siguiente apartado que dará lugar a la definición de cointegración y será de lo cual partamos como motivación para la definición del modelo $VECM$.

Estimados los coeficientes del modelo representados en la matriz \mathbf{A} , debemos realizar la consiguiente estimación sobre la distribución de los residuos. Hemos visto que éstos siguen una distribución K -dimensional de media $\mathbf{0}$ y matriz de covarianzas $\Sigma_{\mathbf{u}}$ la cual debe ser ajustada. Para ello se proponen dos estimadores consistentes y con distribución normal,

$$\hat{\Sigma}_{\mathbf{u}} = \frac{1}{T - Kp} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \quad \text{y} \quad \tilde{\Sigma}_{\mathbf{u}} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t', \quad (2.8)$$

siendo $\hat{\mathbf{u}}_t = \mathbf{y}_t - \hat{\mathbf{A}}_{OLS} \mathbf{Z}_{t-1}$ las componentes de los residuos del modelo (2.6).

Una vez que hemos mostrado cómo ajustar un modelo $VAR(p)$, debemos hallar el orden p más adecuado según nuestros datos, esto es, debemos especificar el modelo. Para ello podemos seguir dos caminos, el segundo de los cuales es una generalización del método mostrado en el capítulo anterior. El primero de ellos se basa en la selección de un orden máximo $p_{\text{máx}}$ para el proceso y con él se crea una sucesión de contrastes con la secuencia de hipótesis nulas dadas por $\{H_0^i \equiv \mathbf{A}_{p_{\text{máx}}-i} = \mathbf{0}, i = 0, \dots, p_{\text{máx}}\}$. Supongamos que el primer rechazo resultante de todos los contrastes se tiene para un cierto i^* , entonces el orden definitivo para el modelo será $p_{\text{máx}} - i^*$. Como en estadística no paramétrica con el parámetro de suavizado h , en este procedimiento es capital una adecuada selección de $p_{\text{máx}}$, ya que si este valor es muy pequeño se producirán errores en la diagnosis posterior, mientras que si $p_{\text{máx}}$ es demasiado grande el error de Tipo I asociado a los contrastes aumentará. Existen ciertas técnicas para la correcta selección de tal parámetro, pero las omitiremos por en la práctica se empleará el segundo método que ahora comentamos.

El otro método se encuadra en la selección del orden p que minimice un cierto criterio de información, como en el caso univariante. La expresión general del criterio vendrá dado por

$$IC_v(p, q) = \delta_v + T^{-1} g_v(p)$$

donde $\delta_v = \log(\det(\hat{\Sigma}_{\mathbf{u}}))$ con $\hat{\Sigma}_{\mathbf{u}}$ la estimación de la matriz de covarianzas de los errores, y $g_v(p)$ es una función que penaliza grandes órdenes del VAR y será ella la que diferencie distintos criterios, entre otros, los más habituales tomando diferentes consideraciones son:

$$\begin{aligned} AIC(p, q) &= \delta_v + T^{-1} pK^2 2(p + q), \\ HQ(p, q) &= \delta_v + T^{-1} \log(\log(T)) pK^2 (p + q), \\ SC(p, q) &= \delta_v + T^{-1} \log(T) pK^2. \end{aligned}$$

Como en el anterior caso, tenemos el de Akaike, el de Hamilton-Quinn y el de Schwarz. Suponiendo el intervalo $\mathcal{I} = [1, T]$ con medida en general T , podemos determinar relaciones entre los órdenes hallados según cada uno de los criterios (denotados por $\hat{p}(IC_v)$) condicionados al tamaño del intervalo considerado,

$$\hat{p}(SC_v) \leq \hat{p}(AIC_v) \quad \text{para} \quad T \geq 8,$$

$$\hat{p}(SC_v) \leq \hat{p}(HQ_v) \quad \text{para todo} \quad T,$$

$$\hat{p}(HQ_v) \leq \hat{p}(AIC_v) \quad \text{para} \quad T \geq 16.$$

Que, en definitiva, para un valor de $T \geq 16$ se tiene $\hat{p}(SC_v) \leq \hat{p}(HQ_v) \leq \hat{p}(AIC_v)$.

Realizada la estimación, el siguiente paso es comprobar si ésta es correcta y válida con ciertos contrastes de diagnosis. Pero el desarrollo de tales test los realizaremos en la Sección 2.3 conjuntamente con los asociados a los $VECM$, ya que muchos se definen indistintamente para cada uno de los modelos. Por ello pasamos a la sección donde motivaremos e introduciremos este nuevo modelo.

2.2. Modelos VECM

En la sección anterior, y en particular en el tercer punto de la Observación 2.4, hemos comentado la posible existencia de una raíz unidad en el proceso de estudio, lo que tendría asociada la existencia de alguna variable integrada entre las consideradas y con ello la falta de estacionariedad del proceso. Esto es, al existir raíz unitaria se asegura existe $\mathbf{y}_t \sim I(m)$ para cierto $m \geq 1$. Con este suceso en mente, se define lo que se entenderá por un conjunto de vectores cointegrados y que dará solución al problema planteado, siendo la idea central de este concepto la existencia de ciertos vectores denotados por $\boldsymbol{\beta}$ que a partir de un proceso no estacionario, el producto con ellos lo transforma en estacionario.

Definición 2.5. Dado un vector de variables $\mathbf{y}_t \in \mathbb{R}^K$ con alguna de sus componente integrada, se dice que es cointegrado si al menos existe un vector K -dimensional $\boldsymbol{\beta}$ tal que $\boldsymbol{\beta}'\mathbf{y}_t$ es estacionario. En tal caso, $\boldsymbol{\beta}$ es llamado el vector de cointegración. Si r es tal que $\boldsymbol{\beta}_i$ para $i = 1, \dots, r$ son vectores de cointegración linealmente independientes, diremos que $\mathbf{y} = \{\mathbf{y}_t\}_t$ es cointegrado con rango de cointegración r .

Como hemos visto en los VAR, este suceso podía acaecer pero no se aportaba ninguna solución, siendo ésto debido a que las posibles relaciones de cointegración existentes en el proceso de estudio estaban presentes de forma implícita y por ello no se podían modelizar. De hecho, el VAR es un procesos suficiente para explicar las variables con tendencia determinística común que es lo que teníamos hasta ahora, pero no es el modelo más adecuado si se quiere reflejar la cointegración, ya que lo deseable para su estudio es una cierta expresión que la muestre de una forma explícita.

Por tanto, con el fin de dar un modelo en base a una expresión donde éstas estén presentes explícitamente y así poder estudiarlas, se definen otros procesos también multivariantes que sí que poseen esta característica, éstos son los modelos de correlación de errores vectorial, VECM. Señalar que éstos no generarán procesos diferentes de los VAR, simplemente serán una reescritura de éstos para explicitar la cointegración en base a una matriz que se denotará como $\boldsymbol{\Pi}$. Veremos primeramente que estos procesos se basan en tomar las primeras diferencias regulares en el VAR, y luego indicaremos como se estimaría y especificaría el modelo, del mismo modo que los modelos anteriormente introducidos. Señalar que en esta definición tampoco incluiremos por simplicidad variables exógenas como ya pasaba en el caso anterior, que de nuevo serán triviales de añadir al modelo. A la hora de la validación recurriremos a la Sección 2.3 como ya indicamos en el caso del VAR.

Definición 2.6. Sea $\mathbf{y}_t = (\mathbf{y}_{1t}, \dots, \mathbf{y}_{Kt})$ la representación de K variables endógenas en el instante $t \in \mathcal{I}$. Se define el modelo vectorial de errores correlados de orden $p-1$, denotándose por VECM($p-1$), como

$$\Delta \mathbf{y}_t = \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{y}_{t-1} + \dots + \boldsymbol{\Gamma}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{u}_t, \quad (2.9)$$

siendo $\boldsymbol{\Pi} = -(\mathbf{I}_K - \sum_{j=1}^p \mathbf{A}_j) \in \mathcal{M}_{K \times K}$ recogiendo los parámetros denominados *long-run*, $\boldsymbol{\Gamma}_i = -\sum_{j=i+1}^p \mathbf{A}_j \in \mathcal{M}_{K \times K}$ para $i = 1, \dots, p-1$ con los parámetros *short-run* y $\mathbf{u}_t \in \mathbb{R}^K$ un proceso de ruido blanco con $\mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\Sigma}_u$, esto es, $\mathbf{u}_t \sim \mathcal{D}_K(\mathbf{0}, \boldsymbol{\Sigma}_u)$. Siendo $\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ la primera diferencia regular del vector \mathbf{y}_t .

Observación 2.7. La razón de por qué se ha definido el VECM para el orden $p-1$ y no para el que parecería más lógico p , es el directo paso de un VAR(p) de la Definición 2.1 a este VECM($p-1$). Ya que en este último concepto se definen las matrices que toman parte en el proceso en función de las matrices del primero de los procesos. Veamos esta justificación partiendo de un VECM($p-1$),

$$\begin{aligned} \Delta \mathbf{y}_t &= \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{y}_{t-1} + \dots + \boldsymbol{\Gamma}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{u}_t, \\ \mathbf{y}_t - \mathbf{y}_{t-1} &= -\left(\mathbf{I}_K - \sum_{j=1}^p \mathbf{A}_j\right) \mathbf{y}_{t-1} - \sum_{j=2}^p \mathbf{A}_j (\mathbf{y}_{t-1} - \mathbf{y}_{t-2}) - \dots - \mathbf{A}_p (\mathbf{y}_{t-p+1} - \mathbf{y}_{t-p}), \\ \mathbf{y}_t - \mathbf{y}_{t-1} &= \mathbf{y}_{t-1} + \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t, \\ \mathbf{y}_t &= \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t. \end{aligned}$$

Del mismo modo podemos dar una forma de expresar las matrices \mathbf{A}_j para $j = 1, \dots, p$ en función de las nuevas $\mathbf{\Pi}$ y $\mathbf{\Gamma}_i$, esta es,

$$\mathbf{A}_j = \begin{cases} \mathbf{\Pi} + \mathbf{\Gamma}_1 + \mathbf{I}_K & \text{si } i = 1, \\ \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1} & \text{si } i = 2, \dots, p-1, \\ -\mathbf{\Gamma}_{p-1} & \text{si } i = p. \end{cases}$$

Así, con estas relaciones, podemos pasar de un modelo a otro sin problemas, e incluso a la hora de especificar algún parámetro como puede ser el orden p , podemos hallarlo en base a uno y aplicárselo al otro directamente. De hecho, en la práctica, el orden del *VECM* será hallado con el modelo *VAR* asociado.

Visto como es el modelo, debemos reflexionar en dónde aparecen de forma explícita las posibles relaciones de cointegración, ya que ésta es la motivación del presente modelo. Supongamos que las variables son, como mucho, integradas de orden uno, esto es, se tiene que $y_k \sim I(m)$ con $m \leq 1$. Por tal suposición, al calcular diferencias regulares de las mismas mediante el operador Δ , la integrabilidad desaparece, entonces $\Delta y_{y-i} \sim I(0)$ para todo $i = 0, \dots, p-1$. Este suceso implica que, $\mathbf{\Pi} \mathbf{y}_{t-1} \sim I(0)$ por ser el resto de elementos de esta naturaleza. Entonces, partiendo de un vector $\mathbf{y}_{t-1} \sim I(1)$ se ha llegado a un $\mathbf{\Pi} \mathbf{y}_{t-1} \sim I(0)$, por lo que, atendiendo a la Definición 2.6, se deduce que en la matriz $\mathbf{\Pi}$ se encierran los vectores de cointegración. Supongamos en este sentido que el *VAR*(p) de (2.1) tiene una raíz unidad, $\det(\mathbf{A}_p(\mathbf{z}^*)) = 0$ para cierto $\mathbf{z}^* \in \mathbb{C}^K$ verificando $|\mathbf{z}^*| = 1$. Por definición, $\mathbf{\Pi} = -\mathbf{A}_{p-1}(\mathbf{1}_K)$, entonces $\mathbf{\Pi}$ es singular, con lo que, sin pérdida de generalidad, definimos $r = \text{rk}(\mathbf{\Pi})$. Lo que implica que

$$\mathbf{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}' \quad \text{tal que} \quad \boldsymbol{\alpha} \in \mathcal{M}_{K \times r}, \quad \boldsymbol{\beta} \in \mathcal{M}_{K \times r} \quad \text{con} \quad \text{rk}(\boldsymbol{\alpha}) = \text{rk}(\boldsymbol{\beta}) = r.$$

Recordemos que $\mathbf{\Pi} \mathbf{y}_{t-1} \sim I(0)$, entonces

$$\mathbf{\Pi} \mathbf{y}_{t-1} = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-1} \sim I(0) \quad \Rightarrow \quad (\boldsymbol{\alpha}' \boldsymbol{\alpha})^{-1} \boldsymbol{\alpha}' \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-1} = \boldsymbol{\beta}' \mathbf{y}_{t-1} \sim I(0).$$

Deduciendo que en la matriz $\boldsymbol{\beta}' \mathbf{y}_{t-1} \in \mathcal{M}_{r \times 1}$ se encuentran las relaciones de cointegración. Por tal razón se puede asegurar que existen r relaciones de cointegración linealmente independientes sobre el vector \mathbf{y}_t . El valor r , como ya se dijo, se denomina rango de cointegración, $\boldsymbol{\beta}$ matriz de cointegración y $\boldsymbol{\alpha}$ matriz de carga.

Observación 2.8. Interpretación del valor del rango r .

- Si se tiene que todo $\mathbf{y}_t \sim I(0)$, entonces el rango de cointegración es igual al número de variables endógenas consideradas, $\text{rk}(\mathbf{\Pi}) = r = K$. Y si ésto se verifica, es indicador de que el proceso es estacionario.
- Si no hubiese vector de cointegración alguno, $r = 0$, implicaría que $\mathbf{\Pi} \mathbf{y}_{t-1} = \mathbf{0}$, con lo que podría deducirse que $\delta \mathbf{y}_t$ admite una representación *VAR* estacionaria.
- Si por otro lado $0 < \text{rk}(\mathbf{\Pi}) = r < K$, se podría demostrar que $\Delta \mathbf{y}_t$ no admitiría una representación *VAR* finita.
- Si existe $k^* \in \{1, \dots, K\}$ tal que $y_{k^*} \sim I(0)$, entonces este suceso se reflejaría en la matriz de cointegración con la fila k^* dada por las Deltas de Kronecker, esto es,

$$\beta_{ik^*} = \begin{cases} 1 & \text{si } i = k^*, \\ 0 & \text{si } i \neq k^*. \end{cases}$$

Debemos señalar que las matrices $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$ no son únicas, ya que si tomamos una matriz $\mathbf{B} \in \mathcal{M}_{r \times r}$ no singular, se tiene que $\mathbf{\Pi} = \boldsymbol{\alpha} \mathbf{B} (\boldsymbol{\beta} \mathbf{B}'^{-1})'$, y bastaría tomar como matriz de carga $\boldsymbol{\alpha} \mathbf{B}$ y como matriz de cointegración $\boldsymbol{\beta} \mathbf{B}'^{-1}$. Para que la unicidad este presente es necesario información adicional a la recogida directamente de la muestra.

2.2.1. Estimación y especificación

Una vez que hemos planteado el modelo genérico, es momento de estimarlo para un cierto rango de cointegración el cual debemos especificar. Las dos tareas las realizaremos en un proceso común, ya que a partir del planteamiento de un contraste para hallar el rango de cointegración se llega a la estimación deseada.

Partimos de una visión matricial de la expresión (2.9) dada por

$$\Delta \mathbf{Y} = \mathbf{\Pi} \mathbf{Y}_{-1} + \mathbf{\Gamma} \mathbf{X} + \mathbf{U} \quad (2.10)$$

siendo $\Delta \mathbf{Y} = (\Delta \mathbf{y}_1, \dots, \Delta \mathbf{y}_T) \in \mathcal{M}_{K \times T}$, $\mathbf{Y}_{-1} = (\mathbf{y}_0, \dots, \mathbf{y}_{T-1}) \in \mathcal{M}_{K \times T}$, $\mathbf{\Gamma} = (\mathbf{\Gamma}_1 : \dots : \mathbf{\Gamma}_{p-1}) \in \mathcal{M}_{K \times K(p-1)}$, $\mathbf{X} = (\mathbf{X}_0 : \dots : \mathbf{X}_{T-1}) \in \mathcal{M}_{K(p-1) \times T}$ con $\mathbf{X}_{t-1} = (\Delta \mathbf{y}_{t-1}, \dots, \Delta \mathbf{y}_{t-p+1})$ y $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_T) \in \mathcal{M}_{K \times T}$. Con esta expresión y por medio del método *OLS*, llegamos a la estimación de la matriz general de los parámetros *short-run*,

$$\hat{\mathbf{\Gamma}}_{OLS} = (\Delta \mathbf{Y} - \mathbf{\Pi} \mathbf{Y}_{-1}) \mathbf{X}' (\mathbf{X} \mathbf{X}')^{-1}. \quad (2.11)$$

Si sustituimos en (2.10) este estimador de la matriz $\mathbf{\Gamma}$, llegamos a,

$$\begin{aligned} \Delta \mathbf{Y} &= \mathbf{\Pi} \mathbf{Y}_{-1} + \mathbf{\Gamma} \mathbf{X} + \mathbf{U} \\ \Delta \mathbf{Y} (\mathbf{I}_T - \mathbf{X}' (\mathbf{X} \mathbf{X}')^{-1} \mathbf{X}) &= \mathbf{\Pi} \mathbf{Y}_{-1} (\mathbf{I}_T - \mathbf{X}' (\mathbf{X} \mathbf{X}')^{-1} \mathbf{X}) + \mathbf{U} \\ \Delta \mathbf{Y} \mathbf{M} &= \mathbf{\Pi} \mathbf{Y}_{-1} \mathbf{M} + \mathbf{U} \end{aligned}$$

siendo $\mathbf{M} = \mathbf{I}_T - \mathbf{X}' (\mathbf{X} \mathbf{X}')^{-1} \mathbf{X} \in \mathcal{M}_{T \times T}$. Ahora en base a las notaciones introducidas definimos las siguiente matrices,

$$\mathbf{S}_{00} = \frac{1}{T} \Delta \mathbf{Y} \mathbf{M} \Delta \mathbf{Y}', \quad \mathbf{S}_{11} = \frac{1}{T} \mathbf{Y}_{-1} \mathbf{M} \mathbf{Y}_{-1}' \quad \text{y} \quad \mathbf{S}_{01} = \frac{1}{T} \Delta \mathbf{Y} \mathbf{M} \mathbf{Y}_{-1}'.$$

Con ellas y sabiendo que $\mathbf{\Pi} = \mathbf{\alpha} \mathbf{\beta}'$, se tiene que el estimador mediante *OLS* de la matriz de carga está dada por,

$$\hat{\mathbf{\alpha}}_{OLS} = \mathbf{S}_{01} \mathbf{\beta} (\mathbf{\beta}' \mathbf{S}_{11} \mathbf{\beta})^{-1}, \quad (2.12)$$

y en base a ella,

$$\hat{\mathbf{\Sigma}}_{\mathbf{u}} = \mathbf{S}_{00} - \hat{\mathbf{\alpha}}_{OLS} \mathbf{\beta}' \mathbf{S}_{01}' = \mathbf{S}_{00} - \mathbf{S}_{01} \mathbf{\beta} (\mathbf{\beta}' \mathbf{S}_{11} \mathbf{\beta})^{-1} \mathbf{\beta}' \mathbf{S}_{01}'. \quad (2.13)$$

En este punto planteamos el contraste que mencionábamos al comienzo. Para cualquier $r \leq K$ se formula la hipótesis nula $H_0 \equiv \text{rk}(\mathbf{\Pi}) \leq r$, para lo cual S. Johansen en [Johansen, S. 1995] propone el estimador de razón de verosimilitud,

$$-2 \log(Q) = -T \sum_{i=r+1}^K \log(1 - \hat{\lambda}_i) \sim \chi^2(m), \quad (2.14)$$

con ciertos grados de libertad m que no especificaremos y siendo los $\hat{\lambda}_i$ los $K - r$ autovalores más pequeños de la matriz $\mathbf{S}_{01}' \mathbf{S}_{00}^{-1} \mathbf{S}_{01}$ con respecto a \mathbf{S}_{11} , esto es, las $K - r$ soluciones más pequeñas de la ecuación

$$|\lambda \mathbf{S}_{11} - \mathbf{S}_{01}' \mathbf{S}_{00}^{-1} \mathbf{S}_{01}| = 0. \quad (2.15)$$

Este contraste se realizará para todo los valores $r \leq K$, en una secuencia de hipótesis nulas que terminará con el primer no rechazo de ésta, y se determina, por tanto, el rango de cointegración r correspondiente. Señalar que si el rechazo se produce cuando $r = 0$ entonces recurrimos al segundo punto de la Observación 2.8, pero si por el contrario resulta que todos los test son rechazados salvo para $r = K$ nos disponemos en el tercer punto de la Observación 2.8.

Una vez hallado el rango de cointegración, continuamos con el proceso de estimación. Retomamos la idea de la expresión (2.15) y definimos el estimador de la matriz de cointegración como $\hat{\beta} = (\mathbf{v}_1, \dots, \mathbf{v}_r)$ siendo $\mathbf{v}_1, \dots, \mathbf{v}_r$ los autovectores de los r más grandes autovalores resultantes, los no empleados en el estadístico (2.14). Señalar que $\hat{\beta}$, por definición, verifica que $\hat{\beta}' \mathbf{S}_{11} \hat{\beta} = \mathbf{I}_r$, también comentar que este proceso de estimación de la matriz de cointegración se denomina de rango reducido, técnica que se basa en la idea de reducir la información relativa a las T observaciones en el espacio K -dimensional a un espacio de menor dimensión r -dimensional, siendo éste el relativo a las r relaciones de cointegración de nuestro conjunto de series temporales. Una vez definida esta estimación, podemos sustituirla en la expresión (2.12),

$$\hat{\alpha}_{OLS} = \mathbf{S}_{01} \hat{\beta} (\hat{\beta}' \mathbf{S}_{11} \hat{\beta})^{-1} = \mathbf{S}_{01} \hat{\beta},$$

y en la de $\hat{\Sigma}_u$, expresión (2.13),

$$\hat{\Sigma}_u = \mathbf{S}_{00} - \mathbf{S}_{01} \hat{\beta} (\hat{\beta}' \mathbf{S}_{11} \hat{\beta})^{-1} \hat{\beta}' \mathbf{S}'_{01} = \mathbf{S}_{00} - \mathbf{S}_{01} \hat{\beta} (\mathbf{S}_{01} \hat{\beta})' = \mathbf{S}_{00} - \hat{\alpha}_{OLS} \hat{\alpha}'_{OLS}.$$

Por último se puede hallar $\hat{\Pi}$ sin más que realizar el producto entre $\hat{\alpha}_{OLS}$ y $\hat{\beta}$, de este modo $\hat{\Pi} = \hat{\alpha}_{OLS} \hat{\beta}' = \mathbf{S}_{01} \hat{\beta} \hat{\beta}'$, y con ello retomando (2.11) obtenemos que $\hat{\Gamma}_{OLS} = (\Delta \mathbf{Y} - \hat{\Pi} \mathbf{Y}_{-1}) \mathbf{X}' (\mathbf{X} \mathbf{X}')^{-1}$. Señalar que bajo condiciones de gaussianidad y tomando la condición dada por los valores necesarios de la premuestra, los estimadores hallados son consistentes y asintóticamente conjuntamente normales, esto es,

$$\sqrt{T} \text{vec} \left(\hat{\Gamma}_1 : \dots : \hat{\Gamma}_{p-1} \right) \stackrel{d}{\sim} \mathcal{N}(\mathbf{\Gamma}, \mathbf{\Sigma}_{\hat{\Gamma}}) \quad \text{y} \quad \sqrt{T} \text{vec} \left(\hat{\Pi} \right) \stackrel{d}{\sim} \mathcal{N}(\mathbf{\Pi}, \mathbf{\Sigma}_{\hat{\Pi}}).$$

Siendo vec el operador que toma los elementos de una matriz por columnas y los dispone en un vector.

Observación 2.9. Sobre las distribuciones mostradas.

- Si la matriz de covarianzas $\mathbf{\Sigma}_{\hat{\Gamma}}$ es no singular entonces la inferencia clásica sobre el estimador correspondiente es válida.
- Se tendrá que la matriz $\mathbf{\Sigma}_{\hat{\Pi}}$ es singular si $\text{rk}(\mathbf{\Pi}) = r \in (0, K)$.
- Si no existen restricciones o condiciones adicionales, el estimador conjunto $\hat{\Pi} = \hat{\alpha} \hat{\beta}$ es consistente, pero por separado $\hat{\alpha}$ y $\hat{\beta}$ no lo son.

Un ejemplo sobre la imposición de restricciones en este tema, es la suposición de que la primera parte de la matriz β es la identidad, esto es, $\beta' = (\mathbf{I}_r : \beta'_{(K-r)})$, donde $\beta_{(K-r)} \in \mathbb{M}_{K-r \times r}$. Esta normalización debe ser realizada con mucho cuidado, pues si se hace de forma arbitraria podría ocurrir que se incurrirá en divisiones cuyos coeficientes pertenecen a la relación de cointegración y que por tanto son en realidad cero. De hecho, una buena organización previa de las variables puede ayudar a analizar la cointegración entre subconjuntos de variables que antes estaban esparcidas por el sistema. También facilitaría una mejor interpretación de las variables desde un punto de vista económico. En cuanto a la estimación de los parámetros $\beta'_{(K-r)}$ decir que éstos tienen una distribución asintótica que es una normal multivariante bajo la normalización adecuada, más concretamente, particionamos \mathbf{y}_t en $\mathbf{y}_t = \left(\mathbf{y}_t^{(1)}, \mathbf{y}_t^{(2)} \right)$ donde $\mathbf{y}_t^{(1)} \in \mathbb{R}^r$ y $\mathbf{y}_t^{(2)} \in \mathbb{R}^{K-r}$, y obtenemos el siguiente resultado

$$\text{vec} \left(\hat{\beta}'_{(K-r)} \right) \approx \mathcal{N} \left(\text{vec} \left(\beta'_{(K-r)} \right), \mathbf{\Omega} \right).$$

Para cierta matriz de covarianzas $\mathbf{\Omega}$ que depende de la matriz de covarianzas de los residuos $\hat{\Sigma}_u$, con lo que si la estimamos por máxima verosimilitud obtenemos $\hat{\mathbf{\Omega}}_{ML}$ y sustituimos en la expresión de la distribución dada, este resultado puede ser empleado para contrastar hipótesis compuestas sobre $\beta'_{(K-r)}$. Para una matriz $\mathbf{R} \in \mathcal{M}_{r \times (K-r)}$ de rango J y $r \in \mathbb{R}^J$, se define el test en base a la hipótesis nula

$$H_0 \equiv \mathbf{R} \text{vec} \left(\beta'_{(K-r)} \right) = r,$$

cuyo estadístico tiene una distribución χ^2 con J grados de libertad y está dado por

$$\lambda_{\mathbf{R}} = \left(\mathbf{Rvec} \left(\hat{\boldsymbol{\beta}}'_{(K-r)} \right) - r \right)' \left(\mathbf{R} \hat{\boldsymbol{\Omega}}_{ML} \mathbf{R}' \right)^{-1} \left(\mathbf{Rvec} \left(\hat{\boldsymbol{\beta}}'_{(K-r)} \right) - r \right) \sim \chi^2(J).$$

Partiendo de lo cual se llegaría a cualquier planteamiento que se desee realizar en cuanto a la matriz de cointegración.

2.3. Diagnósis y predicción de los modelos

Hasta este punto se han planteado sendos modelos de ciertos parámetros para explicar nuestras series temporales, lo que se debe hacer a continuación, como en el caso univariante, es ver si tales modelos son correctos. A lo largo de esta sección mostraremos algunas de las técnicas que se deben emplear a la hora de comprobar que el ajuste realizado sobre nuestros datos con la estimación que hemos desarrollado, es válido. Como hemos comentado, las herramientas que a continuación se exponen pueden ser empleadas tanto para un proceso descrito por un *VAR* como para un *VECM*, siempre y cuando se adapten correctamente a cada uno, debido a que este segundo es, en base a ciertas transformaciones, el primero.

Las comprobaciones que haremos serán referentes a los residuos que nuestro ajuste genera, lo deseable será que éstos se comporten según las hipótesis que se plantean en los modelos, es decir, que se asocien a un proceso de ruido blanco, Definición 1.4, y en particular uno gaussiano. Por ello contrastaremos la incorrelación, normalidad y homocedasticidad de dichos residuos. Esperando que no existan evidencias para el rechazo de dichas hipótesis. Por último, veremos si existe estabilidad estructural en nuestro ajuste, comprobando así que no existe ningún problema en relación con un *break points* que pueda desvirtuar la estimación realizada. Una vez todos estas comprobaciones se hayan realizado con un resultado satisfactorio, se pasará a la etapa de predicción y del análisis de la causalidad que luego comentaremos, de no ser así debería de plantearse un nuevo modelo con su correspondiente estimación.

Correlación de los residuos

Para comprobar la posible existencia de cierto grado de correlación entre los residuos $\hat{\mathbf{u}} = (\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ de nuestro modelo, planteamos dos posibles caminos a seguir. El primero está basado en el test de *Portmanteau* y el siguiente en multiplicadores de Lagrange (*LM*). De ambos, mostraremos el estimador original y luego comentamos una modificación empleada para muestras pequeñas.

Comenzamos por el test de *Portmanteau*, el cual establece las siguientes hipótesis para su realización

$$H_0 \equiv \mathbb{E}(\mathbf{u}_t \mathbf{u}'_{t-i}) \forall i = 0, \dots, h > p,$$

$$H_1 \equiv \exists i = 1, \dots, h > p \quad \text{tal que} \quad \mathbb{E}(\mathbf{u}_t \mathbf{u}'_{t-i}) \neq 0.$$

Definiéndose como estadístico el valor

$$Q_h = T \sum_{j=1}^h \frac{1}{T-j} \text{tr} \left(\hat{\mathbf{C}}'_j \hat{\mathbf{C}}_0^{-1} \hat{\mathbf{C}}_j \hat{\mathbf{C}}_0^{-1} \right) \sim \chi^2(K^2 h - n),$$

siendo $\hat{\mathbf{C}}_i = \frac{1}{T} \sum_{t=i+1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}'_{t-i}$ y n el parámetro que representa el número de coeficientes a estimar en el *VAR(p)* sin contabilizar los términos deterministas. A este tipo de test se le realiza una modificación para ser utilizado con muestras pequeñas, cambios que asemejan el estadístico resultante al definido por Ljung-Box bien conocido en univariante, por ello no lo mostraremos.

La otra vía sería en base a *LM*. Para ello definimos los residuos como

$$\hat{\mathbf{u}}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{C} \mathbf{D}_t + \mathbf{B}_1 \hat{\mathbf{u}}_{t-1} + \dots + \mathbf{B}_h \hat{\mathbf{u}}_{t-h} + \boldsymbol{\varepsilon}_t, \quad (2.16)$$

siendo ε_t un proceso de ruido blanco donde sus estimaciones serán denotadas por $\hat{\varepsilon}_t$ y cuya matriz de covarianzas viene estimada de forma análoga a la dada en la segunda expresión de (2.9) sustituyendo $\hat{\mathbf{u}}_t$ por $\hat{\varepsilon}_t$, obteniendo $\tilde{\Sigma}_\varepsilon$. Con un razonamiento idéntico pero partiendo del modelo (2.16) sin los valores $\hat{\mathbf{u}}_{t-i}$ para todo i y denotando los residuos de dicho modelo por $\hat{\varepsilon}_t^R$, llegamos a obtener la estimación de la matriz de covarianzas $\tilde{\Sigma}_R$. Señalar que si se parte de la expresión (2.16), es análogo a definir un modelo VAR sobre los residuos. Se plantea pues las siguientes hipótesis para contrastar la correlación de los residuos originales

$$H_0 \equiv \mathbf{B}_i = 0 \quad \forall i = 1, \dots, h > p,$$

$$H_1 \equiv \exists i = 1, \dots, h > p \quad \text{tal que} \quad \mathbf{B}_i \neq 0.$$

En base a éstas, Breusch y Godfrey plantean el siguiente estadístico

$$LM_h = T \left(K - \text{tr} \left(\tilde{\Sigma}_\varepsilon \tilde{\Sigma}_R^{-1} \right) \right) \sim \chi^2(hK^2).$$

Que, como antes con el test de *Portmanteau*, es modificado para muestras de tamaño reducido, transformación que fue llevada a cabo por Edgerton y Shukur dando como resultado el siguiente estadístico

$$FLM_h = \left(\left(\frac{|\tilde{\Sigma}_R|}{|\tilde{\Sigma}|} \right)^{1/s} - 1 \right) \cdot \frac{Ns - q}{Km} \sim F(hK^2, Ns - q), \quad (2.17)$$

siendo $q = \frac{1}{2}Km - 1$, $N = T - K - m - \frac{1}{2}(K - m + 1)$, $m = Kh$, n el número total de regresores del sistema original y

$$s = \left(\frac{K^2m^2 - 4}{K^2 + m^2 - 5} \right)^{1/2}.$$

Observación 2.10. Tanto el test de Breusch-Godfrey como el de Edgerton-Shukur basados en LM son empleados para valores de h pequeños, en cambio el contraste de *Portmanteau* es comúnmente utilizado cuando h es grande.

