

Universidad de Huelva

Departamento de Economía



Desarrollo de aplicaciones de sistemas dinámicos y econométricos a problemas económicos

Memoria para optar al grado de doctor
presentada por:

Juan Luis Martín Suárez

Fecha de lectura: 25 de septiembre de 2017

Bajo la dirección de los doctores:

Emilio Congregado Ramírez de Aguilera

Antonio Aníbal Golpe Moya

Huelva, 2017



**Desarrollo de aplicaciones de sistemas
dinámicos y econométricos a problemas
económicos**

Juan Luis Martín Suárez

Universidad de Huelva Departamento de Economía

Directores:

Emilio Congregado

Antonio Aníbal Golpe Moya

Mayo, 2017

A mis padres

Índice

PARTE I. INTRODUCCIÓN

1	Introducción	13
1.1	Introducción	1
1.2	Datos y medidas	16
1.3	Metodología	16
1.4	Estructura capitular	18
1.5	Publicaciones	21

PARTE II. DETECCIÓN DE REDES COMPLEJAS EN SERIES TEMPORALES ECONÓMICAS

2	Una exploración de las series de consumo energético americanas a través del uso de grafos de visibilidad	25
2.1	Introducción	25
2.2	Redes complejas y series temporales	27
2.3	Metodología	29
2.4	Propiedades de las redes	24
2.5	Referencias	48
3	El uso del análisis de redes como forma de caracterizar el ciclo económico. Una aplicación del análisis del ciclo norteamericano	57
3.1	Introducción	58
3.2	Sobre el algoritmo de visibilidad	60
3.3	Metodología	63
3.4	Propiedades topológicas	67
3.5	Referencias	77

PARTE III. UN ENSAYO DE MACROECONOMETRÍA APLICADA

4	The U Shape revisited : testing for structural breaks in the relationship between entrepreneurship and income	83
4.1	Introduction	84
4.2	The relation between Firm size and economic development	86
4.3	Model and data	90
4.4	Empirical framework	94
4.5	Results	98
4.6	Conclusions and policy implications	99
	References	101

PARTE IV. APLICACIONES A LA DOCENCIA DE MACROECONOMÍA DINÁMICA

5	Modelización de problemas económicos a través de sistemas de ecuaciones diferenciales resueltos a través de métodos numéricos	123
5.1	Introducción	123
5.2	El planteamiento del problema	123
5.3	Resolución	127
5.4	Implementación y programación	129
5.5	Análisis	131
6	Un simulador de un modelo dinámico para un curso de macroeconomía intermedia	137
6.1	Introducción	138
6.2	Objetivos	138
6.3	El modelo y su solución gráfica	139
6.4	El simulador	143
6.5	Referencias	144

PARTE V. CONCLUSIONES

7	Conclusiones y agenda de investigación futura	149
---	---	-----

Prólogo

Esta tesis doctoral es un compendio de ensayos de economía cuantitativa, y de macroeconometría aplicada, que se encuadra en las tendencias y en la corriente principal por el que discurre el análisis económico en el presente.

Así, el núcleo centra de esta tesis tiene que ver con el análisis aplicado de series temporales económicas en los que éstas son convertidas unívocamente, y a través de una familia de algoritmos, en un grafo o red, en los que además de generar una aplicación para la aplicación de este problema por parte de usuarios no expertos, se contribuye con una digresión acerca de la falta de sensibilidad de la utilización del índice de Hurst –para el análisis de la persistencia en series temporales– en el caso de muestras grandes. Este análisis se lleva a cabo en dos trabajos aplicados que sirven como forma de exponer la utilidad y limitaciones de este enfoque. Por otra parte, esta tesis incluye tres trabajos más. Dos que se encuadran en el campo de la macroeconometría aplicada en el ámbito de la economía laboral y un último trabajo en el campo de la investigación aplicada a la docencia de Macroeconomía, gracias al desarrollo e implementación de una herramienta para el análisis de modelos económicos dinámicos.

Por tanto, en cierta medida, este trabajo es la suma de diferentes proyectos, llevados a cabo con diferentes coautores, diferentes motivaciones y objetivos, así como con diferentes enfoques metodológicos. Tres de los ensa-

yos son el resultado de un trabajo planificado y ejecutado a lo largo de los últimos tres años con los profesores Emilio Congregado, Antonio A. Golpe and María Luis Vílchez, en los que los grafos de visibilidad son aplicados a series temporales, mientras que el otro es fruto de un trabajo a caballo entre la docencia y la investigación en el que el uso de técnicas avanzadas es puesta al servicio de la enseñanza de la macroeconomía dinámica en contextos en los que el aprendizaje de esta última se ve comprometido por el *background* matemático de los estudiantes.

Durante el proceso de elaboración de esta tesis han sido muchos los que han ejercido una influencia sobre mi trabajo.

Me siento en deuda con mi supervisor, Emilio Congregado por su apoyo e indicaciones. Muchas gracias por haberme proporcionado la oportunidad de abordar este proyecto, por haberme incorporado a su equipo, por su ayuda y por su tiempo.

También estoy en deuda con Antonio Golpe y María Luisa Vílchez quienes me enseñaron la aplicación de técnicas y por sus valiosas sugerencias sobre versiones preliminares de este documento.

Mi trabajo también se ha visto favorecido por los consejos y la ayuda de un buen número de investigadores y profesores esperando no olvidarme de ninguno de ellos. Así, me gustaría agradecer a las doctoras Feria y Carmona y a los doctor Iglesias, y a algunos *referees* por sus valiosos comentarios y sugerencias sobre versiones preliminares de este estudio.

También quisiera agradecer a los compañeros del Departamento de Economía de la Universidad de Huelva y a los profesores del Máster en Eco-

nomía, Finanzas y Computación de la Universidad de Huelva por su apoyo y por el clima de trabajo.

Finalmente, gracias a mi familia y a mis amigos por el apoyo recibido durante todo este proceso, y en especial, a mis padres.

Juan Luis Martín
Huelva, Mayo 2017

Parte I: Introducción

Capítulo 1: Introducción

1.1. Introducción

Dejando a un lado economías que basan su riqueza en la tenencia de un recurso natural –normalmente energético– o en su condición de paraíso fiscal, las economías que lideran los rankings de desarrollo económico comparten esta condición con la de ser líderes tecnológicos, pilar central en el que basar, hoy día, cualquier estrategia de desarrollo ya sea nacional, regional o local. Desde esta perspectiva, las sociedades han de aspirar a convertirse en economías basadas en el conocimiento –a *knowledge-driven economy*– lo que ha de traducirse en estar en disposición de liderar la innovación y transferencia en los sectores tecnológicos.¹

¹ En este punto advertir que liderar estos sectores no sólo es la mejor forma de contribuir al crecimiento económico, al ser los sectores que reportan no sólo mayor valor añadido, y los que mejor contribuyen a la generación de empleo, y a la competitividad empresarial, sino que también son los que mejor resisten las fluctuaciones económicas disminuyendo la volatilidad característica de las economías que se sitúan en los sectores marginales.

En este contexto, convendremos que son las llamadas áreas de lo que se viene a denominar STEM –*Science, Technology, Engineering and Maths*– las que están liderando la innovación tecnológica. En este sector, hay un elemento transversal, y que se está analizando de forma muy especial por empresas, investigadores e instituciones, dado su carácter transversal y su carácter esencial para que otros sectores puedan seguir siendo competitivos.

Nos referimos al diseño de métodos de explotación y captura de las cantidades ingentes de información, que los desarrollos de las tecnologías de la información y de la computación permiten capturar y tratar. Un nicho de mercado especialmente importante, ya que los canales de gestión y distribución comercial estarán, en buena medida, condicionados por el lugar ocupado en él. En la era del Big Data –análisis masivo de datos– el disponer de analistas que puedan prestar este tipo de servicios asociados a la economía digital es una preocupación/necesidad de primer orden.

En este contexto la Economía, está sujeta a Fuertes convulsions. Por un lado, encontramos la necesidad de hibridar con otras disciplinas como la computación, la informática o la matemática teórica y aplicada.

Se trata de importar técnicas y métodos, aunque sin abandonar la modelización y sin caer en el riesgo del medir sin teoría o en el de torturar a los datos hasta que confiesen.

La evolución de la disciplina, se encuentra así sometida a un auténtico *trade-off*, en los que los investigadores han de salir de su zona de confort, explorando este instrumental, mientras que los formadores y académicos han de dar un giro a los estudios de las diferentes disciplinas económicas para dar respuesta a las nuevas demandas de analistas económicos cuyas competencias distan de las tradicionalmente asociadas a este tipo de estudios.

Sobre estas premisas, los trabajos que componen esta tesis tratan de realizar aportaciones que dan respuesta a estas cuestiones, tanto en el terreno de la investigación como en el de la enseñanza, adaptando las nuevas necesidades a la realidad del contexto educativo.

La contribuciones de esta tesis son cuatro.

Primero, la aplicación de la teoría de grafos al análisis de series temporales y el desarrollo de aplicaciones informáticas para su implementación por parte de usuarios no expertos.

Segundo, se proporciona una variante del índice de Hurst para superar los problemas que presenta éste en el caso de grandes muestras.

Tercero, se incluye un trabajo de macroeconometría aplicada a uno de los tópicos de economía laboral que mayor interés despiertan en la actualidad, dados los cambios que se están produciendo en las economías-offshoring, out-sourcing y desarrollo del sector GIG, con efectos directos sobre las formas de organización del trabajo y sobre el desarrollo del sector

del autoempleo, haciendo uso de técnicas recientes de cambio estructural y una muestra de países de la OCDE.

Cuarto, esta tesis proporciona dos desarrollos metodológicos para la enseñanza de la macroeconomía dinámica en contextos de iniciación.

1.2. Datos y medidas

En esta tesis se usan series macroeconómicas agregadas de variables de producción, de empleo y de consumo de energía, procedente de diferentes fuentes.

Las bases de datos de la agencia oficial de estadística norteamericana es la fuente primaria de los datos de PIB y de consume energético de los capítulos 3 y 4.

Los datos de tasas de autoempleo son tomados de las Fuentes de la OCDE mientras que los datos macroeconómicos proceden de la base de datos de la *European Commission's Directorate General for Economic and Financial Affairs (AMECO)*.

1.3. Metodología

En esta tesis se hace uso de series temporales y de diferentes métodos numéricos de resolución de sistemas de ecuaciones dinámicos.

En la segunda parte de esta tesis se usan La teoría de redes complejas ha tenido gran impacto en diferentes campos del saber, dando el salto defini-

tivo de la Física a la Economía en los últimos años. Ello es debido fundamentalmente a que una red compleja resulta útil para comprender el impacto de las estructuras topológicas en la dinámica de un sistema. Algunos estudios han relacionado las series temporales con las redes complejas, permitiendo la aplicación de los métodos de éstas para caracterizar aquéllas. La idea es tratar la serie temporal como un conjunto de datos numéricos y transformar dicho conjunto en un objeto matemático diferente, un grafo, para, a continuación, estudiar qué propiedades de la serie original se conservan, cuáles se transforman o qué nuevas conclusiones pueden extraerse del estudio del grafo. En cualquier caso, se trata de una nueva metodología de estudio de series temporales. En esta línea, se han propuesto varias técnicas basadas en conceptos tales como correlaciones, reconstrucciones del espacio de fases, análisis de recurrencia o el análisis de las probabilidades de transición, entre otras.

Los primeros trabajos que surgieron al respecto hablan de redes donde cada nodo es una serie temporal y el peso de un enlace entre dos nodos viene dado por el coeficiente de correlación de las dos series o por la distancia entre ambas.

Para la realización de este trabajo, se realiza una aplicación en Excel en la que se programa el algoritmo del grafo de visibilidad y en la que es posible trabajar con series con un elevado número de observaciones, que es susceptible de ser usada por usuarios no expertos. Como ya hemos advertido, el grafo de visibilidad asociado a cada una de las series temporales hereda algunos de los aspectos importantes de la serie temporal original, por lo que constituye un método alternativo de análisis de series temporales.

La red resultante se analiza utilizando las técnicas habituales de estudio de redes complejas, calculando para ello una serie de métricas relevantes para evaluar las características estructurales y dinámicas de la red, tanto a nivel global como local.

La tercera parte de esta tesis analiza los cambios estructurales a través del análisis de cointegración, En particular se explora el hecho de que las relaciones de cointegración pueden resultar espúreas en presencia de cambio estructural. Para ello se usan los test de Gregory and Hansen (1996a, 1996b) y Hatemi-J (2008) para testar la presencia de uno o dos cambios estructurales en la relación de cointegración Una vez que esta es establecida se procede a estimar un modelo lineal en cada sub-período o régimen Finalmente, la última parte utiliza el método Runge-Kutta como método numérico para resolver problemas de sistemas de ecuaciones dinámicos.

1.4. Estructura capitular

Esta tesis se organiza en base a tres partes claramente diferenciadas, precedidas de una parte inicial de introducción y de una parte dedicada a resumir las principales conclusiones e hitos de la misma.

Esta tesis es un compendio de trabajos en tres líneas diferentes. La primera dedicada a explorar tanto de manera teórica como empírica el uso de grafos de visibilidad en el análisis de series temporales. Así, y frente a las tradicionales técnicas de análisis utilizadas para el datado de series temporales macroeconómicas con el objetivo de analizar cimas y valles y analizar la profundidad y amplitud de las distintas fases cíclicas, esta tesis desarrolla aplicaciones de la teoría de grafos al análisis de series siguiendo el enfoque abierto por Lacasa et al. a la vez que realiza una aportación de tipo

teórico sobre uno de los índices más comúnmente utilizados en la medición de la resiliencia.

Así, esta aportación nace de la necesidad de arbitrar algún mecanismo que permita evitar la falta de sensibilidad del índice cuando la serie temporal dispone de un número muy alto de observaciones.

Todas estas ideas y algoritmos son aplicados al análisis de dos tipos de series económicas para la economía norteamericana: las del producto interior bruto y las series de consumo de energía.

La elección de estas series no es arbitraria en tanto en cuanto se trata de realizar una aproximación alternativa, de forma que su puesta en valor ha de ser realizada en clave comparada con las técnicas tradicionalmente utilizadas para estas cuestiones. Por ello, optamos por abordar una serie clásica, cuyo datado es realizado por el NBER además de ser objeto de numerosos estudios y análisis, así como las series de consumo de energía norteamericanas que han sido objeto de un profuso análisis en el campo de la Economía de la energía.

El algoritmo del grafo de visibilidad (VGA) permite la transformación de una serie temporal en una red compleja de una manera muy simple. El estudio de las propiedades topológicas de la red aporta conocimiento sobre el comportamiento de la serie, pues el grafo hereda muchas de las propiedades de ésta. En los capítulos 2 y 3 procedemos a transformamos las series temporales correspondientes al consume de energía total y por tipo de fuente y a la tasa de variación trimestral del PIB de USA, en un grafo y aplicamos la teoría de las redes complejas para estudiarlo. La aproximación resultante al aplicarle diferentes métricas muestra unos resultados bas-

tante en línea con los que se obtienen con los algoritmos tradicionales, de forma que cabe concluir que estas aplicaciones aportan representaciones bastante útiles para el la previsión y el seguimiento.

Este bloque se cierra con una cuestión relacionada con el análisis de las series temporales, y más concretamente con la capacidad de absorción de un shock por parte de la serie. En concreto, se trata de analizar como uno de los índices más comúnmente utilizados para medir la resiliencia, muestra una cierta insensibilidad cuando el tamaño de la muestra es grande. Para abordar este problema se realiza una digresión y aportación de tipo teórico que puede corregir este sesgo.

Con este capítulo, se cierra esta segunda parte, y se da paso a un ensayo de macroeconometría aplicada en el ámbito de la economía laboral, y más concretamente en el terreno del *entrepreneurship*. En concreto, se aporta evidencia empírica acerca de un hecho estilizado en algunas economías con respecto a la evolución de las tasas de autoempleo: tras un largo período en el que las tasas de autoempleo han descendido de manera sostenida conforme mayor era el proceso de industrialización, parece que éstas muestran un cierto repunte en las economías más avanzadas en unos casos (U-shape) o al menos una estabilización (L-shape). Para este fin y frente al enfoque tradicional en el que se estima la relación entre la tasa de autoempleo de equilibrio y el PIB per cápita, proponemos el testar la hipótesis U/L-shape utilizando dos métodos alternativos: i) primero, descomponiendo la tasa de autoempleo en sus componentes cíclico y natural, y, ii) utilizando las aproximaciones econométricas de detección de cambio estructural de Kejriwal and Perron (2010). Procediendo de esta manera, se contrasta la hipótesis de la U-shape para 23 países de la OECD. Los resultados avalan parcialmente la hipótesis de la U-shape para 15 de los 23 países.

La cuarta parte de la tesis, aborda cuestiones relacionadas con el uso de métodos numéricos y con el desarrollo de aplicaciones susceptibles de ser utilizadas en la enseñanza de la modelización de la dinámica económica. En concreto, se presentan dos ensayos. El primero de ellos, es un ensayo de modelización y programación de problemas dinámicos mediante sistemas de ecuaciones diferenciales, resueltos a través de métodos numéricos. En particular, se presenta el estudio del comportamiento de una masa bacteriana en un cultivo controlado que presenta analogías a la evolución y supervivencia empresarial.

El segundo presenta dos aplicaciones de un simulador de una versión básica del modelo de oferta y demanda agregada dinámica –del conocido modelo de síntesis neoclásica– programada en Excel que ha de permitir que un alumno de un curso de Macroeconomía Intermedia conozca como se formulan, resuelven y los resultados que proporcionan los modelos dinámicos frente a los tradicionales ejercicios de estática comparativa que suelen componer el grueso de las programaciones de un curso estándar de Macroeconomía de este nivel. Así se pretende que estas aplicaciones sirvan al alumno para apreciar la potencia de los modelos dinámicos y la capacidad de éstos de reproducir ajustes cíclicos en el que el ajuste de las variables a sus valores de largo plazo dista de ser un ajuste gradual y lineal.

1.5. Publicaciones

Los capítulos de esta tesis han sido enviados a revistas académicas para su posible publicación, estando uno de ellos publicado en la revista epubli-

ca. Esta estructura capitular permite la lectura diferenciada e independiente de cada uno de los capítulos.

Una versión inicial del capítulo 6 fue galardonada en las Jornadas de Docencia en Economía, y los capítulos 2 y 3 han sido enviados para publicación en revistas JCR.

Part II: Detección de redes complejas en series temporales económicas

Capítulo 2. Una exploración de las series de consumo energético americanas a través del uso de grafos de visibilidad

2.1. Introducción

La disponibilidad de series de alta frecuencia con un elevado número de observaciones, unido al deseo de mejorar las técnicas de análisis de series temporales para el seguimiento y la predicción, está llevando a importar otros métodos y teorías a la exploración de series temporales.

En particular existe una literatura teórica y empírica reciente, dedicada a aplicar el algoritmo del grafo de visibilidad para transformar series temporales en grafos o redes complejas, a fin de estudiar éstas y deducir propiedades de aquéllas que son heredadas por el grafo. En particular, en este trabajo se trata de aplicar esta técnica a las series de consumo de energía agregado y por tipos para la economía norteamericana.

La elección de estas series no sólo responden al interés que despiertan desde el campo de la economía de la energía, sino por tratarse de series exten-

sivamente exploradas haciendo uso de la econometría tradicional de series temporales o a través del *machine learning*, lo que ha de permitirnos poner en valor los resultados obtenidos a través de este enfoque alternativo. En particular, este trabajo hace uso del algoritmo del grafo de visibilidad horizontal (en adelante, HVG) que permite la transformación de una serie temporal en una red compleja un mapeado de las series que las transforma en un grafo o red, de manera que el estudio de las propiedades topológicas de la red aporta conocimiento sobre el comportamiento de la serie, ya que el grafo hereda la mayor parte de las propiedades de ésta.

Así, y para mostrar la aplicación en este trabajo se elabora una aplicación que es aplicada a la transformamos seis series temporales correspondientes al consumo de energía norteamericano, total y desglosado en sus diferentes fuentes (carbón, gas natural, petróleo, energía nuclear, fuentes renovables) haciendo uso de datos mensuales en el período 1973/2-2016/12. Como resultado, las series originales se transforman en seis grafos o redes a las que aplicamos la teoría de las redes complejas para su comprensión y estudio.

En particular, encontramos que las redes construidas son redes con distribuciones de grado de tipo exponencial, $p(k) \sim e^{-\lambda k}$, con exponente $\lambda \ll \ln(2/3)$, lo que evidencia el carácter caótico de las series analizadas. Además, calculamos para cada una de ellas el coeficiente medio de *clustering*, el cual evidencia la elevada densidad de las mismas. Adicionalmente, la estructura de comunidades es analizada como forma de detectar la existencia de grupos de nodos fuertemente conectados, lo que sugiere dinámicas económicas similares en el tiempo para las series analizadas. Así mismo, calculamos el coeficiente de correlación de Pearson para cada una de las cinco series/redes, buscando determinar si se pueden clasificar como

redes *asortativas* o no, esto es, como forma de explorar las relaciones que guardan con el comportamiento a largo plazo de la series originales.

2.2. Redes complejas y series temporales

La teoría de redes complejas ha tenido gran impacto en diferentes campos del saber, dando el salto definitivo de la Física a la Economía en los últimos años. Ello es debido fundamentalmente a que una red compleja resulta útil para comprender el impacto de las estructuras topológicas en la dinámica de un sistema. Algunos estudios han relacionado las series temporales con las redes complejas, permitiendo la aplicación de los métodos de éstas para caracterizar aquéllas. La idea es tratar la serie temporal como un conjunto de datos numéricos y transformar dicho conjunto en un objeto matemático diferente, un grafo, para, a continuación, estudiar qué propiedades de la serie original se conservan, cuáles se transforman o qué nuevas conclusiones pueden extraerse del estudio del grafo. En cualquier caso, se trata de una nueva metodología de estudio de series temporales. En esta línea, se han propuesto varias técnicas basadas en conceptos tales como correlaciones, reconstrucciones del espacio de fases, análisis de recurrencia o el análisis de las probabilidades de transición, entre otras [14].

Los primeros trabajos que surgieron al respecto hablan de redes donde cada nodo es una serie temporal y el peso de un enlace entre dos nodos viene dado por el coeficiente de correlación de las dos series [1] o por la distancia entre ambas [2], pero son los trabajos de Lacasa et al [3, 4, 5] los que suponen una aportación seminal al desarrollar una nueva herramienta para el análisis de series temporales. Es el llamado *algoritmo del grafo de visibilidad* (VGA, por sus siglas en inglés), quienes muestran como la estruc-

tura de la serie temporal se hereda en el grafo asociado de manera unívoca, de tal forma que una serie periódica se transforma en un grafo regular, una serie aleatoria lo hace en un grafo aleatorio con función de distribución de grado exponencial, y una serie fractal se transforma en una red libre de escala, esto es, con función de distribución de grado potencial, $p_k \sim k^{-\gamma}$.

Para la realización de este trabajo, se realiza una aplicación en Excel en la que se programa el algoritmo y en la que es posible trabajar con series con un elevado número de observaciones, que es susceptible de ser usada por usuarios no expertos. Como ya hemos advertido, el grafo de visibilidad asociado a cada una de las series temporales hereda algunos de los aspectos importantes de la serie temporal original, por lo que constituye un método alternativo de análisis de series temporales.

La red resultante se analiza utilizando las técnicas habituales de estudio de redes complejas, calculando para ello una serie de métricas relevantes para evaluar las características estructurales y dinámicas de la red, tanto a nivel global como local [18, 19, 28].

Haciendo uso de este enfoque, y con el objetivo de poder poner en competencia esta propuesta metodológica con las aproximaciones tradicionales, se analizan las series temporales correspondientes al consumo mensual de energía norteamericana y a su desagregación atendiendo al tipo de energía –carbón, gas natural, nuclear, petróleo y renovables– en el período comprendido entre febrero de 1973 y febrero de 2016. Así pues, con las 527 observaciones de cada serie se realizan las transformaciones de éstas en grafos a los que aplicamos la teoría de redes complejas y aplicar métricas

cuyos resultados son una forma alternativa al tradicional análisis de series temporales.

2.3. Metodología

Como ya hemos avanzado, el análisis al que se someten las series tiene dos grandes hitos. El primero es el de la transformación de las series originales y el segundo el de la aplicación de diferentes métricas a las series transformadas. Pasemos a describir los principales elementos en cada uno de ellos.

2.3.1. El grafo de visibilidad

Como ya se ha comentado en este trabajo, los algoritmos basados en el grafo de visibilidad constituyen un grupo de distintas transformaciones geométricas con las que podemos convertir una serie temporal en un grafo [15].

El *algoritmo del grafo de visibilidad natural* (VGA) definido en Lacasa et al. [3], cuyo origen se sitúa en el campo de la geometría computacional [9], ha sido aplicado en campos tan dispares como el de las Finanzas Cuantitativas, la Economía internacional, al comportamiento humano, a la neurología, a la predicción meteorológica o incluso al turismo [10, 11, 12, 13, 16, 17, 29, 30]. Este algoritmo permite convertir una serie temporal en un grafo o red, donde cada nodo es un dato de la serie y donde dos nodos están conectados si hay “visibilidad” entre ellos, esto es, si la recta que une los dos datos de la serie no corta a ningún dato intermedio –esta ausencia de corte puede observarse en la representación del diagrama de barras de la

misma (fig.1) –. En términos más precisos, podemos definir la visibilidad entre dos datos arbitrarios de la serie temporal denotados por los pares (t_a, y_a) y (t_b, y_b) , que se convierten en dos nodos o vértices conectados en la red asociada si, para cualquier otro dato (t_c, y_c) tal que $t_a < t_c < t_b$, se verifica que:

$$y_c < y_a + (y_b - y_a) \cdot \frac{t_c - t_a}{t_b - t_a} \quad (1)$$

Por su parte, existe una versión simplificada de este concepto de visibilidad [5], denominada *grafo de visibilidad horizontal (HVG)*, que es en realidad un sub-grafo del anterior, donde la condición (1) se relaja de forma que dos pares de datos de la serie (t_a, y_a) y (t_b, y_b) tienen visibilidad horizontal si cualquier dato intermedio y_c satisface la condición $y_c < \inf(y_a, y_b)$, con $a < c < b$.

Es este último algoritmo el utilizado en este trabajo para transformar cada una de las series de consumo mensual de energía norteamericanas, total y desagregadas, en un grafo o red.

2.3.2. La aplicación del algoritmo

En la Figura 1 representamos una serie temporal arbitraria con ocho datos y los correspondientes grafos obtenidos mediante la aplicación de los algoritmos de visibilidad vertical y horizontal, respectivamente.

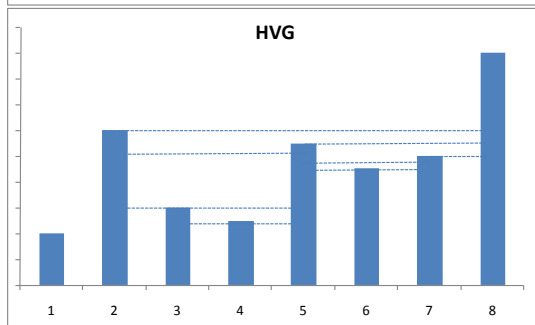
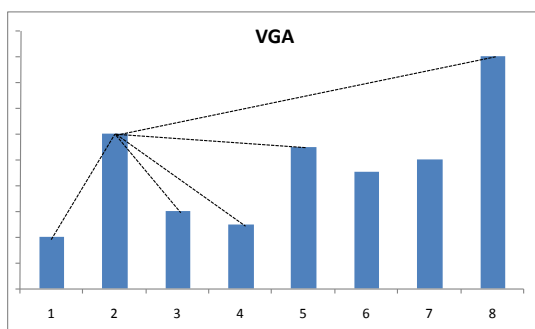
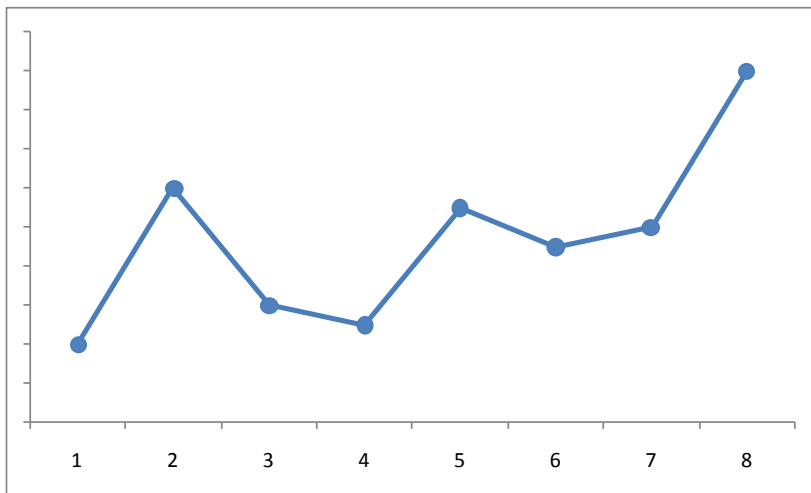
Los grafos construidos a partir de ambos algoritmos son grafos conectados –pues cada nodo se conecta, al menos, a los nodos anterior y posterior–, no dirigidos e invariantes bajo transformaciones afines de los datos de la

serie –re-escalado de ambos ejes, así como traslaciones horizontal o vertical–.

Llegados a este punto, convendría apuntar algo acerca de la conveniencia de utilizar uno u otro método. Optamos en este punto por una solución de compromiso consistente en hacer depender la decisión de qué aspectos de las series o redes construidas queremos analizar.

Así, hay que tener en cuenta que el criterio del *HVG* es más restrictivo que el del *VGA*, por lo que los nodos en el *HVG* tienen menos “visibilidad” que en el *VGA*. Siguiendo a Lacasa et al. [5], a nivel cualitativo esto no tiene ninguna implicación importante, pero a nivel cuantitativo significa que los grafos construidos con el *HVG* tienen menos métrica.

Si quisiéramos, por ejemplo, medir el *exponente de Hurst* a través de su relación con el exponente γ de la distribución del grado [4, 7], es mejor utilizar el *VGA*, dado que el cálculo de γ requiere de mayor número de datos. Si nuestra intención es distinguir un comportamiento caótico de uno estocástico correlado, entonces el *VHG* es el adecuado.



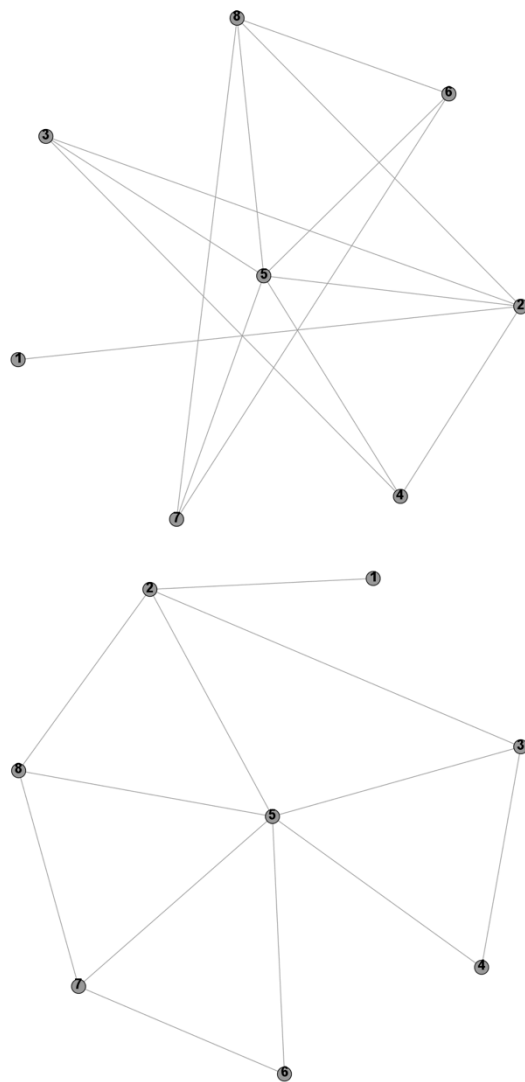


Fig. 1. Serie temporal con ocho datos y los correspondientes grafos asociados según el algoritmo empleado. Las redes constan de ocho nodos o vértices.

2.4. Propiedades de las redes obtenidas por la aplicación del algoritmo de visibilidad horizontal a las series de consumo de energía norteamericanas

Como hemos avanzado anteriormente, en la aplicación contenida en este trabajo se ha utilizado el algoritmo de visibilidad horizontal para transformar las series de consumo de las distintas fuentes de energía en los grafos o redes que representamos en la Figura 2, a partir de los cuales analizaremos qué propiedades tienen así como la aportación informativa que generan no sólo con respecto a la serie original sino también respecto a las técnicas convencionales. Para ello recurrimos a la teoría sobre redes complejas y nos adentramos en el estudio de algunas propiedades topológicas de una red, concretamente la distribución del grado, el coeficiente de *clustering* y la estructura de comunidades, entre otros.