Normalidad

La idea para comprobar la normalidad es transformar la distribución normal conjunta para obtener componentes independientes sobre los residuos y aplicarles el test de Jarque-Bera univariante a ellas. Sean $\hat{\mathbf{u}}_t$ con $t = 1, \dots, T$ las estimaciones de los residuos del VAR, se tiene por tanto en base a (2.16) que

$$\tilde{\Sigma}_u = \frac{1}{T} \sum_{t=1}^T (\hat{\mathbf{u}}_t - \bar{\mathbf{u}})(\hat{\mathbf{u}}_t - \bar{\mathbf{u}})'$$

con $\bar{\mathbf{u}}$ la correspondiente media de los residuos. En este momento se emplea la descomposición de Cholesky sobre $\tilde{\Sigma}_u$ para poder ser calculada, entonces se toma una matriz $\tilde{\mathbf{P}}$ triangular inferior con diagonal principal positiva tal que $\tilde{\Sigma}_u = \tilde{\mathbf{P}}\tilde{\mathbf{P}}'$. Ahora, en base a esa descomposición, se hallan los momentos de orden tres (asimetría) y de orden cuatro (kurtosis) de los residuos estandarizados $\hat{\mathbf{u}}_t^s = (\hat{u}_{1t}^s, \dots, \hat{u}_{Kt}^s) = \tilde{\mathbf{P}}^{-1}(\hat{\mathbf{u}}_t - \bar{\mathbf{u}})$, esto es,

$$\begin{aligned} \mathbf{b}_1 &= (b_{11}, \dots, b_{1K})' \quad \text{con} \quad b_{1j} = \frac{1}{T} \sum_{t=1}^T (\hat{u}_{jt}^s)^3, \text{ y} \\ \mathbf{b}_2 &= (b_{21}, \dots, b_{2K})' \quad \text{con} \quad b_{2j} = \frac{1}{T} \sum_{t=1}^T (\hat{u}_{jt}^s)^4. \end{aligned}$$

Definiéndose con ellos los siguientes estadísticos

$$s_3^2 = \frac{T}{6} \mathbf{b}_1' \mathbf{b}_1 \quad \text{y} \quad s_4^2 = \frac{T}{24} (\mathbf{b}_2 - 3K)' (\mathbf{b}_2 - 3K).$$

Por último, con la suma de ambos se define el estadístico de Jarque-Bera para el caso multivariante,

$$JB_K = s_3^2 + s_4^2 \sim \chi^2(2K).$$

Para una mejor visión de la normalidad de nuestro ajuste, los test univariantes para contrastar la normalidad pueden ser aplicados a los residuos de las expresiones de cada serie de forma individual.

Heterocedasticidad

Para contrastar el efecto de heterocedasticidad en estos modelos multivariantes, existen los test *ARCH* (*Autorregresive Conditional Heteroscedasticity*) tanto univariantes como en más dimensiones. El que mostramos en estos momentos es el *ARCH-LM* multivariante basado en multiplicadores de Lagrange, señalar que el caso particular de una dimensión se obtiene de forma inmediata. Éste está basado en el siguiente modelo de regresión multivariante

$$vech(\hat{\mathbf{u}}_t \hat{\mathbf{u}}_t') = \boldsymbol{\beta}_0 + \mathbf{B}_1 vech(\mathbf{u}_{t-1} \mathbf{u}_{t-1}') + \cdots + \mathbf{B}_q vech(\mathbf{u}_{t-q} \mathbf{u}_{t-q}') + \mathbf{v}_t, \quad (2.18)$$

donde $\boldsymbol{\beta}_0 \in \mathbb{R}^{\frac{1}{2}K(K+1)}$ y $\mathbf{B}_i \in \mathcal{M}_{K(K+1) \times \frac{1}{2}K(K+1)}$ para todo $i = 1, \dots, q$ son matrices de coeficientes constantes a estimar, cada elemento de \mathbf{v}_t es un proceso de ruido blanco y el operador vectorial *vech* es conocida como *half-vectorization*, similar al ya introducido *vec*. En base a este modelo se hace la reflexión de que no existirá una variabilidad condicionada en los residuos \mathbf{v}_t si todas las matrices de coeficientes son nulas. Por tanto se propone el consiguiente contraste definido a través de la hipótesis $H_0 \equiv \mathbf{B}_i = 0$ para todo $i = 1, \dots, q$, contra $H_1 \equiv \exists j = 1, \dots, q$ tal que $\mathbf{B}_j \neq 0$. Cuyo estadístico puede comprobarse, viene dado por

$$VARCH_{LM}(q) = \frac{1}{2}TK(K+1)R_m^2 \sim \chi^2\left(\frac{qK^2(K+1)^2}{4}\right),$$

con $R_m^2 = 1 - 2\frac{1}{K(K+1)}\text{tr}(\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\Omega}}_0^{-1})$, donde $\hat{\boldsymbol{\Omega}}$ representa la estimación de la matriz de covarianzas residual del modelo introducido (2.18) bajo la hipótesis alternativa y $\hat{\boldsymbol{\Omega}}_0$ sobre la nula, con $q = 0$.

Estabilidad Estructural

Para ver si nuestra estimación del modelo *VAR* ha sido correcta en el sentido de poseer estabilidad estructural en ella, es decir no existen *break points* que alteren el comportamiento global de las series, se aplica el test de Chow para cambios estructurales. Para su aplicación suponemos que en un cierto instante T_C se produce un cambio en la estructura y se toman $T_1, T_2 \in \mathcal{I} = [1, T]$ verificando que

$$T_1 < T_C \quad \text{y} \quad T_2 \leq T - T_C.$$

Partiendo de ellos se estima el modelo para tres intervalos de tiempo diferentes, el total $\mathcal{I} = [1, T]$, $[1, T_1]$ y el intervalo final $[T_2 + 1, T]$. Una vez realizadas las sendas estimaciones se obtendrán los residuos $\hat{\mathbf{u}}_t, \hat{\mathbf{u}}_t^{(1)}$ y $\hat{\mathbf{u}}_t^{(2)}$, respectivamente. Siendo sus correspondientes matrices de covarianzas dadas por

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{u}} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t', \quad \tilde{\boldsymbol{\Sigma}}_{(1)} = \frac{1}{T_1} \sum_{t=1}^{T_1} \hat{\mathbf{u}}_t^{(1)} \hat{\mathbf{u}}_t^{(1)'} \quad \text{y} \quad \tilde{\boldsymbol{\Sigma}}_{(2)} = \frac{1}{T_2} \sum_{t=T-T_2+1}^T \hat{\mathbf{u}}_t^{(2)} \hat{\mathbf{u}}_t^{(2)'}$$

Y partiendo de ellas, tomando los diferentes residuos, se definen dos matrices a mayores,

$$\tilde{\boldsymbol{\Sigma}}_{(1,2)} = \frac{1}{T_1} \sum_{t=1}^{T_1} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' + \frac{1}{T_2} \sum_{t=T-T_2+1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \quad \text{y} \quad \tilde{\boldsymbol{\Sigma}}_{1,2} = \frac{1}{T_1 + T_2} \left(\sum_{t=1}^{T_1} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' + \sum_{t=T-T_2+1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \right).$$

Una vez definida esta notación, se introducen dos modificaciones diferentes del test de Chow original, la primera es el llamado *sample-split*, (*SS*) el cual es empleado cuando la matriz de covarianzas del ruido blanco es constante, definiéndose entonces el siguiente estadístico,

$$\lambda_{SS} = (T_1 + T_2) \left(\log \det \tilde{\boldsymbol{\Sigma}}_{1,2} - \log \det \frac{1}{T_1 + T_2} \left(T_1 \tilde{\boldsymbol{\Sigma}}_{(1)} + T_2 \tilde{\boldsymbol{\Sigma}}_{(2)} \right) \right)$$

Por otra parte, está el conocido como *break-point* (*BP*) empleado cuando $\tilde{\Sigma}_{\mathbf{u}}$ no es constante, siendo si correspondiente estadístico dado por

$$\lambda_{BP} = (T_1 + T_2) \log \det \tilde{\Sigma}_{(1,2)} - T_1 \log \det \tilde{\Sigma}_{(1)} - T_2 \log \det \tilde{\Sigma}_{(2)}.$$

Señalar que ambos estadísticos se aproximan mediante una distribución χ^2 con grados de libertad dependiendo de la suposición de ser o no constante la matriz de covarianzas $\tilde{\Sigma}_{\mathbf{u}}$. Una última modificación del test de Chow es el test de predicción (*CF*) en el cual la situación planteada es que el *break-point* en el instante T_C es de interés y $\tilde{\Sigma}_{\mathbf{u}}$ no es constante, con esta consideración, el correspondiente estadístico se define como

$$\lambda_{CF} = \left[\left(\left(\frac{T}{T_1} \right)^K \frac{|\tilde{\Sigma}_{\mathbf{u}}|}{|\tilde{\Sigma}_{(1)}|} \right)^{1/s} - 1 \right] \cdot \frac{Ns - q}{KK^*} \sim F(K^*K, Ns - q), \quad (2.19)$$

siendo $q = \frac{1}{2}KK^* - 1$, $N = T - K - K_1 - \frac{1}{2}(K - K^* + 1)$, $K^* = T - T_1$, K_1 el número total de regresores del sistema original y

$$s = \left(\frac{K^2K^{*2} - 4}{K^2 + K^{*2} - 5} \right)^{1/2}.$$

Observación 2.11. Si ahora nos fijamos en la definición del estadístico introducido por Edgerton y Shukur para el contraste de la correlación de los residuos con *LM* en (2.17), podemos ver grandes similitudes con el estadístico que acabamos de mostrar en (2.19). E incluso realizando las identificaciones adecuadas,

$$\tilde{\Sigma}_{\mathbf{R}} = T^K \tilde{\Sigma}_{\mathbf{u}}, \quad \tilde{\Sigma}_{\boldsymbol{\varepsilon}} = T_1^K \tilde{\Sigma}_{(1)}, \quad m = K^*, \quad n = K_1 \quad \text{y} \quad h = \frac{K^*}{K},$$

se llega a la igualdad de estadísticos y de distribuciones de aproximación. Señalar que esta observación es simplemente a título ilustrativo y como curiosidad, pues no será empleada en la práctica.

Pero decir que el desarrollo de Chow tiene el problema del punto T_C , ya que a priori se desconoce y con ello puede que las estimaciones y conclusiones que se hagan puedan ser erróneas. Por tanto, se plantea otro posible método a la hora de comprobar la estabilidad estructural del proceso estimado, éste es mediante *OLS-CUSUM*, el cual se basa en el cálculo del proceso de fluctuaciones empíricas mediante *OLS*. Sin ahondar más, indicar que en [Ploberger, W. 1992] se puede ver esta herramienta en detalle. Comentar que en la práctica se empleara esta técnica por ser mucho más simple que lo planteado por Chow debido a que no presupone ningún instante de la naturaleza de T_C .

Predicción

Validado nuestro modelo, podemos realizar previsiones en base a él partiendo de la suposición de que queremos hallar la predicción de nuestro modelo a un horizonte h pudiendo variar este valor entre 1 y un cierto número n natural. Entonces, el modelo que se plantea es análogo al descrito en (2.1) pero estimado y h instante hacia adelante con respecto al final del intervalo \mathcal{I} , esto es,

$$\mathbf{y}_{T+h|T} = \hat{\mathbf{A}}_1 \mathbf{y}_{T+h-1|T} + \cdots + \hat{\mathbf{A}}_p \mathbf{y}_{T+h-p|T} + \hat{\mathbf{C}} \hat{\mathbf{D}}_{T+h|T} + \hat{\mathbf{u}}_{T+h|T}.$$

En base a este modelo, podemos calcular la matriz de covarianzas del error de predicción como sigue

$$\text{Cov} \left[\begin{pmatrix} \mathbf{y}_{T+1} + \mathbf{y}_{T+1|T} \\ \vdots \\ \mathbf{y}_{T+h} + \mathbf{y}_{T+h|T} \end{pmatrix} \right] = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \hat{\phi}_1 & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \hat{\phi}_{h-1} & \hat{\phi}_{h-2} & \cdots & \mathbf{I} \end{pmatrix} (\tilde{\Sigma}_{\mathbf{u}} \otimes \mathbf{I}_h) \begin{pmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \hat{\phi}_1 & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \hat{\phi}_{h-1} & \hat{\phi}_{h-2} & \cdots & \mathbf{I} \end{pmatrix}'.$$

donde las matrices $\hat{\phi}_i$ para $i = 1, \dots, h-1$ son las definidas en la expresión (2.4). Si ahora atendemos a la distribución normal que se sigue de los párrafos anteriores sobre los residuos, podemos definir intervalos marginales de predicción para cada serie como en el caso univariante,

$$(y_{k,T+h|T} \pm \mathcal{Z}_{1-\alpha/2} \hat{\sigma}_k(h)) \quad \text{para } k = 1, \dots, K, \quad (2.20)$$

siendo, como es usual, α el nivel de significación, $\mathcal{Z}_{1-\alpha/2}$ el cuantil de la normal de nivel $1 - \alpha$ en un contraste bilateral y $\hat{\sigma}_k(h)$ la desviación típica de la k -ésima variable en la previsión a horizonte h .

Análisis de causalidad

En 1969 C. Granger con el artículo [Granger, C. 1969] introdujo un concepto de causalidad entre variables que ha tomado bastante importancia en la literatura econométrica. En él se toma el vector de variables \mathbf{y}_t y se divide en dos subvectores $\mathbf{y}_{1t} \in \mathbb{R}^{K_1}$ y $\mathbf{y}_{2t} \in \mathbb{R}^{K_2}$ con $K_2 = K - K_1$. Granger define que \mathbf{y}_{2t} es causal en el tiempo a \mathbf{y}_{1t} si la primera ayuda a predecir la segunda, más formalmente.

Definición 2.12. El subvector \mathbf{y}_{2t} es no causal de Granger, si y solo si,

$$\mathbf{y}_{1,t+h|\mathcal{F}_t} = \mathbf{y}_{1,t+h|\mathcal{F}_t \setminus \{\mathbf{y}_{2s} | s \leq t\}} \quad \text{para } h = 1, 2, \dots$$

donde $\mathbf{y}_{1,t+h|\mathcal{F}_t}$ denota la predicción a horizonte h de \mathbf{y}_{1t} con origen t basado en el conjunto de información relevante hasta dicho instante representado por la filtración \mathcal{F}_t .

Esto es, \mathbf{y}_{2t} será no causal de Granger si eliminando la información que éste aporte, no se modifica la predicción de \mathbf{y}_{1t} a cualquier horizonte. Por el contrario, \mathbf{y}_{2t} será causal de Granger si la Definición 2.12 no se verifica para al menos un h .

Tomemos como ejemplo el modelo $VAR(p)$ dado en (2.1) y lo reescribimos separando \mathbf{y}_t en los dos subvectores comentados,

$$\mathbf{y}_t = (\mathbf{y}_{1t}, \mathbf{y}_{2t})' = \sum_{i=1}^p \begin{pmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{pmatrix} (\mathbf{y}_{1,t-i}, \mathbf{y}_{2,t-i})' + \mathbf{CD}_t + (\mathbf{u}_{1t}, \mathbf{u}_{2t}).$$

A la vista de tal expresión vemos que la condición de no causalidad de Granger dada en la Definición 2.12 es equivalente a que se verifique

$$\alpha_{12,i} = 0 \quad \forall i = 1, \dots, p, \quad (2.21)$$

y será causal de Granger si existe al menos un $i = 1, \dots, p$ tal que $\alpha_{12,i} \neq 0$. Ésto construye de forma natural un contraste con las siguientes hipótesis

$$H_0 \equiv \alpha_{12,i} = 0 \quad \forall i = 1, \dots, p,$$

$$H_1 \equiv \exists i = 1, \dots, p \quad \text{tal que } \alpha_{12,i} \neq 0.$$

Cuyo estadístico viene dado por un θ con distribución límite $F(pK_1K_2, KT - n^*)$, siendo n^* el número de parámetros a estimar del $VAR(p)$ incluyendo los términos deterministas.

Observación 2.13. Podemos pensar en un procedimiento similar, pero en este caso dividiendo el vector principal en tres subvectores. Con esta idea el VAR resultaría como sigue,

$$\mathbf{y}_t = (\mathbf{y}_{1t}, \mathbf{y}_{2t}, \mathbf{y}_{3t})' = \sum_{i=1}^p \begin{pmatrix} \alpha_{11,i} & \alpha_{12,i} & \alpha_{13,i} \\ \alpha_{21,i} & \alpha_{22,i} & \alpha_{23,i} \\ \alpha_{31,i} & \alpha_{32,i} & \alpha_{33,i} \end{pmatrix} (\mathbf{y}_{1,t-i}, \mathbf{y}_{2,t-i}, \mathbf{y}_{3,t-i})' + \mathbf{CD}_t + (\mathbf{u}_{1t}, \mathbf{u}_{2t}, \mathbf{u}_{3t}).$$

y no habría problema en establecer una expresión como la descrita en (2.21). Pero ésta, ahora, no es equivalente a la no causalidad de Granger según la Definición 2.12, sino que sería equivalente a la modificación

$$\mathbf{y}_{1,t+1|\mathcal{F}_t} = \mathbf{y}_{1,t+1|\mathcal{F}_t \setminus \{\mathbf{y}_{2s} | s \leq t\}},$$

es decir, tomando el horizonte $h = 1$. Ésto es debido a las conexiones indirectas que podrían existir entre los subvectores. Por lo que para una división mayor de 2 subvectores, la definición de la no causalidad de Granger se modificará en base a restricciones no lineales más complicadas, las cuales no especificaremos en este desarrollo.

Por otro lado, se define otro concepto acerca de la causalidad de un sistema, ésta será una modificación instantánea de la introducida al comienzo.

Definición 2.14. El subvector \mathbf{y}_{2t} se dice causal instantánea en el tiempo a otra \mathbf{y}_{1t} , si y solo si,

$$\mathbf{y}_{1,t+1|\mathcal{F}_t} = \mathbf{y}_{1,t+1|\mathcal{F}_t \cup \mathbf{y}_{2,t+1}},$$

donde los elementos de la expresión ya han sido comentados en la anterior definición.

Una posible interpretación puede ser basada en los residuos del modelo, $\mathbf{u}_t = (\mathbf{u}_{1t}, \mathbf{u}_{2t})'$. Con lo que podremos decir que, \mathbf{y}_{2t} no es causal instantánea para \mathbf{y}_{1t} , si \mathbf{u}_{1t} y \mathbf{u}_{2t} son incorrelados, y si por el contrario verificarán la correlación, se cumple la condición de causalidad instantánea. Consecuentemente la Definición 2.14 es simétrica en el sentido de que si un subvector es causal instantáneo sobre otro, el otro lo es sobre el uno, suceso que no ocurría con la causalidad de Granger. Decir también que, con este concepto será de nuevo modificado con no linealidad si se considera una división mayor que 2 en el vector de partida.

Capítulo 3

Regresión con Series Temporales

Una vez desarrollado en los dos primeros capítulos el estudio de series temporales para poder realizar predicciones de las mismas, entramos en este nuevo capítulo donde nos situaremos en la idea de generar modelos de regresión entre una serie de tiempo que consideraremos respuesta y un conjunto de procesos temporales conformando variables explicativas.

Este capítulo posee dos partes muy diferenciadas, la primera se basa en intentar encontrar relaciones lineales entre la respuesta y una covariable determinada en base a retardos temporales (Sección 3.1). Y la segunda parte es en relación al método de estimación de modelos de regresión por mínimos cuadrados generalizados (*GLS*) dispuesto en la Sección 3.2. Ambos se expondrán por separado por simplicidad, pero como se comprobará en la aplicación práctica (Parte II) su conjunción generará modelos que explican mejor la situación de estudio.

3.1. Modelo dinámico con retardos

En esta sección veremos como hallar, si existen, relaciones lineales entre dos series temporales con respecto al tiempo, para así buscar modelos dinámicos, esto es, trataremos de ver si una serie temporal \mathbf{x} en un instante t se asocia linealmente con otra \mathbf{y} en el instante $t-k$, para cierto entero k . Esto se podría llevar al escenario multivariante donde se tengan K covariables $\{\mathbf{x}_1, \dots, \mathbf{x}_K\}$ y donde se trate de buscar los $\{k_1, \dots, k_K\}$ retardos para sus respectivas asociaciones con \mathbf{y} . Tomaremos como referencia básica [Cryer, J. *et al.* 2008] para este desarrollo.

Comenzamos tomando las series temporales $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ como una cierta variable respuesta y $\mathbf{x} = \{x_t, t \in \mathcal{I}\}$ como covariable. Para indagar sobre la posible estructura de correlación existente entre ellas se suponen dos procesos conjuntamente estacionarios en el sentido de la Definición 1.3 y se define la función de correlaciones cruzadas (*CCF*) con la siguiente expresión,

$$\rho_k(\mathbf{x}, \mathbf{y}) = \text{Corr}(y_t, x_{t-k}) = \text{Corr}(y_{t+k}, x_t) = \frac{\gamma_k}{\sqrt{\text{Var}(y_t)}\sqrt{\text{Var}(x_{t-k})}} \quad \text{para } k \in \mathbb{Z}.$$

siendo $\gamma_k = \text{Cov}(y_t, x_{t-k})$

Observación 3.1. Sobre la función de correlaciones cruzadas:

- El coeficiente ρ_0 mide la relación lineal contemporánea entre ambos procesos.
- En general, se tiene que $\rho_k(\mathbf{y}, \mathbf{x}) \neq \rho_{-k}(\mathbf{y}, \mathbf{x})$. Solo se obtiene la igualdad cuando se verifica $\mathbf{y} = \mathbf{x}$.
- En esta misma situación, la correlación cruzada se convierte en autocorrelación (*ACF*).

En la práctica es evidente que no podremos emplear esta función directamente, por lo que tendremos que estimarla. Para ello introducimos la función de correlaciones cruzadas muestral (*sCCF*),

$$r_k(\mathbf{y}, \mathbf{x}) = \frac{\sum_{t \in \mathcal{I}} (y_t - \bar{\mathbf{y}})(x_{t-k} - \bar{\mathbf{x}})}{\sqrt{(y_t - \bar{\mathbf{y}})^2} \sqrt{(x_t - \bar{\mathbf{x}})^2}}, \quad (3.1)$$

donde T , como ya indicamos, se puede entender como el tamaño de la serie si $\mathcal{I} = [1, T]$. Si además de la estacionariedad conjunta, suponemos que los procesos son independientes entre sí, se puede demostrar que (3.1) se distribuye en base a una distribución normal $\mathcal{N}(0, T^{-1})$. Por lo que, en este escenario un retardo k será significativo para un cierto nivel de significación α para la asociación de las series correspondientes si su estimación en valor absoluto es mayor que $|\mathcal{Z}_{1-\alpha/2} T^{-1/2}|$.

Pero esta normalidad sobre la *sCCF* solo se mantendrá en las hipótesis comentadas, estabilidad conjunta de los dos procesos considerados e independencia de los mismos, ya que si no se verifican éstas, y solo se tiene la estacionariedad de una de ellas e independencia entre ambas, puede demostrarse que

$$\text{Var}(r_k(\mathbf{y}, \mathbf{x})) \approx \frac{1}{T} \left(1 + 2 \sum_{k=1}^{\infty} \rho_k(\mathbf{y}) \rho_k(\mathbf{x}) \right).$$

En base a tal situación, si suponemos que al menos una de nuestras series temporales, por ejemplo \mathbf{y} , es un proceso de ruido blanco y que por tanto su $\rho_k(\mathbf{y}) = 0$ se tiene que $\text{Var}(r_k(\mathbf{y}, \mathbf{x})) \approx T^{-1}$. Lo cual tiene sentido pues si uno de los procesos es ruido blanco se verifica la estacionariedad conjunta trivialmente. Por otro lado, si ambas series son no estacionarias, estas expresiones para la varianza no son válidas ya que una fuerte autocorrelación en los datos es un problema para hallar dependencias entre dos procesos. Esta circunstancia derivará en lo que se conoce como correlación espuria, que a continuación mostramos como corregir. Para realizar esta corrección debemos aportar el siguiente resultado que introduce la noción de filtro de preblanqueado y que es esencial en nuestro objetivo.

Teorema 3.2. *Sea una serie temporal $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ representada por un modelo $ARIMA(p, d, q)$, si verifica la condición de estabilidad, admite una representación AR de la forma*

$$\tilde{y}_t = (1 - \pi_1^{\mathbf{y}} B - \pi_2^{\mathbf{y}} B^2 - \dots) y_t = \pi^{\mathbf{y}}(B) y_t,$$

donde $\tilde{\mathbf{y}} = \{\tilde{y}_t, t \in \mathcal{I}\}$ es un proceso de ruido blanco. El proceso de transformación de \mathbf{y} en $\tilde{\mathbf{y}}$ mediante $\pi^{\mathbf{y}}(B) = 1 - \pi_1^{\mathbf{y}} B - \pi_2^{\mathbf{y}} B^2 - \dots$, es conocido como preblanqueado.

Ahora podemos estudiar la *CCF* entre \mathbf{y} y \mathbf{x} a través del preblanqueado de ambas empleando el filtro $\pi^{\mathbf{y}}(B)$ y calculando las nuevas series $\tilde{\mathbf{y}}$ y $\tilde{\mathbf{x}} = \pi^{\mathbf{y}}(B)\mathbf{x}$ ya que una de ellas es ruido blanco y de este modo podemos emplear el punto de corte $|\mathcal{Z}_{1-\alpha/2} T^{-1/2}|$. Debemos señalar que $\tilde{\mathbf{x}}$ es un abuso del lenguaje en el sentido de que ésta serie no tendrá porque ser un ruido blanco ya que el filtro $\pi^{\mathbf{y}}(B)$ fue hallado como desarrollo autorregresivo en base al proceso \mathbf{y} sin tener en cuenta \mathbf{x} .

Se podrían hacer más observaciones a este cálculo del filtro e incluso darle una interpretación en base a coeficientes de una cierta regresión, pero para lo que la práctica requiere es suficiente con las aclaraciones dadas. Por lo que seguimos con el desarrollo del trabajo y pasamos a introducir el método de estimación *GLS*.

3.2. Generalized Least Squares

En la anterior sección del presente capítulo se ha expuesto cómo poder llegar a definir un modelo de regresión dinámico en base a retardos temporales que reflejan relaciones lineales entre la respuesta y la correspondiente variable explicativa considerada. Por otro lado, lo que se expondrá en el actual apartado es referente al método de estimación por mínimos cuadrado generalizados (*Generalized Least Squares, GLS*). Ésta será una herramienta que generaliza la técnica de mínimos

cuadrados ordinaria (*Ordinary Least Squares, OLS*) clásica, y cuya principal ventaja es que podremos suponer varianzas para los errores que difieran de una matriz diagonal constante. En particular se podrá introducir cualquier proceso estacionario de orden dos sobre tales residuos. Para poder completar el desarrollo que se muestra a continuación, se puede tomar como referencia principal el [Hamilton, J. D. 1994], o [Buse, A. 1973], [Kariya, T. *et al.* 2004] y [Kuan, C. 2004].

Comenzaremos por la suposición del clásico modelo lineal de regresión, a partir del cual realizaremos ciertas suposiciones en base a la matriz de varianzas-covarianzas de los errores, en particular, en cuanto a su posible estructura diagonal o referente al conocimiento de ésta. Entonces, partimos de tomar el modelo lineal de regresión dado por la expresión

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (3.2)$$

donde $\mathbf{y} = (y_1, \dots, y_T)' \in \mathbb{R}^T$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_K) \in \mathcal{M}_{T \times K}$ no estocástica de rango $K < T$ con $\mathbf{x} = (x_{i1}, \dots, x_{iT})' \in \mathbb{R}^T$ para $i = 1, \dots, K$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)' \in \mathbb{R}^K$, y $\mathbf{u} = (u_1, \dots, u_T)' \in \mathbb{R}^T$ verificando $\mathbb{E}(\mathbf{u}) = \mathbf{0}$ y $\text{Var}(\mathbf{u}) = \boldsymbol{\Sigma}_0$. Por lo que se tiene $\mathbf{y} \sim \mathcal{D}_T(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_0)$, que si suponemos el no rechazo de la gaussianidad por parte de los errores, se tiene $\mathbf{y} \sim \mathcal{N}_T(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_0)$.

En este punto se pueden hacer dos consideraciones acerca de la estructura de la matriz $\boldsymbol{\Sigma}_0$. En primer lugar podríamos suponer $\boldsymbol{\Sigma}_0 = \sigma^2 \mathbf{I}_T$, esto es, reflejando homocedasticidad e incorrelación en los errores, que sumándole la normalidad antes supuesta se llega a un modelo donde el método *OLS* clásico se puede aplicar sin problemas. Pero caso aparte sería si $\boldsymbol{\Sigma}_0 \neq \sigma^2 \mathbf{I}_T$, ya que *OLS* no sería el mejor camino a seguir pues el *t*-test o el *F*-test no serían válidos, por esa razón se aplica una nueva técnica como generalización de dicha herramienta, el método *GLS*.

Nos situamos en el escenario expuesto en el párrafo anterior y suponemos que $\boldsymbol{\Sigma}_0$ es una matriz conocida. Por tal hipótesis, aseguramos que existe $\mathbf{V} \in \mathcal{M}_{T \times T}$ simétrica y definida positiva tal que $\boldsymbol{\Sigma}_0 = \sigma^2 \mathbf{V}$. Nuevo elemento \mathbf{V} que, por su definición, asegura la existencia de otra matriz $\mathbf{P} \in \mathcal{M}_{T \times T}$ no singular tal que $\mathbf{V} = \mathbf{P}'\mathbf{P}$, o equivalentemente, para $\mathbf{L} = \mathbf{P}^{-1}$ verificándose $\mathbf{V}^{-1} = \mathbf{L}'\mathbf{L}$. En función de tal matriz \mathbf{L} definimos el siguiente modelo de regresión,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{u}}, \quad (3.3)$$

con $\tilde{\mathbf{y}} = \mathbf{L}\mathbf{y}$, $\tilde{\mathbf{X}} = \mathbf{L}\mathbf{X}$ y $\tilde{\mathbf{u}} = \mathbf{L}\mathbf{u}$. De este modo se tienen las siguientes propiedades referentes a los errores,

$$\begin{aligned} \mathbb{E}(\tilde{\mathbf{u}}) &= \mathbb{E}(\mathbf{L}\mathbf{u}) = \mathbf{L}\mathbb{E}(\mathbf{u}) = \mathbf{0}, \\ \mathbb{E}(\tilde{\mathbf{u}}\tilde{\mathbf{u}}') &= \mathbb{E}(\mathbf{L}\mathbf{u}(\mathbf{L}\mathbf{u})') = \mathbf{L}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{L} = \sigma^2 \mathbf{L}\mathbf{V}\mathbf{L}' \stackrel{(i)}{=} \sigma^2 \mathbf{I}_T. \end{aligned} \quad (3.4)$$

Debido la igualdad (i) a que $\mathbf{L}\mathbf{V}\mathbf{L}' = \mathbf{L}(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}' = \mathbf{L}\mathbf{L}^{-1}(\mathbf{L}')^{-1}\mathbf{L}' = \mathbf{I}_T$. Con lo que verificamos $\tilde{\mathbf{y}} \sim \mathcal{N}_T(\tilde{\mathbf{X}}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T)$, y por tanto podemos calcular el estimador clásico para el vector de parámetros del modelo,

$$\tilde{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}}, \quad (3.5)$$

cuya esperanza verifica que el estimador es insesgado y cuya varianza viene dada por,

$$\begin{aligned} \mathbb{E}(\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}') &= \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}' \mathbb{E}(\tilde{\mathbf{y}}\tilde{\mathbf{y}}') \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \\ &= \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}' \sigma^2 \mathbf{L}\mathbf{V}\mathbf{L}' \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \stackrel{(i)}{=} \sigma^2 \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1}. \end{aligned}$$

Para la misma igualdad (i) de la expresión (3.4). Definiendo, pues, su distribución

$$\tilde{\boldsymbol{\beta}} \sim \mathcal{N}_T \left(\boldsymbol{\beta}, \sigma^2 \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right)^{-1} \right).$$

Este proceso de obtención del estimador es el conocido como método *GLS*, razón por la cual podemos identificar $\boldsymbol{\beta}_{GLS} = \tilde{\boldsymbol{\beta}}$. Para este estimador se introduce el teorema de Gauss-Markov,

el cual retoma alguna salida ya vista en relación a los momentos del estadístico y asegura que es el mejor de entre todos los estimadores de β . Señalar que la demostración puede verse en [Kariya, T. *et al.* 2004].

Teorema 3.3 (Gauss-Markov). *Dado el modelo (3.2) verificándose que $\text{Var}(\mathbf{y}) = \Sigma_0$ es una matriz definida positiva. Entonces $\hat{\beta}_{GLS}$ es el mejor estimador lineal insesgado para β_0 que verifica que $\text{Var}(\hat{\beta}_{OLS}) - \text{Var}(\hat{\beta}_{GLS})$ es una matriz semidefinida positiva. Además su matriz de varianzas-covarianzas viene dada por la expresión $(\mathbf{X}'\Sigma_0^{-1}\mathbf{X})^{-1}$.*

También, y en base a la gaussianidad supuesta al inicio, podemos llegar a comprobar que el método *GLS* sobre el modelo inicial es equivalente al de máxima verosimilitud. Suponiendo que $\mathbf{y} \sim \mathcal{N}_T(\mathbf{X}\beta, \sigma^2\mathbf{V})$, la función de log-verosimilitud asociada viene dada por,

$$L(\beta) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log |\sigma^2\mathbf{V}| - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

Expresión que puede ser reescrita en base a las igualdades definidas en el modelo (3.3) como,

$$L(\beta) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \log |\det(\mathbf{L})| - \frac{1}{2\sigma^2} (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\beta)' (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\beta).$$

Con lo que llegamos a la afirmación que bajo gaussianidad el estimador mediante *GLS* es equivalente al alcanzado por máxima verosimilitud.

Señalar que en base a tales consideraciones de normalidad, podrían plantearse ciertos contrastes sobre el vector β . De una forma general pensaríamos en un test definido a través de la hipótesis nula dada por $H_0 \equiv \mathbf{R}\beta = \mathbf{r}$, donde $\mathbf{R} \in \mathcal{M}_{J \times J}$ es una matriz de rango J que representa ciertas restricciones para los parámetros del modelo y $\mathbf{r} \in \mathbb{R}^J$ el vector resultante de tales condiciones. De este contraste de Wald se deriva el siguiente estadístico con el estimador *GLS*,

$$\sigma^2 (\mathbf{R}\beta_{GLS} - \mathbf{r})' (\mathbf{R}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\mathbf{R}')^{-1} (\mathbf{R}\beta_{GLS} - \mathbf{r}) \sim \chi^2(J),$$

con grados de libertad dados por el rango de la matriz \mathbf{R} .

3.2.1. Estimación *GLS* en base a los datos

Hasta este punto, hemos asumido el conocimiento de la matriz Σ_0 referente a la varianza del vector de errores \mathbf{u} , pero en este punto daremos la idea de cuando ésta no es conocida, esto es, se está en un entorno práctico en el cual debemos estimar sus componentes para poder llegar a los estadísticos requeridos. Como antes se realizó, suponemos $\Sigma_0 = \sigma^2\mathbf{V}$ para cierta matriz \mathbf{V} que deberá ser estimada a partir de los datos. Para solventar este requerimiento se realiza una estimación previa de ella por medio de *GLS* y se define el siguiente estadístico en base a la expresión (3.5) desarrollada,

$$\beta_{FGLS} = (\mathbf{X}'\mathbf{V}_{GLS}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}_{GLS}^{-1}\mathbf{y}.$$

Como se observa a través el subíndice del estimador, este método es lo que se conoce como mínimos cuadrados generalizados factibles, o abreviadamente *FGLS*. En la literatura se suele abusar de notación y se identifican $\beta_{FGLS} \equiv \beta_{GLS}$. En ocasiones, podríamos suponer que $\mathbf{V} = \mathbf{V}(\theta_1, \dots, \theta_p)$ para ciertos parámetros $\theta_1, \dots, \theta_p$ que deberán ser estimados, con lo que $\mathbf{V}_{GLS} = \mathbf{V}(\theta_{1,GLS}, \dots, \theta_{p,GLS})$.

La idea del desarrollo que habría de ser tomado lo explicamos en base a un ejemplo en el cual deberemos de estimar un parámetro de un modelo autorregresivo que es asociado a los errores, amén del vector de parámetros β . Entonces, tomemos el modelo (3.2) de regresión y supongamos que los residuos de éste siguen un proceso autorregresivo de orden 1, cuya definición se particulariza a partir de la expresión (1.3), esto es,

$$u_t = \phi u_{t-1} + \varepsilon_t,$$

con $|\phi| < 1$ para la estabilidad del proceso y $\varepsilon = \{\varepsilon_t\}_t$ un proceso de ruido blanco gaussiano de varianza σ^2 . De este modelo, puede comprobarse que,

$$\mathbb{E}(\mathbf{u}\mathbf{u}') = \frac{\sigma^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \dots & \phi^{T-1} \\ \phi & 1 & \dots & \phi^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \dots & 1 \end{pmatrix} = \sigma^2 \mathbf{V}(\phi).$$

Con lo que en este caso se ha de estimar tanto el vector de parámetros β , como el coeficiente ϕ y σ^2 . Para ello podríamos definir la función de log-verosimilitud denotada por $L(\beta, \phi, \sigma^2)$, que si conociésemos ϕ , puede comprobarse que se maximiza con el β_{GLS} asociado a la regresión de $(y_t - \phi y_{t-1})$ frente a $(x_t - \phi x_{t-1})$ para $t = 2, \dots, T$. Usualmente se inicia la optimización partiendo de $\phi = 0$. Una vez β_{GLS} ha sido obtenido se sustituye en $L(\beta, \phi, \sigma^2)$ y se estima ϕ a través de GLS , obteniendo así ϕ_{GLS} mediante la regresión $(y_t - \mathbf{x}'_t \beta_{GLS})$ frente a $(y_{t-1} - \mathbf{x}'_{t-1} \beta_{GLS})$ para $t = 2, \dots, T$. Y se actuaría de este modo sucesivo actualizando las correspondientes estimaciones, hasta converger a un mínimo local de la función $L(\beta, \phi, \sigma^2)$ de la cual partíamos al comienzo de este párrafo.

Este proceso es conocido como el Método de Cochrane-Orcutt, cuyo desarrollo en detalle puede verse en [Cochrane, D. *et al.* 1949]. En él se muestra cómo los términos de error involucrados en formulaciones económicas poseen correlaciones muy altas. En este sentido se demuestra que cuando las estimaciones de propiedades de autocorrelación de error son basadas en cálculos de los residuos, existe un gran sesgo hacia la aleatoriedad. Y por último, en tal artículo se desarrolla el proceso iterativo que se ha dejado entrever con anterioridad.

Señalar para finalizar que, aquí se ha mostrado como se actuaría ante un proceso autorregresivo de orden uno para los errores, aunque sería sencillo adaptarlo para un orden p genérico. De hecho, se podría plantear cualquier proceso estacionario de forma débil, como un $ARMA(p, q)$, ya que éste podría reescribirse como un autorregresivo de orden infinito, como se mostró en el apartado correspondiente.

Parte II

Aplicación a datos reales

Capítulo 4

Comentarios previos

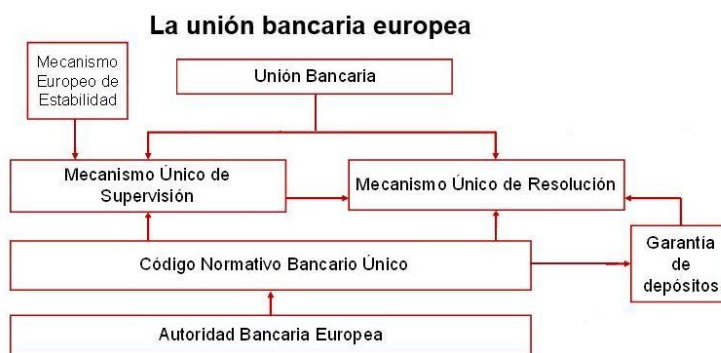
4.1. Motivación financiera y bancaria

La motivación central de la aplicación a datos reales de la metodología mostrada en la Parte I referente al estudio de series temporales, se establece partiendo de la idea de que la gestión de una entidad financiera se basa en su capacidad para generar capital a futuro. Tal proyección de generación se realiza en función de ciertos balances de capital, como los aportados en el Informe de Adecuación de Capital (IAC), pero para los cuales es necesaria una primera predicción del comportamiento del entorno macroeconómico en función de diferentes escenarios en los que se encuadra tal entidad. Para, de esta manera, comprobar cómo ésta respondería a varias situaciones financieras, tanto favorables como adversas (Stress test). Una vez estos balances han sido presentados a los organismos competentes o supervisores, como pueden ser un estado o el banco central del mismo, son examinados para determinar si las proyecciones que la entidad ha propuesto contemplan de forma adecuada todas las condiciones y escenarios que son requeridos.

Todas estas propuestas entregadas y pruebas realizadas sobre ellas vienen recogidas en las denominadas Directivas o Reglamentos realizadas a nivel europeo, el más actual, y sobre la cual haremos las posteriores reflexiones, es la Directiva del Parlamento y Consejo Europeo del 15 de Mayo de 2014 [Directiva Europea 2014], transpuesta a la legislación española el pasado mes de Junio. En la que se establece un marco para la reestructuración y resolución de entidades de crédito y empresas de servicios de inversión. Siendo la reestructuración y resolución una serie de paquetes planteados por la propia entidad y por el organismo competente o supervisor, respectivamente, conteniendo medidas a tomar por la entidad si ésta presenta problemas financieros. A continuación exponemos de forma muy breve la idea sobre la cual se desarrolla la directiva citada sin entrar en profundidad sobre ella, simplemente exponiendo el entorno económico europeo que la precede y el que propone alcanzar.

La crisis financiera ha puesto de manifiesto la falta en el global de la Unión Europea de instrumentos adecuados para hacer frente de forma eficaz a problemas de solidez e inviabilidad de entidades de crédito e inversión. Dichos instrumentos son particularmente necesarios para evitar llegar a una situación de insolvencia, o en último caso, minimizar las repercusiones negativas que ésta pueda conllevar garantizando el pago a los acreedores de la entidad. Durante la crisis, estos sucesos obligaron a los estados miembros a rescatar entidades privadas con dinero de los contribuyentes, lo que estableció como objetivo de un marco creíble de reestructuración y resolución el eliminar una futura acción de ese tipo en la mayor medida de lo posible. Para lo cual se ha generado un Mecanismo Único de Resolución (MUR), cuyo principal objetivo es garantizar que las quiebras bancarias que puedan suceder en un futuro en la unión bancaria se gestionen de forma eficiente, con costes mínimos para el contribuyente y la economía real.

Para hacer frente a una crisis se necesitan medidas destinadas a asegurar el acceso a la financia-



ción por parte de las entidades de crédito, a la cual éstas no han tenido acceso estos últimos años de recesión económica. Para que esto no suceda, produciéndose consecuencias para la economía en general, se han de plantear ciertas medidas que lo aseguren, las cuales requieren de apoyo referente a la liquidez de los bancos centrales y garantías de los productos emitidos por las entidades de crédito solventes. También y debido a la interconexión existente entre los mercados financieros de la UE y entidades que operan en el exterior de dicho marco, la inviabilidad de una entidad transfronteriza podría afectar a la estabilidad tales mercados, por lo que una estabilidad de estos estados extranjeros en los que se opera será así mismo esencial. La revisión del marco regulador referente al refuerzo de capital de seguridad y liquidez, así como unas mejores herramientas para las políticas macroprudenciales (mitigar el riesgo del sistema financiero como un todo), reforzará la resistencia de las entidades y empresas de inversión a las tensiones generadas por futuras crisis. Con ello, los estados miembros han de estar preparados y disponer de instrumentos adecuados para hacer frente tanto a crisis globales como a inviabilidades de entidades individuales. Si el problema surge en una entidad o empresa de servicios de inversión concreta y el resto del sistema financiero no resulta afectado, el ejercicio por parte de las autoridades competentes de tales medidas han de poder ejercerse sin que ello suscite gran preocupación por sus efectos de contagio. Pero en otra situación, se ha de prestar una gran atención a evitar la desestabilización de los mercados financieros.

Los estados miembros exigirán que las entidades preparen y actualicen regularmente planes de reestructuración donde se establezcan medidas que las entidades en cuestión deben adoptar con miras al restablecimiento de su situación financiera tras un deterioro importante. Exigirán asimismo que los planes de reestructuración sean minuciosos y contemplen un abanico de escenarios basados en hipótesis realistas reflejando la importancia global e interconexiones de dicha entidad con otras empresas. Estos planes considerarán escenarios en base a diferentes grados de tensión financiera y macroeconómica, como por ejemplo fenómenos que afecten a todo el sistema, situaciones de tensión específica individual, o trastornos a nivel del grupo. Las entidades deben presentar estos planes a las autoridades competentes, que realizarán una evaluación completa para determinar si éstas permitirían restaurar la viabilidad de una entidad en momentos tensión financiera grave. Para ello se establece el Mecanismo Único de Supervisión (MUS), ubicado en el seno del BCE, se ocupará de la supervisión directa de los bancos europeos de mayor tamaño, mientras que las autoridades nacionales seguirán al resto de bancos, aunque bajo la responsabilidad final del BCE. Si tal supervisión produjese un resultado negativo hacia la entidad, estas autoridades plantearían un plan de resolución para dicha entidad de modo que las premisas requeridas fuesen cumplidas. Mostramos asociado a estos comentarios un esquema para entender de una forma sencilla este desarrollo.

Con estas breves nociones, queda clara la necesidad de realizar proyecciones del marco económico en el que una entidad en particular se establece, en nuestro caso concreto, la entidad bancaria con la que se ha trabajado ha sido ABANCA. Siendo éste un banco establecido casi en su totalidad en la comunidad autónoma de Galicia (España) originado por la compra del extinto Novagalicia Banco en 2013 por parte del multinacional Grupo Banesco. Por tal razón, las consiguientes previsiones a realizar se basarán en determinar el comportamiento a futuro de la economía española con sus

respectivos indicadores.

Éstas se realizarán generando modelos para explicar variables de negocio tales como la Formalización de Créditos para el Consumo en Hogares o el Total de Dudosos OSR (Otros Sectores Residentes), en función de indicadores macroeconómicos al alcance del público en general como puedan ser el Paro registrado o la Demanda nacional. Las primeras se recogerán de bases publicadas por el Banco de España, y las segundas procederán de ubicaciones como el INE (Instituto Nacional de Estadística) o el Ministerio de Economía.

Como hemos dicho, con todas ellas se crearán ciertos modelos econométricos y se realizarán predicciones en base a técnicas de análisis de series temporales, todo lo desarrollado con anterioridad en el trabajo. Por tanto, sin extendernos más, pasamos a la aplicación práctica en cuestión, comenzando por comentar los lugares de obtención de los datos a utilizar y siguiendo con el empleo de la teoría mostrada.

4.2. Referencias de las bases de datos

Una vez sabemos el porqué del trabajo, debemos exponer con que elementos vamos a trabajar, esto es, las variables de evolución del negocio que expondrán el comportamiento del entorno económico español y en función de qué indicadores macroeconómicos éstos se generarán. Éstas las expondremos en sendos cuadros en lo que sigue, puntualizando su frecuencia de observación y su periodo de estudio. Pero antes, mostraremos la lista de instituciones a las que se ha recurrido para la selección y actualización de todas ellas, y un breve comentario de su utilidad.

1. [Ministerio de Economía y Competitividad]

Ésta ha sido la principal referencia para la selección de todas las variables macroeconómicas, en ella se han ido actualizado todas las variables según el Gobierno ha ido subiendo los datos a su base de datos. Señalar que en el correspondiente sitio web se puede crear un usuario para así guardar las consultas que se desean, en nuestro caso en base a indicadores económicos mensuales y trimestrales. Indicar que en la tabla de las variables se asocia a cada una de ellas un código referente a la codificación que tienen en esta dirección.

2. [Instituto Nacional de Estadística] (INE)

Esta institución se ha empleado para el contraste de valores de los obtenidos en las bases del Ministerio de Economía y para llegar a una mejor aclaración de lo que cada variable representa. Por ejemplo, acerca del desempleo en España existe el Total de Paro Registrado, Total de Desempleados o Paro Total, y debemos asegurarnos de cuál es la que queremos seleccionar.

3. [Instituto Galego de Estadística] (IGE)

Ésta es empleada en el mismo sentido que la anterior pero referente a las variables del entorno económico de Galicia para poder ver su relación con el nivel español. En este trabajo no se ha incluido un estudio asociado a Galicia que fuera de éste si ha tenido lugar.

4. [Banco Central Europeo] (BCE)

Para ciertas variables seleccionadas en el Ministerio de Economía, como son los tipos de interés a 3, 6 y 12 meses, el BCE proporciona previsiones en ciertos escenarios base o adversos a un horizonte de cinco años. Por tanto, cuando tengamos que evaluar modelos en previsiones y ciertos tipos de interés tengan presencia, echaremos mano de tales predicciones.

5. Bases de datos propias de la entidad

Debemos señalar que las anteriores referencias fueron utilizadas para los indicadores macroeconómicos, pero para las de tipo de negocio no las emplearemos. Para estas últimas tomaremos bases de datos internas de la entidad bancaria. Estas bases se nutren de la información generada por el Banco de España que internamente se estructura para su correcta utilización.

VARIABLES EMPLEADAS

Una vez que tenemos claro de dónde hemos obtenido todos los datos que se utilizan, damos a conocer en las siguientes tablas todas estas variables empleadas. En ellas puntualizaremos cuáles son de negocio y cuáles no, sus respectivas frecuencias, mensuales o trimestrales, y el periodo de estudio en el que se enmarca los correspondientes valores. Para más información sobre ellas recurriremos al [Instituto Nacional de Estadística], [Instituto Galego de Estadística] o [Ministerio de Economía y Competitividad]. Señalar que a la parte de las macroeconómicas se le añade una columna llamada Código que representa la codificación que la respectiva variable posee para ser descargada y actualizada en las bases de datos del Ministerio de Economía, como ya comentamos.

Señalar que de muchas de las variables, en particular las mensuales, están registrados valores muy antiguos, de hace casi cuarenta años. Por tal razón y dando por supuesto que no es extraño pensar que el comportamiento actual de cierta variable no se corresponde con datos tan antiguos, suprimiremos valores anteriores al año 2000 y trabajaremos con el resto, y que de las trimestrales tomaremos todas las posibles por ser series de tamaños no demasiado grandes. Señalar que en las variables que introduzcamos haremos referencia a ciertas iniciales, las cuales reflejan los siguientes términos,

- OSR ≡ Otros Sectores Residentes.
- SS ≡ Seguridad Social.
- IPI ≡ Índice de Producción Industrial.
- PIB ≡ Producto Interior Bruto.
- FBC ≡ Formación Bruta de Capital.
- FBCF ≡ Formación Bruta de Capital Fijo.
- VAB ≡ Valor Agregado Bruto.

A continuación exponemos la lista de las variables a la que hacíamos referencia. Señalar que una vez que las aportemos, estaremos en disposición de pasar a la aplicación de datos reales propiamente dicha, en la cual intentaremos crear modelos de predicción de las variables que hemos denominado de negocio en función de las macroeconómicas. Para ello desarrollaremos la teoría vista en la metodología mostrada en la primera parte de este trabajo asociada a técnicas de econometría.

VARIABLES DE NEGOCIO

Variable	Frecuencia	Periodo de Estudio
Formalización del Crédito al Consumo en Hogares		
Formalización del Crédito a la Vivienda	Mensual	6/2003 – 1/2015
Depósitos de Minoristas		
Total de Dudosos OSR		12/1999 –
Crédito a Empresas	Trimestre	12/2014

Variables Macroeconómicas Mensuales

Código	Variable	Periodo de Estudio
170000	Paro Registrado Total	1/1996 – 3/2015
190000	Afiliados a la SS	1/1982 – 3/2015
229000	IPI General	1/1975 – 1/2015
233600	Compra-Venta de Viviendas	1/2007 – 1/2015
233610	Compra-Venta de Viviendas Nuevas	1/2007 – 1/2015
236000	Consumo de Cemento	1/1964 – 1/2015
255100	Consumo de Energía Eléctrica	1/1958 – 3/2015
258011	Hipotecas inmobiliarias	1/2003 – 1/2015
25M100	Ventas de Grandes Empresas	1/1996 – 3/2015
271200	Matriculación de Turismos	1/1960 – 2/2015
272701	Índice de Comercio al por menor	1/1995 – 3/2015
300000	Índice de Clima Industrial	1/1993 – 3/2015
330000	Indicador de Confianza de Consumidores	6/1986 – 3/2015
340000	Indicador de Confianza del Comercio	9/1988 – 3/2015
867203	Tipo de Interés a 3 meses	1/1994 – 3/2015
867204	Tipo de Interés a 6 meses	1/1994 – 3/2015
867205	Tipo de Interés a 12 meses	1/1994 – 3/2015

Variables Macroeconómicas Trimestrales

Código	Variable	Periodo de Estudio
449990	Precio Medio del m^2	3/1995 – 12/2014
940000	PIB bruto volumen	
941110	Gasto en Consumo de los Hogares volumen	
942000	Formación Bruta de Capital (FBC) volumen	
942110	FBCF de Viviendas volumen	
942112	FBCF de Construcción volumen	
942120	FBCF de Edificios volumen	
942130	FBCF de Maquinaria volumen	
944000	Demanda Nacional volumen	
944001	Demanda Externa volumen	
945200	VAB Industria volumen	
945210	VAB Manufacturación volumen	
945400	VAB Servicios volumen	
972041	Renta Neta de Hogares	
972051	Ahorro Neto de Hogares	3/2000 – 12/2014
990000	PIB bruto precio	
991110	Gasto en Consumo de los Hogares precio	
992000	FBC precio	
992110	FBCF de Viviendas precio	
992112	FBCF de Construcción precio	
992120	FBCF de Construcción precio	
992130	FBCF de Maquinaria precio	
994000	Demanda Nacional precio	
995200	VAB Industria precio	
995210	VAB Manufacturación precio	
995400	VAB Servicios precio	
996100	Remuneración de Asalariados precios	

Capítulo 5

Aplicación de la metodología

Mostrada la motivación del trabajo y con qué variables vamos a contar para atender a dichos objetivos, encuadrando éste en un entorno económico-financiero cuyo fin es entender la evolución de tal entorno, es momento de aplicar la metodología desarrollada en la primera parte del proyecto (Parte I) a tales datos. Aunque antes de entrar en esta tarea en profundidad, describiremos el esquema que se seguirá a lo largo de esta etapa, así como citar y dar referencias de las herramientas de carácter informático que nos ayudarán en dicho proceso.

Debemos empezar diciendo que este desarrollo se divide en tres fases. La primera es en la cual se llegarán a obtener las expresiones de los modelos que nos interesan (Sección 5.1), en ella es donde seleccionaremos las mejores variables en cuanto a la explicación sobre las respuestas que estamos considerando, y se echará mano del método *Generalized Least Squares (GLS)* de estimación para así poder incluir cierta componente de correlación en los residuos. A continuación, pensando en que queremos predecir las respuestas con los susodichos modelos, debemos modelizar primero las covariables y predecir éstas (Sección 5.2), para lo cual tomaremos la visión univariante del Capítulo 1 y la vectorial del Capítulo 2. Y una vez que tengamos esas previsiones a cierto futuro, llegaríamos a la última etapa que sería la sustitución de dichas evoluciones futuras en los modelos, para llegar así a predicciones de las respuestas (Sección 5.3).

Por último indicar que para los cálculos realizados se emplea como herramienta informática el *software* estadístico R v.3.1.1 (Apéndice B) en base a ciertas librerías como son `nlme` [Pinheiro J. 2014], `TSA` [Chan, K-S. *et al.* 2012], `tseries` [Trapletti, A. *et al.* 2013] y el paquete `vars` [Pfaff, B. 2008b]. También, de forma puntual se tomará un *software* comercial llamado IBM SPSS Modeler v.16 [IBM Corporation 2013] (Apéndice C) para una mayor comodidad a la hora de implementar todos los modelos creados y por la habitual utilización de éste en la entidad bancaria. Señalar que con R emplearemos [Cowpertwait, P. *et al.* 2009] para el tratamiento clásico univariante de series de tiempo y luego echaremos mano de los paquetes ya comentados y algún apunte adicional.

5.1. Creación de los modelos econométricos

Comenzaremos por la generación de los modelos econométricos, los cuales explicarán las variables de evolución del negocio, respuestas, en función de los indicadores macroeconómicos, explicativas, ya expuestos. Recordemos que las respuestas que vamos a considerar se disponen en la tabla titulada Variables de Negocio de la Sección 4.2. Pero de ellas no vamos a desarrollar todos los detalles de cada modelo en particular, sino que profundizaremos en el primero de ellos, el relativo a la Formalización de Crédito al Consumo en Hogares (FCCH) y del resto solo indicaremos finalmente la expresión analítica que lo define.

5.1.1. Selección y transformación de variables

Primero, seleccionaremos las variables más adecuadas para dicho modelo y éstas las transformaremos con el objetivo de alcanzar la homocedasticidad en ellas o simplemente para que, posteriormente en la estimación del modelo, se asocien a coeficientes sencillos de analizar, es decir, una reducción de la escala. También en este momento se tomarán retardos de las variables en relación con la respuesta, para poder crear modelos dinámicos y no solo contemporáneos, ya que por esta segunda vía podría perderse cierta información.

Denotemos por $\mathbf{y}^c = \{y_t^c, t \in \mathcal{I}_c\}$ la FCCH, siendo \mathcal{I}_c en intervalo de estudio de tal respuesta (para ver el correspondiente periodo, acudir a tabla Variables de Negocio de la Sección 4.2), una primera batería de variables asociadas a ella y que se plantean para la construcción del modelo correspondiente, está dada en base a unas vinculaciones previas consideradas internamente por el departamento de Planificación y Estudios dentro de la propia entidad. En nuestro caso, asociadas al consumo se plantean las siguientes $K = 9$ variables mensuales,

$$\begin{aligned} \mathbf{x}_1 &= \{x_{1t}\}_t \equiv \text{Total de Paro Registrado,} \\ \mathbf{x}_2 &= \{x_{2t}\}_t \equiv \text{Total de Trabajadores Afiliados a la SS,} \\ \mathbf{x}_3 &= \{x_{3t}\}_t \equiv \text{Consumo de Energía Eléctrica,} \\ \mathbf{x}_4 &= \{x_{4t}\}_t \equiv \text{Ventas Totales de Grandes Empresas,} \\ \mathbf{x}_5 &= \{x_{5t}\}_t \equiv \text{Total de Matriculación de Turismos,} \\ \mathbf{x}_6 &= \{x_{6t}\}_t \equiv \text{Índice de Comercio al Pormenor,} \\ \mathbf{x}_7 &= \{x_{7t}\}_t \equiv \text{Indicador de Confianza de los Consumidores,} \\ \mathbf{x}_8 &= \{x_{8t}\}_t \equiv \text{Indicador de Confianza del Comercio Minorista,} \\ \mathbf{x}_9 &= \{x_{9t}\}_t \equiv \text{Tipo de Interés de Depósitos a 6 meses.} \end{aligned}$$

Donde denotaremos por \mathcal{I}_i para $i = 1, \dots, 9$ los correspondientes intervalos de estudio de las variables explicativas (para ver los correspondientes períodos, acudir a la tabla Variables Macroeconómicas Mensuales de la Sección 4.2). En la práctica no siempre coincidirán dichos intervalos, de hecho se tiene

$$\begin{aligned} \mathcal{I}_c &= (6/2003 - 1/2015) \Rightarrow T_c = 140, \\ \mathcal{I}_i &= (1/2000 - 3/2015) \text{ para } i = \{1, \dots, K\} \setminus \{4, 6\} \Rightarrow T_i = 183, \text{ y} \\ \mathcal{I}_j &= (1/2000 - 1/2015) \text{ para } j = \{4, 6\} \Rightarrow T_j = 181, \end{aligned}$$

denotando por T el instante final y, si $\mathcal{I} = [1, T]$, el tamaño de los periodos de estudio. Pero esta falta de coincidencia no será problema a la hora del desarrollo del modelo, como ya se observará. Entonces, teniendo como respuesta \mathbf{y}^c y como covariables $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_K\}$, lo que se realizará ahora serán las transformaciones que en el primer párrafo de este apartado se citaban.

Para alcanzar la primera de las metas, homocedasticidad o disminución de la escala, emplearemos la transformación de Box-Cox que se ha introducido en la Sección 1.3, siendo, simplemente, una transformación logarítmica por si se desea reducir la escala. De este modo se obtienen las siguientes nuevas variables,

$$\mathbf{y}^{c*} = \log \mathbf{y}^c, \quad \mathbf{x}_i^* = \begin{cases} \mathbf{x}_i & \text{si } i = 4, 6, 7, 8, 9, \text{ y,} \\ \log \mathbf{x}_i & \text{si } i = 1, 2, 3, 5. \end{cases}$$

donde $\mathbf{x}_i^* = \{x_{it}^*, t \in \mathcal{I}_i\}$ para todo $i = 1, \dots, K$. Señalar que cuando se transforma por medio de Box-Cox, muchas veces el λ óptimo es cercano a cero y por tanto bastaría con hallar el logaritmo natural como consecuencia de la continuidad de la expresión que define a tal función, y así no indagar demasiado en dicho λ . Para el desarrollo del modelo por simplificar la escritura, abusaremos de notación y reescribiremos $\mathbf{y}^c = \mathbf{y}^{c*}$ y $\mathbf{x}_i = \mathbf{x}_i^*$ para todo i . Al final, cuando visualicemos el resto de modelos, se especificarán las transformaciones realizadas para que no exista ningún tipo de confusión.

La segunda de las modificaciones es a la hora de hallar posibles relaciones lineales entre cada una de las variables explicativas con la respuesta en base a ciertos retardos temporales. Lo que

buscaremos será un vector $\mathbf{s} = \{s_i \in \mathbb{Z}, i = 1, \dots, K\}$ tal que la serie $\{x_{i,t-s_i}, t \in \mathcal{I}_i\}$ para $i = 1, \dots, K$ se relacione de forma lineal con $\{y_t^c, t \in \mathcal{I}_c\}$. El correspondiente cálculo se realizará según lo indicado en la Sección 3.1, obteniendo de esta forma el siguiente vector,

$$\mathbf{s} = \{s_i, i = 1, \dots, K\} = \{0, 4, -1, 0, 0, 12, 0, 0, 0\}.$$

Cuya interpretación puede ejemplificarse aclarando que, entre \mathbf{x}_1 y \mathbf{y}^c se supondrá una relación contemporánea, en cambio, lo que ocurre con \mathbf{x}_2 hace 4 instantes de tiempo con respecto al presente afecta a \mathbf{y}^c en la actualidad, análogamente se interpretan los demás retardos. Señalar que en ciertas relaciones la suposición de que existe relación contemporánea, $s_i = 0$ para cierto i , es debido a dos posibles circunstancias. La primera y lógica es deducida a través de que el retado $s_i = 0$ resulte ser significativamente distinto de cero en el cálculo de $sCCF(\mathbf{y}^c, \mathbf{x}_i)$, en el sentido explicado. Y la segunda es en la cual, a pesar de que s_i no resulta ser significativo para ningún valor, la variable correspondiente \mathbf{x}_i no se quiere eliminar del modelo por ser importante en relaciones no lineales, requerir su presencia en el correspondiente estudio o por no querer disminuir en exceso la cantidad K de covariables consideradas en el modelo posterior.