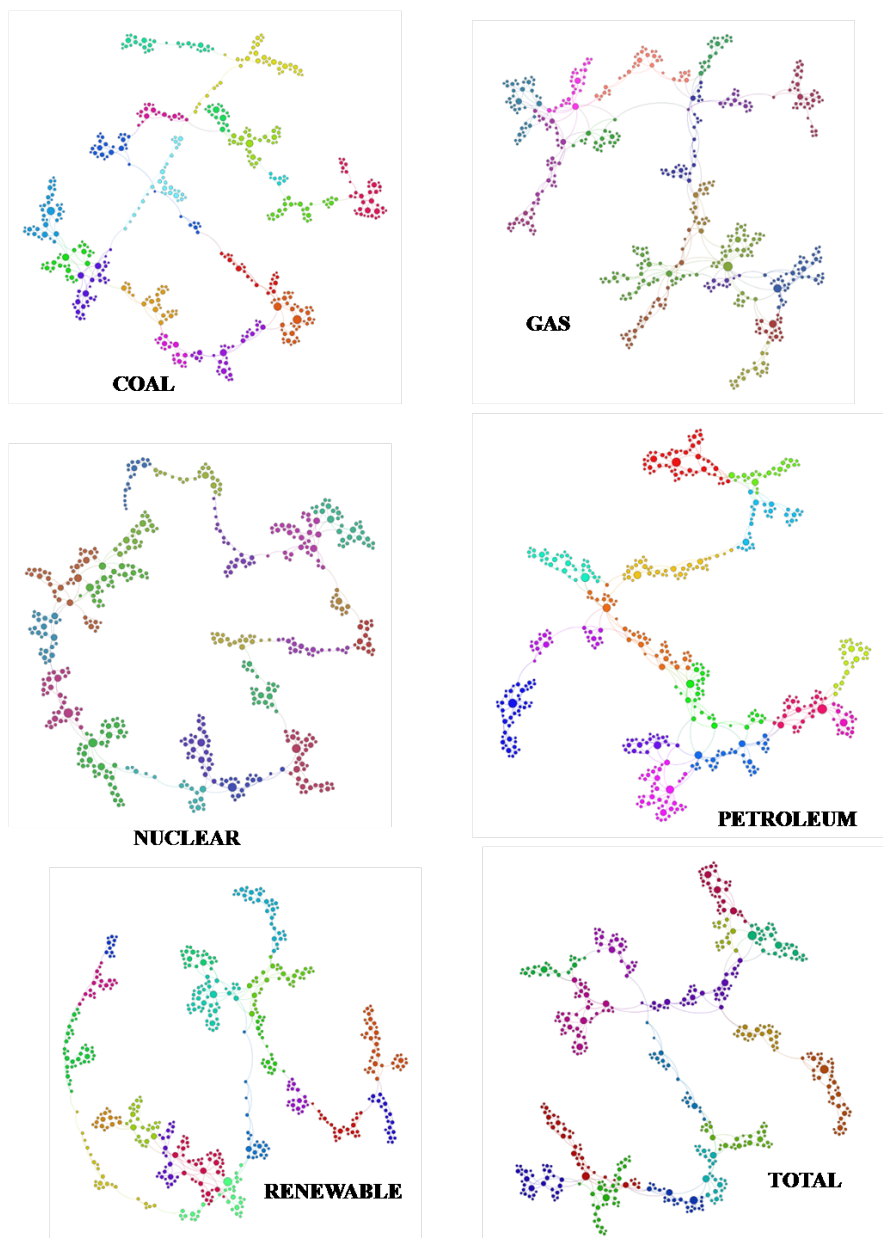


Fig. 2. Redes construidas al aplicar el HVG a las series del consumo mensual de las distintas fuentes de energía en USA. Los distintos colores representan diferentes comunidades y el tamaño de los nodos es proporcional al grado de los mismos.

2.4.1. Propiedades topológicas de las redes complejas

Grado, distribución del grado y relaciones

El grado de un nodo se define como el número de enlaces (conexiones) que inciden en él y, de alguna manera, mide cómo de grande es su influencia sobre los otros nodos de la red.

Si representamos conjuntamente las series originales y el grado de cada nodo (Fig. 3), podemos observar que los nodos más conectados (en adelante, *hubs*) se corresponden con picos de las series originales. Esto es lógico dado que los picos tienen mayor visibilidad que aquellos que no lo son. Los valles de la serie, por el contrario, se corresponden con nodos que, según el *HVG*, solamente se conectan con los datos anterior y posterior.

Conociendo la distribución del grado de la red pueden derivarse interesantes propiedades sobre la estructura de la misma. Muchas redes complejas muestran una distribución del grado especial, la cual sigue una ley de potencias, $p(k) \sim k^{-\gamma}$, siendo $p(k)$ la probabilidad de que un nodo elegido al azar tenga grado k . Se dice que son redes *libres de escala* pues no hay una escala que permita comparar los grados de los nodos, contrariamente a lo que sucede para las redes aleatorias donde la mayoría de los nodos tiene grado comparable al grado medio, $\langle k \rangle$, siendo la distribución del grado una distribución de *Poisson*. La consecuencia de una distribución del grado de tipo potencial es inmediata: hay unos pocos nodos (llamados *hubs*) que tienen un gran número de conexiones mientras que la inmensa mayoría tiene un número muy limitado de ellas.

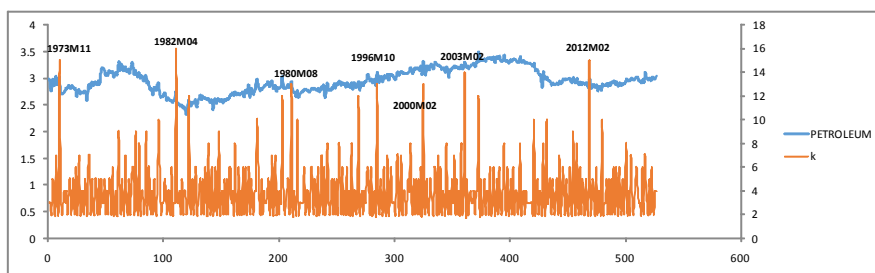
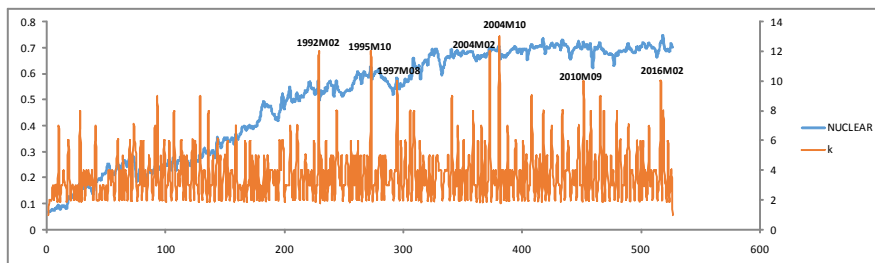
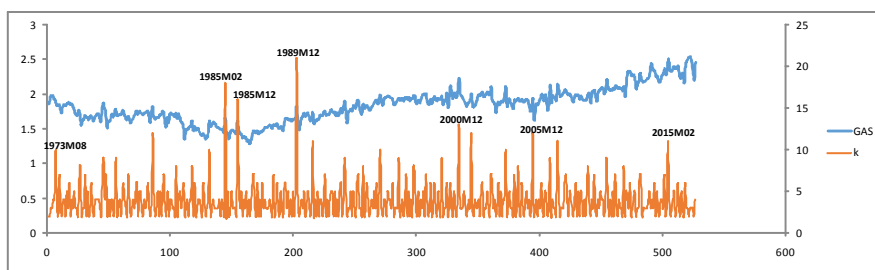
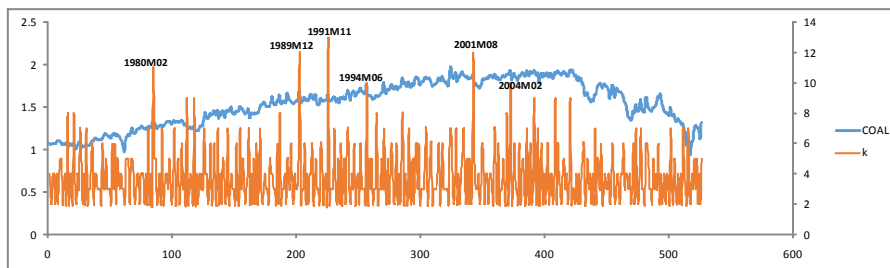
Conocer la distribución del grado de la red permite además distinguir [5, 6] si una serie temporal presenta un comportamiento caótico o si, por el contrario, se trata de un proceso estocástico (correlado o no). En efecto, si la distribución del grado es exponencial, $p(k) \sim e^{-\lambda k}$, existe un valor de λ que supone un umbral de separación entre los sistemas con comportamiento caótico y aquéllos cuya dinámica es fruto de un proceso estocástico. Dicho valor es $\lambda_c = \ln(2/3) = 0.405$, de manera que un valor del exponente λ claramente por debajo del mismo indica un comportamiento caótico de la serie, mientras que valores por encima significan que el proceso que origina la serie es un proceso estocástico.

Para el caso de las series que nos ocupan, las distribuciones del grado de cada una de las redes en las que se transforman mediante el algoritmo de visibilidad horizontal, se ajustan a una exponencial, $p(k) \sim e^{-\lambda k}$, siendo $p(k)$ la probabilidad de que un nodo elegido al azar tenga grado k , y donde el exponente λ toma los valores recogidos en la Tabla 2.

En la figura 4 representamos las distribuciones de grado y sus pendientes. Como puede observarse, los valores del exponente λ en todos los casos está muy por debajo del valor crítico λ_c , pudiendo confirmar así el carácter caótico de las series estudiadas.

Grafo de la serie	$\lambda (p(k) \sim e^{-\lambda k})$
Coal	0.22±5.2%
Gas	0.13±12.7%
Nuclear	0.2±3.2%
Petroleum	0.16±4.2%
Renewable	0.18±5.7%
Total	0.17±2.9%

Tabla 2. Exponente de la distribución de grado.



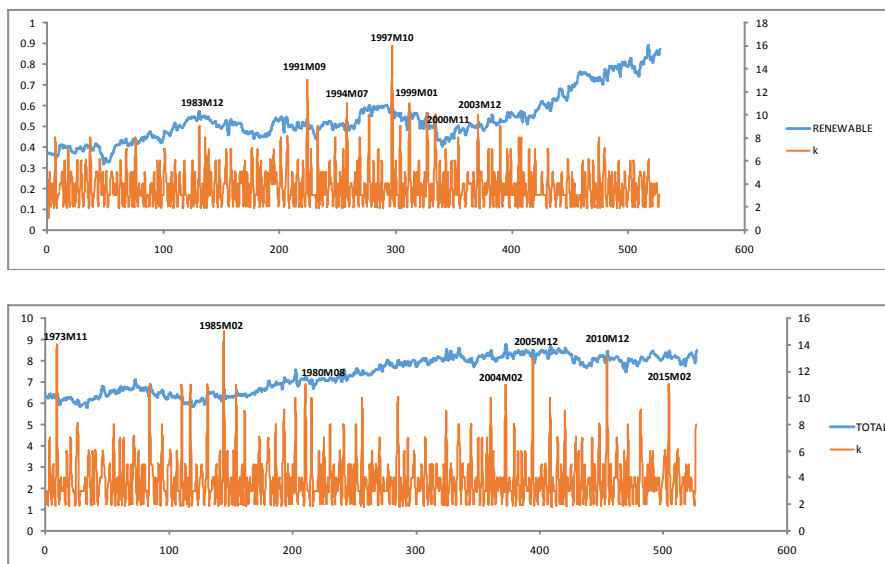


Fig. 3. Consumo de las distintas fuentes frente al grado. Señalados los *hubs*.

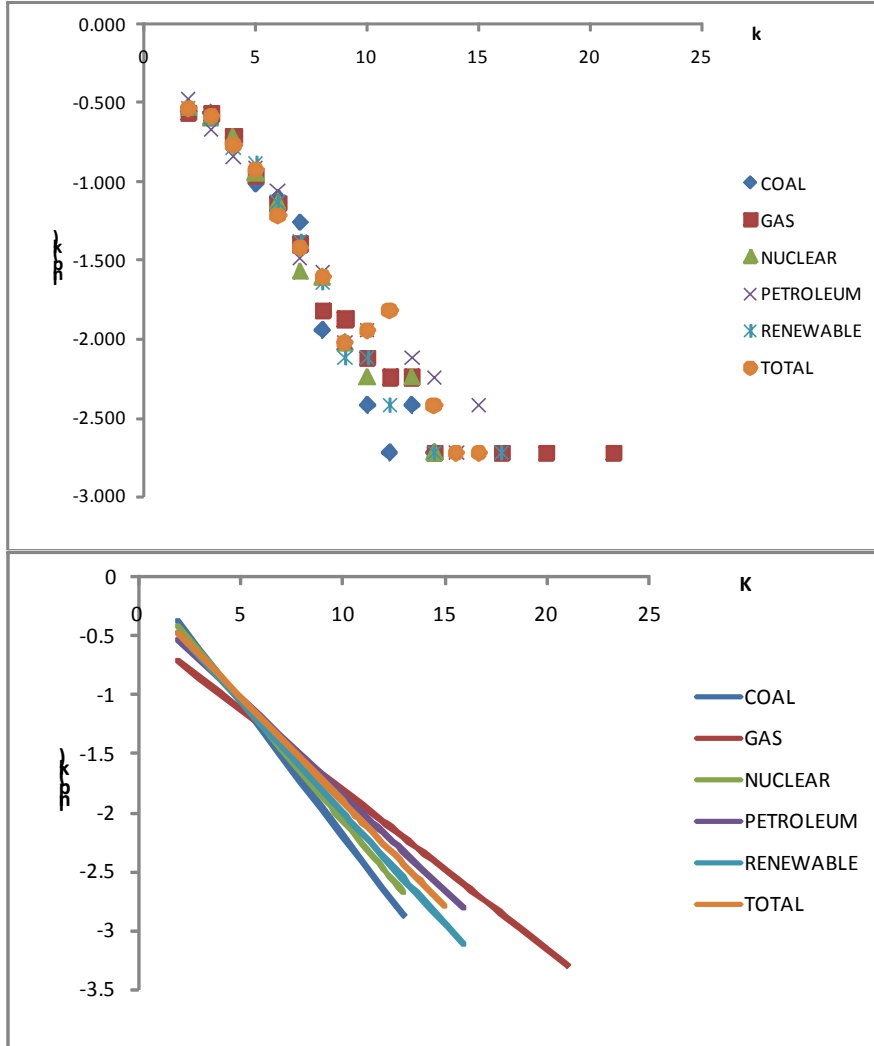


Fig 4. Distribuciones de grado y pendientes.

Como ya hemos indicado en otro apartado, la forma de la distribución de grado, tiene influencia directa en las propiedades de la red. Sin embargo, no puede aspirar a transmitir toda la información sobre la estructura de la misma; dos redes pueden tener distribuciones similares y exhibir diferentes características estáticas o dinámicas. Esta información adicional puede ser revelada mediante el estudio de las correlaciones entre los grados (véase

Gallos et al. [20]). Esta relación entre los nodos que nos proporciona la correlación juega un papel fundamental a la hora de determinar, por ejemplo, cómo se propaga por la red una perturbación, pues si ésta comienza en un *hub*, afectará con cierta probabilidad a su primer vecino, al segundo e incluso a otros vecinos más alejados. Y aún más, tal y como sugiere Newman [8], la *resiliencia* de una red –esto es, su capacidad para volver al equilibrio tras soportar shocks internos o externos– es muy sensible a las correlaciones de grado.

Una manera de calcular la correlación entre los grados de vértices diferentes es mediante el coeficiente de correlación de Pearson de los grados en ambos extremos de un enlace [8], el cual viene dado por:

$$r = \frac{\frac{1}{L} \sum_{j>i} k_i k_j - \left[\frac{1}{L} \sum_{j>i} \frac{1}{2} (k_i + k_j) A_{ij} \right]^2}{\frac{1}{L} \sum_{j>i} \frac{1}{2} (k_i^2 + k_j^2) A_{ij} - \left[\frac{1}{L} \sum_{j>i} \frac{1}{2} (k_i + k_j) A_{ij} \right]^2} \quad (2)$$

donde L es el número total de enlaces de la red, $\mathbf{A} = (A_{ij})$ es la matriz de adyacencia y k_i y k_j son los grados de los nodos en los extremos del i -ésimo enlace, $i = 1, 2, \dots, L$. En la terminología de las redes este coeficiente mide lo que se denomina la *asortatividad* de la red, que puede interpretarse como la preferencia que muestran los nodos a conectarse con otros nodos de grado similar. Si $r > 0$ se dice que la red es asortativa, por ejemplo, la mayoría de las redes sociales lo son, mientras que si $r < 0$ se dice que la red es *no asortativa*, siendo éste el caso de las redes tecnológicas y biológicas en las cuales nodos con grado elevado tienden a conectarse con nodos poco co-

nectados. Si $r = 0$, entonces no hay correlación entre los grados de los nodos. Serrano et al. [21] relacionan la asortatividad de una red con la resiliencia de la misma, concluyendo que cuanto más asortativa es una red, mayor es su resiliencia.

Para el caso de las redes que nos ocupan, hemos calculado el coeficiente de Pearson, que mostramos en la Tabla 3, concluyendo que en las redes correspondientes al consumo de energía no hay apenas correlación entre los grados de los nodos.

Serie	r
Coal	0.011
Gas	0.081
Nuclear	0.088
Petroleum	0.029
Renewable	0.071
Total	0.007

Tabla 3. Coeficientes de correlación de Pearson

El coeficiente de clustering

El coeficiente medio de *clustering* de una red mide la concentración de conexiones entre los vecinos de un nodo. Se define como la media entre los coeficientes de *clustering* locales de todos los nodos de la red, es decir:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i \quad (3)$$

donde N es el número de nodos de la red y $C_i = \frac{2L_i}{k_i(k_i - 1)}$, siendo L_i el

número de enlaces entre los nodos vecinos del nodo i -ésimo y k_i el grado

del mismo. C_i es el coeficiente de clustering local del nodo i -ésimo; representa la probabilidad de que dos vecinos de dicho nodo estén conectados. De alguna manera, C_i mide la densidad local de la red: cuanto más interconectados estén los vecinos del nodo i , mayor será C_i (comprendido entre 0 y 1). En términos probabilísticos, $\langle C \rangle$ es la probabilidad de que dos nodos elegidos aleatoriamente estén conectados.

Hemos calculado el coeficiente de *clustering* medio para cada una de las redes construidas a partir de las series de consumo de las distintas fuentes de energía en Estados Unidos, obteniendo los valores que recogemos en la Tabla 4. Observamos que el coeficiente medio de *clustering* es elevado para todas las redes, lo que significa que hay una alta probabilidad de que dos nodos elegidos al azar estén conectados. Este parámetro nos da una idea de la elevada densidad de nuestras redes.

Así mismo, el coeficiente de *clustering* puede utilizarse para revelar la estructura jerárquica de una red [31] si se demuestra que su dependencia respecto del grado es del tipo $\langle C \rangle \sim k^{-\alpha}$. En la Tabla 4 mostramos los valores hallados para el exponente α , constatando que, en todos los casos, $\langle C \rangle \sim 1/k$.

Serie	$\langle C \rangle \sim k^{-\alpha}$	α
Coal	0.618	$1 \pm 0.3\%$
Gas	0.624	$1 \pm 0.2\%$
Nuclear	0.60	$0.99 \pm 0.3\%$
Petroleum	0.646	$1 \pm 0.1\%$
Renewable	0.621	$0.99 \pm 0.1\%$
Total	0.634	$1 \pm 0.2\%$

Tabla 4. Coeficiente medio de clustering

Comunidades

Una característica común a la mayoría de las redes reales es la presencia en las mismas de grupos de nodos fuertemente conectados entre ellos y con conexiones menos densas con nodos externos al grupo. La identificación de estos grupos, llamados *comunidades* o *clusters*, es importante ya que los nodos que pertenecen a la misma comunidad son más propensos a compartir las mismas propiedades topológicas y la misma dinámica. Existen numerosos algoritmos para detectar comunidades en una red compleja, así como para medir su tamaño [22].

Todos ellos tratan de identificar la subdivisión óptima de la red maximizando un índice de calidad para la partición, denominado *índice de modularidad*, el cual fue introducido originariamente por Newman y Girvan [23] como criterio de parada del algoritmo y adoptado rápidamente por otros autores [24, 25].

Este criterio, se basa en la idea de que en un grafo aleatorio no se espera una estructura de grupos, por lo que la posible existencia de comunidades es revelada por la comparación entre la densidad real de aristas en un sub-grafo y la densidad que se esperaría en el sub-grafo si los nodos se unieran independientemente de la estructura de comunidad. El índice de modularidad se define como:

$$Q = \sum_{c=1}^{n_c} \left(\frac{l_c}{L} - \left(\frac{d_c}{2L} \right)^2 \right) \quad (4)$$

donde n_c es el número de *clusters* de la red, l_c es el número total de enlaces en el cluster c , y d_c es la suma total de los grados de los nodos del *cluster* c . El primer término de cada sumando representa la fracción de enlaces del grafo dentro del *cluster*, mientras que el segundo término representa la fracción esperada de enlaces que debería haber si el grafo fuera un grafo aleatorio con el mismo grado esperado para cada nodo.

Siguiendo a Fortunato [22], de acuerdo a la expresión anterior de Q , un sub-grafo se considerará una comunidad si su correspondiente contribución a la modularidad en la suma anterior es positiva.

Cuanto mayor sea la diferencia entre las dos fracciones que aparecen en cada sumando, mejor definida estará la comunidad. De esta manera, valores positivos y cercanos a 1 de la modularidad indicarán buenas particiones. La modularidad del grafo completo, esto es, considerado como una única comunidad, es igual a cero, pues los dos términos del único sumando de Q son iguales en este caso.

La modularidad es siempre menor que la unidad y puede ser también negativa. Por ejemplo, la partición en la que cada vértice se considera como una comunidad es siempre negativa, pues cada uno de los n_c sumandos en la expresión de Q es negativo, ya que la primera fracción que aparece vale 0. Esta es una buena característica de la medida, implicando que, si no hay particiones con modularidad positiva, el grafo no tiene estructura de comunidad. Por el contrario, la existencia de particiones con valores grandes negativos para la modularidad puede indicar la existencia de subgrupos con muy pocos enlaces internos y gran cantidad de enlaces entre ellos [27].

La modularidad máxima de un grafo crece generalmente si aumenta el tamaño y/o el número de *clusters* [26]. Por lo tanto, la modularidad no debe utilizarse para comparar la calidad de la estructura de comunidad de grafos que sean muy diferentes en tamaño.

Para una red construida a partir de una serie temporal de datos económicos como las nuestras, mediante el *HVG*, la interpretación de los diferentes *clusters* o comunidades es inmediata: los nodos que pertenecen a la misma comunidad se corresponden con períodos de la serie temporal con la misma dinámica económica.

Para identificar las comunidades en las redes creadas a partir de las series temporales del consumo de las distintas fuentes de energía en EE.UU, hemos seguido el algoritmo definido en [24] por su precisión y velocidad al tratar redes con elevado número de nodos. En la Tabla 5 mostramos el número de comunidades encontradas en cada grafo, así como el valor de la modularidad correspondiente. Como puede observarse, en todos los casos, el valor de Q es elevado, lo que es indicativo de la buena calidad de las particiones halladas.

En la Figura 2 cada comunidad aparece representada en un color diferente. Así mismo, en las figuras 5 a 10, sobre las series de datos originales, hemos indicado la división en comunidades, haciendo notar en cada caso los puntos de inflexión que marcan el paso de un grupo a otro. Hemos dividido el gráfico en dos para hacer más clara la representación. En algunos de los gráficos, las líneas de un mismo color, hacen referencia a períodos que pertenecen a la misma comunidad. Por ejemplo, en la Figura 5, para el consumo de carbón, encontramos que los nodos (datos) comprendidos en

los períodos 2000M02 a 2000M11 y 2004M02 a 2005M08 (en rojo), pertenecen a la misma comunidad. Lo mismo sucede en la Figura 6 para los períodos señalados en rojo, azul y verde, respectivamente. Este hecho, como hemos indicado con anterioridad, sugiere una misma dinámica económica para dichos períodos.

Serie	N ^a Com.	Q
COAL	18	0.9
GAS	18	0.88
NUCLEAR	20	0.9
PETROLEUM	15	0.87
RENEWABLE	19	0.89
TOTAL	16	0.89

Tabla 5. Número de comunidades y valor de la modularidad

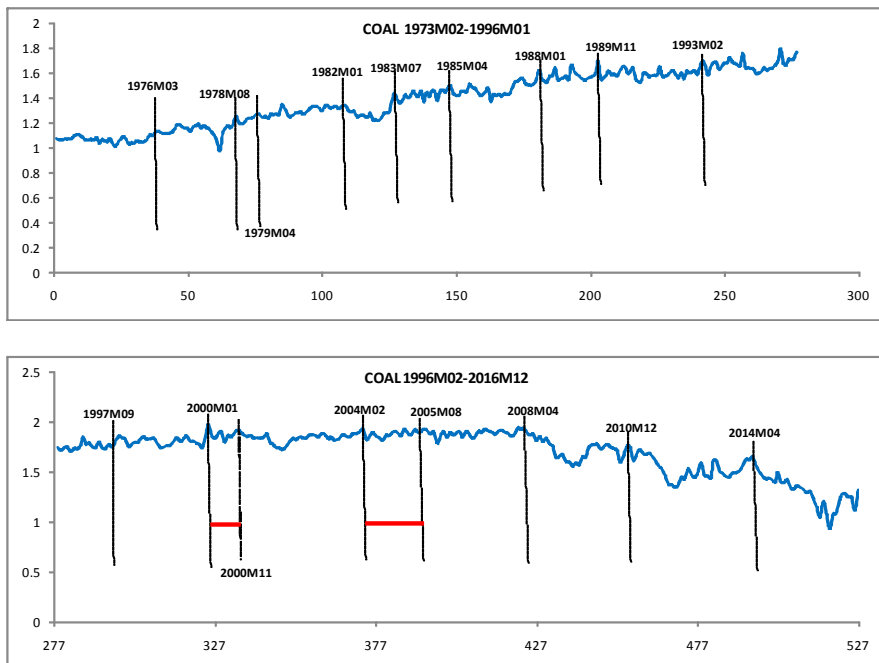


Figura 5. División en comunidades. Consumo de carbón.

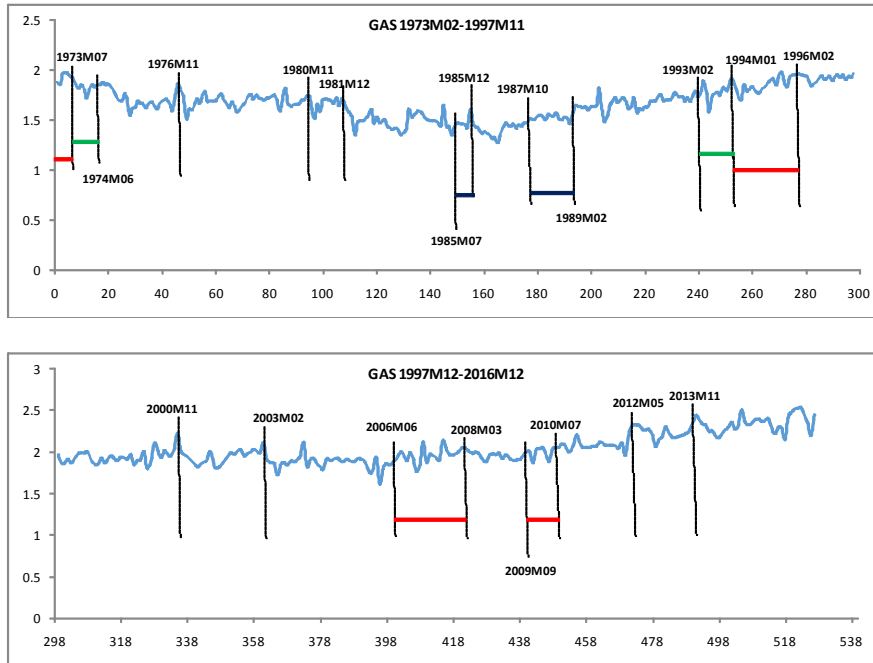


Figura 6. División en comunidades. Consumo de gas.

2.5. Referencias

- [1] Plerou, V. et al (1999), Universal and non universal properties of cross-correlations in financial time series. *Physical Review Letters*, 83, 1471-1474.
- [2] Onnela, J.P. et al (2002), Dynamic asset trees and portfolio analysis. *Eur. Phys. J. B* 30, 285-288.
- [3] Lacasa, L. et al (2008), From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA* 105, 4972-4975.
- [4] Lacasa, L. et al (2009), The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys. Lett.* 86, 30001-30004.
- [5] Luque, B. et al (2009), Horizontal visibility graphs: exact results for random time series. *Phys. Rev.* 80, 046103.

- [6] Lacasa, L., Toral, R. (2010), Description of stochastic and chaotic series using visibility graphs. *Physical Review E*, 82, art. 036120.
- [7] Wang, N. et al (2012), Visibility graph analysis on quarterly macroeconomic series of China Based on complex network theory. *Physica A* 391, 6543-6555.
- [8] Newman, M. E. J. (2002), Assortative mixing in networks. *Phys. Rev. Lett.* 89, 208701.
- [9] M. de Berg et al (2000), *Computational geometry: Algorithms and Applications*, 2nd ed., Berlin, New York: Springer-Verlag.
- [10] Y. Yang et al (2009), Visibility graph approach to exchange rate series. *Physica A: Statistical Mechanics and its applications*, 388(20), pp. 4431-4437.
- [11] Liu C. et al (2009), Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. Arxiv.org/abs/0905.1831.
- [12] Elsner, J. B. et al (2009), Visibility Network of United States Hurricanes. *Geophysical research letters*, 36.
- [13] Zhuang, E. et al (2014), Time series analysis of the developed financial markets integration using visibility graphs. *Physica A* 410, pp. 483-495
- [14] Donner et al (2010), Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcations and Chaos*, 21, 4, pp. 1019-1046.
- [15] Núñez, A. M. et al (2012), Visibility algorithms: A short review. En Zhang, Y. (Ed.) *New Frontiers in Graph Theory*, InTech, Rijeka, Croatia, pp. 119-152.
- [16] Da Fontoura Costa, L. and Baggio, R. (2009), The web of connections between tourism companies: structure and dynamics. *Physica A*, 388, pp. 4286-4296.

- [17] Baggio, R. (2014), Complex tourism systems: a visibility graph approach. *Kybernetes*, 43(3/4), pp. 445-461.
- [18] Da Fontoura Costa, L. and Baggio, R. (2007), Characterization of complex networks: a survey of measurements. *Advances in Physics*, 46, 1, pp. 167-242.
- [19] Newman, M. E. J. (2010), *Networks-An introduction*. Oxford University Press, Oxford.
- [20] Gallos, L. K., Song, C. and Makse, H. A. (2008), Scaling of degree correlations and its influence on diffusion in scale-free networks. *Physical review Letters*, vol. 100, 248701
- [21] M. A. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani (2007). Correlations in complex networks, in *Large Scale Structure and Dynamics of Complex Networks: from Information Technology to Finance and Natural Science*, G. Caldarelli and A. Vespignani editors, World Scientific, pp. 35-66.
- [22] Fortunato, S. (2010), Community detection in graphs. *Physics Reports*, 486 No. 3-5, pp. 75-174.
- [23] Newman, M. E. J. and Girvan, M. (2004), Finding and evaluating community structure in networks. *Physical Review E*, vol. 69, p. 026113. [arXiv:cond-mat/0308217](https://arxiv.org/abs/cond-mat/0308217) [cond-mat.stat-mech].
- [24] Blondel, V. D. et al (2008), Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008 P10008. arXiv: 0803.0476 [physics.soc-ph].
- [25] Clauset, A. et al (2004), Finding community structure in very large networks. *Physical Review E*, vol. 70, p. 066111.
- [26] Good, B. H. et al (2009), The performance of modularity maximization in practical contexts. arXiv:0910.0165 [physics.data-an].
- [27] Newman, M. E. J., (2006a), Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74(3), 036104.
- [28] Barabási, A. L. , *Network Science Book*, available for download on the [web page](http://barabasilab.neu.edu/networksciencebook/downloadPDF.html) <http://barabasilab.neu.edu/networksciencebook/downloadPDF.html>.

[29] Baggio, R. and Sainaghi, R. (2016), Mapping time series into networks as a tool to assess the complex dynamics of tourism systems. *Tourism Management* 54, pp. 23-33.

[30] Zhang Bo et al (2015), Volatility behavior of visibility graph EMD financial time series from Ising interacting system. *Physica A*, 432, pp. 301-314.

[31] Ravasz, E. and Barabási, A. L. (2003), Hierarchical organization in complex networks. *Physical Review E*, 67, p.026112.

El algoritmo de visibilidad para una serie temporal donde una variable y depende del tiempo t se basa en una comprobación secuencial ordenada donde para cada punto de la serie (t_i, y_i) se busca si “es visible” por el resto de puntos que le suceden. Para realizar dicha comprobación, para cada pareja de puntos i - j -ésimos con $j > i$, se determinará que están conectados y son visibles entre sí si para cada punto k -ésimo con $i < k < j$ se cumple la condición expresada por la siguiente fórmula simplificada:

$$y_k < y_i + \frac{t_k - t_i}{t_j - t_i} \cdot (y_j - y_i)$$

Para realizar esta operación fácilmente se va a programar una hoja de cálculo Excel en VBA (Visual Basic Applicated) que realice la búsqueda de conexiones.

Este algoritmo genera una tabla de conexiones que se podría asimilar a la matriz de adyacencia de una red basada en grafos. Por tanto, a partir de este punto se pueden calcular parámetros relacionados con redes. En este punto los parámetros que se calculan son los grados de los nodos (número de conexiones directas que tiene cada nodo), los coeficientes de clustering individuales (un parámetro que determina la fracción de tripletes a los que pertenece un nodo con respecto al máximo de tripletes a los que podría pertenecer) y el coeficiente global de clustering (que determina la fracción de tripletes existentes con respecto al máximo que podrían haber en la red).

Para realizar el rastreo punto a punto buscando conexiones se utiliza este código:

```
For i = 1 To fin - 1
  For j = i + 1 To fin
    vis = True
    a = (Cells(j, 2).Value - Cells(i, 2).Value) /
(Cells(j, 1).Value - Cells(i, 1).Value)
    b = Cells(i, 2).Value - a * Cells(i, 1).Value
    For k = i + 1 To j - 1
      yc = a * Cells(k, 1).Value + b
```

```

        If yc <= Cells(k, 2).Value Then
            vis = False
            Exit For
        End If
    Next
    If vis Then
        ReDim X(1 To cont, 1 To 2)
        X(cont, 1) = Cells(i, 1).Value
        X(cont, 2) = Cells(j, 1).Value
        Km(i) = Km(i) + 1
        Km(j) = Km(j) + 1
        If inci Then
            sh.Cells(cont, 1).Value = Cells(i,
1).Value
            sh.Cells(cont, 2).Value = Cells(j,
1).Value
        End If
        cont = cont + 1
    End If
Next
Next

```

Con este algoritmo se comprueban las conexiones y se calcula el grado de cada nodo al mismo tiempo, pudiéndose copiar en la Hoja 'sh' una referencia a las conexiones.