Comentar también que tenemos dos formas de actuar en este momento acerca de los intervalos \mathcal{I}_i . En base a los retardos hallados se podría tomar un intervalo intersección de estudio para \mathbf{y}^c , esto es, $\mathcal{I} = (\cap_{i=1}^K \mathcal{I}_i) \cap \mathcal{I}_c$, pero de esta manera se emplearía la mínima cantidad de datos posible y no es un camino deseado. Por lo que se desarrollará una segunda vía donde toda la información que se pueda obtener de los intervalos será empleada, mediante el desplazamiento de éstos a derecha o izquierda según sea conveniente. Una vez empleada esta técnica, seleccionamos el mismo número de observaciones para cada variable y con ello reescribimos un mismo intervalo de estudio para todas ellas con la notación general $\mathcal{I} = \{1, \dots, T\}$.

5.1.2. Modelos en base a la estimación por GLS

Siguiendo con la particularización del modelo referente a FCCH, o con la notación \mathbf{y}^c , el objetivo de este apartado es el de generar un modelo de regresión mediante el método *GLS* (introducido en la Sección 3.2) que explique dicha variable en función de las $K = 9$ covariables que en el anterior apartado han resultado para dicha variable de negocio. Recordemos que estamos trabajando con

$$\mathbf{y}^c = \{y_t^c\}_t \quad \text{y} \quad \mathbf{x}_i = \{x_{i,t-s_i}, s_i \in \mathbf{s}\}_t \quad \text{para} \quad i = 1, \dots, K \quad \text{y} \quad t \in \mathcal{I}.$$

Para crear el modelo requerido vamos a realizar un proceso en dos etapas, en la línea de lo desarrollado en [Fox, J. *et al.* 2010]. Este artículo establece un primer paso donde se define un modelo lineal con las variables consideradas y de éste se toman los residuos de su estimación para modelizarlos mediante un cierto *ARMA*. Y un segundo, donde tomamos el modelo lineal pero indicando que los errores siguen el *ARMA* antes hallado, y, de este modo, empleando *GLS*, es estimado. Para un correcto modelo de regresión comprobaremos si los residuos que genera dicho ajuste se comportan según el *ARMA* indicado.

Explicado el proceso a seguir, comenzamos con la definición del modelo de regresión lineal dado por

$$\mathbf{y}^c = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_c^{LM}$$

donde $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ es un vector que posee los coeficientes reales asociados a las variables de \mathbf{X} , con

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_K) = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{K,1} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1,T} & \dots & x_{K,T} \end{pmatrix} \in \mathcal{M}_{T \times K+1},$$

y $\mathbf{e}_c^{LM} = \{e_{c,t}^{LM}, t \in \mathcal{I}\}$ los errores de tal expresión. Entonces, si ajustamos este modelo y tomamos los residuos $\hat{\mathbf{e}}_c^{LM}$, y a su vez éstos los modelizamos mediante Box-Jenkins, obtenemos que $\hat{\mathbf{e}}_c^{LM} \sim AR(1)$. Modelo que podría ser explicado más en detalle a través de sus coeficientes estimados, pero no lo haremos porque en este momento lo capital será decir que los residuos siguen

un autorregresivo de orden 1, y no su resultado numérico particular. Pasamos ahora a la segunda etapa tomando el modelo lineal de regresión dado por

$$\mathbf{y}^c = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_c^{GLS}, \quad \text{verificando } \mathbf{e}_c^{GLS} \sim AR(1).$$

Éste es estimando mediante *GLS* y una vez ajustado seleccionamos las variables significativas del mismo, esto es, que sus *p*-valores asociados sean mayores que un cierto nivel de significación, en general, $\alpha = 0.05$. Si esto se realiza, se llega a la expresión final para cierto $t \in \mathcal{I}$,

$$y_t^c = (x_{1,t}, x_{2,t}, x_{4,t}, x_{5,t}) \boldsymbol{\beta}_{GLS} + \hat{e}_{c,t}^{GLS}, \quad \text{con } \hat{e}_{c,t}^{GLS} \sim AR(1), \quad (5.1)$$

siguiendo la notación de que $\boldsymbol{\beta}_{GLS}$ es la estimación de $\boldsymbol{\beta}$ por medio de *GLS*, como ya se comentó en la Sección 3.2. Siendo $\boldsymbol{\beta}_{GLS} = (\hat{a}_0, \dots, \hat{a}_4)'$, con

$$\begin{aligned} \hat{a}_0 &= -15.24 \quad (6.00), & \hat{a}_1 &= -1.33 \quad (0.16), & \hat{a}_2 &= 2.68 \quad (0.69), \\ \hat{a}_3 &= -0.01 \quad (0.00) & \text{y } \hat{a}_4 &= 0.72 \quad (0.04), \end{aligned}$$

siendo el valor entre paréntesis las desviaciones típicas de las estimaciones, las cuales son empleadas en los contrastes de significación de los respectivos coeficientes. Se verifica que todos los parámetros mostrados son significativos a un nivel de 0.05.

Una vez que el ajuste ha sido realizado, debemos comprobar que es correcto, para lo cual tomamos $\hat{e}_{c,t}^{GLS}$ y comprobamos si estos se comportan como un proceso *AR*(1), esto es, $\hat{e}_{c,t}^{GLS} = \phi_1 \hat{e}_{c,t-1}^{GLS} + \varepsilon_{c,t}^{GLS}$ para $\varepsilon_{c,t}^{GLS} = \{\varepsilon_{c,t}^{GLS}, t \in \mathcal{I}\}$ un proceso de ruido blanco. Y será esto último lo que intentaremos verificar, si $\varepsilon_{c,t}^{GLS}$ es ruido blanco. Con tal premisa, tomamos la Figura 5.1 (para la heterocedasticidad y la normalidad gráfica) y la Tabla 5.1 (para la media nula, normalidad e incorrelación), en las cuales se resumen los contrastes necesarios para tal comprobación, gráfica y analítica, respectivamente.

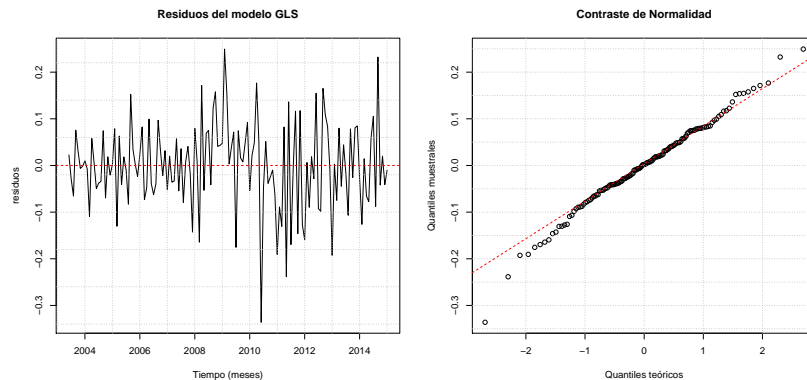


Figura 5.1: A la izquierda representación de la serie $\hat{e}_{c,t}^{GLS}$ para comprobar su homocedasticidad y media cero. Y a la derecha el *qq*-plot correspondiente a los mismos residuos. Lo deseado es estar en presencia de un proceso de ruido blanco.

	<i>t</i> -test	shapiro-wilks	Jarque-Bera
Estadístico	-0.05	0.99	6.95
<i>p</i> -valor	0.96	0.29	0.03
	Ljung-Box (retardos)		
	1	5	10
Estadístico	3.03	9.13	13.25
<i>p</i> -valor	0.08	0.10	0.21

Cuadro 5.1: Resumen de los contrastes de media nula con el *t*-Test, normalidad mediante Jarque-Bera y shapiro-wilks, y de incorrelación de los residuos por medio de Ljung-Box para retardos 1, 5 y 10, en base a los estadísticos y *p*-valores correspondientes. Todos ellos aplicados a los residuos $\hat{e}_{c,t}^{GLS}$.

A la vista de tales resultados no existen evidencias significativas para el rechazo de la hipótesis que define $\varepsilon_{c,t}^{GLS}$ como un proceso de ruido blanco, con lo que podemos concluir que el modelo *GLS*

propuesto para la variable y^c en la expresión (5.1) no es erróneo, y por tanto podremos emplearlo para realizar predicciones de la misma variable respuesta. En este punto representamos la Figura 5.2, en la cual podemos visualizar cómo se comporta la estimación hallada a través del método *GLS* creado (línea roja) frente a los valores observados de la serie (línea negra). Vemos que en los valores centrales correspondientes al año 2008, inicio de la crisis financiera actual, se estima una caída más pronunciada que la que en realidad hubo, pero éste es un suceso crítico que no se asemeja al comportamiento del resto de la serie, en ella tanto el comienzo como el final se ajustan de una forma correcta. Señalar que esta imagen está acompañada por los coeficientes de determinación $R^2 = 0.90$ y $R_{aj}^2 = 0.90$, lo cual sugiere una buena bondad de ajuste.

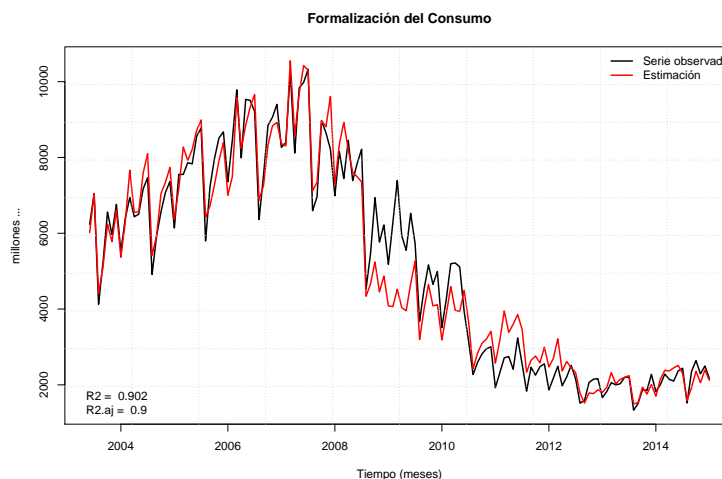


Figura 5.2: Comparación de los valores observados de la serie de la Formalización del Crédito para el Consumo en Hogares, línea negra, y su estimación mediante el modelo *GLS*, línea roja. Se acompaña tal ajuste de los coeficientes de determinación R^2 y R_{aj}^2 .

Dado el modelo definitivo para esta primera variable de evolución del negocio, podemos darle una interpretación en el aspecto macroeconómico. Para empezar, el coeficiente asociado al Total del Paro Registrado ($\hat{a}_1 = -1.3$) tiene sentido que sea negativo, pues cuanto mayor número de parados, menor cantidad de gente pedirá crédito para su consumo personal y por tanto menor número de créditos serán aprobados. También tiene sentido que tanto el parámetro asociado al Total de Afiliados a la SS ($\hat{a}_2 = 2.7$) como el referente al Total de Matriculación de Turismos ($\hat{a}_4 = 0.7$) sean positivos, pues si ambas variables aumentan, es un reflejo de que la gente podrá invertir más en sí misma y pedirá mayor número de créditos para consumo propio. El coeficiente asociado a las Ventas Totales de Grandes Empresas ($\hat{a}_3 = -0.01$) parece razonable que fuese positivo, pero podría existir algún efecto confusor y de correlación con alguna variable, en el cual no indagaremos.

Consideración de otras respuestas

Visto como elaborar un modelo correcto para la FCCH, se muestran los obtenidos para las otras cuatro respuestas consideradas en el proyecto. De cada uno de los modelos visualizaremos la expresión analítica resultante a partir del método *GLS*, los contrastes correspondientes para los residuos generados del *ARIMA* referente a los errores derivados del *GLS* planteado, y por último, la comparación gráfica entre lo estimado y lo observado. Resumiendo, de este modo, los pasos dados en los anteriores párrafos. Señalar que todo ello se mostrará de forma esquemática dado que el desarrollo es análogo al elaborado para el modelo (5.1). Aunque antes, introducimos la siguiente notación para simplificar la escritura con respecto a las respuestas,

$$\begin{aligned}
 \mathbf{y}^v &= \{y_t^v\}_t \equiv \text{Formalización del Crédito para Vivienda,} \\
 \mathbf{y}^d &= \{y_t^d\}_t \equiv \text{Total de Depósitos Minoristas,} \\
 \mathbf{y}^{du} &= \{y_t^{du}\}_t \equiv \text{Total de Dudosos OSR.} \\
 \mathbf{y}^{cr} &= \{y_t^{cr}\}_t \equiv \text{Crédito a Empresas sin construcción ni inmobiliario,}
 \end{aligned}$$

y para las covariables significativas empleadas en los modelos,

$$\begin{aligned}
 \mathbf{x}_{10} &= \{x_{10,t}\}_t \equiv \text{Total de Compra-Venta de Viviendas,} \\
 \mathbf{x}_{11} &= \{x_{11,t}\}_t \equiv \text{Tipo de Interés de Depósito a 12 meses,} \\
 \mathbf{x}_{12} &= \{x_{12,t}\}_t \equiv \text{Renta neta disponible en los Hogares,} \\
 \mathbf{x}_{13} &= \{x_{13,t}\}_t \equiv \text{FBC en precio,} \\
 \mathbf{x}_{14} &= \{x_{14,t}\}_t \equiv \text{VAB Servicios en precio,} \\
 \mathbf{x}_{15} &= \{x_{15,t}\}_t \equiv \text{Demanda Nacional en volumen.}
 \end{aligned}$$

Introducida esta notación, mostramos el esquema resumen de cada uno de los modelos.

$$\begin{aligned}
 \log y_t^v &= b_0 + b_1 \log x_{2,t} + b_2 \log x_{10,t+2} + b_3 x_{11,t} + e_{v,t}^{GLS}, \quad \text{para } e_v^{GLS} \sim AR(1), \\
 \log y_t^d &= c_0 + c_1 \log x_{1,t} + x_{12,t}^{-0.5} (c_2 + c_3 \log x_{2,t} + c_4 x_{9,t} + c_5 \log x_{2,t} x_{9,t}) + e_{d,t}^{GLS}, \\
 &\quad \text{para } e_d^{GLS} \sim MA(1), \\
 \log y_t^{du} &= d_0 + \log x_{1,t+1} (d_1 + d_2 \log x_{1,t+1}) + \log x_{2,t+1} (d_3 + d_4 \log x_{2,t+1}) + \\
 &\quad x_{11,t+2} (d_5 + d_6 x_{11,t+2}) (d_7 + d_8 x_{9,t+1}) + e_{du,t}^{GLS}, \quad \text{para } e_{du}^{GLS} \sim MA(1), \\
 \log y_t^{cr} &= e_0 + e_1 \log x_{15,t-2} + e_2 \log x_{13,t-2} + e_3 \log x_{14,t-2} + e_{cr,t}^{GLS}, \\
 &\quad \text{para } e_{cr}^{GLS} \sim MA(1).
 \end{aligned} \tag{5.2}$$

		t-Test	shapiro-wilks	Jarque-Bera	Ljung-Box (retardos)		
					2	5	10
ε_v	Estadístico	0.00	0.98	4.76	0.01	1.72	8.05
	p-valor	0.99	0.21	0.09	0.91	0.89	0.62
ε_d	Estadístico	0.04	0.97	2.66	0.33	7.14	12.05
	p-valor	0.97	0.36	0.26	0.56	0.21	0.28
ε_{du}	Estadístico	0.16	0.99	0.07	1.83	4.63	8.28
	p-valor	0.88	0.69	0.97	0.18	0.46	0.60
ε_{cr}	Estadístico	-0.49	0.98	1.23	0.15	3.00	4.61
	p-valor	0.63	0.54	0.54	0.70	0.72	0.92

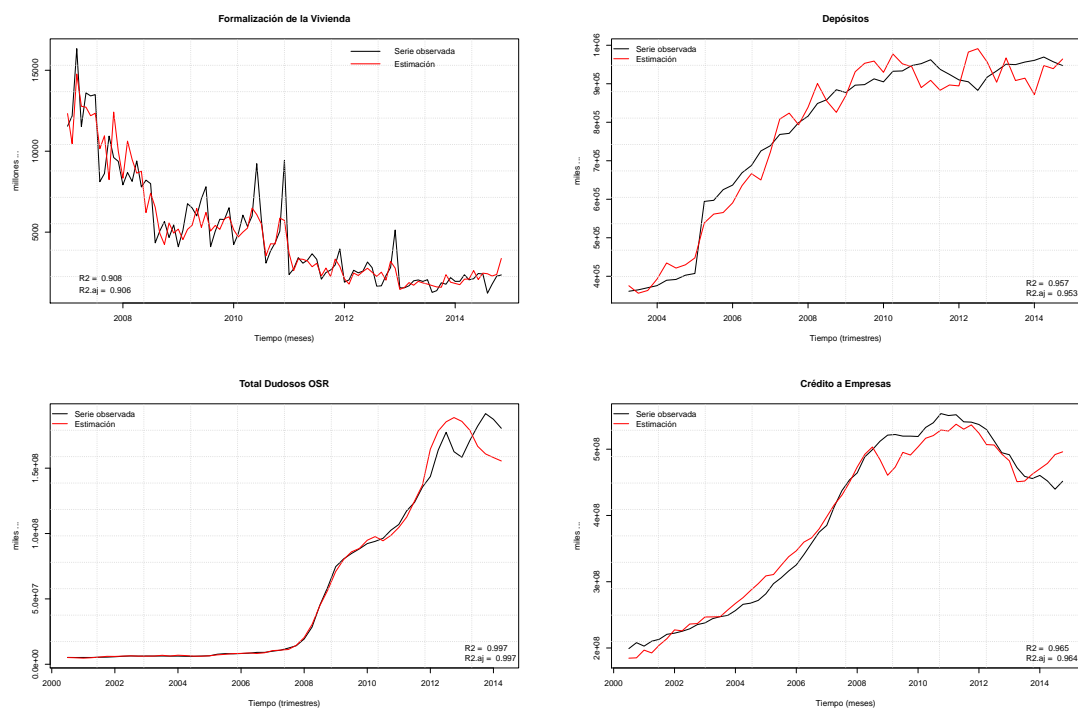
Cuadro 5.2: Resumen de los contrastes de media nula, normalidad e incorrelación para los residuos generados del ajuste realizado para los errores e_v^{GLS} , e_d^{GLS} , e_{du}^{GLS} y e_{cr}^{GLS} con sus respectivos modelos que se muestran en (5.2), en base a los estadísticos y p-valores. Lo deseado es estar en presencia de procesos de ruido blanco.

En base a los resultados mostrados en la Tabla 5.2 los modelos que se han mostrado en (5.2) son adecuados para modelizar sus correspondientes respuestas. Por ello mostramos ahora en las figuras recogidas en la Tabla 5.3 una comparativa entre lo estimado por dichos modelos y las series observadas.

Análisis de sensibilidad de los modelos

Una vez planteados los modelos econométricos referentes a las variables de evolución del negocio consideradas, y antes de comenzar la modelización de las covariables correspondientes para la futura predicción de las respuestas, es adecuado realizar un análisis de sensibilidad de ellos. Para tal objetivo, lo que queremos hallar es el efecto que un solo indicador macroeconómico tiene sobre cierta variable de negocio, efecto que se expresará en función de tasas de variación. Señalar que éstas se calcularán de Diciembre de un año frente al mismo periodo del año siguiente, salvo en la FCCH y en la Formalización del Consumo en la Vivienda que serán calculadas de forma acumulada en el año.

El proceso a seguir comienza por fijar un año base (bY), éste será 2014 debido a que los datos de nuestras variables de negocio son observados hasta comienzos del año 2015, y un año a variar (vY), que será el propio 2015. Esto quiere decir que, para el análisis de sensibilidad se tomarán los valores



Cuadro 5.3: Comparación de las series observadas (negro) de las variables de negocio y su estimación con los diferentes modelos (rojo). De izquierda a derecha y de arriba a abajo. Formalización del Crédito para Viviendas. Total de Depósitos Minoristas. Total de Dudosos OSR. Créditos a Empresas sin construcción ni inmobiliario. Todos ellos acompañados por sus coeficientes de determinación.

de los indicadores explicativos durante vY con los valores de bY , salvo la covariable de interés, que de un año para otro la haremos variar cierta tasa. De este modo, aislaremos el efecto que la variable tiene sobre el modelo completo, sin tener en cuenta una evolución en el resto. Señalar que este análisis no es coherente con la teoría económica, ya que cuando un indicador macroeconómico varía, otros también lo hacen debido a estar relacionados en la economía. Por ejemplo, si se modificasen los afiliados de un país, la cantidad de parados de tal estado no permanecería constante, sino que también variaría ciertas unidades.

Comentado cómo realizaremos tal proceso, mostramos la Tabla 5.4, de la cual damos a continuación ciertos apuntes, y donde recogemos todas las salidas numéricas que se han hallado para este análisis de sensibilidad. En ella las columnas son referidas a las variables de evolución del negocio y las filas a los indicadores macroeconómicos que toman parte en los diferentes modelos generados (ambas con la notación ya introducida). Estas últimas se dividen en filas donde se reflejan las variaciones interanuales que se les han aplicado.

Comentar que x_1 y el PIB están dispuestas de forma separada al resto de covariables, debido a que éstas se identificarán con variables exógenas en los futuros modelos *VAR* y *VECM*, y por tanto su variación afectará al resto de variables. Por lo que, aunque en las expresiones (5.1) y (5.2) no se reflejen estos dos indicadores, sí tendrán repercusión en las previsiones futuras de todos los modelos. También señalar que, en las celdas correspondientes a variables que no tienen efecto en un respectivo modelo, se indicará un guión reflejando que no tendremos valor de tal análisis, exceptuando, claro está, x_1 y PIB.

Para explicar los valores de la tabla, tomemos como ejemplo la variable x_2 (Total de Afiliados a la SS) y el modelo de la FCCH (y^c). Si de un año para otro los afiliados disminuyen un 5%, esto es, se tiene $(1 - 0.05) \cdot x_2$ afiliados ese nuevo año, entonces para tal año se tendrá una FCCH de $(1 - 0.529) \cdot y^c$. Si, por el contrario, aumentan un 3%, implicará que tendremos en ese año siguiente $(1 + 0.402) \cdot y^c$. De forma análoga se interpretan y desgranar el resto de valores que en la tabla

		y^c	y^v	y^d	y^{du}			y^d	y^{cr}
x_2	-5	-5.29	-19.14	-13.22	2.86	x_{12}	-10	-1.69	
	-1	-0.74	2.37	-2.74	-0.30		5	0.76	-
	3	4.02	30.22	8.51	3.36		20	2.78	
x_4	-90	0.08				x_{13}	-1		1.98
	0	-0.31	-	-	-		1	-	-1.92
	90	-0.50					2		-3.78
x_5	-20	-2.62				x_{14}	-1		-0.67
	0	1.64	-	-	-		1	-	0.67
	10	5.81					2		1.34
x_{10}	-20		-1.81			x_{15}	-1		-6.04
	0		5.10		-		1	-	6.36
	20		12.04		-		2		13.06
PIB	-5	-3.43	3.52	5.91	-10.68	PIB	-5		-1.84
	5	-1.40	10.78	9.31	-9.50		5		-1.14
x_1	-5	3.81	12.37	4.09	-24.32				
	5	-7.83	2.37	10.78	5.46				

Cuadro 5.4: Resumen del análisis de sensibilidad basado en tasas de variación (%) de los modelos generados anteriormente (5.1) y (5.2). La notación x_i e y^j para ciertos i y j , son asociadas a las introducidas al principio de la sección.

referida se indican.

Está claro que este análisis podría ser realizado con más rigurosidad desde el punto de vista económico, seleccionando escenarios macroeconómicos completos para que fuesen coherentes con la realidad, pero lo que se quiere aportar es una idea estadística de cómo afectan las variables de forma individual, y ese objetivo, en cierta medida, ha sido alcanzado. Ahora nos marcamos como meta realizar previsiones con los modelos creados, pero para ello debemos modelizar las variables explicativas de cada modelo.

5.2. Modelización de series de tiempo

Una vez que los modelos para explicar el entorno económico español han sido creados, lo que se pretende ahora es emplearlos para hallar cómo este comportamiento evoluciona a futuro, en particular, lo deseado será un estudio hasta Diciembre de 2017. Pero para tal predicción es necesario modelizar todas las variables explicativas que en dichos modelos toman parte. Tal objetivo puede ser basado en una concepción univariante acorde a lo expuesto en el Capítulo 1, corresponderá a la Sección 5.2.1 donde se propondrán modelos *ARIMA* para las covariables, o vectorial según lo descrito en el Capítulo 2, desarrollado en la Sección 5.2.2 y que generará modelos *VAR* y *VECM* para dichas variables.

De entre todas las variables que han resultado significativas en los modelos generados, debemos señalar que de algunas variables no será necesaria su modelización debido a estar dadas externamente por organismos oficiales. Éstas serán el Total del Paro Registrado, el Producto Interior Bruto y los Tipos de Interés de Depósito a 6 y 12 meses. Siendo las dos primeras tomadas como variables exógenas en el contexto vectorial, y los tipos de interés dados por el [Banco Central Europeo] serán empleados únicamente en cuanto se sustituyan las predicciones en el modelo correspondiente.

5.2.1. Análisis univariante

Como ya hicimos a la hora de mostrar los modelos, donde solo desarrollamos uno de ellos y del resto dimos las expresiones que los caracterizaba, en este momento no modelizaremos todas las variables explicativas que hemos empleado en las expresiones (5.1) y (5.2), sino que explicaremos la obtención de un modelo para una de ellas y el resto serán análogas. Decir que este apartado se concluirá con la realización de previsiones del indicador macroeconómico de estudio y de todo el resto de variables hasta el horizonte marcado. Señalar que la variable que seleccionaremos para el estudio será el Total de Afiliados a la Seguridad Social, x_2 , dado que ésta forma parte del modelo de la Formalización del Crédito al Consumo en Hogares, desarrollado con anterioridad, por lo que

parece razonable su selección. Podríamos haber escogido sin ningún problema alguna de las otras covariables referentes a tal modelo, descartando, claro está, el Total del Paro Registrado.

De acuerdo con el Capítulo 1, lo primero que debemos hacer es transformar la correspondiente serie en un proceso estacionario de orden dos. Partimos de su gráfico secuencial dispuesto en la Figura 5.3 (a), en la cual vemos que esta variable no parece tener problemas en cuanto al nivel de varianza se trata, con lo que no estaríamos en presencia de heterocedasticidad por lo que no habría que realizar transformación Box-Cox alguna. Sin embargo, en cuanto a la media, sí se tendría que actuar debido a que un efecto de tendencia está presente. Hecho que podemos comprobar si hallamos la función de autocorrelaciones simples (*ACF*), Figura 5.3 (b), ya que en ella observamos la significación de retardos lejanos asegurando entonces la existencia de tendencia. Por tal razón debemos diferenciar la serie de forma regular. Denotando \mathbf{x}_2 a los afiliados, aplicando diferencias regulares de primer orden, obtenemos $\mathbf{x}_2^1 = (1 - B)\mathbf{x}_2$ cuyo gráfico se encuentra en la Figura 5.3 (c). Comprobamos que la tendencia ha desaparecido. Si calculamos y graficamos la *ACF* y la función de autocorrelaciones parciales (*PACF*) en las Figuras 5.3 (d) y (e), respectivamente, vemos que dichos valores en su mayoría se encuentran dentro de los intervalos marcados en líneas azules, con la salvedad de la existencia de periodicidad a 12 instantes de tiempo, los cuales si son significativos. Pero éstos no alteran la estacionariedad de orden dos necesaria por no existir efecto de estacionalidad, lo único es que debemos indicar dicho periodo y no habrá problema.

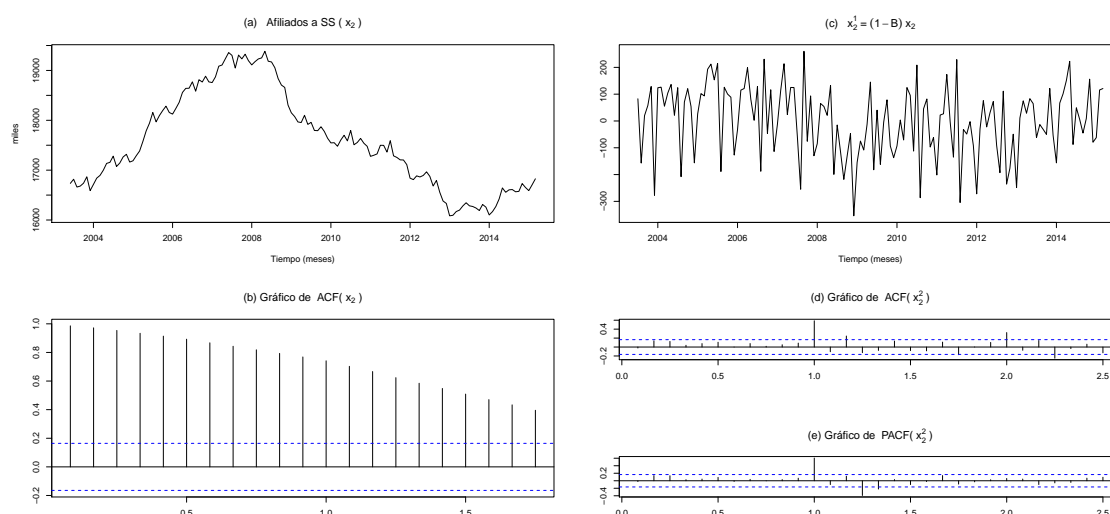


Figura 5.3: (a) Representación del gráfico secuencial de la variable del Total de Afiliados a la Seguridad Social en miles, \mathbf{x}_2 . (b) Representación de la *ACF* correspondiente. (c) Representación de la serie diferenciada regularmente para la eliminación de la tendencia, $\mathbf{x}_2^1 = (1 - B)\mathbf{x}_2$. (d) Representación de la *ACF* de \mathbf{x}_2^1 . (e) Representación de la *PACF* de \mathbf{x}_2^1 .

Ahora debemos saber qué modelo *ARIMA* se ajusta a dicha serie sabiendo que $d = 1$ (diferencias regulares) y $D = 0$ (diferencias estacionales). Para ello y atendiendo a la teoría expuesta con anterioridad, tomamos diferentes parámetros p , q , D y Q , y estimamos los modelos resultantes. Una vez realizadas dichas estimaciones, no las mostramos en este momento, calculamos ciertos criterios de información asociados a ellas y los tabulamos. En la siguiente tabla tomamos los criterios expuestos por Akaike, su corrección y el de Bayes, para los tres modelos que menores valores de los criterios calculan.