También se calcula el coeficiente de asortatividad de Pearson relacionado con una propiedad de las redes mediante la fórmula:

$$r = \frac{\frac{1}{M} \cdot \sum_{j>i} k_i k_j - \left[\frac{1}{M} \cdot \sum_{j>i} \frac{1}{2} (k_i + k_j) \right]^2}{\frac{1}{M} \cdot \sum_{j>i} \frac{1}{2} (k_i^2 + k_j^2) - \left[\frac{1}{M} \cdot \sum_{j>i} \frac{1}{2} (k_i + k_j) \right]^2}$$

Donde M es el número total de conexiones de la red, k_i y k_j son los grados de cada par de nodos conectados. Se puede programar mediante el algoritmo:

```

For i = 1 To fin - 1
    For j = i + 1 To fin
        vis = True
    
```

```

    a = (Cells(j, 2).Value - Cells(i, 2).Value) /
(Cells(j, 1).Value - Cells(i, 1).Value)
    b = Cells(i, 2).Value - a * Cells(i, 1).Value
    For k = i + 1 To j - 1
        yc = a * Cells(k, 1).Value + b
        If yc <= Cells(k, 2).Value Then
            vis = False
            Exit For
        End If
    Next
    If vis Then
        Skk = Skk + Km(i) * Km(j)
        S2k = S2k + 0.5 * (Km(i) + Km(j))
        Sk2 = Sk2 + 0.5 * (Km(i) ^ 2 + Km(j) ^ 2)
        sL = sL + 1
    End If
Next
Next
rM = (Skk / sL - (S2k / sL) ^ 2) / (Sk2 / sL - (S2k /
sL) ^ 2)

```

También para calcular los coeficientes de clustering individuales mediante la fórmula:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

Donde L es el número de enlaces entre los nodos vecinos al nodo i y k su grado. A partir de aquí se puede calcular el coeficiente de clustering individual promedio haciendo la media geométrica entre todos los coeficientes individuales.

Y el coeficiente global de clustering se calcula a través de la fórmula:

$$C = \frac{3 \cdot \text{Número de triángulos}}{\text{Número de triplete}}$$

Todo se puede programar como:

```
For i = 1 To num
  For k = 1 To num
    M(k) = 0
  Next
  cl = 0
  ck = 0
  For j = 1 To num
    If Not (i = j) Then
      vis = True
      paso = Math.Sgn(j - i)
      If lineahorizontal.Value Then
        For k = i + paso To j - paso Step
          If Cells(i + 2, 2).Value <=
Cells(k + 2, 2).Value Or Cells(j + 2, 2).Value <=
Cells(k + 2, ex).Value Then
            vis = False
            Exit For
          End If
        Next
      ElseIf linearecta.Value Then
        a = (Cells(i + 2, 2).Value - Cells(j
+ 2, 2).Value) / (Cells(i + 2, 1).Value - Cells(j + 2,
1).Value)
        b = Cells(i + 2, 2).Value - a *
Cells(i + 2, 1).Value
        For k = i + paso To j - paso Step
          yc = a * Cells(k + 2, 1).Value +
b
          If yc <= Cells(k + 2, 2).Value
Then
            vis = False
            Exit For
          End If
        Next
      End If
      If vis Then
        M(j) = 1
        ck = ck + 1
      End If
    End If
  Next

  For j = 1 To num
    If M(j) = 1 Then
```

```
        For k = 1 To num
            If Not (j = k) And Not (i = k) Then
                vis = True
                paso = Math.Sgn(k - j)
                a = (Cells(j + 2, 2).Value -
Cells(k + 2, 2).Value) / (Cells(j + 2, 1).Value -
Cells(k + 2, 1).Value)
                b = Cells(j + 2, 2).Value - a *
Cells(j + 2, 1).Value
                For o = j + paso To k - paso Step
paso
                    yc = a * Cells(o + 2,
1).Value + b
                    If yc <= Cells(o + 2,
2).Value Then
                        vis = False
                        Exit For
                    End If
                Next
                If vis Then
                    ctriple = ctriple + 1
                    If M(k) = 1 Then
                        ctriang = ctriang + 1
                        cl = cl + 1
                    End If
                End If
            End If
        Next
    End If
Next
    If ck > 1 Then
        Cm(i) = cl / (ck * (ck - 1))
    Else
        Cm(i) = 0
    End If
Next
s = 0
For i = 1 To num
    Cells(i + 2, 6).Value = Cm(i)
    s = s + Cm(i)
Next
Cmedio = s / num
Cglobal = ctriang / ctriple
```

Capítulo 3. El uso de análisis de redes como forma de caracterizar el ciclo económico: Una aplicación al análisis del ciclo norteamericano.

Existe una literatura cada vez más abundante que aplica la teoría de grafos y redes al análisis económico y financiero. En este trabajo se aplican desarrollos recientes de algoritmos de visibilidad orientados a la detección de los puntos de inflexión como forma de datar el ciclo económico.

3.1. Introducción

El interés por el ciclo económico es en sí mismo cíclico (Mullineux & Dickinson, 1992), de forma que el interés por el análisis del mismo se asocia a períodos recesivos como el recientemente experimentado a nivel internacional. Entender estas fluctuaciones y diseñar métodos de datado han sido desde los trabajos seminales de Burns & Mitchell (1946) y de Bry & Boschan (1971) tradicionalmente las dos grandes cuestiones del análisis de los ciclos económicos.

En el momento actual el estudio de los ciclos económicos es un tema de gran actualidad e interés no sólo para analizar cómo y cuando impactó la Gran Recesión y cuando se inició el despegue sino también para evaluar el grado de integración económica entre las economías.

Así pues, el datado, se ha convertido en temas de especial interés. En este contexto, este trabajo pretende aplicar nuevas herramientas analíticas en el análisis de ciclos económicos, enfocadas en la detección de sobre la ocurrencia de un cambio en la fase del ciclo económico.

De lo expuesto hasta ahora convendremos que la medición de los ciclos es clave para el análisis macroeconómico. Para medirlos, requerimos de una definición del ciclo que nos permita detectarlo tras lo cual estaremos en disposición de determinar los puntos de inflexión. La elección de un procedimiento adecuado de análisis es crucial para medir las características del ciclo.

En este sentido y como ponen de manifiesto Harding & Pagan (2003) existe una falta de un consenso sobre la mejor forma de medirlo. Desde el enfoque tradicional de Bry & Boschan (1971) para el datado del ciclo con el objetivo de emular la metodología propuesta por el NBER, han sido múltiples las alternativas propuestas –véase por ejemplo por Pagan (1997a,b), Harding & Pagan (2002, 2003, 2005), Engel, Haugh & Pagan (2005) o más recientemente Stock & Watson (2014) entre otros–.

Pero, no solo debemos de centrarnos en la obtención de los “turning points” sino también en las distintas características de dichas situaciones de expansión (definidas de valle a pico) o recesión (de pico a valle), tales

como la sincronía entre los distintos ciclos (Harding & Pagan, 2002; McDermott & Scott, 2000; Artis, Kontolemis & Osborn, 1995; Camacho, Pérez-Quirós & Saiz, 2006 o Krolzig & Toro, 2005), las duraciones de las expansiones y recesiones, la amplitud o el incremento/decremento medio de las expansiones/recesiones o el exceso (Camacho, Pérez-Quirós & Saiz, 2008) o la forma cóncava/convexa, (por encima o por debajo del crecimiento medio lineal).

La identificación e interpretación de la existencia de patrones en los datos ha sido siempre un reto para un gran número de disciplinas, pero en especial para la Economía. El hallazgo de estos patrones en series temporales económicas y la provisión de explicaciones de estos hechos estilizados han sido el núcleo de la investigación teórica y empírica dedicada al estudio de los ciclos económicos. El desarrollo de diferentes metodologías para datar e identificar ciclos económicos y de crecimiento, el estudio de las relaciones entre ciclos de distintas regiones, sectores o variables, y el desarrollo de teorías que lo expliquen, ha sido el denominador común en distintos campos de investigación de Economía. Sobre estas premisas, este trabajo trata de avanzar en la aplicación de algoritmos alternativos a los tradicionales que serán aplicados al análisis del ciclo norteamericano. En este punto, tras la recesión, se reaviva el interés por la medición de las fluctuaciones de la actividad económica y por capturar patrones entre diferentes economías dependiendo de su mayor o menor grado de integración. En cualquiera de los dos casos, el control de dimensiones tales como la duración y profundidad o el datado del comienzo y final de regímenes expansivos o recesivos son cuestiones esenciales no sólo por proporcionar información relevante sobre el estado de la economía, que es relevante para la toma de decisiones por parte de los diferentes agentes económicos, sino

también porque proporcionan información clave para la previsión y el seguimiento. Dentro del amplio abanico existente dentro de dicho estudio sobre el ciclo económico, podríamos señalar, principalmente la necesidad u objetivo de identificar “turning points” (máximos y mínimos locales, en otras palabras, picos y valles) y establecer el datado de fases.

Con este objetivo este trabajo aplica el algoritmo del grafo de visibilidad para transformar la serie temporal de la tasa de variación trimestral del PIB norteamericano en el período 1947 a 2015, en una red compleja, a fin de estudiar ésta y deducir propiedades de aquélla que son heredadas por el grafo.

3.2. Sobre el algoritmo de visibilidad

El algoritmo del grafo de visibilidad (VGA) permite la transformación de una serie temporal en una red compleja de una manera muy simple. El estudio de las propiedades topológicas de la red aporta conocimiento sobre el comportamiento de la serie, pues el grafo hereda muchas de las propiedades de ésta. Transformamos la serie temporal correspondiente a la tasa de variación trimestral del PIB de USA durante el período que va de enero de 1947 a enero de 2016, en un grafo y aplicamos la teoría de las redes complejas para estudiarlo. En particular, encontramos que la red construida es una red *libre de escala*, con distribución de grado de tipo potencial, $p_k \sim k^{-\gamma}$, lo que es habitual en la mayoría de las redes complejas reales. Además, calculamos para el grafo construido la longitud de camino medio, el coeficiente medio de clustering y la estructura de comunidades, evidenciando con estos datos que se trata de una red de mundo pequeño. También

determinamos el coeficiente de Hurst de la serie, el cual está relacionado con la memoria a largo plazo de la misma y con el exponente γ . Así mismo, calculamos el coeficiente de correlación de Pearson de la red, determinando así si se trata de una red asortativa o no, lo que guarda relación con el comportamiento a largo plazo de la serie original. La estructura de comunidad de esta red sugiere cambios en la dinámica de la serie de la que procede.

Como sabemos, de un tiempo a esta parte, la teoría de redes complejas ha tenido gran impacto en diferentes campos del saber, dando el salto definitivo de la Física a la Economía en los últimos años. Ello es debido fundamentalmente a que una red compleja resulta útil para comprender el impacto de las estructuras topológicas en la dinámica de un sistema. En los últimos años, algunos estudios han relacionado las series temporales con las redes complejas, permitiendo la aplicación de los métodos de éstas para caracterizar aquéllas. La idea es tratar la serie temporal como un conjunto de datos numéricos y transformar dicho conjunto en un objeto matemático diferente, un grafo, para, a continuación, estudiar qué propiedades de la serie original se conservan, cuáles se transforman o qué nuevas conclusiones pueden extraerse del estudio del grafo. En cualquier caso, se trata de una nueva metodología de estudio de series temporales. En esta línea, se han propuesto varias técnicas basadas en conceptos tales como correlaciones, reconstrucciones del espacio de fases, análisis de recurrencia, probabilidades de transición, etc. [14]. Los primeros trabajos que surgieron al respecto hablan de redes donde cada nodo es una serie temporal y el peso de un enlace entre dos nodos viene dado por el coeficiente de correlación de las dos series [1] o por la distancia entre ambas [2], pero es el tra-

bajo de Lacasa et al [3,4,5] el que atrae la atención sobre una nueva herramienta para el análisis de series temporales llamada *el algoritmo del grafo de visibilidad* (VGA, por sus siglas en inglés). Los autores demuestran que la estructura de la serie temporal se hereda en el grafo asociado, de tal forma que una serie periódica se transforma en un grafo regular, una serie aleatoria lo hace en un grafo aleatorio con función de distribución de grado exponencial, y una serie fractal se transforma en una red libre de escala, esto es, con función de distribución de grado de tipo potencial, $p_k \sim k^{-\gamma}$.

El algoritmo es fácil de programar y funciona incluso con series de muchos datos. El grafo de visibilidad asociado a una serie temporal hereda algunos aspectos importantes de la serie temporal original, por lo que constituye un nuevo método de análisis de series temporales. La red resultante se analiza utilizando las técnicas habituales de estudio de redes complejas, calculando para ello una serie de métricas relevantes para evaluar las características estructurales y dinámicas de la red, tanto a nivel global como local [18, 19]. En este trabajo analizamos mediante esta metodología la serie temporal correspondiente a la tasa de variación trimestral del PIB de USA, durante los años comprendidos entre enero de 1947 y enero de 2016. Se trata de una serie de 276 datos, la cual transformamos en un grafo al que aplicamos la teoría de redes complejas, obteniéndose interesantes conclusiones.

3.3. Metodología

3.3.1. El grafo de visibilidad

Como ya se ha comentado en este trabajo, los algoritmos basados en el grafo de visibilidad constituyen un grupo de distintas transformaciones geométricas con las que podemos convertir una serie temporal en un grafo [15].

El *algoritmo del grafo de visibilidad natural* (VGA) definido en Lacasa et al. [3], cuyo orígenes se sitúan en el campo de la geometría computacional [9], ha sido aplicado en campos tan dispares como el de las Finanzas Cuantitativas, la Economía internacional, al comportamiento humano, a la neurología, a la predicción meteorológica o incluso al turismo [10, 11, 12, 13, 16, 17, 29, 30]. Este algoritmo permite convertir una serie temporal en un grafo o red, donde cada nodo es un dato de la serie y donde dos nodos están conectados si hay “visibilidad” entre ellos, esto es, si la recta que une los dos datos de la serie no corta a ningún dato intermedio –esta ausencia de corte puede observarse en la representación del diagrama de barras de la misma (fig.1) –. En términos más precisos, podemos definir la visibilidad entre dos datos arbitrarios de la serie temporal denotados por los pares (t_a, y_a) y (t_b, y_b) , que se convierten en dos nodos o vértices conectados en la red asociada si, para cualquier otro dato (t_c, y_c) tal que $t_a < t_c < t_b$, se verifica que:

$$y_c < y_a + (y_b - y_a) \cdot \frac{t_c - t_a}{t_b - t_a} \quad (1)$$

En la Figura 1 representamos una serie temporal con seis datos y el correspondiente grafo obtenido mediante la aplicación del algoritmo de visibilidad.

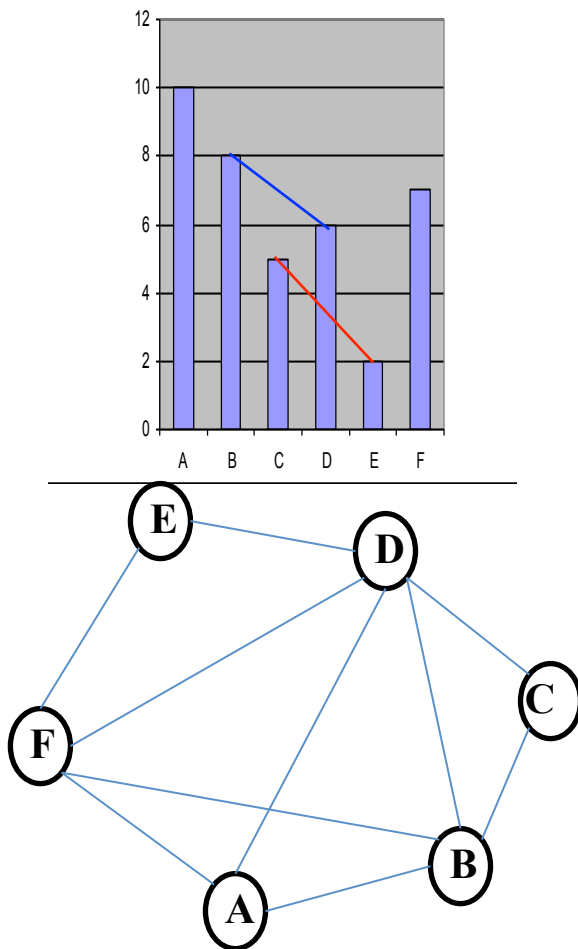


Fig. 1. Serie temporal con seis datos y el correspondiente grafo asociado. La red consta de seis nodos o vértices. Obsérvese cómo, por ejemplo, los nodos B y D están conectados, pero no sucede así con los nodos C y E, ya que el segmento que une los extremos de las barras correspondientes en el gráfico de la serie temporal corta al dato intermedio, D.

El grafo construido a partir del VGA es un grafo conectado (pues cada nodo se conecta, al menos, a los nodos anterior y posterior), no dirigido e invariante bajo transformaciones afines de los datos de la serie (re-escalado de ambos ejes, así como traslaciones horizontal o vertical).

Los algoritmos basados en el grafo de visibilidad constituyen un grupo de distintas transformaciones geométricas con las que podemos convertir una serie temporal en un grafo [15]. El algoritmo descrito se conoce como *algoritmo de visibilidad natural*. Existe una versión simplificada del mismo [5], el *grafo de visibilidad horizontal* (HVG), que es en realidad un sub-grafo del anterior, donde la condición (1) se relaja y dos datos de la serie (t_a, y_a) y (t_b, y_b) tienen visibilidad horizontal si cualquier dato intermedio y_c satisface la condición $y_c < \inf(y_a, y_b)$, $a < c < b$.

En la Figura 2 mostramos el grafo asociado a la serie temporal en estudio.

Llegados a este punto, nos preguntamos qué aspectos importantes de la serie original hereda el grafo construido a partir del VGA. Uno de ellos es la relación existente entre la distribución del grado de las redes construidas y el exponente de Hurst de la serie temporal original, relación ampliamente investigada por Lacasa et al [4]. Este exponente, H , se usa como una medida del comportamiento o memoria a largo plazo de la serie. Su valor oscila entre 0 y 1 con un punto crítico intermedio en $H = 0.5$. Un valor $H = 0.5$ indica que la serie original sigue un movimiento browniano estándar, siendo el suceso que la origina independiente, aleatorio y no correlacionado. Si $0.5 < H < 1$, la serie muestra correlación positiva y persistencia, mientras que si $H < 0.5$, entonces la serie muestra anti-persistencia y anti-

correlación. Lacasa et al [4] demuestran que el exponente de Hurst de una serie puede ser estimado a partir de la distribución del grado, $p(k) \sim k^{-\gamma}$, de la red asociada mediante el VGA. En efecto, prueban que la relación entre γ y H es lineal, concretamente, $\gamma \approx 3 - 2H$. Ni et al [6] encuentran también una relación lineal, aunque diferente, entre ambos parámetros. Así mismo, Wang et al [7] determinan una relación no lineal entre H y el exponente β de la distribución de grado exponencial para un pseudo brownian motion series. Así mismo, Zu-Guo et al [31] relacionan H linealmente con la dimensión fractal media de los grafos de visibilidad horizontal construidos a partir de fractional Brownian motions. Todos estos estudios ponen en evidencia que los algoritmos del grafo de visibilidad constituyen otro método de análisis de series temporales.

Para la serie del PIB de USA analizada, encontramos que $H = 0.644$ con un error del 1.5%, lo que nos lleva a afirmar que la serie muestra correlación positiva y persistencia.

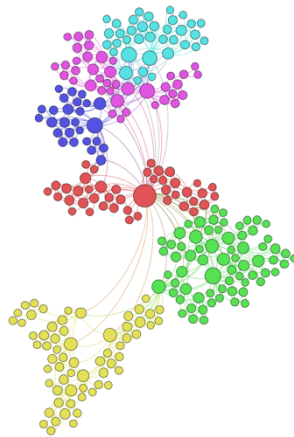


Fig. 2. Grafo construido a partir de la serie temporal analizada

A continuación nos cuestionamos qué otras propiedades de la red construida aportan nuevo conocimiento sobre la serie temporal de origen. Para ello recurrimos a la teoría sobre redes complejas y nos adentramos en el estudio de las propiedades topológicas de una red, concretamente la distribución del grado, el coeficiente de clustering, la longitud media de camino más corto y la estructura en comunidades.

3.4. Propiedades topológicas de las redes complejas

Las características topológicas más importantes de las redes complejas son el grado de un vértice, la longitud media de camino más corto y el coeficiente de clustering.

El grado, la distribución de grado y relaciones

El grado de un nodo se define como el número de enlaces (conexiones) que inciden en él y, de alguna manera, mide cómo de grande es su influencia sobre los otros nodos de la red. Conociendo la distribución del grado de la red, pueden derivarse interesantes propiedades sobre la estructura de la misma. Muchas redes complejas muestran una distribución de grado especial, la cual sigue una ley de potencias, $p_k \sim k^{-\gamma}$. Se dice que son redes *libres de escala* pues no hay una escala que permita comparar los grados de los nodos, contrariamente a lo que sucede para las redes aleatorias donde la mayoría de los nodos tiene grado comparable al grado medio, $\langle k \rangle$, siendo la distribución de grado una distribución de Poisson. La consecuencia de una distribución del grado de tipo potencial es inmediata: hay unos pocos

nodos (llamados *hubs*) que tienen un gran número de conexiones mientras que la inmensa mayoría tiene un número muy limitado de ellas.

Para el caso de la serie que nos ocupa, hemos representado en la Figura 3 la distribución del grado de la red en la que se transforma mediante el VGA. Obsérvese que se trata de una distribución de tipo potencial, donde el exponente γ obtenido es $\gamma = 1.686 \pm 10.6\%$. Se comprueba entonces la relación lineal entre γ y H establecida por Lacasa et al [5].

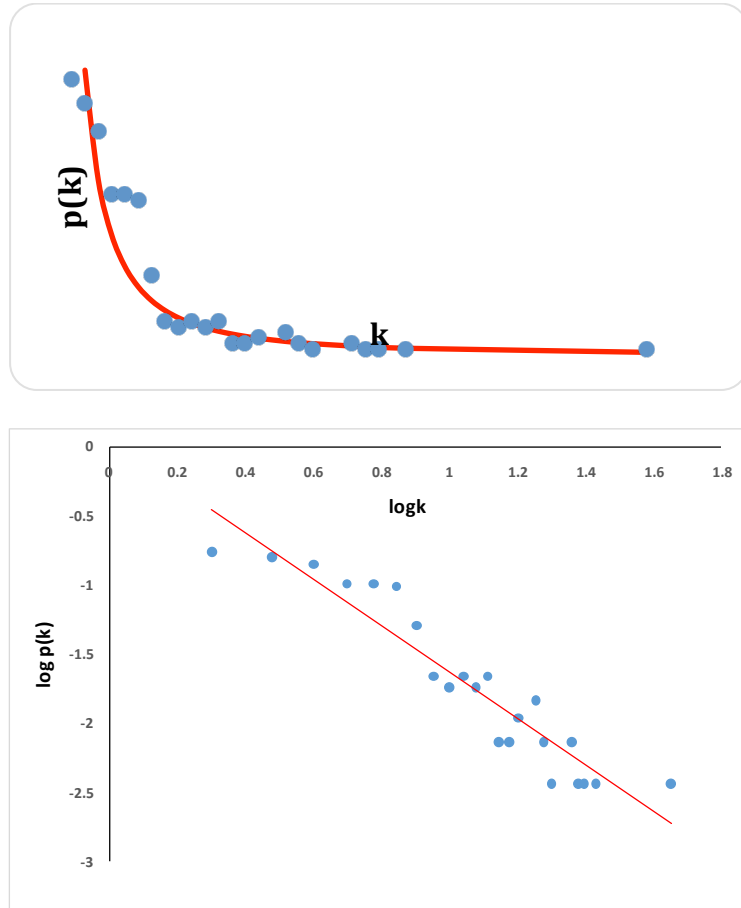


Fig. 3. Distribución de grado y pendiente de la misma.

Si representamos conjuntamente la serie original y el grado de cada nodo (Figura 4), podemos observar que los *hubs*, esto es, los nodos más conectados, se corresponden con picos de la serie original. Concretamente, estos hubs se sitúan en el segundo trimestre de 1978 ($k = 45$), segundo trimestre de 1996 ($k = 27$), primer trimestre de 1971 ($k = 25$) o el cuarto trimestre de 1952 ($k = 24$).

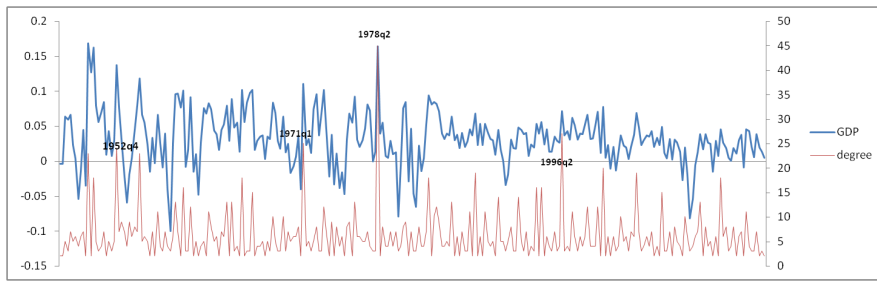


Fig. 4. PIB y grado.

La forma de la distribución de grado, como hemos comentado, tiene influencia directa en las propiedades de la red, sin embargo no puede transmitir toda la información sobre la estructura de la misma; dos redes pueden tener distribuciones similares y exhibir diferentes características estáticas o dinámicas. Esta información adicional puede ser revelada mediante el estudio de las correlaciones entre los grados (Gallos et al [20]). Esta relación entre los nodos que nos proporciona la correlación juega un papel fundamental a la hora de determinar, por ejemplo, cómo se propaga por la red una perturbación, pues si ésta comienza en un hub, afectará con cierta probabilidad a su primer vecino, al segundo e incluso a otros vecinos más alejados. Y aún más, tal y como sugiere Newman [8], la resiliencia de una

red (esto es, su capacidad para soportar shocks internos o externos y recuperarse en un tiempo razonable) es muy sensible a las correlaciones de grado.

Una manera de calcular la correlación entre los grados de vértices diferentes es mediante el coeficiente de correlación de Pearson de los grados en ambos extremos de un enlace [8], el cual viene dado por:

$$r = \frac{\frac{1}{L} \sum_{j>i} k_i k_j - \left[\frac{1}{L} \sum_{j>i} \frac{1}{2} (k_i + k_j) A_{ij} \right]^2}{\frac{1}{L} \sum_{j>i} \frac{1}{2} (k_i^2 + k_j^2) A_{ij} - \left[\frac{1}{L} \sum_{j>i} \frac{1}{2} (k_i + k_j) A_{ij} \right]^2} \quad (2)$$

donde L es el número total de enlaces de la red, $A = (A_{ij})$ es la matriz de adyacencia y k_i y k_j son los grados de los nodos en los extremos del i -ésimo enlace, $i = 1, 2, \dots, L$. En la terminología de las redes este coeficiente mide lo que se denomina la *asortatividad* de la red que puede interpretarse como la preferencia que muestran los nodos a conectarse con otros nodos de grado similar. Si $r > 0$ se dice que la red es asortativa, por ejemplo, la mayoría de las redes sociales lo son, mientras que si $r < 0$ se dice que la red es *disassortative*, siendo éste el caso de las redes tecnológicas o biológicas en las cuales nodos con grado elevado tienden a conectarse con nodos poco conectados. Si $r = 0$, entonces no hay correlación entre los grados de los nodos. Serrano et al [21] relacionan la asortatividad de una red con la resiliencia de la misma, concluyendo que cuanto más asortativa es una red, mayor es su resiliencia.

Para el caso de la red que nos ocupa, hemos calculado que el coeficiente de Pearson resulta ser igual a 0.13, lo que confirma la asortatividad de la red.

El coeficiente de clustering

El coeficiente medio de *clustering* de una red mide la concentración de conexiones entre los vecinos de un nodo. Se define como la media entre los coeficientes de *clustering* locales de todos los nodos de la red, es decir:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i \quad (3)$$

donde N es el número de nodos de la red y $C_i = \frac{2L_i}{k_i(k_i - 1)}$, siendo L_i el número de enlaces entre los nodos vecinos del nodo i -ésimo y k_i el grado del mismo. C_i es el coeficiente de clustering local del nodo i -ésimo; representa la probabilidad de que dos vecinos de dicho nodo estén conectados. De alguna manera, C_i mide la densidad local de la red: cuanto más interconectados estén los vecinos del nodo i , mayor será C_i (comprendido entre 0 y 1). En términos probabilísticos, $\langle C \rangle$ es la probabilidad de que dos nodos elegidos aleatoriamente estén conectados.

Hemos calculado el coeficiente de clustering medio para la red construida a partir de la serie del PIB de USA, obteniendo que $\langle C \rangle = 0.756$, es decir, hay un 75% de probabilidad de que dos nodos elegidos al azar estén conectados. Este parámetro nos da una idea de la elevada densidad de nuestra red. Así mismo, el coeficiente de clustering puede utilizarse para revelar la estructura jerárquica de una red [32] si se demuestra que su de-

pendencia respecto del grado es del tipo $\langle C \rangle \sim k^{-\alpha}$. En la figura 5 representamos $\langle C \rangle$ frente a k , constatando que la relación entre ambas es del tipo anterior, con exponente $\alpha = 0.71 \pm 6.6\%$.

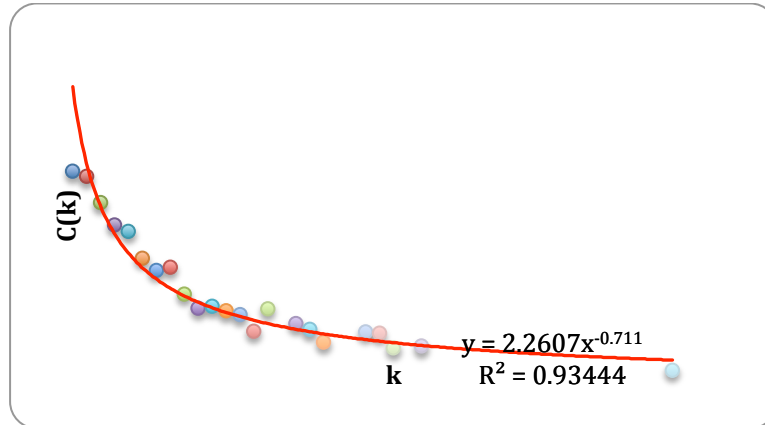


Fig. 5. Coeficiente de clustering en función del grado

La longitud media de camino más corto

La longitud media de camino se define como la distancia media entre todos los pares de nodos. Se calcula mediante la fórmula:

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{i,j=1,\dots,N} d_{ij} \quad (4)$$

donde d_{ij} denota la distancia entre los nodos i -ésimo y j -ésimo, entendida ésta como el número de enlaces que hay que recorrer para ir del nodo i al nodo j por el camino más corto posible. Para el caso de nuestra red, hemos obtenido una longitud media de camino igual a 4.20, es decir, para ir de un nodo a otro de esta red, hay que visitar por término medio, unos 4 enlaces. En la Figura 6 representamos la distribución de las distancia en

nuestra red. Hemos calculado también el *diámetro* de la red, esto es, la distancia entre los dos nodos más alejados, el cual resulta ser igual a 9.

La mayoría de las redes reales son *redes de mundo pequeño* [28], es decir, redes donde la longitud de camino medio es pequeña en comparación con el tamaño de la red, lo que garantiza que para desplazarnos de un nodo a otro de la red tenemos que visitar unos pocos enlaces nada más. Obsérvese que la distribución de las distancias más cortas (Figura 6) se aproxima a una Normal, de manera que las distancias son comparables a $\langle d \rangle$. Existe entre un 25% y un 30% de probabilidad de que las distancias más cortas entre pares de nodos sea del orden de $\langle d \rangle$, pequeña si la comparamos con el tamaño de la red. Adicionalmente, en una red de mundo pequeño, el coeficiente de clustering es elevado, evidenciando la existencia de grupos de nodos fuertemente conectados entre sí. Estas dos características se cumplen en nuestra red, por lo que estamos ante una red de mundo pequeño.

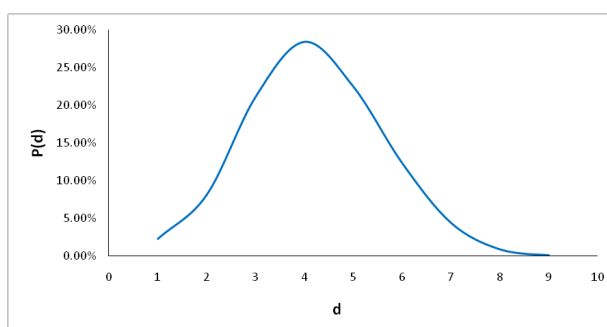


Fig. 6. Distribución de camino más corto

Comunidades

Una característica común a la mayoría de las redes reales es la presencia en las mismas de grupos de nodos fuertemente conectados entre ellos y con conexiones menos densas con nodos externos al grupo. La identificación de estos grupos, llamados *comunidades* o *clusters*, es importante ya que los nodos que pertenecen a la misma comunidad son más propensos a compartir las mismas propiedades topológicas y la misma dinámica. Existen numerosos algoritmos para detectar comunidades en una red compleja, así como para medir su tamaño [22]. Todos ellos tratan de identificar la subdivisión óptima de la red maximizando un índice de calidad para la partición, denominado *índice de modularidad*, el cual fue introducido originariamente por Newman and Girvan [23] como criterio de parada del algoritmo y adoptado rápidamente por otros autores [24, 25].