	<i>AIC</i>	<i>AICC</i>	<i>BIC</i>
$ARIMA(1, 1, 2) \times (1, 0, 1)_{12}$	1690.17	1690.80	1704.92
$ARIMA(2, 1, 2) \times (1, 0, 1)_{12}$	1692.13	1692.97	1709.83
$ARIMA(2, 1, 2) \times (1, 0, 0)_{12}$	1699.59	1700.66	1714.34

Lo que en la parte teórica se comentó implica que estos criterios de información sean calculados para los modelos cuya diagnosis sea correcta en el sentido que ya se indicó. Por tal razón

obviamos los dos primeros modelos mostrados en la tabla anterior, y seleccionamos el tercero, $ARIMA(1, 1, 2) \times (1, 0, 1)_{12}$. El cual estimaremos y corroboraremos que los respectivos contrastes son satisfactorios. Mostramos entonces la expresión que define a tal modelo,

$$(1 - \phi_1 B - \phi_2 B^2) (1 - \Phi_1 B^{12}) (1 - B)x_{2,t} = (1 + \theta_1 B + \theta_2 B^2)u_t,$$

cuya correspondiente estimación viene dada por,

$$\begin{aligned} \hat{\phi}_1 &= 1.07 \quad (0.19), \quad \hat{\phi}_2 = -0.29 \quad (0.19), \quad \hat{\theta}_1 = -1.24 \quad (0.16), \\ \hat{\theta}_2 &= 0.63 \quad (0.15), \quad \hat{\Phi}_1 = 0.70 \quad (0.06). \end{aligned}$$

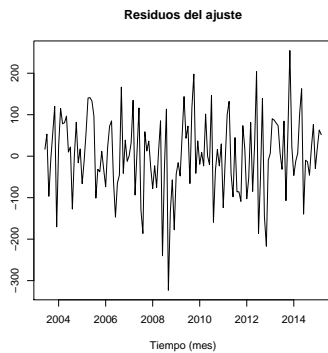
siendo los valores entre paréntesis las desviaciones típicas de las respectivas estimaciones, empleadas para el contraste de la significación de cada uno de los parámetros. Con ellas se comprueba que el parámetro ϕ_2 no es significativo a nivel de significación 0.05, con lo que lo eliminaríamos del modelo. Por lo que resultaría el modelo $ARIMA(1, 1, 2) \times (1, 0, 0)_{12}$,

$$(1 - \phi_1 B) (1 - \Phi_1 B^{12}) (1 - B)x_{2,t} = (1 + \theta_1 B + \theta_2 B^2)u_t,$$

con estimación,

$$\hat{\phi}_1 = 0.81 \quad (0.08), \quad \hat{\theta}_1 = -1.03 \quad (0.10), \quad \hat{\theta}_2 = 0.40 \quad (0.10), \quad \hat{\Phi}_1 = 0.70 \quad (0.06),$$

donde ahora todos los parámetros son significativamente distintos de cero con nivel 0.05. Para tal estimación se tiene que la desviación típica tiene un valor de $\sigma^2 = 8977$. Ahora, para comprobar que dicha modelización es correcta se deben de realizar ciertos contrastes, de forma análoga a los desarrollados en la anterior sección, esto es, se debe verificar que los residuos generados por el ajuste realizado \hat{u} se comportan como un proceso de ruido blanco. Con tal fin mostramos la figura siguiente, para la media nula y la homocedasticidad, y la tabla que recoge los respectivos contrastes de media nula, normalidad e incorrelación.



	t-test	shapiro-wilks	Jarque-Bera
Estadístico	0.31	0.99	4.00
p-valor	0.76	0.54	0.14
Ljung-Box (retardos)			
	1	5	10
Estadístico	0.40	10.71	14.37
p-valor	0.51	0.06	0.16

A la luz de dicha gráfica y tabla no existen evidencias suficientes para el rechazo de las hipótesis correspondientes a la suposición de ruido blanco por parte de los residuos \hat{u} , por ello, se toma como no rechazable el modelo planteado para los afiliados y podremos pasar a la siguiente y última etapa, la de predicción de la variable x_2 a cierto horizonte h . Recordar que éstas se originarán con la consideración de dicho modelo para un instante futuro $T + h$, siendo T el instante final observado, y que el horizonte deseado es Diciembre de 2017 (sabiendo que x_2 está observado hasta Marzo de 2015, el modelo se empleará para $h = 1, \dots, 33$), esto es,

$$(1 - \hat{\phi}_1) (1 - \hat{\Phi}_1 B^{12}) (1 - B)x_{2,T+h} = (1 + \hat{\theta}_1 B + \hat{\theta}_2 B^2)\hat{u}_{T+h}.$$

También decir que en base al no rechazo de la normalidad de la variable de estudio, podemos tomar la expresión (1.12) y proponer intervalos de predicción de dicho indicador macroeconómico. Si éstas previsiones e intervalos son calculados como se indicó en teoría y se representan, se obtiene la Figura 5.4 donde en negro se establece la serie de valores observada, en línea roja se encuentran las predicciones hasta el último periodo de 2017 y de color verde discontinuo los intervalos de



Figura 5.4: Representación de la serie observada de los Afiliados a la Seguridad Social, en línea negra, junto con sus predicciones, en rojo, a través de la teoría de series de tiempo univariantes y sus intervalos de predicción por poseer residuos a los que no rechazamos la normalidad, en verde.

predicción correspondientes al 95%. Como reflexión económica podemos decir que hasta el 2017 existirá una tendencia positiva de los afiliados a la SS recuperando valores que no se tenían desde 2009. En relación con esta interpretación y para aportar una salida más cómoda de analizar, se establece el cálculo de tasas de variación interanual de dichas predicciones, las cuales nos dirán cómo de Diciembre de un año a mismo mes del año siguiente dicho indicador ha variado. Si las calculamos en porcentaje obtenemos los valores 3.53 %, 2.54 % y 1.74 %, indicándonos el progresivo aumento de afiliados en los próximos tres años.

Consideración de otras covariables

Realizamos ahora procesos análogos sobre las otras variables consideradas en los modelos y calculamos sus predicciones hasta el horizonte establecido. Señalar que no todas las variables que estamos considerando necesitan el mismo horizonte h , por ejemplo, teníamos que \mathbf{x}_2 necesitaba $h = 33$, pero \mathbf{x}_{10} termina de observarse en Enero de 2015 con lo que $h = 35$, y si consideramos las trimestrales que se observan hasta Diciembre de 2014, su horizonte es de $h = 12$. Con todo ello, y si tomamos las tasas de variación interanual en porcentajes, mostramos la Tabla 5.5 donde se recogen el resto de variables explicativas a modelizar presentes en los modelos definidos. Señalar que de éstas también podrían darse intervalos de predicción si verificasen la gaussianidad o, en último caso, unos intervalos orientativos por si la normalidad no se cumpliera, aunque en la tabla referida de tasas no están reflejadas.

Variable	Modelo	2015	2016	2017
\mathbf{x}_4	$ARIMA(1, 1, 0)$	21.65	8.35	3.04
\mathbf{x}_5	$ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$	43.79	0.00	0.00
\mathbf{x}_{10}	$ARIMA(0, 1, 1) \times (1, 1, 0)_{12}$	15.80	0.02	0.02
\mathbf{x}_{12}	$ARIMA(1, 1, 0) \times (1, 1, 0)_4$	2.49	2.88	2.81
\mathbf{x}_{13}	$ARIMA(1, 1, 0)$	0.70	0.13	0.03
\mathbf{x}_{14}	$ARIMA(0, 2, 1)$	0.88	0.87	0.86
\mathbf{x}_{15}	$ARIMA(0, 2, 1)$	2.35	2.30	2.25

Cuadro 5.5: Tasas de variación interanual hasta 2017 de todas las variables explicativas consideradas en los modelos expuestos anteriormente en (5.1) y (5.2). Éstas están halladas en base a las modelizaciones $ARIMA$ expuestas en la segunda columna de la tabla. La notación es la acorde a la introducida en la anterior sección.

Señalar que los valores antes mostrados están calculado mediante el *software* estadístico R v. 3.1.1 y específicamente modelos $ARIMA$, pero podrían considerarse otros modelos más generales de corte exponencial como podrían ser el de tendencia lineal de Holts o de Browns, de tendencia suavizada, o los aditivos y multiplicativos de Winters, para las cuales emplearíamos el programa IBM

SPSS Modeler v.16 citado al inicio e introducida su noción en el Apéndice C, el cual selecciona el que crea más oportuno de todos ellos mediante un ajuste experto. Aunque no ahondaremos en sus resultados, simplemente saber que se podría tomar este camino más genérico.

5.2.2. Análisis vectorial

En el anterior apartado hemos modelizado mediante modelos *ARIMA* univariantes las variables explicativas que resultaban significativas en los modelos de regresión hallados mediante el método *GLS*, pero en este momento tomaremos el Capítulo 2 y estudiaremos tales covariables en función de modelos vectoriales como el *VAR* y el *VECM*. Señalar que el desarrollo que a continuación mostramos es referente a las variables mensuales ($\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_{10}$) para no recargar demasiado el trabajo también con las trimestrales ($\mathbf{x}_{12}, \mathbf{x}_{13}, \mathbf{x}_{14}, \mathbf{x}_{15}$). Distinción realizada para no disminuir en exceso la cantidad de datos mensuales.

Como ya se apuntaba en el análisis de sensibilidad de los modelos en el Apartado 5.1.2, consideramos, para todos los modelos vectoriales, dos variables exógenas, las cuales serán \mathbf{x}_1 (Total del Paro Registrado) y $\mathbf{x}_{16} = \{x_{16,t}\}_t$ (PIB). Aquí ya se empieza a entender el porqué de que en la Tabla 5.4 apareciesen efectos de ellas en todos los modelos. En este punto tampoco se considerarán, como ocurría en el caso univariante, los Tipo de Interés de Depósito a 6 y 12 meses, ya que las previsiones de estos pueden ser obtenidas externamente.

Modelo *VAR*

Tomemos las siguientes dos matrices de datos para las variables endógenas y exógenas, respectivamente, en cierto instante de tiempo t ,

$$\mathbf{X}_t^{end} = (x_{2,t}, x_{4,t}, x_{5,t}, x_{10,t})' \in \mathcal{M}_{K \times 1} \quad \text{y} \quad \mathbf{X}_t^{ex} = (x_{1,t}, x_{16,t})' \in \mathcal{M}_{M \times 1},$$

siendo $K = 4$ y $M = 2$. Y con ellas, recurriendo a la expresión del *VAR* dada en (2.1), escribimos nuestro correspondiente modelo para cada instante $t \in \mathcal{I}$,

$$\mathbf{X}_t^{end} = \mathbf{A}_1 \mathbf{X}_{t-1}^{end} + \cdots + \mathbf{A}_p \mathbf{X}_{t-p}^{end} + \mathbf{C} \mathbf{D}_t + \mathbf{\Xi} \mathbf{X}_t^{ex} + \mathbf{u}_t,$$

siendo $\mathbf{\Xi} \in \mathcal{M}_{K \times M}$ la matriz asociada a las variables exógenas. En el desarrollo teórico no la mostrábamos por simplicidad, pero su adición no requiere mayor complicación. Señalar debido a un estudio más exhaustivo, que no mostramos por no recargar el presente estudio, los elementos determinísticos no serán significativos, con lo que no tendremos en cuenta $\mathbf{C} \mathbf{D}_t$ para la estimación que sigue.

Planteado el modelo para un p genérico, debemos especificarlo para unos ciertos valores de éste, en particular lo haremos variando su valor de 1 a 4, y una vez estimados los modelos correspondientes, seleccionamos *a priori* el que minimice los criterios de información que en teoría se mostraban. Si realizamos este proceso llegamos a la Tabla 5.6 análoga a la del caso univariante donde se muestran los valores de los criterios de información de Akaike, de Hamilton-Quinn y de Swartz, para cada orden seleccionado.

	Orden p			
	1	2	3	4
$AIC(p)$	-14.99	-15.17	-15.11	-15.31
$HQ(p)$	-14.72	-14.73	-14.49	-14.51
$SC(p)$	-14.33	-14.09	-13.59	-13.35

Cuadro 5.6: Valores de los diferentes criterios de información para diferentes órdenes de p considerados.

A la luz de dicha tabla podemos ver que los valores mínimos se alcanzan en $p = 4$, $p = 2$ y $p = 1$, respectivamente para cada criterio considerado. Por tal razón y pensando que el orden 1 puede ser

demasiado escaso, tomamos $p = 2$, y por consiguiente un modelo $VAR(2)$. Luego el proceso vendrá dado por

$$\mathbf{X}_t^{end} = \mathbf{A}_1 \mathbf{X}_{t-1}^{end} + \mathbf{A}_2 \mathbf{X}_{t-2}^{end} + \Xi \mathbf{X}_t^{ex} + \mathbf{u}_t. \quad (5.3)$$

Cuya estimación se explicita mediante las siguientes matrices,

$$\hat{\mathbf{A}}_1 = \begin{pmatrix} 0.63 & (0.12) & 0.00 & (0.01) & 0.00 & (0.00) & 0.01 & (0.00) \\ -3.86 & (2.39) & 0.33 & (0.11) & 0.00 & (0.01) & -0.00 & (0.08) \\ 29.77 & (47.02) & 1.17 & (2.13) & 0.51 & (0.11) & 0.65 & (1.47) \\ 4.78 & (3.56) & 0.34 & (0.16) & 0.01 & (0.01) & 0.36 & (0.11) \end{pmatrix},$$

$$\hat{\mathbf{A}}_2 = \begin{pmatrix} 0.22 & (0.11) & 0.01 & (0.01) & 0.00 & (0.00) & 0.00 & (0.00) \\ 4.26 & (2.19) & 0.03 & (0.11) & 0.00 & (0.01) & 0.17 & (0.07) \\ -67.83 & (43.19) & 0.19 & (2.13) & 0.28 & (0.10) & 2.05 & (1.44) \\ -6.40 & (3.27) & 0.01 & (0.16) & -0.01 & (0.01) & -0.07 & (0.11) \end{pmatrix},$$

$$\hat{\Xi} = \begin{pmatrix} -0.01 & (0.01) & 0.03 & (0.01) \\ -0.36 & (0.18) & 0.09 & (0.26) \\ -1.72 & (3.52) & 8.23 & (5.02) \\ -0.50 & (0.27) & 0.58 & (0.38) \end{pmatrix}.$$

Donde los valores entre paréntesis representan, de nuevo, las desviaciones típicas de las correspondientes estimaciones y con las cuales podemos ver que muchos de los coeficientes estimados no son significativos, pero para no complicar el desarrollo del modelo no los eliminaremos. Pasamos entonces a la validación del modelo alcanzado, donde el primer paso será comprobar si este proceso es estacionario, objetivo para el cual recurrimos a la Observación 2.2 donde se comprobaba ésta en función de una matriz \mathbf{A} dada por $\hat{\mathbf{A}}_1$ y $\hat{\mathbf{A}}_2$. En dicha reflexión se establece la estacionariedad, si y solo si, los autovalores de \mathbf{A} son menores que 1. Si los calculamos se obtiene,

$$\lambda = (0.89, 0.89, 0.61, 0.44, 0.44, 0.398, 0.29, 0.29)$$

por lo que se comprueba que $\lambda_i < 1$ para todo i y se asegura así la estacionariedad del proceso estimado.

Vista la estabilidad, realizamos la batería de contrastes que se expusieron en la Sección 2.3 y los resumimos en la Tabla 5.7 donde mostramos el test de incorrelación en base a lo expuesto por Breusch-Godfrey (por trabajar con muestras de tamaños no muy amplias), el de heterocedasticidad basado en modelos $ARCH$, el Jarque-Bera multivariante para la normalidad acompañado de la curtosis y la asimetría.

	Estadístico	p -valor
Incorrelación (BG)	378.73	0.05
$ARCH$ VAR	517.50	0.29
Jarque-Bera VAR	10.26	0.25
Kurtosis	7.59	0.11
Asimetría	2.67	0.62

Cuadro 5.7: Resumen de contrastes multivariantes para el modelo $VAR(2)$ ajustado. Se realizan los test de incorrelación según Breusch-Godfrey, el llamado $ARCH$ -test para la heterocedasticidad, Jarque-Bera para la normalidad, y curtosis y asimetría en relación a esta última.

En tal tabla podemos ver que no existen evidencias para el rechazo de ninguna de las hipótesis básicas que se han de cumplir, por lo que el modelo está correctamente planteado. Por último comprobamos la estabilidad estructural del proceso ajustado empleando la herramienta $OLS-CUSUM$ [Ploberger, W. 1992] análoga al test de Chow que en la teoría fue citado. Ésta se basará en el cálculo del proceso de fluctuación empírico para observar grandes cambios de nivel en el modelo. De forma gráfica, tomamos la Figura 5.5 con cada una de las variables endógenas consideradas y observamos que las cuatro se comportan de una forma adecuada en cierto intervalo mostrado con línea roja, por lo que aseguramos la estabilidad estructural del ajuste.

Modelo $VECM$

Alcanzado el $VAR(2)$ adecuado para los valores considerados, debemos pasar al siguiente de los modelos, el $VECM$. Éste se define para un orden $p - 1$, como se dijo en teoría, por tanto se tomará

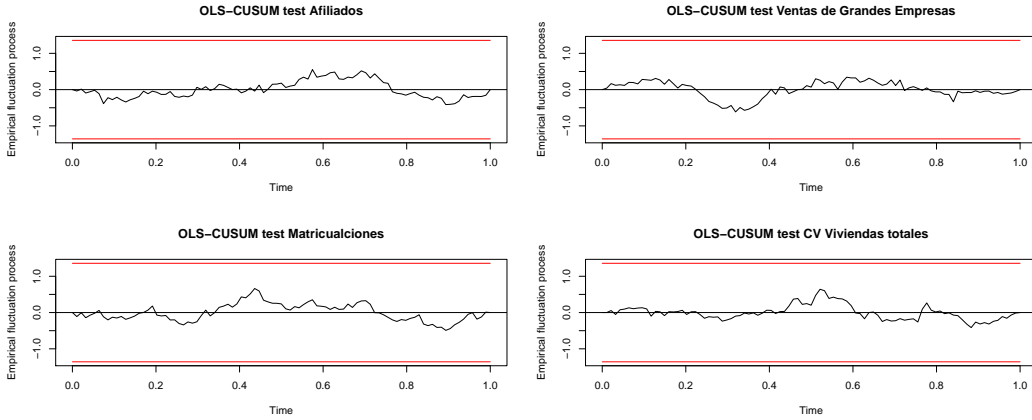


Figura 5.5: Representación del ajuste por medio de *OLS-CUSUM* para comprobar la estabilidad estructural de cada una de las variables endógenas utilizadas con respecto al modelo vectorial *VAR(2)* ajustado.

un *VECM(1)* originado la siguiente expresión,

$$\Delta \mathbf{X}_t^{end} = \mathbf{C} + \mathbf{\Pi} \mathbf{X}_{t-1}^{end} + \mathbf{\Gamma}_1 \Delta \mathbf{X}_{t-1}^{end} + \mathbf{\Upsilon} \mathbf{X}_t^{ex} + \mathbf{u}_t. \quad (5.4)$$

Añadimos una constante $\mathbf{C} \in \mathbb{R}_K$ porque en la elaboración que sigue puede comprobarse su significación. Decir que la matriz $\mathbf{\Upsilon} \in \mathcal{M}_{K \times M}$, como antes $\mathbf{\Xi}$, se asocia a las variables exógenas que en la parte teórica habíamos omitido.

Como se ha comentado en el capítulo correspondiente, la gran virtud de esta expresión es que las relaciones de cointegración aparecen de forma explícita dadas a través de la matriz $\mathbf{\Pi}$ de rango r . Para hallar tal parámetro r , realizamos para ciertos valores de éste, una sucesión de contrastes dados en base al estadístico de la expresión (2.14), alcanzando los correspondientes estadísticos recogidos en la Tabla 5.8. Con la cual, basándonos en los cuantiles, determinamos que $r = 2$ para todo nivel de significación usual.

	Estadístico	Quantiles		
		0.01	0.05	0.10
$r \leq 3$	7.06	10.49	12.25	16.26
$r \leq 2$	12.50	16.85	18.96	23.65
$r \leq 1$	29.06	23.11	25.54	30.34
$r = 0$	68.86	29.12	31.46	36.65

Cuadro 5.8: Estadísticos originados al plantear los contrastes para las hipótesis nulas acerca del rango de cointegración r . A la derecha se especifica el valor de los cuantiles teóricos en cada situación.

Si ahora estimamos todas las componentes del modelo resultante, sabiendo que $\hat{\mathbf{\Pi}} = \hat{\boldsymbol{\alpha}} \hat{\boldsymbol{\beta}}'$ para $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}} \in \mathcal{M}_{K \times r}$, obtenemos,

$$\hat{\mathbf{C}} = \begin{pmatrix} 6.08 & (0.00) \\ -13.62 & (28.05) \\ 381.28 & (598.35) \\ 30.41 & (43.28) \end{pmatrix}, \quad \hat{\mathbf{\Gamma}}_1 = \begin{pmatrix} -0.18 & (0.11) & 0.00 & (0.01) & 0.00 & (0.00) & 0.00 & (0.00) \\ -3.90 & (2.06) & -0.00 & (0.11) & -0.01 & (0.01) & -0.17 & (0.07) \\ 37.96 & (43.93) & -0.00 & (2.27) & -0.37 & (0.11) & -0.83 & (1.42) \\ 7.33 & (3.18) & 0.18 & (0.16) & 0.00 & (0.01) & 0.10 & (0.10) \end{pmatrix},$$

$$\hat{\mathbf{\Upsilon}} = \begin{pmatrix} -0.01 & (0.01) & -0.14 & (0.03) \\ -0.37 & (0.13) & 1.15 & (0.64) \\ 2.73 & (2.84) & -11.78 & (13.7) \\ -0.51 & (0.21) & -0.11 & (0.99) \end{pmatrix}, \quad \hat{\mathbf{\Pi}} = \hat{\boldsymbol{\alpha}} \hat{\boldsymbol{\beta}}' = \begin{pmatrix} -0.02 & -0.01 \\ -2.50 & -0.74 \\ 5.99 & 0.88 \\ -1.36 & 0.01 \end{pmatrix} \cdot \begin{pmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \\ -0.01 & 0.00 \\ 0.53 & -1.97 \end{pmatrix}'.$$

Como antes, existen muchos valores no significativos en las estimaciones, pero no los eliminaremos por simplicidad en el desarrollo. Podría ahora realizarse la conversión de este *VECM* al *VAR* con las relaciones entre las matrices que en el Capítulo 2 se muestran, pero no la realizamos por

resultar unos valores similares a los mostrados con anterioridad para el *VAR*, como era de esperar. Notar que la matriz $\hat{\beta}$ está transformada para que su matriz superior de rango r sea la identidad, con lo que por lo dicho en teoría al respecto, daríamos con un estimador consistente para β .

Una vez estimado, y como procedimos con el anterior modelo, realizamos los contrastes adecuados, siendo éstos resumidos en la Tabla 5.9. En ella comprobamos que no existen evidencias para el rechazo de las hipótesis consideradas, con lo que podremos tomar este modelo como válido y pasar a la siguiente y última etapa de predicción.

	Estadístico	p -valor
Incorrelación (BG)	380.00	0.05
<i>ARCH VAR</i>	518.66	0.27
Jarque-Bera <i>VAR</i>	6.78	0.56
Kurtosis	5.61	0.23
Asimetría	1.17	0.88

Cuadro 5.9: Resumen de contrastes multivariantes para el modelo ajustado *VECM*(1) con $r = 2$. Se realizan los test de incorrelación según Breusch-Godfrey, *ARCH*-test para la heterocedasticidad, Jarque-Bera para la normalidad y curtosis y asimetría en relación a ella.

Predicción

Una vez han sido hallados los dos respectivos modelos para las variables mensuales, lo que se requiere ahora es hallar previsiones con ellos. Pero primero debemos de aportar diferentes escenarios donde se establezcan los comportamientos de las variables exógenas que estamos considerando (paro y PIB), también aparecerán los futuros de los tipos de interés que en las sustituciones posteriores de los modelos tendrán su relevancia. Los escenarios propuestos son los mostrados en la Tabla 5.10, el primero será el Base (*B*) que supone un paulatina recuperación de la economía española y el otro será el Adverso (*Ad*) que prevé una nueva recesión en dicha economía.

Tasas de Variación (%)	Escenarios					
	Base (<i>B</i>)			Adverso 1 (<i>Ad</i>)		
	2015	2016	2017	2015	2016	2017
Total del Paro Registrado (\mathbf{x}_1)	-4.00	-4.00	-4.00	4.00	2.00	-1.00
PIB (\mathbf{x}_{16})	2.00	2.10	2.20	-0.40	0.40	0.50
Tipo de Interés a 6 meses (\mathbf{x}_9)	0.15	0.19	0.50	0.13	0.13	0.13
Tipo de Interés a 12 meses (\mathbf{x}_{11})	0.32	0.38	0.78	0.25	0.25	0.25

Cuadro 5.10: Tasas de variación interanual en % para el Total de Paro Registrado y PIB, y valor de los Tipos de Interés a 6 y 12 meses, en los escenarios *B* y *Ad*, respectivamente, para los años 2015, 2016 y 2017.

Planteados los escenarios a considerar en base a las variables exógenas, vamos a tomar éstas e introducirlas en las expresiones (5.3) y (5.4), y calcularemos los futuros hasta 2017 de las covariables mensuales consideradas. Si esto es realizado, obtenemos las salidas gráficas dispuestas en las Figuras de 5.6 hasta 5.9, donde se establecen las previsiones de las covariables mensuales para el modelo *VAR* bajo los escenarios *B* y *Ad*, y para el modelo *VECM* de forma análoga, respectivamente. Decir que, en ellas, en líneal azul se representan las previsiones correspondientes y en rojo los intervalos marginales a nivel 95 % de confianza, acorde con lo expuesto en la expresión (2.20) teórica.

Vistas las salidas gráficas, éstas podemos resumirlas con tasas de variación interanual de cada una de las variables consideradas. En Tabla 5.11 y Tabla 5.12 recogemos tales valores para el *VAR* y *VECM* en ambos escenarios considerados, respectivamente. Señalar que también incluimos en estos cuadros las tasas referentes a las variables trimestrales, de las cuales no mostramos los modelos generados pero si estas últimas salidas numéricas. Solo comentar que estos valores han resultado de modelos *VAR*(2) y *VECM*(1) con 2 relaciones de cointegración para las trimestrales.

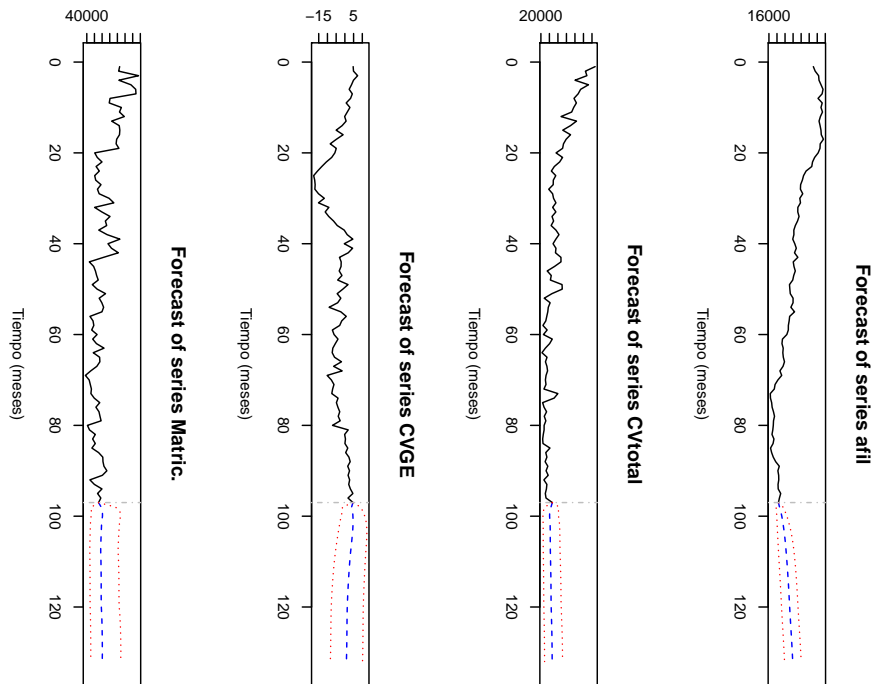


Figura 5.6: Representación de las predicciones de las variables endógenas (Total de Afiliados a la SS, Total de Compra-Venta de Viviendas, Ventas Totales de Grandes Empresas y Total de Matriculación de Turismos) en función del modelo *VAR* bajo el escenario Base. En negro las series observadas, en azul la estimación de las previsiones y en rojo los intervalos marginales al 95% de cada uno de los indicadores macroeconómicos considerados.

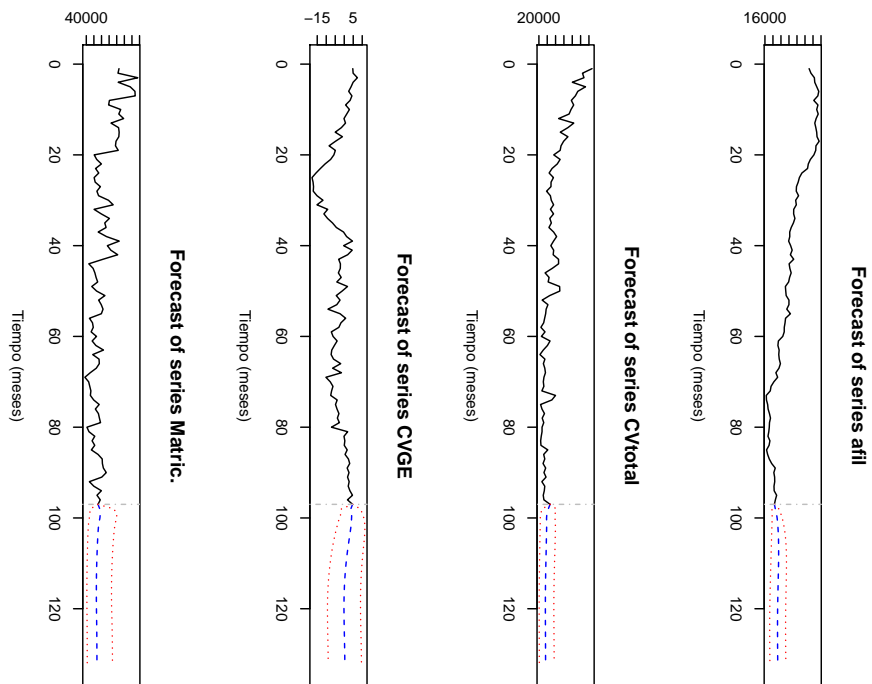


Figura 5.7: Representación de las predicciones de las variables endógenas (Total de Afiliados a la SS, Total de Compra-Venta de Viviendas, Ventas Totales de Grandes Empresas y Total de Matriculación de Turismos) en función del modelo *VAR* bajo el escenario Adverso. En negro las series observadas, en azul la estimación de las previsiones y en rojo los intervalos marginales al 95% de cada uno de los indicadores macroeconómicos considerados.

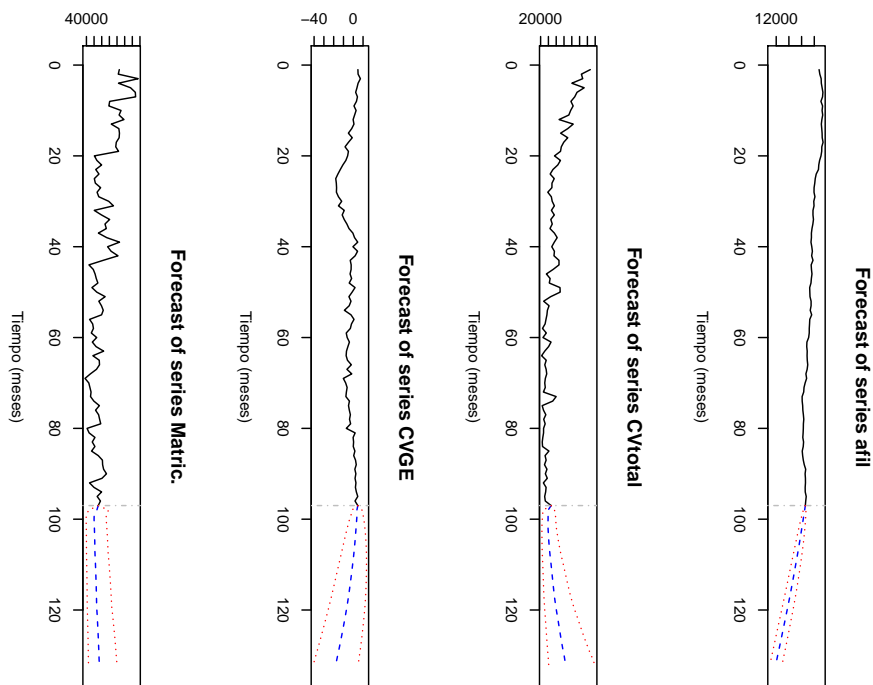


Figura 5.8: Representación de las predicciones de las variables endógenas (Total de Afiliados a la SS, Total de Compra-Venta de Viviendas, Ventas Totales de Grandes Empresas y Total de Matriculación de Turismos) en función del modelo *VECM* bajo el escenario Base. En negro las series observadas, en azul la estimación de las previsiones y en rojo los intervalos marginales al 95% de cada uno de los indicadores macroeconómicos considerados.

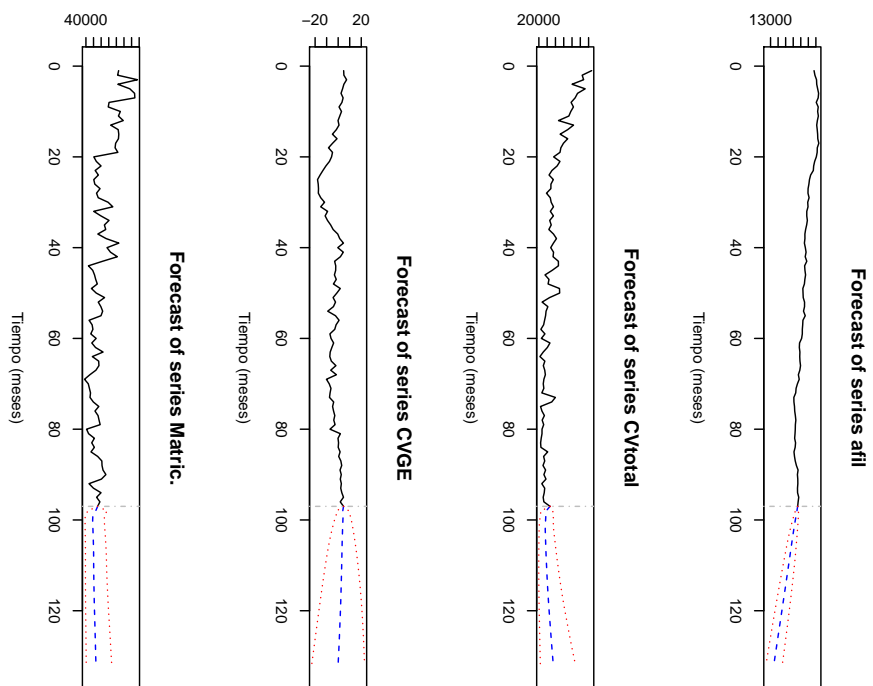


Figura 5.9: Representación de las predicciones de las variables endógenas (Total de Afiliados a la SS, Total de Compra-Venta de Viviendas, Ventas Totales de Grandes Empresas y Total de Matriculación de Turismos) en función del modelo *VECM* bajo el escenario Adverso. En negro las series observadas, en azul la estimación de las previsiones y en rojo los intervalos marginales al 95% de cada uno de los indicadores macroeconómicos considerados.

		Modelo <i>VAR</i>							
		x_2	x_4	x_5	x_{10}	x_{12}	x_{13}	x_{14}	x_{15}
<i>B</i>	2015	2.24	19.38	111.77	0.04	-11.13	7.97	2.02	2.29
	2016	1.37	2.91	-58.84	1.21	1.23	1.42	1.18	0.16
	2017	1.12	3.95	-33.83	4.01	1.18	-2.67	1.09	-0.21
<i>Ad</i>	2015	1.207	10.85	25.76	-9.92	-11.33	8.20	1.96	2.28
	2016	-0.31	-3.64	-104.75	-3.75	0.93	1.20	0.98	-0.06
	2017	-0.04	0.76	-362.06	2.19	0.672	-3.70	0.60	-0.70

Cuadro 5.11: Tasas de variación interanual hasta 2017 de todas las variables explicativas consideradas en los modelos expuestos en (5.1) y (5.2), que no son dadas por organismos oficiales. Éstas están halladas en base a un modelo *VAR*(2) con variables exógenas el PIB y el Total del Paro Registrado. Esta última solo es considerada para las variables mensuales (las cuatro primeras columnas).

		Modelo <i>VECM</i>							
		x_2	x_4	x_5	x_{10}	x_{12}	x_{13}	x_{14}	x_{15}
<i>B</i>	2015	-6.50	19.45	-26.49	-19.17	-6.73	3.840	-2.15	-0.24
	2016	-10.02	23.83	-518.98	7.25	-0.70	-1.18	-6.11	-2.30
	2017	-12.70	30.39	188.09	9.52	-6.10	-5.426	-15.04	-5.87
<i>Ad</i>	2015	-5.54	9.00	87.44	-21.98	-7.65	3.53	-1.20	0.11
	2016	-6.67	11.51	-34.09	3.48	0.09	-0.13	-2.667	-0.90
	2017	-7.34	15.64	-86.57	6.14	-2.24	-2.17	-5.86	-2.35

Cuadro 5.12: Tasas de variación interanual hasta 2017 de todas las variables explicativas consideradas en los modelos expuestos en (5.1) y (5.2), que no son dadas por organismos oficiales. Éstas están halladas en base un modelo *VEC*(1) para 2 relaciones de cointegración con variables exógenas el PIB y el Total del Paro Registrado. Esta última solo es considerada para las variables mensuales (las cuatro primeras columnas).

5.3. Predicción a través de los modelos econométricos

Hasta este punto se han desarrollado ciertos modelos econométricos mediante el método *GLS*, definidos en (5.1) y (5.2), para los cuales se han tomado los indicadores macroeconómicos que resultaron significativos y se han modelizado, individual y vectorialmente, para obtener predicciones con horizonte Diciembre de 2017. Lo que ahora se plantea es tomar esas predicciones y juntar éstas a las que los organismos oficiales nos aportan y son significativas en las expresiones de los modelos citados, éstas son, el paro y los tipos de interés mostrados en la Tabla 5.10. No tomaremos el PIB por no explicitarse en ninguna expresión, ya se ha tenido en cuenta como variable exógena. Una vez que tengamos todas las previsiones de covariables, las sustituiremos en los modelos y obtendremos, así, la evolución a futuro de las variables de negocio.

Señalar que, como cabe esperar, tendremos seis predicciones diferentes de cada modelo, esto es, una para cada modelo empleado sobre las covariables (*ARIMA*, *VAR* o *VECM*) y uno para cada escenario considerado (*B* o *Ad*). Apuntar que en todas las gráficas que mostremos, representaremos respectivamente en negro, verde y rojo, lo observado, lo previsto bajo el escenario *B* y bajo el *Ad*. Una vez visualicemos todos ellos, mostraremos como resumen numérico las tasas de variación interanual de las respuestas. Como ya comentamos, las de la FCCH y la de la Formalización del Crédito a la Vivienda serán acumuladas anualmente.

En la Figura 5.10 podemos observar las previsiones de forma gráfica de las cinco variables de negocio consideradas a través de modelizaciones *ARIMA* de las covariables en los dos marcos definidos, de izquierda a derecha tenemos la FCCH, Formalización del Crédito a la Vivienda, Total de Depósitos de Minoristas, Total de Dudosos OSR y Créditos a Empresas. La siguiente, Figura 5.11, muestra las predicciones de las respuestas respectivas en base al modelo *VAR*(2) para las explicativas, y de nuevo con los dos escenarios considerados. Y por último, la Figura 5.12 representa las variables de negocio con predicciones de las covariables con un *VECM*(1) con 2 relaciones de cointegración para los respectivos entornos.

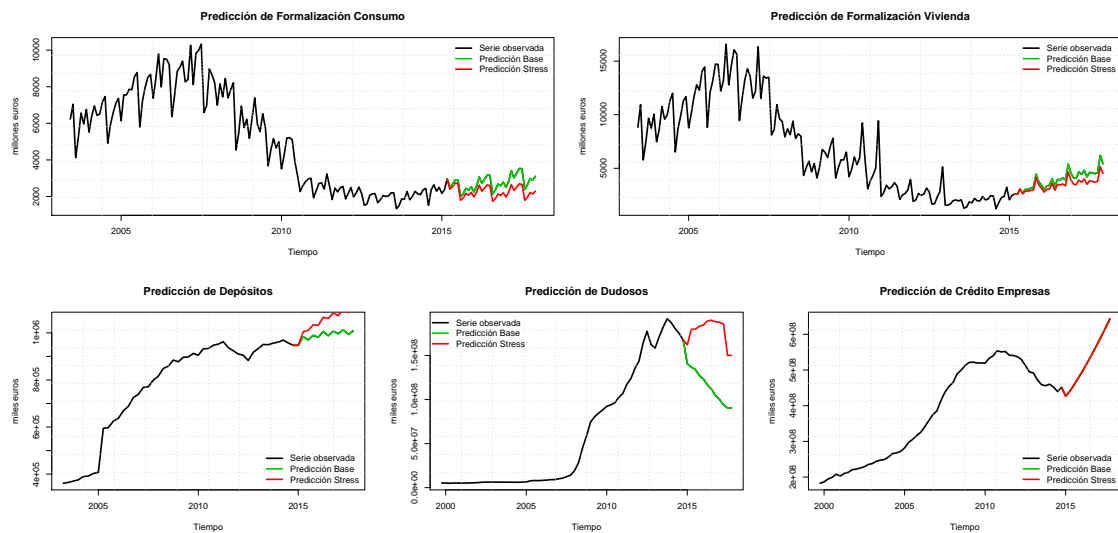


Figura 5.10: Comparativa de la previsión de las variables de negocio los próximos tres años, hasta 2017, bajo un escenario base *B* (verde) y bajo uno adverso *Ad* (rojo), empleando para la predicción de explicativas modelos *ARIMA*. En negro se representan los valores observados de cada variable. De izquierda a derecha y de arriba a abajo. Formalización del Crédito al Consumo en Hogares, Formalización del Crédito a la Vivienda, Total de Depósitos de Minoristas, Total de Dudosos OSR y Créditos a Empresas.

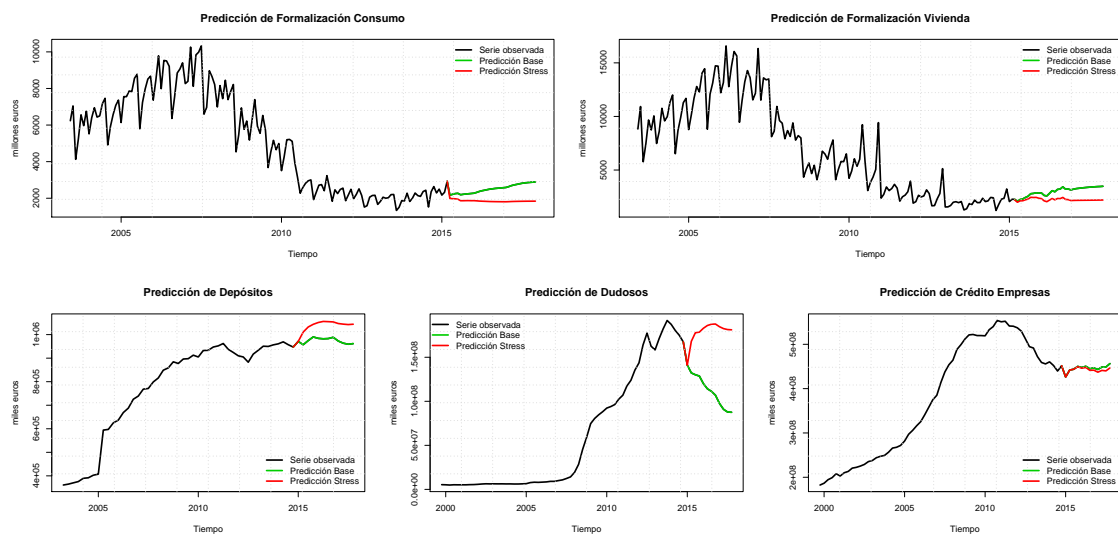


Figura 5.11: Comparativa de la previsión de las variables de negocio los próximos tres años, hasta 2017, bajo un escenario base *B* (verde) y bajo uno adverso *Ad* (rojo), empleando para la predicción de explicativas modelos *VAR*. En negro se representan los valores observados de cada variable. De izquierda a derecha y de arriba a abajo. Formalización del Crédito al Consumo en Hogares, Formalización del Crédito a la Vivienda, Total de Depósitos de Minoristas, Total de Dudosos OSR y Créditos a Empresas.

Comentar que para el modelo del Crédito a Empresas se ha realizado una corrección en la constante de éste, debido a que el comportamiento futuro, en tasa, es como cabía esperar atendiendo a lo apuntado por los organismos oficiales, pero partiendo de un primer valor predicho que no daba continuidad a la serie. Por tal razón, y con un simple cambio de localización, el comportamiento futuro se mantiene y se observa cierta continuidad.

En la Tabla 5.13 se recogen las tasas de forma porcentual para todos los modelos definidos con anterioridad, con diferentes estructuras para las explicativas, en columnas, y los dos escenarios considerados, en filas. Tanto en ella como en las figuras citadas, podemos ver grandes diferencias entre los resultados alcanzados entre *ARIMA* y *VAR* frente al *VECM*. Este último, tanto gráfica

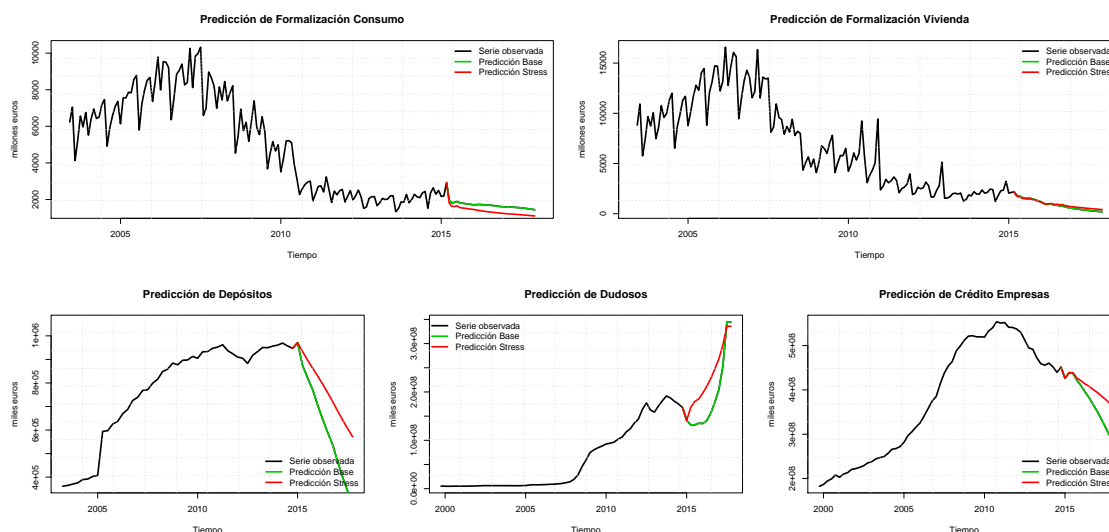


Figura 5.12: Comparativa de la previsión de las variables de negocio los próximos tres años, hasta 2017, bajo un escenario base B (verde) y bajo uno adverso Ad (rojo), empleando para la predicción de explicativas modelos $VECM$. En negro se representan los valores observados de cada variable. De izquierda a derecha y de arriba a abajo. Formalización del Crédito al Consumo en Hogares, Formalización del Crédito a la Vivienda, Total de Depósitos de Minoristas, Total de Dudosos OSR y Créditos a Empresas.

como numéricamente, para ambos escenarios, devuelve valores demasiado extremos para lo que puede suponerse para un futuro, lo cual puede ser sospechoso. Esta diferencia entre las tasas calculadas para las tres situaciones planteadas, podría ser explicada a través de la falta de idoneidad en los escenarios planteados para su sustitución en los modelos generados.

		ARIMA			VAR(2)			VECM(1), $r = 2$		
		2015	2016	2017	2015	2016	2017	2015	2016	2017
B	y^c	13.27	8.18	11.20	3.81	7.44	12.94	-10.51	-14.78	-8.82
	y^v	37.76	28.58	21.10	12.80	23.10	11.32	-22.24	-49.87	-58.91
	y^d	4.54	1.77	0.10	4.54	-0.15	-2.72	-18.48	-30.25	-46.27
	y^{du}	-24.09	-17.66	-13.48	-23.17	-16.93	-18.12	-19.21	31.90	93.22
	y^{cr}	5.12	16.69	16.19	-0.33	-0.79	2.21	-6.81	-12.90	-18.60
Ad	y^c	6.09	-4.43	2.31	-6.88	-10.93	0.03	-18.58	-25.51	-12.69
	y^v	32.12	18.38	14.40	3.00	0.19	-2.99	-25.01	-42.89	-41.58
	y^d	9.24	4.79	2.10	10.20	0.98	-0.92	-8.89	-16.35	-20.82
	y^{du}	9.41	2.82	-20.34	6.62	5.13	-3.49	10.83	33.32	35.61
	y^{cr}	5.12	16.69	16.19	-0.50	-1.72	1.13	-5.66	-5.95	-7.57

Cuadro 5.13: Tasas de variación interanual en % de las variables respuesta consideradas en el proyecto. Todas ellas en base a los dos escenarios que se han especificado en la Tabla 5.10. Como se indicó con notación, de arriba hacia abajo, Formalización del Crédito para el Consumo en Hogares, Formalización del Crédito para la Vivienda, Total de Depósitos, Total de Dudosos OSR y Créditos a Minoristas.

Para comprobar como de alejados son nuestros datos futuros con respecto a los observados que han servido para la creación de las expresiones comentadas, la herramienta más común es el cálculo de la distancia de Mahalanobis [Mahalanobis P. 1936]. Radicando su utilidad en que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales, diferenciándose de la distancia euclídea en que tiene en cuenta la correlación entre las variables aleatorias. Puede demostrarse que para dos muestras K -dimensionales de tamaño n , normales con vector de medias y matrices de covarianzas significativamente no diferentes, la distancia de Mahalanobis se distribuye aleatoriamente siguiendo una χ^2 con K grados de libertad. Con tal premisa suponemos esas hipótesis en nuestros datos y calculamos las distancias de los datos estimados de futuro con respecto a los observados y hallamos los p -valores asociados a ellas. De esta forma determinaremos si tiene sentido la utilización de tales datos en nuestros modelos.

En la Figura 5.13 representamos los p -valores para cada una de las situaciones contempladas en las Tabla 5.13, separando los modelos mensuales de los trimestrales. En ellas, en rojo discontinuo dibujamos la línea que marca el valor de un nivel de significación, 0.05. A la luz de dichas gráficas,

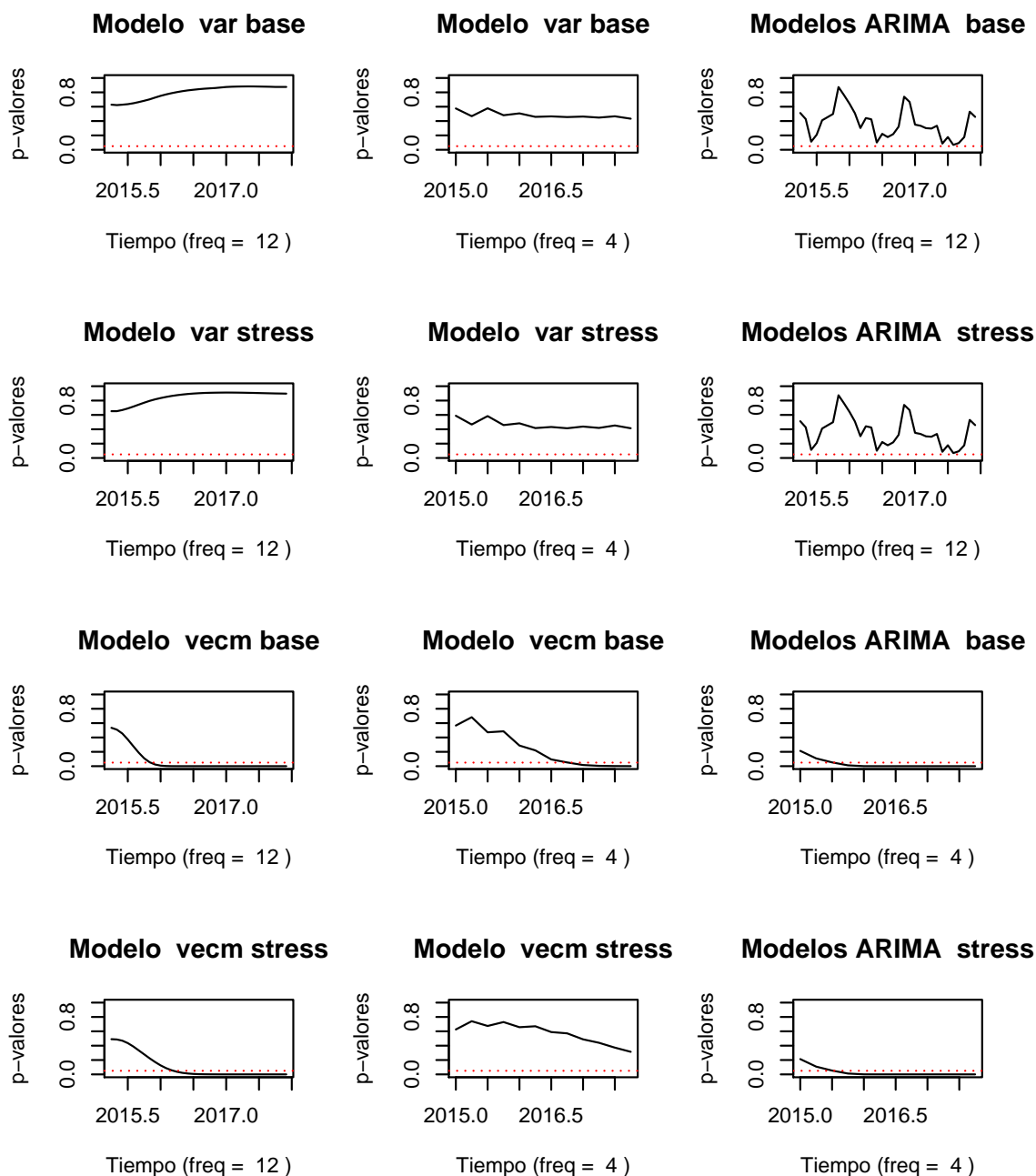


Figura 5.13: Representación de los p -valores derivados de la suposición de la distancia de Mahalanobis como una χ^2 entre muestras similares. Asociados éstos a las medidas entre los valores predichos por cada modelo y los valores empleados para tal expresión. En rojo línea marcando el valor 0.05. Entre los títulos de las gráficas y la respectiva aclaración en los ejes x , tanto los modelos como el escenario y la frecuencia de las variables seleccionadas, quedan claros.

los valores hallados con el modelo VECM están significativamente alejados de los observados y con ello no se puede asegurar la correcta predicción de éstas, por lo que no las tendremos en cuenta en lo que sigue. Lo mismo ocurre con las predicciones de los indicadores trimestrales con modelos *ARIMA* por lo que tampoco los consideraremos, siendo adecuadas las respectivas mensuales. Por otro lado, las previsiones basadas en el modelo *VAR*, sí que son más adecuadas, resultando sus p -valores asociados superiores a 0.05 para todo horizonte. Por tal observación, nos quedaremos con las predicciones procedentes del uso del modelo *VAR* para los indicadores macroeconómicos, siendo éstos los valores que se incluirían en un posible Informe de Adecuación de Capital (IAC).

Como hemos indicado que el *VAR* es el modelo con el cual mejores predicciones son alcanzadas en cuanto a su sustitución en las expresiones obtenidas mediante *GLS* se refiere. Exponemos en la Figura 5.14 la previsión de éstos acorde con la Figura 5.11 pero incluyendo intervalos de predicción. Señalar que como se indica en la Tabla 5.7 la normalidad multivariante no es rechazada por lo que podemos tomar los intervalos marginales y sustituirlos adecuadamente en los modelos.

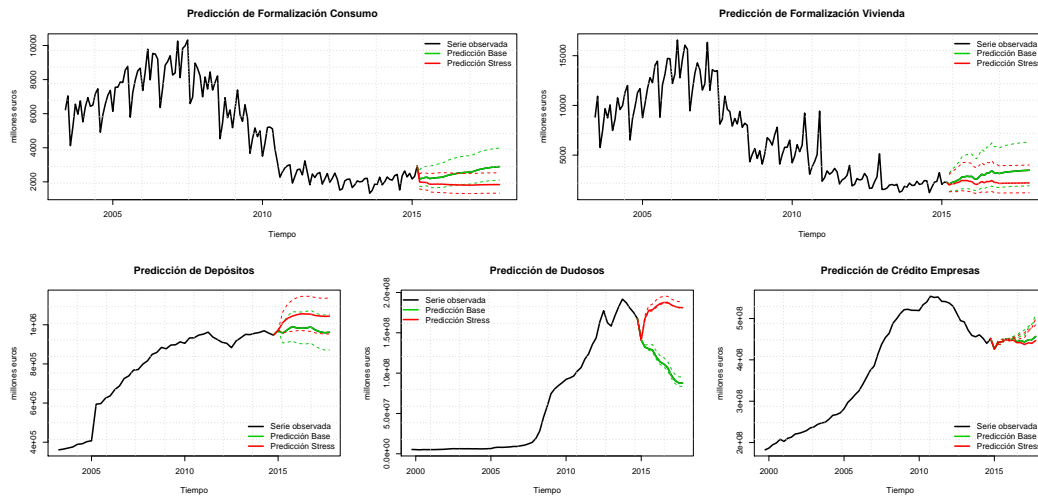


Figura 5.14: Representación de la Figura 5.11 con adición de los intervalos de predicción al 95 % generados a partir de la sustitución en los modelos ya referenciados de los intervalos marginales de predicción de las covariables respectivas. Los colores siguen el patrón ya comentado.

Por último, de forma sencilla, representamos la evolución de las tasas de variación desde el año 2010 para las mensuales y desde 2013 para las trimestrales hasta 2017. Éstas pueden observarse en las gráficas de la Figura 5.15, donde en negro se representan las tasas observadas, en verde las originadas a partir del escenario Base y tomando el Adverso en rojo.

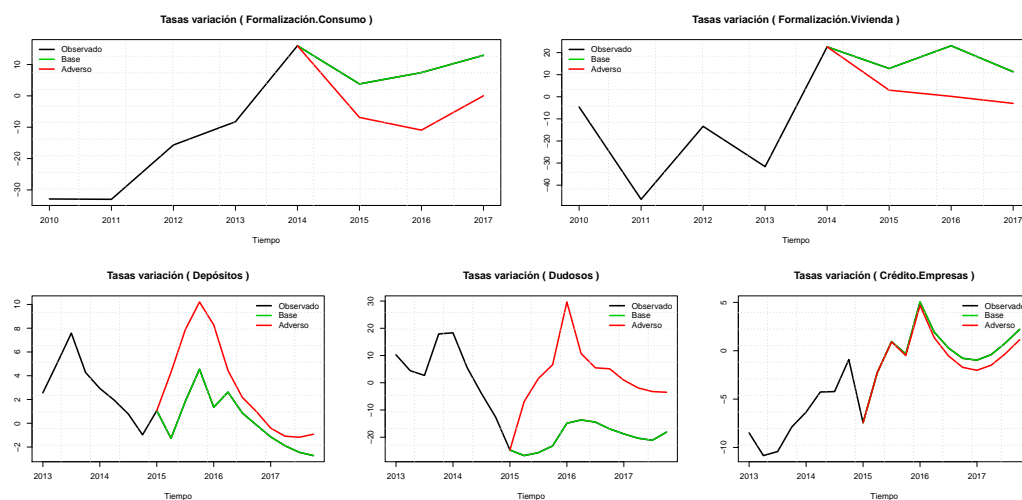


Figura 5.15: Evolución de las tasas de variación acumuladas e interanuales, respectivamente, de las variables de negocio a las cuales les hemos creado modelos en base a indicadores macroeconómicos. Comparamos el comportamiento bajo el escenario Base considerado (línea verde) y en presencia del Adverso (roja). De izquierda a derecha y de arriba a abajo. Formalización del Crédito al Consumo en Hogares, Formalización del Crédito a la Vivienda, Total de Depósitos de Minoristas, Total de Dudosos OSR y Créditos a Empresas.

Conclusiones

En este último punto del trabajo vamos a realizar algunas reflexiones acerca de cómo han resultado los modelos y técnicas estadísticas empleadas para generar el estudio econométrico del presente trabajo. Así como concluir cómo se comportará a futuro, según las variables de evolución del negocio, el entorno económico español.

Como apunte a las técnicas estadísticas empleadas hacemos hincapié en la diferencia de uso de los modelos *ARIMA*, *VAR* y *VECM*. Los primeros veíamos que no daban malos resultados con las variables mensuales, mientras que con las trimestrales fallaban más, esto podría ser debido a que se está intentando modelizar los indicadores macroeconómicos de forma individual sin atender a posibles relaciones entre ellos. Con los *VECM* se obtenían valores muy extremos lo cual se intentó explicar en base a la distancia de Mahalanobis, y que se puede explicar por la poca flexibilidad que se obtiene al fijar cierto parámetro r de cointegración. Por otra parte, los modelos *VAR*, tanto en su diagnosis al crearlos como con los resultados que se han obtenido con él, no han tenido ningún problema, por tal razón son las previsiones a futuro que se seleccionarán para entender el comportamiento de la economía a cierto horizonte.

En cuanto al comportamiento del entorno económico español se refiere, podemos distinguir dos posibles visiones, acorde con cada uno de los escenarios planteado. Señalar que estas reflexiones serán basadas en los resultados alcanzados con los modelos *VAR* sobre los indicadores macroeconómicos. Si nos situamos bajo el escenario optimista que hemos llamado Base, podemos decir:

- ↪ Las tendencias tanto de la Formalización del Crédito para el Consumo en Hogares como de la Vivienda son positivas. En la primera de forma ligera al principio, llegando posteriormente a un crecimiento cercano al 10%. Y en la segunda con variaciones superando con creces dicho porcentaje.
- ↪ Los Depósitos de Minoristas comenzarán con un crecimiento del 4% pero disminuirán de forma paulatina en los siguientes dos años.
- ↪ Durante los tres años de estudio, los Dudosos OSR disminuirán tasas superiores al -15%, incluso cercanas al -20%.
- ↪ El comportamiento del Crédito a Empresas nos indica un decrecimiento hasta el año 2017 con tasas cercanas al -0.5%, momento en el cual se prevé un ligero repunte positivo del 2%.

Supongamos ahora un cambio de escenario hacia uno adverso, en el que se estresa el entorno económico a través de una nueva recesión en España:

- ↪ En este caso la tendencia de la Formalización del Consumo en Hogares es negativa hasta llegar a 2017, con tasas de bajada del 5%, pero a partir de ese instante se predice un casi nulo crecimiento.
- ↪ Para la Formalización del Crédito a la Vivienda el crecimiento es mucho menor que bajo el escenario Base e incluso en 2017 disminuye casi un 3% del total.
- ↪ Los Depósitos de Minoristas comienzan con una subida del 10%, pero en los años siguientes se mantienen bastante estables con tasas del 1% y del -1%, respectivamente.
- ↪ Por su parte los Dudosos OSR aumentarán en los dos primeros años casi un 5% y al tercer año disminuirán un 3%.
- ↪ Por último, el comportamiento del Crédito a Empresas es muy similar al comentado para el escenario Base, con bajadas hasta 2017 y repunte positivo en ese año.

Podría reflexionarse más a fondo acerca del comportamiento futuro que se ha expuesto para las variables de negocio consideradas, pero por no complicar demasiado las observaciones ni introducirnos en el área donde los economistas son los especialistas, éstas las dejaremos con lo mostrado. Dejando así a tarea del lector interesado la búsqueda de repercusiones económicas y financieras que estas previsiones del entorno español traen consigo.

Apéndice A

Teorema de la descomposición de Wold

En este primer apéndice mostramos el enunciado el Teorema de Wold, el cual propone una cierta descomposición de una serie temporal en función de sus errores pasados, sirviendo ésta de motivación para el planteamiento de los modelos *ARIMA* expuestos en la primera parte del trabajo. En particular, este resultado se tendrá en cuenta en la expresión (1.7) de la Sección 1.2 donde un proceso descrito por un *ARMA* se interpreta como un caso particular de la descomposición que mostramos en este apéndice. Después de su exposición reflexionamos de forma breve sobre las componentes de tal descomposición.