Se basa en la idea de que en un grafo aleatorio no se espera una estructura de grupos, por lo que la posible existencia de comunidades es revelada por la comparación entre la densidad real de aristas en un subgrafo y la densidad que se esperaría en el subgrafo si los nodos se unieran independientemente de la estructura de comunidad. El índice de modularidad se define como:

$$Q = \sum_{c=1}^{n_c} \left(\frac{l_c}{L} - \left(\frac{d_c}{2L} \right)^2 \right) \quad (5)$$

donde n_c es el número de clusters de la red, l_c es el número total de enlaces en el cluster c , y d_c es la suma total de los grados de los nodos del cluster c . El primer término de cada sumando representa la fracción de enlaces del grafo dentro del cluster, mientras que el segundo término repre-

senta la fracción esperada de enlaces que debería haber si el grafo fuera un grafo aleatorio con el mismo grado esperado para cada nodo.

Siguiendo a Fortunato [22], de acuerdo a la expresión anterior de Q , un subgrafo se considerará una comunidad si su correspondiente contribución a la modularidad en la suma anterior es positiva. Cuanto mayor sea la diferencia entre las dos fracciones que aparecen en cada sumando, mejor definida estará la comunidad. De esta manera, valores positivos y cercanos a 1 de la modularidad indicarán buenas particiones. La modularidad del grafo completo, esto es, considerado como una única comunidad, es igual a cero, pues los dos términos del único sumando de Q son iguales en este caso. La modularidad es siempre menor que la unidad y puede ser también negativa. Por ejemplo, la partición en la que cada vértice se considera como una comunidad es siempre negativa, pues cada uno de los n_c sumandos en la expresión de Q es negativo, ya que la primera fracción que aparece vale 0. Esta es una buena característica de la medida, implicando que, si no hay particiones con modularidad positiva, el grafo no tiene estructura de comunidad. Por el contrario, la existencia de particiones con valores grandes negativos para la modularidad puede indicar la existencia de subgrupos con muy pocos enlaces internos y gran cantidad de enlaces entre ellos [27].

La modularidad máxima de un grafo crece generalmente si aumenta el tamaño y/o el número de clusters [26]. Por lo tanto, la modularidad no debe utilizarse para comparar la calidad de la estructura de comunidad de grafos que sean muy diferentes en tamaño.

Para una red construida a partir de una serie temporal de datos económicos como la nuestra, mediante el VGA, la interpretación de los diferentes clusters o comunidades parece inmediata: los nodos que pertenezcan a la misma comunidad se corresponderán con períodos de la serie temporal con la misma dinámica en la economía –o pertenecientes al mismo ciclo económico–.

Para identificar las comunidades en la red creada a partir de la serie temporal del PIB de EE.UU, hemos seguido el algoritmo definido en [24] por su precisión y velocidad al tratar redes con elevado número de nodos.

Como puede intuirse a partir de la visualización de la red mostrada en la Figura 2, al realizar el análisis de la modularidad se han detectado 6 grupos de nodos densamente conectados. La división en estos 6 grupos o comunidades puede considerarse una buena división, dado que el índice de modularidad obtenido es igual a 0.72. El hecho de encontrar 6 comunidades supone la existencia de 5 puntos de inflexión en la serie del PIB, correspondientes a los períodos 1966q1, 1972q3, 1981q1, 1996q3 y 2005q2 (Figura 7). Estos puntos separan períodos con dinámicas económicas diferentes.

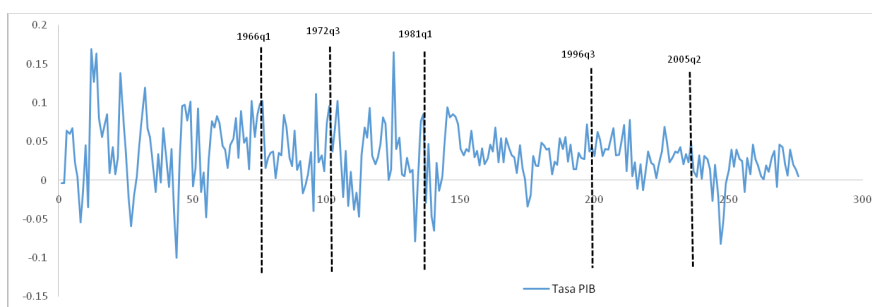


Fig 7. Puntos de inflexión y división en comunidades.

3.5. Referencias

- [1] Plerou, V. et al (1999), Universal and non universal properties of cross-correlations in financial time series. *Physical Review Letters*, 83, 1471-1474.
- [2] Onnela, J.P. et al (2002), Dynamic asset trees and portfolio analysis. *Eur. Phys. J. B* 30, 285-288.
- [3] Lacasa, L. et al (2008), From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA* 105, 4972-4975.
- [4] Lacasa, L. et al (2009), The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys. Lett.* 86, 30001-30004.
- [5] Luque, B. et al (2009), Horizontal visibility graphs: exact results for random time series. *Phys. Rev.* 80, 046103.
- [6] Ni, X. H. et al (2009), Degree distributions of the visibility graph mapped from fractional Brownian motions and multifractal random walks. *Phys. Lett. A* 373, 3822-3826.
- [7] Wang, N. et al (2012), Visibility graph analysis on quarterly macroeconomic series of China Based on complex network theory. *Physica A* 391, 6543-6555.
- [8] Newman, M. E. J. (2002), Assortative mixing in networks. *Phys. Rev. Lett.* 89, 208701.
- [9] M. de Berg et al (2000), *Computational geometry: Algorithms and Applications*, 2nd ed., Berlin, New York: Springer-Verlag.
- [10] Y. Yang et al (2009), Visibility graph approach to exchange rate series. *Physica A: Statistical Mechanics and its applications*, 388(20), pp. 4431-4437.
- [11] Liu C. et al (2009), Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. [Arxiv.org/abs/0905.1831](https://arxiv.org/abs/0905.1831).

[12] Elsner, J. B. et al (2009), Visibility Network of United States Hurricanes. *Geophysical research letters*, 36.

[13] Zhuang, E. et al (2014), Time series analysis of the developed financial markets integration using visibility graphs. *Physica A* 410, pp. 483-495

[14] Donner et al (2010), Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcations and Chaos*, 21, 4, pp. 1019-1046.

[15] Núñez, A. M. et al (2012), Visibility algorithms: A short review. En Zhang, Y. (Ed.) *New Frontiers in Graph Theory*, InTech, Rijeka, Croatia, pp. 119-152.

[16] Da Fontoura Costa, L. and Baggio, R. (2009), The web of connections between tourism companies: structure and dynamics. *Physica A*, 388, pp. 4286-4296.

[17] Baggio, R. (2014), Complex tourism systems: a visibility graph approach. *Kybernetes*, 43(3/4), pp. 445-461.

[18] Da Fontoura Costa, L. and Baggio, R. (2007), Characterization of complex networks: a survey of measurements. *Advances in Physics*, 46, 1, pp. 167-242.

[19] Newman, M. E. J. (2010), *Networks-An introduction*. Oxford University Press, Oxford.

[20] Gallos, L. K., Song, C. and Makse, H. A. (2008), Scaling of degree correlations and its influence on diffusion in scale-free networks. *Physical review Letters*, vol. 100, 248701

[21] M. A. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani (2007). Correlations in complex networks, in *Large Scale Structure and Dynamics of Complex Networks: from Information Technology to Finance and Natural Science*, G. Caldarelli and A. Vespignani editors, World Scientific, pp. 35-66.

[22] Fortunato, S. (2010), Community detection in graphs. *Physics Reports*, 486 No. 3-5, pp. 75-174.

[23] Newman, M. E. J. and Girvan, M. (2004), Finding and evaluating community structure in networks. *Physical Review E*, vol. 69, p. 026113. [arXiv:cond-mat/0308217](https://arxiv.org/abs/cond-mat/0308217) [cond-mat.stat-mech].

[24] Blondel, V. D. et al (2008), Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008 P10008. arXiv: 0803.0476 [physics.soc-ph].

[25] Clauset, A. et al (2004), Finding community structure in very large networks. *Physical Review E*, vol. 70, p. 066111.

[26] Good, B. H. et al (2009), The performance of modularity maximization in practical contexts. arXiv:0910.0165 [physics.data-an].

[27] Newman, M. E. J., (2006a), Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74(3), 036104.

[28] Barabási, A. L. , *Network Science Book*, available for download on the [web](http://barabasilab.neu.edu/networksciencebook/downloadPDF.html) page <http://barabasilab.neu.edu/networksciencebook/downloadPDF.html>.

[29] Baggio, R. and Sainaghi, R. (2016), Mapping time series into networks as a tool to assess the complex dynamics of tourism systems. *Tourism Management* 54, pp. 23-33.

[30] Zhang Bo et al (2015), Volatility behavior of visibility graph EMD financial time series from Ising interacting system. *Physica A*, 432, pp. 301-314.

[31] Zu-Guo Yu et al (2016), Multifractality and Laplace spectrum of horizontal visibility graphs constructed from fractional Brownian motions. *J. Stat. Mech.* 033206.

[32] Ravasz, E. and Barabási, A. L. (2003), Hierarchical organization in complex networks. *Physical Review E*, 67, p.026112.

[33] Artis, M.J., Kontolemis, Z.G., & Osborn, D.R. (1995), *Classical Business Cycles for G7 and European Countries*, CEPR Discussion Paper No 1137, Centre for Economic Policy Research, London.

[34] Bry, G., & Boschan, C. (1971). Cyclical analysis of time series: Selected procedures and computer programs. National Bureau of Economic Research.

[35] Burns, A.F. and Mitchell, W.C. (1946). *Measuring Business Cycles*, National Bureau of Economic Research, Cambridge, M.A.

[36] Harding, D. & A. Pagan (2002). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics* 49, 365-381.

[37] Harding, D., & Pagan, A. (2003). A comparison of two business cycle dating methods. *Journal of Economic Dynamics & Control*, 27(9), 1681-1690.

[38] Krolzig H-M, & Toro J. (2005), Classical and modern business cycle measurement: the European case. *Spanish Economic Review*, 7(1), 1–21

[39] McDermott, C.J. & Scott A. (2000). *Concordance in business cycles*. Technical report, IMF.

[40] McQueen, G. & Thorley, S. (1993). Asymmetric Business Cycle turning points. *Journal of Monetary Economics*, 31(3), 341-362.

[41] Pagan, A.R. (1997), *Policy, Theory and the Cycle*, *Oxford Review of Economic Policy*, 13, 19-33.

[42] Pagan, A.R. (1997). Towards an understanding of some business cycle characteristics. *Australian Economic Review*, 30, 1-15.

[43] Škare, M. & Stjepanović, S. (2015). *Measuring Business Cycles: A Review*. *Contemporary Economics*, 10(1), 83-94.

[44] Stock, J.H., & M. W. Watson (2014). Estimating turning points using large data sets. *Journal of Econometrics* 178, 368-381.

**Parte III: Un ensayo de macroeconomía
aplicada**

Capítulo 4. The U-shape revisited: testing for structural breaks in the relationship between entrepreneurship and income

Carree et al. (2002, 2007) provided empirical evidence supporting a stylized fact observed in some economies with regard to the business ownership rates development: after a long-decline in the rate of self-employment and since 1978, self-employment rates appeared to increase in many industrialized countries in such a way that the trend of self-employment rates seemed to show a structural shift in terms of a revival (U-shape) or at least a stabilization (L-shape). For this end, they developed an empirical model in which estimates of different functional forms of the relationship between the ‘equilibrium’ self-employment rate and the GDP per capita allowed them to infer the shape of this relationship for 23 OECD countries. In their empirical model, the equilibrium self-employment rate was obtained using some assumptions about the relationship between self-employment, unemployment, labour incomes and some lags structures. Opposite to this approach, we propose to test the U/L-shape hypothesis using statistical methods: i) first, decomposing the self-employment rate into their two components –i.e. the cyclical and natural components-, and, ii) using a recent econometric approach for detecting the presence of structural breaks (Kejriwal and Perron, 2010). In this way, the U-shape hypothesis is tested for 23 OECD countries, using data on GDP per capita and the natural self-employment rate component (a proxy for the ‘equilibrium’ self-employment rate) over the period 1972-2008. Our results only provide

a partial support for the U-shape hypothesis: for 15 out of 23 countries we find a significantly positive relation between GDP per capita and the natural self-employment rate. Our results suggest that, notwithstanding the rise of self-employment observed in many countries over the last few decades, economies of scale and scope continue to play an important role in many advanced economies.

Keywords: firm size distribution, economic development, Lucas hypothesis

JEL classification: L11, L16, O11

4.1. Introduction

The size distribution of private firms is an important topic for policy makers in modern economies. Economies dominated by large firms require different economic policies compared to economies dominated by small firms (Audretsch, Grilo and Thurik, 2007). Therefore it is interesting to make predictions about future developments of size distributions. In his seminal article, Lucas (1978) predicted that the average size of firms would continue to increase with progressive economic development. This would be the case because higher capital per capita ratios raise the opportunity costs of managing a firm (i.e., wages) relative to the marginal managerial rents. This, in turn, would induce ‘marginal’ managers (entrepreneurs) to become employees, thereby increasing average firm size.¹ As an

¹ In the Lucas model individuals are homogeneous with respect to productivity in paid-employment, but they differ with respect to managerial ability in entrepreneurship (Parker, 2004, p. 55). Individuals freely choose between becoming an entrepreneur with an expected return or becoming a wage-worker earning a fixed wage. Entrepreneurs maximize profits which are an increasing function of

empirical test of his occupational choice model, Lucas (1978) estimated a regression model where average firm size depends on GNP per capita and a time trend. Using time series data for the United States over the period 1900-1970, Lucas found strong support for a positive relation between average firm size and GNP per capita, consistent with his theoretical model.

However, since the publication of Lucas' paper in 1978, self-employment rates have increased in many industrialized countries. For instance, for a set of 23 OECD countries, the business ownership rate, defined as the number of owner-managers of unincorporated and incorporated businesses (outside agriculture) as a fraction of total labour force, increased from 0.100 in 1972 to 0.107 in 2008 (Van Stel, Cieslik and Hartog, 2010). The developments for individual countries are in some cases even more spectacular. The business ownership rate of the Netherlands increased from 0.078 in 1984 to 0.121 in 2008 while that of Germany rose from 0.068 to 0.097 over the same period. On the other hand, even though the business ownership rate of the United States rose from 0.082 in 1972 to 0.108 in 1990, it fell back to 0.098 in 2008 (Van Stel, Cieslik and Hartog, 2010).

managerial ability. In the solution of the model, the more able entrepreneurs run the largest firms. Concerning the role of capital in determining the distribution of the workforce between wage-workers and entrepreneurs, assuming that the elasticity of factor substitution between capital and labor is less than unity (which is generally found in empirical work, see Hamermesh, 1993), entrepreneurs benefit less from an increase in capital stock than wage-workers do (Parker, 2004, p. 56). This will cause 'marginal' entrepreneurs (those who are indifferent –in terms of income– between entrepreneurship and paid-employment) to become employees, thereby increasing the wage-earners to entrepreneurs ratio.

These developments in self-employment rates raise the question to what extent the U-shape hypothesis is true at country level. In this paper we estimate Lucas' model, using data on GDP per capita and the employment to self-employment ratio (a proxy for average firm size) for 23 OECD countries over the period 1972-2008. Using modern econometric techniques which allow for the presence of structural breaks, we find considerable support for the Lucas hypothesis: for 15 out of 23 countries we find a significantly positive relation between GDP per capita and the employment to self-employment ratio. Our results suggest that, notwithstanding the rise of self-employment observed in many countries over the last few decades, economies of scale and scope continue to play an important role in many advanced economies.

The paper is organized as follows. In Section 2 we discuss the relation between average firm size and economic development. In Section 3 we describe the regression model and the data base used to estimate the model. In Section 4 we describe the econometric approach, while Section 5 presents the estimation results. Finally, Section 6 concludes.

4. 2. The relation between average firm size and economic development

In recent decades the relation between average firm size and economic development in advanced economies has been subject to change. Throughout the first three quarters of the 20th century, economic development meant an ever increasing importance of economies of scale and scope in areas such as production, distribution and management (Chandler,

1990). Also, the switch in production from agriculture to manufacturing and increasing wage levels influencing occupational choice implied a positive relation between economic development and average firm size (Wennekers et al., 2010).

However, from the 1970s onwards, self-employment levels started to increase in many advanced economies, first in the United States. Also, in terms of firm size distributions, after a long decline the employment share of SMEs started to increase in several OECD countries (Loveman and Sengenberger, 1991). Driving forces of this shift towards smallness include “the rapidly growing services sector with its smaller scale and lower entry barriers, an increasing differentiation of consumer preferences, declining transactions costs, and a trend in occupational preferences toward more autonomy and self-realization. Additionally, globalization in concert with the spread of ICT (information and communication technologies) enables solo entrepreneurs and small firms to reap the fruits of scale economies through loosely organized networks. And last but not least new technologies create opportunities for new technology-based business startups.” (Wennekers et al., 2010, p. 169).