Supongamos un proceso estacionario de orden dos, pudiendo ser modelizado mediante un *ARMA*, el cual es entendido como una combinación lineal de los valores pasados de la serie y los errores generados de la misma. Pero también puede reescribirse como una suma infinita de los antiguos valores de los errores, como hemos comentado. Aunque existe otra posibilidad, separar el proceso entre su parte determinista y estocástica, para ello Wold introduce el teorema que a continuación se muestra.

Teorema A.1 (Teorema de Wold). *Supongamos un proceso estacionario de segundo orden dado por $\mathbf{y} = \{y_t, t \in \mathcal{I}\}$ con $\mathbb{E}(\mathbf{y}) = 0$ y $\gamma_j = \mathbb{E}(y_t y_{t-j})$ para todo j . Entonces*

$$y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j} + \eta_j \quad \text{para } t \in \mathcal{I}, \quad (\text{A.1})$$

donde $\mathbf{u} = \{u_t, t \in \mathcal{I}\}$ es un proceso de ruido blanco independiente del proceso $\boldsymbol{\eta} = \{\eta_t, t \in \mathcal{I}\}$,

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty \quad \text{para } \psi_0 = 1,$$

y $\mathcal{P}(\eta_{t+s} | x_{t-1}, x_{t-2}, \dots)$ con $s \geq 0$. Siendo \mathcal{P} la representación del operador proyección.

La primera parte de la expresión (A.1), el sumatorio, es idéntica a la dada por un $MA(\infty)$, lo cual puede comprobarse tomando en la ecuación (1.5) el parámetro q y haciéndolo tender hacia infinito. Pero la parte novedosa recae en el segundo de los sumandos de tal expresión, siendo éste el elemento que encierra la parte determinista de la serie. Señalar que no podrá contener componente de tendencia, ya que si fuese así, el proceso \mathbf{y} no sería estacionario de orden dos, como se ha impuesto en las hipótesis iniciales del teorema.

De hecho, esta representación tiene diferencias notables con respecto de un $MA(\infty)$, una de ellas es con respecto a la estacionariedad. Un proceso *ARMA* basa ésta en que el polinomio asociado a la

parte autorregresiva no tenga raíces dentro del círculo unidad complejo, como se comprueba en la sección a la que hacíamos referencia. Por el contrario, la descomposición de Wold no establece ninguna restricción de este tipo para alcanzar la estacionariedad. Lo cual deja intuir la generalización de la que hablábamos antes del teorema.

Como se ha dicho, la razón de este apéndice es que el resultado sirve de motivación para que una serie temporal o proceso pueda explicarse mediante un sumatorio de errores pasados. Y serán los coeficientes ψ_j para $j = 0, 1, \dots$ asociados a tal sumatorio los denominados coeficientes de Wold, como ya se pueda ver en el desarrollo teórico.

Apéndice B

Utilización del *software* estadístico R

Paquetes empleados

A continuación mostramos la lista de las librerías que ha sido necesario instalar y cargar antes de la implementación de los códigos asociados a la aplicación práctica desarrollada en el Capítulo 5 del presente trabajo. Entre ellas citamos algunas que pertenecen al paquete base que se instala por defecto en el *software* al ser descargado e instalado [R Development Core Team 2008], éstas serán identificadas con la etiqueta *Paquete base* para una mayor aclaración. Señalar que de todas ellas mostraremos una breve descripción y, en ciertos casos, daremos el nombre de las principales funciones que se han utilizado para el desarrollo del trabajo y la referencia en la cual se presenta el correspondiente paquete.

- `leaps`: Selección de subconjuntos de regresión a través de una búsqueda exhaustiva.
- `lmtest`: Colección de contrastes, conjuntos de datos y ejemplos para la diagnosis en modelos de regresión lineal. A mayores, carga algunas herramientas genéricas para la inferencia en modelos paramétricos.
- `locfit`: Estimación por regresión local, verosimilitud y densidad.
- `MASS`: (*Paquete base*) Funciones y bases de datos para apoyar lo desarrollado en *Modern Applied Statistics with S* de Venables y Ripley, Cuarta edición, 2002.
- `methods`: (*Paquete base*) Métodos y clases definidas formalmente para objetos de R, más otras herramientas de programación.
- `mgcv`: Rutinas para *GAM*'s y otras regresiones *ridge* generalizadas con selección de parámetros suavizados múltiples mediante *GCV*, *REML* o *UBRE/AIC*. También incluye *GAMM*'s.
- `nlme`: Ajuste y comparación de linealidad y no linealidad Gaussiana en modelos de efectos mixtos [Pinheiro J. 2014].

`gls` y `corARMA`.

- `sandwich`: Estimadores estándar de modelos robustos para selección-cruzada, series temporales y datos longitudinales.
- `stats`: (*Paquete base*) Contiene funciones para cálculos estadísticos y generación de números aleatorios.

`agreggate.ts`, `ARMAtoMA`, `decompose`, `filter`, `Box.test`, `KalmanForecast`,
`lm`, `predict`, `shapiro.test`, `t.test`, `ts.plot` y `ts`.

- **strucchange**: Contrastar, monitorizar y señalar cambios estructurales en modelos de regresión. Métodos procedentes de contrastes de fluctuaciones generalizadas como el F -test (contraste de Chow). Esto incluye métodos de ajuste, gráficas y contrastes de procesos de fluctuaciones (*OLS-CUSUM*) y estadísticos F .

- **TSA**: Contiene funciones de R y bases de datos detallados asociados a las series de tiempo [Cryer, J. *et al.* 2008].

`prewhiten` y `arimax`.

- **tseries**: Paquete con funciones para el análisis de series temporales y de finanzas computacionales [Trapletti, A. *et al.* 2013].

`jarque.bera.test`.

- **urca**: Contraste de la raíz unidad y la cointegración en aplicaciones de análisis econométrico.

`ca,jo` y `cajorls`.

- **vars**: Estimación, selección de retardos, contrastes de diagnosis, predicción, análisis de causalidad, descomposición de la varianza de error de predicción y funciones de respuesta de impulso en los modelos *VAR* y *VECM* [Pfaff, B. 2008b].

`arch.test`, `fanchart`, `normality.test`, `roots`, `serial.test`, `stability`, `VAR`,
`VARselect` y `vec2var`.

- **zoo**: Métodos para el orden total de las observaciones. Es particularmente propicio para series de tiempo irregulares de vectores, matrices y factores.

Funciones programadas

A mayores de todas las funciones que se han mostrado en la anterior sección y que se detallan en las referencias indicadas, se han programado otras para una mayor eficiencia y agilidad a la hora de ejecutar ciertos *scripts*. Éstas, con su descripción de parámetros de entrada y de salida, son mostradas a continuación por orden alfabético.

- **best.arima.TSA**: Función que ordena modelos *ARIMA* con sus correspondientes órdenes en base al valor obtenido para el criterio de información que se indique.

Uso

```
best.arima.TSA (x = x, p.max = 5, q.max = 5, d = 0, P.max = 0, Q.max = 0,
               D = 0, p.min = 0, q.min = 0, P.min = 0, Q.min = 0, period
               = 1, xreg = NULL, include.mean = TRUE, criterio = "BIC",
               dist.max.crit = 2)
```

Argumentos

- `x` serie de tiempo univariante la cual quiere modelizarse con modelos Box-Jenkins.
- `p.min` mínimo orden para tomar en la parte autorregresiva del modelo. Por defecto, 0.
- `q.min` mínimo orden para tomar en la parte de medias móviles del modelo. Por defecto, 0.
- `p.max` máximo orden para tomar en la parte autorregresiva del modelo. Por defecto, 5.
- `q.max` máximo orden para tomar en la parte autorregresiva del modelo. Por defecto, 5.
- `d` número de diferencias regulares aplicadas para eliminar la tendencia de la serie temporal. Por defecto, 0.
- `P.min` mínimo orden para tomar en la parte autorregresiva estacional del modelo. Por defecto, 0.

- Q.min** mínimo orden para tomar en la parte de medias móviles estacional del modelo. Por defecto, 0.
- P.max** máximo orden para tomar en la parte autorregresiva estacional del modelo. Por defecto, 0.
- Q.max** máximo orden para tomar en la parte de medias móviles estacional del modelo. Por defecto, 0.
- D** número de diferencias estacionales aplicadas para la eliminación en la estacionalidad de la serie temporal. Por defecto, 0.
- period** es el valor que refleja el periodo de estacionalidad de la serie. Por defecto, `period = 1`.
- xreg** opcionalmente, un vector o matriz de regresores externos, los cuales deben tener el mismo número de filas que `x`. Por defecto, `xreg = NULL`.
- include.mean** parámetro lógico que indica si el modelo *ARIMA* debe incluir una media o término constante. Por defecto, `include.mean = TRUE`.
- criterio** tipo de criterio aplicado al modelo resultante para comprobar la información que puede explicar. Existe la opción de *AIC*, *AICC* y *BIC* (por defecto).
- dist.max.crit** distancia máxima a partir del mejor modelo, en el sentido del criterio, que se quiere mostrar por pantalla. Por defecto, `dist.max.crit = 2`.

Valores

- p** orden de autoregresión del modelo mostrado.
- q** orden de medias móviles del modelo mostrado.
- P** orden de autoregresión de la parte estacional del modelo mostrado.
- Q** orden de medias móviles de la parte estacional del modelo mostrado.
- criterio** valor del criterio de información asociado al respectivo modelo.

```
best.arima.TSA <- function(x = x, p.max = 5, q.max = 5, d = 0, P.max = 0,
  Q.max = 0, D = 0, p.min = 0, q.min = 0, P.min = 0,
  Q.min = 0, period = 1, xreg = NULL, include.mean = TRUE,
  criterio = "BIC", dist.max.crit = 2)
{
  library(TSA)
  if (is.ts(x)) period <- frequency(x)
  num.x.perdidos <- d + period*D
  T <- length(x) - num.x.perdidos
  # Creamos una matriz que posteriormente contendrá todas las combinaciones
  # de los órdenes considerados y los valores de la función criterio para
  # los ARIMAs correspondientes
  A <- matrix(0, ((p.max-p.min)+1)*((q.max-q.min)+1)*((P.max-P.min)+1)*
    ((Q.max-Q.min)+1), 5)
  # La penalización que sobre la cantidad de parámetros imponen las
  # funciones
  #criterio AIC o BIC depende del factor definido abajo
  if (criterio=="AIC") factor <- 2
  else if (criterio=="BIC") factor <- log(T)
  fila <- 0
  for (p in p.min:p.max)
  for (q in q.min:q.max)
  for (P in P.min:P.max)
  for (Q in Q.min:Q.max) {
    #optim.control=list(maxit=500)
    fila <- fila + 1
    ajuste <- try(arimax(x=x, order=c(p,d,q),
```

```

                                seasonal=list(order=c(P,D,Q), period=period),
                                xreg=xreg, include.mean=include.mean), silent=TRUE)
if (class(ajuste)=="try-error") {
  A[fila, ] <- c(p, q, P, Q, NaN)
  next
}
k <- length(ajuste$coef)
# La penalización que sobre la cantidad de parámetros impone la función
#criterio AICC depende del factor definido abajo
if (criterio=="AICC") criterio.ajuste <- ajuste$aic +2*(k+1)*(k+2)/
                                (T-k-2)

else criterio.ajuste <- -2*ajuste$loglik + factor*k
A[fila, ] <- c(p, q, P, Q, criterio.ajuste)
}
min.ic <- min(A[,5], na.rm=TRUE)
A <- A[(A[,5]-min.ic)<=dist.max.crit,]
if (!is.matrix(A)) A <- t(as.matrix(A))
A <- A[!is.na(A[,1]),]
if (!is.matrix(A)) A <- t(as.matrix(A))
A <- A[order(A[,5]),]
if (!is.matrix(A)) A <- t(as.matrix(A))
A <- as.data.frame(A)
if (criterio=="AIC") names(A) <- c("p", "q", "P", "Q", "AIC")
else if (criterio=="AICC") names(A) <- c("p", "q", "P", "Q", "AICC")
else names(A) <- c("p", "q", "P", "Q", "BIC")
return(A[,c(1*(p.max!=0), 2*(q.max!=0), 3*(P.max!=0), 4*(Q.max!=0), 5)])
}

```

- **Box.Cox:** Función para buscar el mejor parámetro λ para la transformación de los datos de una serie de tiempo para conseguir que desaparezca su posible efecto de heterocedasticidad. Teóricamente esta transformación se define como se presenta en la Sección 1.3.

Uso

Box.Cox (x, la = seq(-1, 1, length = 10), lb = NULL, l.len = NULL)

Argumentos

- x serie de tiempo univariante que queremos transformar.
- la rango de valores de λ para tomar en la transformación definida. Por defecto, se tomará `seq(-1, 1, length = 10)`. Podría ser un único valor, pero conlleva la definición de los dos siguientes valores.
- lb extremo superior del intervalo de valores de λ . Por defecto, tomaremos NULL.
- l.len longitud para el intervalo de λ 's. Por defecto, tomaremos NULL.

Valores

- lambda valores de λ empleados, menos el 1 y el 0, por ser estos dos los correspondientes a la serie sin transformar (solo se aplica un cociente) y la serie con el logaritmo natural.
- R2 valores de R^2 para los diferentes ajustes realizados a diferentes λ 's.
- p.val p -valores para los diferentes ajuste definidos.
- lambda.h λ óptimo asociado al ajuste con menor R^2 .


```

Box.Cox <- function(asd, as)
{
  ## Verificaciones a realizar para los parámetros introducidos
  if (length(la) > 1 && length(lb)) lb = NULL
  if (!length(lb) && length(l.len) || length(lb) && !length(l.len)) {
    return("Definir bien el intervalo de lambda's")
  }
  if (!length(lb))
    l = la
  else l = seq(from = la[1], to = lb, length = l.len)
  l = l[l != 0 & l != 1]
  if(!length(l)) {
    return("lambda sólo contine 0 y/o 1, no puede ser calculado")
  }
  ## Regresión
  R2 = p.val = NULL
  for (i in 1:length(l)){
    x.t = (x^l[i] - 1)/l[i]
    media.anual <- aggregate.ts(x.t, FUN = "mean")
    desv.tip.anual <- aggregate.ts(x.t, FUN = "sd")
    log.media.anual <- log(media.anual)
    log.desv.tip.anual <- log(desv.tip.anual)
    aj <- lm(log.desv.tip.anual ~ log.media.anual)
    R2 = c(R2, sum((mean(log.desv.tip.anual) - aj$fitted.values)^2)/
           sum((log.desv.tip.anual - mean(log.desv.tip.anual))^2))
    p.val = c(p.val, summary(aj)$coef[2, 4])
  }
  lambda.h = round(l[which.min(R2)], 4)
  return(list(lambda = l, R2 = R2, p.val = p.val, lambda.h = lambda.h))
}

```

- **calc.tx.ct**: Función encargada de calcular predicciones de valores en base a tasas de variación y extenderla durante los periodos intermedios de esas tasas con el método de desagregación temporal. Si se tienen los valores de las predicciones se extienden con desagregación temporal y se calculan las respectivas tasas.

Uso

```

calc.tx.ct (serie, per.end = 3, freq = 12, start = c(2000, 3), tasas = NULL,
           n.años = 3, cant = NULL, do.pred = FALSE, cum = FALSE, type =
           'additive', años = 2014:2017)

```

Argumentos

- serie** serie temporal a introducir para el cálculo de tasas o valores de predicción.
- per.end** indicador del periodo en el que el estudio finaliza. Si acaba en Marzo y la frecuencia de la serie es 12, entonces **per.end=3** y si la frecuencia es 4, entonces **per.end=3**. Por defecto, 3.
- freq** frecuencia de medición de la serie. Por defecto, 12.
- start** vector bidimensional donde se recoge el año y el mes de comienzo de estudio de la serie. Por defecto, **start = c(2003, 3)**.
- tasas** tasas de variación interanual. Por defecto, NULL.
- n.años** número de años a predecir. Si el parámetro **tasas** no es NULL, éste **n.años** es la longitud de **tasas**.
- cant** valores en predicción. Por defecto, NULL.

- `do.pred` parámetro lógico que incida con `TRUE` que se quieren hacer predicciones con tasas de variación. Y con `FALSE` (por defecto) que no hace falta predecir, simplemente se quiere calcular tasas.
- `cum` parámetro lógico que indica con `TRUE` si se quiere calcular tasas de la serie de forma acumulada. Y con `FALSE` (por defecto) si no.
- `type` tipo de descomposición de la serie que se desea realizar, ' `additive`' (por defecto) o `"multiplicative"`.
- `años` vector de años para renombrar las columnas de la salida por pantalla de la función. Por defecto `años = 2014:2017`.

Valores

- `serie.orig` serie de tiempo original, la que se introduce en la función.
- `serie.compl` concatenación de la `serie.orig` y de los valores predichos. Si `do.pred=FALSE`, este valor es la serie introducida.
- `tasas.100` tasas de variación en porcentaje.
- `tabla` tabla con valores de los años que se consideran de predicción y sus respectivas tasas interanuales.

```

calc.tx.ct <- function(serie, per.end = 3, freq = 12,
  start = c(2000, 3), tasas = NULL, n.años = 3,
  cant = NULL, do.pred = FALSE, cum = FALSE,
  type = "additive", años = c("2014", "2015", "2016", "2017"))
{
  serie = ts(serie, start = start, freq = freq)
  ## Si no queremos hacer predicciones
  if ( !do.pred ) {
    if (cum) {
      serie1 = serie[ - (1 : (freq + 1 - per.end)) ]
      val = rowSums(matrix(serie1, ncol = freq, byrow = TRUE))
      tasas1 = as.vector(t(val[-1] / val[1:(length(val) - 1)] - 1) * 100)
      valores = t(tail(cbind(val[-1], tasas1), n.años))
      return(valores)
    } else {
      if (length(dim(serie)) == 0) {
        l.pred = n.años * freq + 1
        val = tail(serie, l.pred)[seq(1, l.pred, by = freq), ]
        tasas1 = as.vector(t(val[-1] / val[1:(length(val) - 1)] - 1) * 100)
        valores = t(tail(cbind(val[-1], tasas1), n.años))
        return(valores)
      } else {
        l.pred = n.años * freq + 1
        val = tail(serie, l.pred)[seq(1, l.pred, by = freq), ]
        tasas1 = t(val[-1, ] / val[1:(dim(val)[1] - 1), ] - 1) * 100
        return(list(cantidad = t(val[-1, ]), tasas = tasas1))
      }
    }
  }
  ## Si falta algun parámetro de entrada
  ind1 = length(c(1, tasas))
  ind2 = length(c(1, cant))
  if (ind1 + ind2 == 2) {
    return("Tasas o Cantidades necesarias")
  } else if (ind1 > 1 && ind2 > 1) {
    return("Tasas y Cantidades no a la vez")
  }
}

```

```

    }
    serie1 = serie
    if (ind1 > 1) {
      # Cálculo de cantidades
      decomp = decompose(serie, type = type)
      month = start[2]; year = start[1]
      seas.ind = freq + 2 - month
      baba = decomp$seasonal[seas.ind : (seas.ind + freq - 1)]
      for (i in 2:ind1) {
        tasa = tasas[i - 1]
        if (per.end != freq && i == 2) {
          dic.ind = per.end + 1
          tasa1 = tail(serie1, dic.ind)[1] * (1 + tasa) / tail(serie1, 1) - 1
          delete.ind = -(1 : per.end)
        } else {
          tasa1 = tasa
          delete.ind = -99999
        }
        as = (cumsum(abs(baba)/sum(abs(baba)))) * tasa1[delete.ind]
        serie1 = c(serie1, tail(serie1, 1) * (1 + as))
      }
      serie1 = ts(serie1, start = start, freq = freq)
    } else {
      # Cálculo de tasas
      tasa1 = NULL
      for (j in 2:ind2) {
        if (per.end != freq && j == 2) {
          dic.ind = per.end + 1
          compl.ind = freq - per.end
        } else {
          dic.ind = 1
          compl.ind = freq
        }
        tasa1 = c(tasa1, (cant[j - 1]/tail(serie1, dic.ind)[1] - 1))
        serie1 = c(serie1, seq(tail(serie1, 1), cant[j - 1], length =
                               compl.ind))
      }
      serie1 = ts(serie1, start = start, freq = freq)
      tasas = tasa1
    }
    l.pred = freq * (ind1 + ind2 - 2) + 1
    v.tab = tail(serie1, l.pred)[seq(1, l.pred, by = freq)]
    tab = rbind(c(v.tab), c(NA, tasas * 100))
    colnames(tab) = años
    rownames(tab) = c("cantidades", "tasas")
    l = list(serie.orig = serie, serie.compl = serie1, tasas.100 = tasas * 100,
            tabla = tab)
    return(l)
  }
}

```

- **contrast**: Función que comprueba qué parámetros del modelo ajustado Box-Jenkins no son significativos, y de no existir ninguno, se realizan los test de normalidad (Shapiro Wilk, Jarque Bera), media cero (t -Test) e independencia (Ljung-Box) de los residuos de dicho modelo. Hay la opción de la visión gráfica de los dos primeros contrastes.

Uso

```
contrast (fit, lag.max = 12, elim = TRUE, plotit = TRUE)
```

Argumentos

- fit** objeto de tipo `arima`. Se corresponde con el ajuste Box-Jenkins al cual queremos realizar los respectivos contrastes.
- lag.max** retardo máximo con el que aplicar el test de Ljung-Box de independencia.
- elim** parámetro lógico que indica si por alguna circunstancia interna o externa no se quiere eliminar ningún parámetro del ajuste `fit`. `TRUE` si se quieren eliminar los parámetros no significativos (por defecto), `FALSE` si no se quiere eliminar e ir directamente a los contrastes.
- plotit** parámetro lógico indicando si se quieren las representaciones gráficas de la normalidad y de la media nula de los residuos correspondientes. Por defecto `TRUE`.

Valores

- sig.text** indicaciones de si se ha forzado a la función a no rechazar la significación de los parámetros (primera línea) e indicación de la existencia de parámetros en el ajuste (segunda línea).
- tabla.LB** tabla que contiene los estadísticos y *p*-valores del test de Ljung-Box para la incorrelación hasta el retardo `lag.max`.
- nor.t** tabla que contiene los estadísticos y *p*-valores del *t*-test, test de Shapiro-Wilk y el Jarque-Bera.
- atipA0** existencia de valores atípicos aditivos.
- atipI0** existencia de valores atípicos innovativos.

```
contrast <- function(fit, lag.max = 12, elim = TRUE, plotit = TRUE)
{
  ## No significación
  cond.elim = as.logical(length(fit$coef)) && elim
  if (cond.elim) {
    a = sort(abs(fit$coef[fit$coef != 0]) / (1.96 *
                                             sqrt(diag(fit$var.coef))))[1]

    if (a < 1) {
      ind = which(names(a) == names(fit$coef))
      l1 = list(param = a, pos = ind)
      return(l1)
    } else {
      l1 = list(param = "No eliminás")
    }
  }
  if (!cond.elim || is.character(l1$param)) {
    cond1 = paste("(*) Puede que se forzase a ello: ", !elim)
    cond2 = paste("(**) No hay parámetros: ", !as.logical(length(fit$coef)))
    cond = rbind(cond1, cond2)
    colnames(cond) = ""
    ## Contrastos
    tab = matrix(0, 2, lag.max)
    cn = NULL
    for (i in 1:lag.max) {
      ind.test = Box.test(fit$res, type = "Ljung-Box", lag = i)
      tab[, i] = round(c(ind.test$statistic, ind.test$p.value), 4)
      cn = c(cn, paste("Lag ", i))
    }
    colnames(tab) = cn
  }
}
```

```

rownames(tab) = c("Statistic", "P.value")
t.t = t.test(residuals(fit), mu = 0)
sh.t = shapiro.test(residuals(fit))
jb.t = jarque.bera.test(residuals(fit))
nor.t = round(rbind(c(t.t$stat, t.t$p.value), c(sh.t$stat, sh.t$p.value),
                  c(jb.t$stat, jb.t$p.value)), 5)
colnames(nor.t) = c("Estadístico", "p-valor")
rownames(nor.t) = c("t.test", "Shapiro-Wilk", "Jarque-Bera")
## Datos atípicos
atipI0 = detectI0m(fit)$value
atipA0 = detectA0m(fit)$value
## Representación
if (plotit) {
  windows()
  par(mfrow = c(2, 1))
  plot(residuals(fit), type = "o",
       xlab = paste("Time", "\n", "(p-valor t.test = ", nor.t[1, 2],
                    ")"), main = "Residuos del fit")
  abline(h = 0, lty = 2, col = 2)
  qqnorm(residuals(fit),
        xlab = paste("Theoretical Quantiles", "\n", "(p-valores: SW = ",
                    nor.t[2, 2], ";", " JB = ", nor.t[3, 2], ")"),
        main = paste("Normal Q-Q Plot", "\n", "Gráfico y anáilico"))
  qqline(residuals(fit), lty = 2, col = 2)
}
return(list(sig.text = cond, nor.t = nor.t, tabla.LB = tab,
          atipI0 = atipI0, atipA0 = atipA0))
}
}

```

- **detectA0m**: Función encargada de detectar si hay o no algún *outlier* aditivo (*AO*). Es una modificación de la función **detectA0** del paquete **TSA**, con el objetivo de la no impresión por pantalla de los valores.

Uso

```
detectA0m (object, alpha = 0.05, robust = TRUE)
```

Argumentos

object modelo ARIMA ajustado.

alpha nivel de significación (5% por defecto). Se emplea la regla de Bonferroni para controlar el ratio de error.

robust parámetro lógico que indica si la desviación típica del ruido es estimada mediante residuos absolutos medios (**TRUE**, por defecto), o si es estimada mediante el ajuste *ARIMA* (**FALSE**).

Valores

ind índice temporal del potencial *AO*.

lambda2 correspondiente estadístico de contraste.

```

detectA0m <- function (object, alpha = 0.05, robust = TRUE)
{
  resid = residuals(object)
  piwt = ARMAtoMA(ar = -object$mod$theta, ma = -object$mod$phi,
                 lag.max = length(resid) - 1)

```

```

piwt = c(1, piwt)
omega = filter(c(0 * resid[-1], rev(resid)), filter = piwt, sides = 1,
              method = "convolution")
omega = omega[!is.na(omega)]
rho2 = 1/cumsum(piwt^2)
omega = omega * rho2
if (robust)
  sigma = sqrt(pi/2) * mean(abs(resid), na.rm = TRUE)
else sigma = object$sigma2^0.5
lambda2T = omega/sigma/sqrt(rho2)
lambda2T = rev(lambda2T)
cutoff = qnorm(1 - alpha/2/length(lambda2T))
out = abs(lambda2T) > cutoff
ind = seq(lambda2T)[out]
lambda2 = lambda2T[out]
if (length(ind) != 0)
  value = rbind(ind, lambda2)
else value = "No AO detected"
invisible(list(value))
}

```

- **detectIOm**: Función encargada de detectar si hay o no algún *outlier* innovativo (*IO*). Es una modificación de la función **detectIO** del paquete **TSA**, con el objetivo de la no impresión por pantalla de los valores.

Uso

```
detectIOm (object, alpha = 0.05, robust = TRUE)
```

Argumentos

object modelo ARIMA ajustado.

alpha nivel de significación (5% por defecto). Se emplea la regla de Bonferroni para controlar el ratio de error.

robust parámetro lógico que indica si la desviación típica del ruido es estimada mediante residuos absolutos medios (**TRUE**, por defecto), o si es estimada mediante el ajuste *ARIMA* (**FALSE**).

Valores

ind índice temporal del potencial *IO*.

lambda1 correspondiente estadístico de contraste.

```

detectIOm <- function (object, alpha = 0.05, robust = TRUE)
{
  resid = residuals(object)
  if (robust)
    sigma = sqrt(pi/2) * mean(abs(resid), na.rm = TRUE)
  else sigma = object$sigma2^0.5
  lambda1T = resid/sigma
  cutoff = qnorm(1 - alpha/2/length(lambda1T))
  out = abs(lambda1T) > cutoff
  ind = seq(lambda1T)[out]
  lambda1 = lambda1T[out]
  if (length(ind) > 0)
    value = rbind(ind, lambda1)
  else value = "No IO detected"
}

```

```
invisible(list(value = value))
}
```

- **evalmod**: Función con la cual se realiza la predicción con modelos econométricos en base a la sustitución de previsiones de las variables explicativas de dicho modelo halladas mediante diferentes métodos y para distintos escenarios. Específica para el proyecto desarrollado.

Uso

```
evalmod (status, modelo = " arima", val = , n.ahead = 33, m.trim = 12)
```

Argumentos

- status** elemento que indica el escenario a tomar para las predicciones. Éste puede ser "base", para una evolución positiva de la economía, o "stress", para una nueva recesión de la misma.
- modelo** indica el modelo empleado para la modelización de las covariables de los modelos. Puede ser " arima", por defecto, "var" o "vecm".
- val** predicción que se desea obtener, esto es, puede desearse el valor estimado o los extremos de su intervalo de predicción. Puede ser para obtener la estimación numérica (por defecto), "_inf" para el extremo inferior y "_sup" para el superior.
- n.ahead** número natural que indica el horizonte de previsión de los modelos mensuales. Como es una función específica para el proyecto, si no se tienen nuevos valores para las variables, debe mantenerse la opción por defecto, **n.ahead** = 33.
- m.trim** número natural que indica el horizonte de previsión de los modelos trimestrales. Como es una función específica para el proyecto, si no se tienen nuevos valores para las variables, debe mantenerse la opción por defecto, **m.trim** = 12.

Valores

- val.consumo** predicción de la Formalización del Crédito al Consumo en Hogares en las condiciones indicadas a través de los argumentos de la presente función.
- val.vivienda** predicción de la Formalización del Crédito a la Vivienda en las condiciones indicadas a través de los argumentos de la presente función.
- val.depositos** predicción de los Depósitos Minoristas en las condiciones indicadas a través de los argumentos de la presente función.
- val.dudosos** predicción del Total de Dudosos OSR en las condiciones indicadas a través de los argumentos de la presente función.
- val.creditos** predicción del Crédito a las Empresas en las condiciones indicadas a través de los argumentos de la presente función.
- tasas** Tasas de variación acumulada anual (para las formalizaciones) e interanual (para el resto) hasta el año 2017.

```
evalmod <- function(status, modelo = "arima", val = "", n.ahead = 33,
                    m.trim = 12)
{
  if (status != "base" && status != "stress") {
    return("status equals base or stress")
  }
  if (modelo != "arima" && modelo != "var" && modelo != "vecm") {
    return("modelo equals arima, var or vecm")
  }
  if (modelo != "arima") {
    if (val != "_inf" && val != "" && val != "_sup") {
      return("val equals _inf, empty or _sup")
    }
  }
}
```

```

    }
  }
  ## Variables a emplear
  if (modelo == "arima") {
    preds = read.csv2(paste("Predicciones/Univariante/Series_Pred_",
                           status, val, ".csv", sep = ""))
    preds[, 16:18] = read.table(paste("Covariables/Tipos/Tipos_Pred_", status,
                                     ".txt", sep = ""), header = TRUE,
                              sep = "\t")[-(176:179), -(1:2)]
    preds.trim = preds[seq(1, dim(preds)[1], by = 3), ]
    resp = read.table("Respuestas/respuestas.txt", header = TRUE, sep = "\t")
  } else {
    exo = ts(read.table(paste("Predicciones/Vectorial/exogenas_", status,
                              ".txt", sep = "")), start = c(2000, 3),
             freq = 12)
    men = ts(read.table(paste("Predicciones/Vectorial/new", modelo, "men_",
                              status, val, ".txt", sep = ""),
                              header = TRUE)[, -(1:2)], start = c(2000, 3),
             freq = 12)
    tri = ts(read.table(paste("Predicciones/Vectorial/new", modelo, "tri_",
                              status, val, ".txt", sep = ""),
                              header = TRUE)[, -(1:2)], start = c(2000, 1),
             freq = 4)
    tipos = read.table(paste("Covariables/Tipos/Tipos_Pred_", status, ".txt",
                              sep = ""), sep = "\t", header = TRUE)
    resp = ts(read.table("Respuestas/respuestas.txt", header = TRUE,
                        sep = "\t"), start = c(1999, 12), freq = 12)
  }
  ## Consumo
  ind.resp = which(colnames(resp) == "Formalización.Consumo")
  resp.c = resp[!is.na(resp[, ind.resp]), ind.resp]
  if (modelo == "arima") {
    preds.con = cbind(preds[-(1:4), 5],
                      preds[-((dim(preds)[1]-3):dim(preds)[1]), 7],
                      c(preds[-(1:5), 9], 2.772), preds[-(1:4), 10])
  } else {
    preds.con = cbind(exo[-c(1:4, (dim(exo)[1]-1):dim(exo)[1]), 1],
                      men[-((dim(men)[1]-5):dim(men)[1]), 1],
                      men[-c(1:5, dim(men)[1]), 3],
                      men[-c(1:4, (dim(exo)[1]-1):dim(exo)[1]), 4])
  }
  res.c = read.table("Residuos de modelos/res_consumo.txt")
  coef.mod.c = as.numeric(t(read.table("Coeficientes de modelos/
                                       coef consumo.txt")))

  preds.con.t = preds.con
  preds.con.t[, c(1:2, 4)] = log(preds.con.t[, c(1:2, 4)])
  aa.c = exp(cbind(rep(1), as.matrix(tail(preds.con.t, n.ahead + 2))) %*%
            coef.mod.c + tail(res.c, n.ahead + 2))
  aa.c = ts(c(resp.c, aa.c[, 1]), start = c(2003, 6), freq = 12)
  ## Vivienda
  ind.resp = which(colnames(resp) == "Formalización.Vivienda")
  resp.v = resp[!is.na(resp[, ind.resp]), ind.resp]
  if (modelo == "arima") {
    preds.viv = cbind(preds[-c(1:43, (dim(preds)[1]-1):dim(preds)[1]), 7],
                      preds[-c(1:45), 8], preds[-c(1:43, (dim(preds)[1]-1):

```



```

                                                                 dim(preds)[1]), 18])
} else {
  preds.viv = cbind(men[-c(1:80, (dim(men)[1]-1):dim(men)[1]), 1],
                   men[-c(1:82), 2],tipos[-c(1:39, (dim(tipos)[1]-5):
                                                                 dim(tipos)[1]), 5])
}
res.v = read.table("Residuos de modelos/res_vivienda.txt")
coef.mod.v = as.numeric(t(read.table("Coeficientes de modelos/
                                   coef vivienda.txt")))

coef.mod.v[1] = -93.62
preds.viv.t = preds.viv
preds.viv.t[, 1:2] = log(preds.viv.t[, 1:2])
aa.v = exp(cbind(rep(1), as.matrix(tail(preds.viv.t, n.ahead + 2)))) %%%
        coef.mod.v + tail(res.v, 35))
aa.v = ts(c(resp.v, aa.v[, 1]), start = c(2003, 6), freq = 12)
## Depósitos
ind.resp = which(colnames(resp) == "Depósitos")
resp.d = resp[!is.na(resp[, ind.resp]), ind.resp]
resp.d = resp.d[seq(1, length(resp.d), by = 3)]
if (modelo == "arima") {
  preds.dep = cbind(preds.trim[, 5], preds.trim[, 7],preds.trim[, 17],
                   preds.trim[, 13])
} else {
  preds.dep = cbind(exo[seq(40, dim(exo)[1], 3), 1],
                   men[seq(40, dim(men)[1], 3), 1],
                   tipos[seq(1, dim(tipos)[1]-4, 3), 4],
                   tri[-c(1:13, (dim(tri)[1]-1):dim(tri)[1]), 4])
}
res.d = read.table("Residuos de modelos/res_depositos.txt")
coef.mod.d = as.numeric(t(read.table("Coeficientes de modelos/
                                   coef depositos.txt")))

preds.dep.t = cbind(
  log(preds.dep[, 1]),
  preds.dep[, 4]^(-.5),
  log(preds.dep[, 2])*preds.dep[, 4]^(-.5),
  preds.dep[, 3]*preds.dep[, 4]^(-.5),
  log(preds.dep[, 2])*preds.dep[, 3]*preds.dep[, 4]^(-.5)
)
aa.d = exp(cbind(rep(1), as.matrix(tail(preds.dep.t, m.trim)))) %%%
        coef.mod.d + tail(res.d, m.trim))
aa.d = ts(c(resp.d, aa.d[, 1]), start = c(2002, 6), freq = 4)
## Dudosos
ind.resp = which(colnames(resp) == "Dudosos")
resp.du = resp[!is.na(resp[, ind.resp]), ind.resp]
if (modelo == "arima") {
  preds.dud = cbind(c(preds.trim[-1, 5], 3700),
                   c(preds.trim[-1, 6], 11830),
                   c(preds.trim[-1, 17], 0.74),
                   c(preds.trim[-(1:2), 18], 0.95, 1)      )
} else {
  preds.dud = cbind(c(exo[seq(40, dim(exo)[1], 3), 1]),
                   c(men[seq(40, dim(men)[1], 3), 1]),
                   c(tipos[seq(1, dim(tipos)[1]-6, 3), 4], 0.5),
                   c(tipos[seq(4, dim(tipos)[1]-2, 3), 5], 0.8))
}

```

```

res.du = read.table("Residuos de modelos/res_dudosos.txt")
coef.mod.du = as.numeric(t(read.table("Coeficientes de modelos/
                                coef_dudosos.txt")))

preds.dud.t = cbind(
  poly(log(preds.dud[, 1]), degree = 2, raw = TRUE),
  poly(log(preds.dud[, 2]), degree = 2, raw = TRUE),
  poly(preds.dud[, 4], degree = 2, raw = TRUE),
  preds.dud[, 3]*preds.dud[, 4],
  preds.dud[, 3]*preds.dud[, 4]^2
)
aa.du = exp(cbind(rep(1), as.matrix(tail(preds.dud.t, m.trim))) %*%
            coef.mod.du + tail(res.du, m.trim))
aa.du = ts(c(resp.du, aa.du[, 1]), start = c(1999, 4), freq = 4)
aa.du[73] = aa.du[72] - 300000
## Crédito empresas
ind.resp = which(colnames(resp) == "Crédito.Empresas")
resp.cr = resp[!is.na(resp[, ind.resp]), ind.resp]
if (modelo == "arima") {
  preds.emp = cbind(preds.trim[-((dim(preds.trim)[1]-1):
                                dim(preds.trim)[1]), 12],
                    preds.trim[-((dim(preds.trim)[1]-1):
                                dim(preds.trim)[1]), 14],
                    preds.trim[-(1:2), 15])
} else {
  preds.emp = cbind(tri[-((dim(tri)[1]-3):dim(tri)[1]), 1],
                    tri[-((dim(tri)[1]-3):dim(tri)[1]), 3],
                    tri[-c(1:2, (dim(tri)[1]-1):dim(tri)[1]), 4])
}
res.e = read.table("Residuos de modelos/res_credito.txt")
coef.mod.e = as.numeric(t(read.table("Coeficientes de modelos/
                                coef_credito.txt")))

coef.mod.e[1] = 4.83
preds.emp.t = log(preds.emp)
aa.e = exp(cbind(rep(1), as.matrix(tail(preds.emp.t, m.trim))) %*%
            coef.mod.e + tail(res.e, m.trim))
aa.e = ts(c(resp.cr, aa.e[, 1]), start = c(1999, 4), freq = 4)
## Recopilación de tasas
a1 = calc.tx.ct(aa.c, per.end = 6, start = c(2003, 6), cum = TRUE)
a2 = calc.tx.ct(aa.v, per.end = 6, start = c(2003, 6), cum = TRUE)
a3 = calc.tx.ct(aa.d, per.end = 2, start = c(2003, 2), freq = 4)
a4 = calc.tx.ct(aa.du, per.end = 4, start = c(1999, 4), freq = 4)
a5 = calc.tx.ct(aa.e, per.end = 4, start = c(1999, 4), freq = 4)
aa = rbind(a1, a2, a3, a4, a5)[c(2, 4, 6, 8, 10), ]
rownames(aa) = colnames(resp)[-(1:2)]
colnames(aa) = c("2015", "2016", "2017")
return(list(val.consumo = aa.c, val.vivienda = aa.v, val.depositos = aa.d,
            val.dudosos = aa.du, val.creditos = aa.e, tasas = aa))
}

```

- **mod.sig**: Función que sustrae de un cierto modelo, podría ser un *LM* o un *GLS*, las variables explicativas no significativas a un nivel de significación α que será dado como entrada.

Uso

`modelo.sig (v.resp, v.expl, lineal = FALSE, p = 0, q = 0, alpha = 0.05)`

Argumentos

- v.resp** variable respuesta que queremos introducir en nuestro modelo.
- v.expl** variables explicativas que queremos emplear en el modelo.
- lineal** parámetro lógico que indica si queremos realizar un *LM* (TRUE) o un *GLS* (FALSE, por defecto).
- p** parámetro que indica el orden de la parte autorregresiva que explica los residuos y que formará parte de la correlación que se le pueda indicar al modelo *GLS*. Por defecto, $p = 0$.
- q** parámetro que indica el orden de la parte de medias móviles que explica los residuos y que formará parte de la correlación que se le pueda indicar al modelo *GLS*. Por defecto, $q = 0$.
- alpha** nivel de significación que asegura la significación de las variables en nuestro modelo. Por defecto, $\alpha = 0.05$.

Valores

modelo modelo con todos los coeficientes significativamente diferentes de 0 a un nivel **alpha**.

```

modelo.sig <- function(v.resp, v.expl, lineal = FALSE, p = 0,
                      q = 0, alpha = 0.05)
{
  ## Definición del modelo
  v.expl = as.matrix(v.expl)
  if (lineal) {
    modelo = lm(v.resp ~ v.expl)
    p.valores = as.numeric(summary(modelo)$coef[, 4])[-1]
  } else {
    modelo = gls(v.resp ~ v.expl, correlation = corARMA(p = p, q = q),
                method = "REML")
    p.valores = as.numeric(summary(modelo)$tTable[, 4])[-1]
  }
  ## Eliminación de las no significativas mientras sus p-valores < alpha
  while (sum(p.valores > alpha) > 0) {
    iii = as.numeric(which.max(p.valores))
    v.expl <- v.expl[, -iii]
    if (lineal) {
      modelo = lm(v.resp ~ v.expl)
      p.valores = as.numeric(summary(modelo)$coef[, 4])[-1]
    } else {
      modelo = gls(v.resp ~ v.expl, correlation = corARMA(p = p, q = q),
                  method = "REML")
      p.valores = as.numeric(summary(modelo)$tTable[, 4])[-1]
    }
    modelo = modelo
  }
  return(modelo)
}

```

- **sens**: Función encargada de realizar el análisis de la sensibilidad del modelo que queramos, modificando una sola variable una cierta tasa, si es posible.

Uso

```

sens (v.expl, v.resp, ind.var, tasas = 0, f1 = log, f2 = exp, start
     = c(2003, 6), per.end = 1, freq = 12, acum = TRUE, auto.ef =
     FALSE, x.elim = mod.sigcoef[ind.var + 1])

```

Argumentos

- `v.expl` columnas correspondientes a las variables explicativas consideradas en el modelo tratado.
- `v.resp` variable respuesta de dicho modelo.
- `ind.var` índice que indica que columna de `v.expl` está asociada a la variable a la cual se realiza la variación.
- `tasas` tasas de incremento para el año siguiente de la explicativa. Por defecto, `tasas = 0`.
 - `f1` transformación a realizar a la explicativa. Por defecto, `f1 = log`.
 - `f2` transformación inversa a realizar a la explicativa. Por defecto, `f2 = exp`.
- `start` vector de dos componentes indicando el año y el mes de comienzo de las series de estudio. Por defecto, `start = c(2003, 6)`.
- `per.end` periodo de fin de las series de estudio. Por defecto, `per.end = 1`. Si `freq = 4`, la numeración de los periodos va desde 1 a 4.
- `freq` frecuencia de las series a estudio. Por defecto, `freq = 12`.
- `acum` indicador lógico para saber si las tasas a calcular son acumuladas (`TRUE`) o no (`FALSE`). Por defecto, `acum = TRUE`.
- `auto.ef` valor lógico para indicar que en los efectos que son eliminados del modelo para el análisis de sensibilidad no se tiene la propia variable explicativa (`FALSE`) o si (`TRUE`). Por defecto, `auto.ef = TRUE`.
- `x.elim` valor para eliminar al último dato observado de la explicativa para la realización del análisis de sensibilidad.

Valores

- `tx.resp` correspondientes tasas de variación de la respuesta como resultado de la variación de la explicativa con el argumento `tasas`.

```
sens <- function(v.expl = v.expl2, v.resp = v.resp2, ind.var = 1, tasas = 0,
  f1 = log, f2 = exp, start = c(2003, 6), per.end = 1, freq = 12,
  acum = TRUE, auto.ef = FALSE, x.elim = mod.sig$coef[ind.var + 1])
{
  ## explicativa original a estudio
  x = f2(v.expl2[, ind.var]) ## ind
  aa = NULL
  for (i in tasas) {
    ## aplicación de tasas
    v = calc.tx.ct(x, per.end = per.end, start = start, freq = freq,
      tasas = i, do.pred = TRUE, años = 2014:2015)
    if (per.end == freq) per.end = 0
    ## último valor de la respuesta transformada
    y1 = tail(v.resp, 1)
    ## eliminación del efecto de la explicativa transformada
    ## en la respuesta transformada
    yx1 = y1 - tail(f1(x), 1) * x.elim
    ## transformación de nuevos valores
    sel = freq - per.end
    s1 = f1(tail(v$serie.compl, sel))
    ## efecto de la transformación de nuevos valores
    if (auto.ef) {
      x.elim1 = mod.sig$coef[ind.var + 2] + mod.sig$coef[ind.var + 3] * s1
      s11 = s1 * x.elim1
    } else {
      s11 = s1 * x.elim
    }
  }
}
```

```

## suma entre la respuesta sin efecto x1 y los efectos s11
y2 = yx1 + s11
## nuevos valores de la respuesta
yn = exp(c(tail(v.resp, per.end), y2))
## valores antiguos de la respuesta
ya = exp(tail(v.resp, freq + per.end)[1:freq])
## tasa de variación
if (acum) {
  aa1 = sum(yn)/sum(ya) - 1
} else {
  aa1 = tail(yn, 1)/tail(ya, 1) - 1
}
aa = c(aa, aa1)
}
tx.resp = round(aa, 5) * 100
return(tx.resp)
}

```

- **var.retard4**: Función para la obtención de nuevas variables explicativas y respuesta a partir de las originales que se han introducido a las cuales se les aplican ciertos retardos temporales para posteriormente trabajar con modelos dinámicos. Es una generalización de la función `zlag` del paquete `TSA`.

Uso

```
var.retard4 (v.resp, v.expl, ret, ay, freq = 12, ind.año = NULL,
            n.lin.cod = 0.5)
```

Argumentos

- v.resp** variable respuesta del modelo que queremos ajustar.
- v.expl** variables explicativas del modelo, deben estar dispuestas en columnas. Se puede incluir delante de ellas dos columnas indicando el año y el mes de cada valor, en este caso no modificamos `ind.año`. Si éstas no son añadidas, indicamos `ind.año`.
- ret** retardos que se quieren aplicar a cada una de las variables explicativas. Éste será un vector en el cual, la incorrelación se codifica con el parámetro `n.lin.cod`.
- ay** vector 2-dimensional con el primer elemento el año de comienzo del estudio de la variable respuesta, y segundo elemento el correspondiente mes.
- freq** frecuencia anual que se toma en el estudio. Por defecto, `freq = 12`.
- ind.año** índice en el que existe la coincidencia entre `ay` y nuestras covariables. Por defecto `ind.año = NULL`, esto indica que se añaden las dos columnas en `v.expl` que se han comentado.
- n.lin.cod** valor que codifica la incorrelación entre variables. Por defecto, `n.lin.cod = 0.5`.

Valores

- v.resp** nueva variable respuesta para nuestro modelo, con la aplicación de los retardos a las covariables.
- v.expl** nuevas variables explicativas en nuestro modelo, con la aplicación de sus correspondientes retardos.
- n.com** vector 2-dimensional con elementos, el año de comienzo del estudio con las variables modificadas y el mes asociado.

```

var.retard4 <- function(v.resp, v.expl, ret, ay, freq = 12, ind.año = NULL,
                       n.lin.cod = 0.5)
{
  n = length(v.resp)
  p = dim(v.expl)[2] - 2
  ## Variables no declaradas independientes
  p1 = (1:p)[ret != n.lin.cod] + 2
  ## Índice de comienzo de la respuesta con respecto a las explicativas
  ind.año = c(ind.año, which(v.expl[,1] == ay[1] & v.expl[,2] == ay[2]))
  ## Bucle para mover las variables según los retardos
  v.expl1 = ind2.n.mes = nom.var = NULL
  for (i in p1) {
    indx = ind.año - ret[i-2]
    x = as.vector(v.expl[, i])
    if (ret[i-2] == 0) {
      v.ex.add = x[ind.año:(n + ind.año - 1)]
    } else if (ret[i-2] < 0) {
      v.ex.add = x[indx:(n + indx - 1)]
    } else {
      r.na = sum((indx:(n + indx - 1)) < 1)
      x = c(rep(NA, r.na), x)
      v.ex.add = x[1:n]
    }
    v.expl1 = cbind(v.expl1, as.numeric(v.ex.add))
    ind1.n.mes = which(!is.na(v.expl1[, which(p1 == i)]))
    ind2.n.mes = c(ind2.n.mes, ind1.n.mes[1])
    nom.var = c(nom.var, paste(c("v", i - 2), collapse = ""))
  }
  ## Si existen NA
  ind.na = rowSums(is.na(v.expl1))
  if (sum(ind.na) > 0) {
    v.expl1 = v.expl1[-which(ind.na > 0), ]
    v.resp = v.resp[-which(ind.na > 0)]
  }
  n.mes = ay[2] + max(ind2.n.mes) - 1
  colnames(v.expl1) = nom.var
  v.resp = ts(v.resp, start = c(ay[1], n.mes), freq = freq)
  l = list(v.resp = v.resp, v.expl = v.expl1, n.com = c(ay[1], n.mes))
  return(l)
}

```

Apéndice C

Implementación con IBM SPSS Modeler v16

El programa IBM SPSS Modeler es un *software* comercial empleado para el análisis predictivo y diseñado para aportar mejoras predictivas a decisiones desarrolladas por grupos de personas, sistemas o empresas. Éste proporciona un buen número de algoritmos y técnicas avanzadas, como podrían ser el análisis de texto, el análisis de entidad o de gestión y optimización de decisiones, para ayudar a un mejor resultado en la selecciones de ciertas acciones.

Para completar el Modeler y sus habilidades de minería de datos, los programadores de *IBM* han introducido al entorno de trabajo de este *software* nodos que permiten a los usuarios expertos de R implementar sus propios *scripts* de R en dicho programa para el procesamiento de datos, construcción de modelos y puntuación de los mismos. Éstos se encuentran a partir de la versión v16, donde recurrimos a [IBM Corporation 2013] para la ampliación tanto de la teoría que se programa como de la misma ejecución de lo programado.

En esta sección no hablaremos del entorno específico de todo este programa, como implementar todas las opciones que están a disposición del usuario, sino que citaremos los nodos empleados en nuestro trabajo y los incorporados a la nueva versión que permiten introducir programación en lenguaje de R, así como concretar con un sencillo ejemplo cómo se han empleado éstos. Decir que en la parte práctica de este trabajo, ésta posibilidad ha sido de gran utilidad debido a que se ha podido unir la modelización flexible que con *scripts* de R se puede alcanzar, en nuestro caso de modelos *GLS* de regresión y *ARIMA*, *VAR* y *VECM* para series temporales, con la potencia que el IBM SPSS Modeler aporta.

Los nodos básicos del Modeler que se han implementado son los que a continuación se citan y comentan. Para mayor información se puede recurrir al manual de la aplicación del programa que antese se ha citado [IBM Corporation 2013].

- **Lectura:** Introducción de los datos que se deseen.
- **Tipo:** Declarar las variables empleadas para saber con qué valores se está trabajando.
- **Filtro:** Seleccionar las variables que se deseen de entre todas las introducidas en el nodo.
- **Intervalos de tiempo:** Conversión de una variable en serie de tiempo, se introduce la frecuencia, el inicio y el horizonte al que se desea realizar previsiones de la misma.
- **Fundir:** Permite fusionar registros de diversos conjuntos de datos, de distintos orígenes, utilizando una ordenación, un campo clave o una condición sobre los registros.

- **Modelo temporal:** Generación de un modelo temporal, puede ser el que el programa estime por su cuenta o puede ser introducido a mano, como unos *ARIMA* para nuestro caso particular.
- **Gráfico temporal:** Representación de la serie de tiempo que se le indique.

Y los nodos referentes a R que se han implementado en el *software* comercial citado son los siguientes. De nuevo, para más información se puede consultar [IBM Corporation 2013].

- **R Transform:** éste permite tomar datos del *Modeler* y modificarlos empleando nuestros propios *scripts* de R personalizados. A continuación los datos modificados son devueltos al entorno de trabajo del *Modeler* y se sigue con la ruta programada.
- **R Building:** con éste se pueden construir modelos personalizados en R para luego puntuarlos en el *Modeler*, ya hemos dicho que es una de sus funciones.
- **R model:** éste define otro nodo contenedor del modelo generado para poder ser usado cuando sea añadido al entorno del *Modeler*.
- **R Output:** éste permite analizar datos y resultados procedentes de una puntuación de modelos empleando nuestros propios *scripts* de R. También se pueden analizar textos y gráficos.

Una vez que se han introducido los nodos más importantes que en este trabajo se han empleado cuando se ha manejado este *software*, mostramos una pequeña parte de la ruta que se ha programado. Ésto será a título informativo para dar una idea de cómo es el entorno de trabajo en dicho programa.

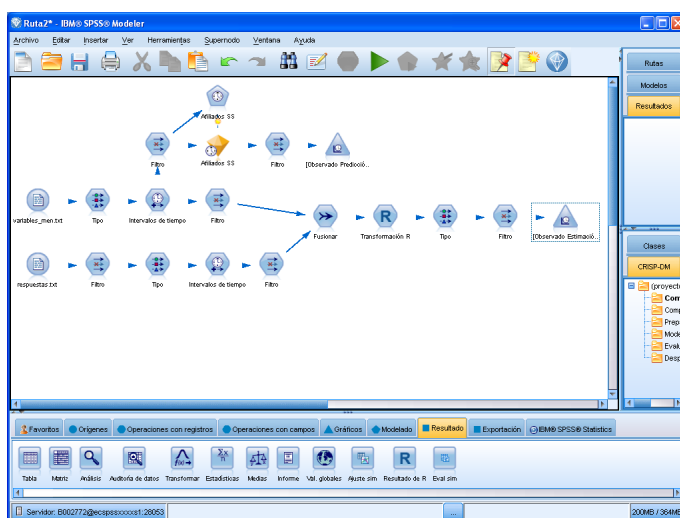


Figura C.1: Parte de la ruta programa en el entorno del programa IBM SPSS Modeler. Se realiza la modelización *GLS* de la Formalización del Consumo, y la modelización *ARIMA* para los Afiliados a la Seguridad Social.

En ella lo que se realizará será la modelización con *GLS* de la Formalización del Crédito al Consumo en función de las variables que resultaron significativas para ella en el Capítulo 5. Esta estimación se comparará con lo observado para ver la calidad del ajuste. Por otra parte se modelizará con modelos Box-Jenkins el Total de Afiliados a la Seguridad Social, como ya se mostró en la Sección 5.2.1. Este proceso se muestra en la Figura C.1, lo explicamos a continuación.

En un primer momento leemos el archivo `respuestas.txt` y definimos sus valores con el nodo *Tipo*, luego las convertimos en series de tiempo a través de *Intervalos de tiempo* y seleccionamos la Formalización del Consumo con *Filtro*. Por otro lado, leemos el archivo `variables_men.txt`, definimos sus valores con *Tipo*, las convertimos en series de tiempo (*Intervalos de tiempo*) y seleccionamos las significativas en relación a la Formalización del Consumo con *Filtro*.

Ambos grupos de variables son juntadas en función de sus años y meses con el nodo **Fundir**, y con el nodo **Transformación R** se define el modelo *GLS*, del cual se comprueban sus valores (Tipo) seleccionando su estimación y lo observado (**Filtro**) para representarlos con [**Observado Estimación**]. Esta gráfica puede visualizarse en la Figura C.2, la cual es semejante a la dada en la Figura 5.2, como podía esperarse.

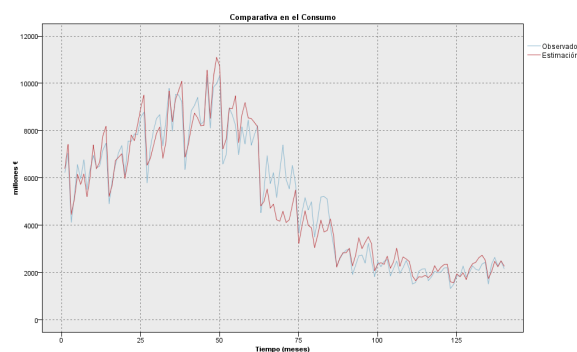


Figura C.2: Comparativa en el entorno IBM SPSS Modeler entre lo estimado en rojo y lo observado en azul de la Formulación del Consumo en base al modelo *GLS* creado.

Para acabar, a partir de la definición de series temporales de la covariables, se selecciona únicamente la correspondiente a los Afiliados a la Seguridad Social con un nodo **Filtro** y se modeliza mediante el nodo modelo **Afiliados SS**. Del cual representamos su estimación con intervalos de predicción y lo observado [**Observado Predicción**] en la Figura C.3.

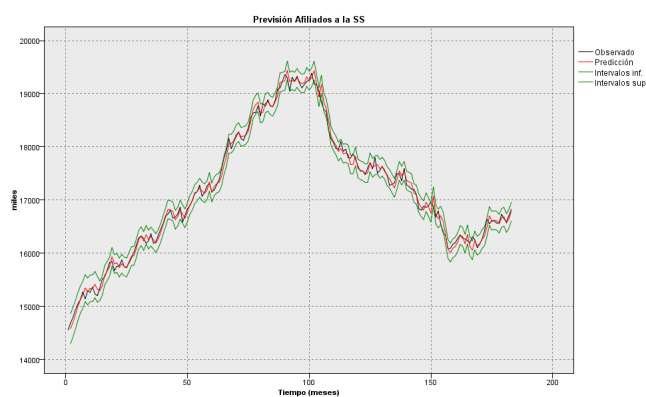


Figura C.3: Comparativa en el entorno IBM SPSS Modeler entre lo estimado en rojo con sus intervalos de predicción en verde y lo observado en negro del Total de Afiliados a la Seguridad Social en base a un modelo *ARIMA*.

Bibliografía

- [Bhattacharya, R. *et al.* 2009] Bhattacharya R., Waymire E. (2009) *Stochastic Processes with Applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
- [Banco Central Europeo] Banco Central Europeo (BCE). <https://www.ecb.europa.eu/ecb/html/index.es.html>. Accedido 1 de mayo de 2015.
- [Box, G. *et al.* 2008] Box, G., Jenkins, G., Reinsel, G. (2008) *Time Series Analysis. Forecasting and Control*. John Wiley & Sons, Hoboken, New Jersey.
- [Buse, A. 1973] Buse, A. (1973) Goodness of Fit in Generalized Least Squares Estimation. *The American Statistician* **27**, 3:106 – 108.
- [Chan, K-S. *et al.* 2012] Chan, K-S. y Ripley, B. (2012) *TSA: Time Series Analysis*. R package version 1.0.1. <http://CRAN.R-project.org/package=TSA>. Accedido 1 de mayo de 2015.
- [Cochrane, D. *et al.* 1949] Cochrane, D. y Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, **44**(245):32 – 61.
- [Cowpertwait, P. *et al.* 2009] Cowpertwait, P., Metcalfe, A. (2009) *Introductory Time Series with R*. Springer, Iowa City, Iowa.
- [Cryer, J. *et al.* 2008] Cryer, J., Chan, K-S. (2008) *Time Series Analysis. With Applications in R*. Springer.
- [Granger, C. 1969] Granger, C. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37**:424 – 438.
- [Hamilton, J. D. 1994] Hamilton, J. D. (1994) *Time Series Analysis*. Princeton University Press, Princeton.
- [Directiva Europea 2014] Directiva 2014/59/UE del Parlamento y Consejo Europeo de 15 de Mayo de 2014. <http://www.boe.es/doue/2014/173/L00190-00348.pdf>
- [Fox, J. *et al.* 2010] Fox, J., Weisberg, S. (2010) Time-Series Regression and Generalized Least Squares in R. *An R Companion to Applied Regression*. Segunda edición, SAGE.
- [IBM Corporation 2013] IBM Corporation: Software Group (2013) *IBM SPSS Modeler 16: Algorithms Guide & Applications Guide*, Chicago, Estados Unidos. <http://www-01.ibm.com/support/docview.wss?uid=swg27038316#en> Accedido 1 de mayo de 2015.
- [Instituto Galego de Estadística] Instituto Galego de Estadística (IGE). <https://www.ige.eu/web/index.jsp?paxina=001&idioma=gl> Accedido 1 de mayo de 2015.
- [Instituto Nacional de Estadística] Instituto Nacional de Estadística (INE). <http://www.ine.es/> Accedido 1 de mayo de 2015.
- [Johansen, S. 1995] Johansen, S. (1995) Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, *Advanced Texts in Econometrics*. Oxford University Press, Oxford.

- [Kariya, T. *et al.* 2004] Kariya, T., Kurata, H. (2004) *Generalized Least Squares*. John Wiley & Son, Hoboken, New Jersey, pp 25 – 66.
- [Kuan, C. 2004] Kuan, C. (2004) *Statistics: Concepts and Methods*. Taipei: Hua-Tai Publisher, China, pp 77 – 110.
- [Lütkepohl, H. *et al.* 2004] Lütkepohl, H., Krätzig, M. (2004) *Applied Time Series Econometrics*. Cambridge University Press, Cambridge.
- [Mahalanobis P. 1936] Mahalanobis P. (1936) On the generalised distance in statistics, *Proceedings of the National Institute of Science of India* 12. India, pp 49 – 55.
- [Ministerio de Economía y Competitividad] Ministerio de Economía y Competitividad. Dirección General de Análisis Macroeconómico y Economía Internacional. Indicadores e Informes Macroeconómicos. <http://serviciosweb.meh.es/APPS/DGPE/BDSICE/HomeBDSICE.aspx>
- [Peña, D. 2005] Peña, D. (2005) *Análisis de Series Temporales*. Alianza Editorial, Madrid.
- [Pfaff, B. 2008a] Pfaff, B. (2008) *Analysis of Integrated Series with R and Cointegrated Time*. Springer, Segunda edición, Nueva York.
- [Pfaff, B. 2008b] Pfaff, B. (2008) VAR, SVAR and SVEC Models: Implementation within R package vars. *Journal of Statistical Software*, **27** 4. <http://www.jstatsoft.org/v27/i04/>
- [Pinheiro J. 2014] Pinheiro J., Bates, D., DebRoy, S., Sarkar, D. y R Development Core Team (2014) *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-117. <http://CRAN.R-project.org/package=nlme>. Accedido 1 de mayo de 2015.
- [Ploberger, W. 1992] Ploberger, W., Krämer, W. (1992) The Cusum Test with Ols Residuals, *Econometrica*. The Econometric Society, Marzo, **60** 2: 271 – 285.
- [R Development Core Team 2008] R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>. Accedido 1 de mayo de 2015.
- [Shumway, R. *et al.* 2011] Shumway, R. y Stoffer, D. (2011) *Time Series Analysis and Its Applications with R examples*. Springer, New York.
- [Trapletti, A. *et al.* 2013] Trapletti, A. y Hornik, K. (2013) *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-32. <http://CRAN.R-project.org/package=tseries>. Accedido 1 de mayo de 2015.