The developments described above raise the question to what extent the Lucas hypothesis, stating that average firm size and economic development are positively related, is still true in modern times. Indeed, based on an analysis of industry data on average firm size for a set of European countries in the early 1990s, Kumar, Rajan and Zingales (2001) report that “once we correct for institutional development, there is little evidence that richer countries or countries with higher average human capital have larger firms”. Nevertheless, even though the developments described in the pre-

vious paragraph seem to point at a negative relation between average firm size and economic development in recent decades (thus contradicting the Lucas hypothesis), several empirical studies using data over a recent and long period of time do not support such a negative relation. First, using self-employment data in OECD countries, Carree et al. (2002, 2007) investigate whether the relation between economic development and self-employment rates is U-shaped (implying a positive relation from a certain level of per capita income onwards) or L-shaped (implying a continuing decline of self-employment but at an ever decreasing pace towards an asymptote).² In both studies the difference in model fit between the two specifications was not statistically significant. Second, Van Praag and Van Stel (2010) find evidence for a negative relation between education levels and the ‘optimal’ self-employment rate in OECD countries. Assuming that education levels increase with economic development, their finding is consistent with the Lucas model where the more able entrepreneurs run the largest firms (Parker, 2004, p. 55). Third, using data on occupation categories for 115 countries worldwide, Eeckhout and Jovanovic (2010) find a positive relation between GDP per capita and the ‘proportion of managerial jobs’ (where a higher proportion indicates a higher presence of large firms).

² Since in general the self-employment rate and average firm size are negatively related, a U-shape in terms of the relation between self-employment and economic development may be thought of as an inverse U-shape between average firm size and economic development, implying a decreasing average firm size for the most highly developed economies (contradicting the Lucas hypothesis). On the other hand, an L-shaped relation between self-employment and economic development implies a positive relation between average firm size and economic development (consistent with Lucas) but with an elasticity decreasing in value.

In the Lucas model, average firm size is explained by GDP per capita and a time trend. In order to judge the validity of the Lucas hypothesis in present times, an important question is which developments associated with the recent rise in self-employment are related with economic development (GDP per capita), and which are related with global trends benefiting business activity by small firms. One may argue that many of the driving forces of the rise in self-employment are not necessarily or at least not only related to economic development. For instance, the recent trends of globalization and the diffusion of ICT have created many business opportunities as the scale necessary to survive was reduced in many sectors of economy (Jensen, 1993). But these trends are not necessarily related to economic development as these business opportunities arose in high and low developed countries. Also, in many countries recent government policies, including privatization efforts, have been supportive of economic activity by new and small enterprises (Wennekers et al., 2010).

For many of the developments influencing self-employment and average firm size it is hard to disentangle to what extent they are related to economic development and to what extent they are related with general time trends influencing small business activity. However, as far as increasing economies of scale and scope are concerned, the main driver of average firm size in Lucas' model, this phenomenon is clearly related to economic development. Increasing GDP per capita implies a higher capital per capita ratio, which implies more opportunities for exploiting economies of scale and scope, consistent with the Lucas model. Hence, it seems safe to assume that the phenomenon of increasing economies of scale and scope influencing average firm size is captured by GDP per capita and not by general time trends.

For simplicity, we will assume in this paper that GDP per capita captures economies of scale and scope (i.e. we will consider GDP per capita to be a proxy for per capita capital, as in Lucas, 1978), whereas the time trend captures all other developments influencing firm size in modern economies in modern times. Among other things, these developments include technological developments favoring small business activity and a recent trend in occupational preferences toward more autonomy and self-realization. As will be explained in the next section, in our empirical work we allow the relative importance of GDP per capita and general time trends influencing average firm size, to change over time.

4.3. Model and data

4.3.1 Model

In Lucas' theoretical occupational choice model the ratio of employees to managers is an increasing function of the marginal entrepreneurial ability rate, which, in turn, is an increasing function of the capital-labor ratio. Hence, Lucas' theoretical model predicted a positive relation between the ratio of employees to managers and the capital-labor ratio. In his empirical model Lucas used average firm size (the ratio of workers to the number of firms) as a proxy of the ratio of employees to managers and GNP per capita as a proxy of the capital-labor ratio. He also included a time trend in this model. Using U.S. data for the period 1900-1970, Lucas indeed found empirical support for a positive relation between average firm size and GNP per capita.

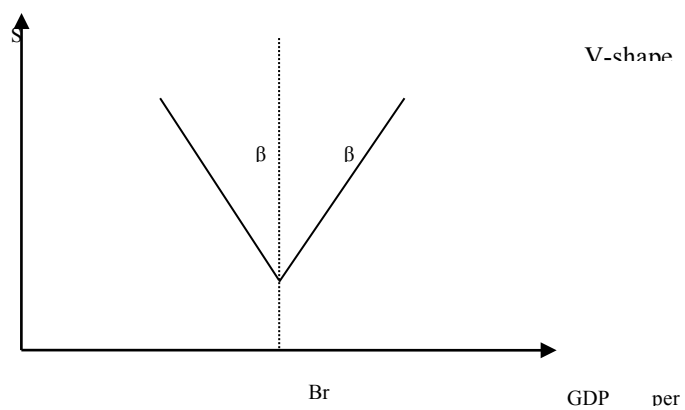
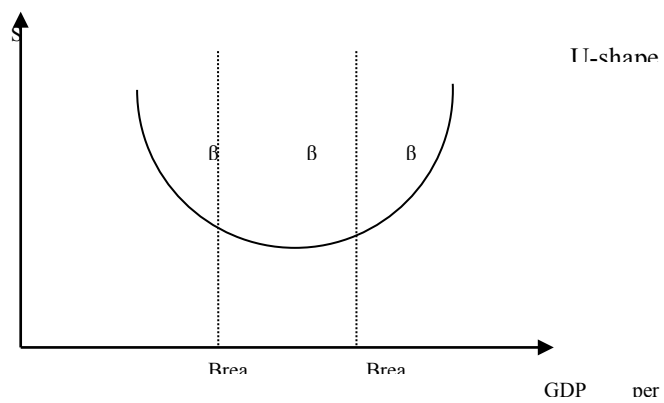
In this paper we will estimate the following model, which closely follows Carree et al. (2002, 2007):

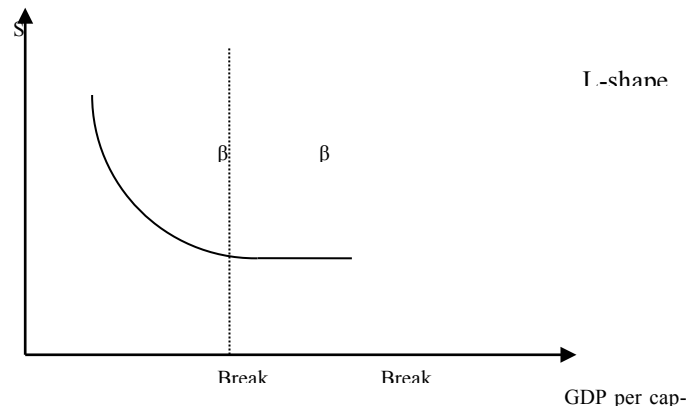
$$S_t^* = \alpha + \beta \ln y_t + \varepsilon_t \quad (1)$$

where, S_t is the number of self-employed in year t , and y_t is real GDP per capita.³ It is important to note that Lucas used the term ‘manager’ in a loose way. Apart from the resources under control of the manager, he actually equated a manager with a firm (Lucas, 1978, p. 510). Hence, it may be argued that the ‘manager’ in the Lucas model is actually a self-employed individual. Given the occupational choice character of Lucas’ theoretical model, we therefore use the ratio of workers to self-employed as dependent variable in the model, instead of the ratio of workers to firms.

Our parameter of interest is β and our objective is to estimate, for each country in our data base, and for each subperiod –obtained by means of the statistical detection of structural breaks– the elasticity between the ‘equilibrium’ self-employment rate and GDP per capita. Interestingly, if empirical estimates report a positive elasticity in the first regime, a negative elasticity in the second and again a positive estimate in the last one a U-shaped relationship would be present. By contrast, a sequence consisting of negative-positive-negative estimates would imply that business ownership will decrease structurally.

³ Following the usual convention in macro-economic studies of self-employment, we exclude the agricultural sector from the analysis. Agriculture is structurally different from the rest of the economy in that self-employment is the natural labor force status in this sector. Also, the sector is characterized by heavy subsidies and a relatively high proportion of unpaid family workers (Parker and Robson, 2004).





followed by a positive one and followed by a negative one for each subperiod report a sequence positive elasticity and given that capital per capita tends to grow over time, a prediction emerges: average firm size will inexorably increase over time (Parker, 2004, p. 56).

4.3. Data

In our empirical analysis we use data of 23 OECD countries including the fifteen countries of the EU-15, Australia, Canada, Iceland, Japan, New Zealand, Norway, Switzerland and the United States for the period 1972-2007.

Self-employment is defined as the total number of unincorporated and incorporated self-employed outside the agriculture, hunting, forestry and fishing industries, who carry out self-employment as their primary employment activity, see Van Stel (2005, p. 108). Unpaid family workers are excluded. These data are taken from EIM's COMPENDIA data base

(version 2008.1).⁴ In this data base, self-employment numbers as published in OECD Labour Force Statistics are corrected for measurement differences across countries and over time and thus harmonized.⁵

The other variable, real GDP per capita is expressed in purchasing power parities per U.S. \$ in 2000 prices. Data on real GDP are taken from OECD National Accounts while data on population are taken from OECD Labour Force Statistics.

4.4. Empirical framework

As mentioned earlier, our aim is to establish, for each country in our data base, the long-run relationship between (the log of) self-employment and (the log of) GDP per capita. Once having checked that these two variables are non-stationary, we will estimate linear cointegration relations. However, since we are considering a long period of time (1972-2007), it is possible that the relationship between the two variables changes over time, i.e., it is possible that estimation of linear cointegration relations yields spurious inference results due to the presence of one or more structural breaks in the relation (Perron, 1997). In this article we employ one of the most up-to-date econometric techniques regarding the detection of structural breaks. In particular we use the tests developed by Gregory and Hansen (1996a, 1996b) and Hatemi-J (2008) to test for the presence of one or two structural breaks in the cointegration relation. After having established

⁴ COMPENDIA is an acronym for COMParative ENtrepreneurship Data for International Analysis. See <http://www.ondernemerschap.nl> for the data and Van Stel (2005) for a justification of the harmonization methods. This database has been used and acknowledged widely (see, among other studies, Armour and Cumming, 2008, Carree et al., 2002, 2007, Davis, 2008 (p. 54), Koellinger and Thurik, 2009, Nyström, 2008, and Parker, Congregado and Golpe, 2011).

⁵ Data taken directly from the OECD Labour Force Statistics suffer from a lack of comparability across countries and over time. In particular, owner-managers of incorporated businesses (OMIBs) are counted as self-employed in some countries, and as employees in other countries. Also, the raw OECD data suffer from many trend breaks relating to changes in self-employment definitions (Van Stel, 2005).

these breaks, we estimate linear models for each country and each regime (subperiod) associated with the structural breaks (if present). This allows us to analyse, for each country, if and how the relation between the employment/self-employment ratio and GDP per capita changes over time. In the remainder of this section we describe our methodological approach to test for structural breaks. All test statistics discussed in the present section are available upon request from the authors.

4.1 Testing for unit roots

Since estimation of a linear cointegration model requires the series to be nonstationary, we start with testing for a unit root in self-employment and GDP per capita, using the full data period 1972-2007 for each country. We apply the class of unit root tests developed by Ng and Perron (2001) which solve several statistical problems associated with more ‘conventional’ unit root tests.⁶ Concerning self-employment, the null hypothesis of nonstationarity is not rejected for all 23 countries in our data base, i.e. the evidence is consistent with a unit root. For GDP per capita the results of the Ng-Perron tests are consistent with a unit root for all 23 countries.

⁶ In general, the majority of the conventional unit root tests such as the Dickey-Fuller tests and the Phillips-Perron tests suffers from three problems. First, many tests have low power when the root of the autoregressive polynomial is close to but less than unity (DeJong et al., 1992). Second, most tests suffer from severe size distortions when the moving-average polynomial of the first-differenced series has a large negative autoregressive root (Schwert, 1989; Perron and Ng, 1996). Third, the implementation of unit root tests often requires the selection of an autoregressive truncation lag k ; however, as discussed in Ng and Perron (1995) there is a strong association between k and the severity of size distortions and/or the extent of power loss. Ng and Perron (2001) solved these problems and we refer to their article for further details.

4.2 Testing for linear cointegration

In the previous subsection we showed that, when considering the full data period 1972-2008, all countries in our data sample have a unit root in at least one of the two series. Hence, for all countries the unit root requirement for the existence of a (linear) cointegration relation is met. In this subsection we test, for each country, whether linear cointegration between GDP per capita and the employment/self-employment ratio indeed takes place. For this purpose we apply the well-known Johansen maximal eigenvalue and trace tests (Johansen, 1988, 1995). When applying these tests it can be concluded that for 21 out of 23 countries, there is evidence for a cointegration relation.⁷ For Ireland and Norway, the Johansen tests indicate there is no (linear) cointegration.

4.3 Testing for non-linear cointegration

Testing for structural breaks

Accounting for parameter shifts is crucial in cointegration analysis. Our data covers thirty six years, during which time the relationship have probably changed. Therefore, as we argued before, it is important to account for structural breaks in the cointegration relationship. In this subsection we test for the existence of non-linear cointegration between GDP per capita and self-employment. First, we use Kejriwal and Perron (2010) ap-

⁷ For 16 countries the null hypothesis of zero cointegration relations is rejected at the 5% significance level. In addition, for Denmark, Sweden, Switzerland, the UK and Canada the null hypothesis is rejected at the 10% level only.

proach to test for structural breaks in the cointegration relationship. The Gregory and Hansen tests are based on the study of the residuals from the long-run regression model of Engle and Granger (1987), in which a break in the model is included with an a priori unknown date, which is endogenously determined by the data. There are different alternatives to account for structural breaks in the standard Engle and Granger's cointegration model, although the null hypothesis in all these alternatives is that the series are not cointegrated. The alternative hypothesis in the Gregory-Hansen approach is that the series are cointegrated with one structural break. We refer to Gregory and Hansen (1996a, 1996b) for further details. The Gregory and Hansen tests allow both to detect the existence of cointegration under the presence of time discontinuities in the parameters of the long-run cointegration relation and to efficiently test for the breakpoint date (T_b). For five countries in our data base, the Gregory and Hansen (1996a, 1996b) test statistics point at the existence of a non-linear cointegration relation with one structural break. These are Denmark, Ireland, Italy, The Netherlands and Norway.

Second, we use the Hatemi-J (2008) approach to test for two structural breaks in the cointegration relationship. In this approach the timing of each shift is unknown a priori and is determined endogenously. These new tests build on the Gregory and Hansen tests for cointegration in the presence of one shift. The Hatemi-J test uses test statistics and critical values which are adjusted for the possibility of two regime shifts. In our application of the Hatemi-J test we find evidence of a non-linear cointegration relation with two structural breaks for Austria, Denmark, Finland, Greece, Ireland, Italy and the United States. Table 1 summarizes the results of our cointegration tests.

5. Results

The parameter estimates of Model (1) are presented in Table 2. We dispose of data for 23 OECD countries over the period 1972-2008. For each country we estimated the model for one, two or three subperiods, depending on the results of the tests for structural breaks (see Table 1). The years corresponding to the structural breaks are also reported in Table 2. In case no break was detected, the model was estimated for the full period 1972-2008.

Table 2 shows that for some countries the parameters change considerably over time, illustrating the importance of allowing for structural breaks in the relation. As we are mainly interested in the relation at the present time, we will focus on the estimates for the most recent subperiod for each country (including the full sample estimates for countries without a break in the relation).

Considering parameter β describing the relation between GDP per capita and the employment/self-employment ratio, the estimation results are summarized in Table 3. We see that 15 out of 23 countries have a positive sign, supporting the Lucas hypothesis. Although most of these countries also show a significant negative time trend, pointing at general developments favoring small business activity (see Section 2), the positive sign for GDP per capita suggests that exploitation of scale advantages is still important in the present time.

Five countries show a non-significant relation between GDP per capita and the employment/self-employment ratio while Italy, Portugal and Japan show a negative relation. For these countries average firm size decreases with progressive economic development. It is hard to explain why

these three countries are different, but it seems to be the case that small firm production is an inherent part of the organization of industry. Specific specialization patterns in these countries could play a role (e.g. the fashion industry in Italy). Also, the institutional environment in Italy, Portugal and Japan seems to favor small firm production more than in other OECD countries (see, e.g., Okamuro, Van Stel and Verheul, 2011, p. 40, Figure D3). Nevertheless, the positive estimates for parameter δ in these three countries point at some autonomous developments towards larger scale production as well.

6. Conclusions

In his seminal article, Lucas (1978) predicted that the average size of firms would continue to increase with progressive economic development. However, since 1978, self-employment rates have increased in many industrialized countries. This raises the question to what extent the Lucas hypothesis is still true in modern times. In this paper we test the Lucas hypothesis for 23 OECD countries, using data on GDP per capita and the employment to self-employment ratio over the period 1972-2008. Using modern econometric techniques which allow for the presence of structural breaks, we find considerable support for the Lucas hypothesis: for 15 out of 23 countries we find a significantly positive relation between GDP per capita and the employment to self-employment ratio. Our results suggest that, notwithstanding the rise of self-employment observed in many countries over the last few decades, economies of scale and scope continue to play an important role in many advanced economies.

Several possible explanations can be brought forward for the continued importance of scale economies in modern times. First, although entrepreneurial activity in the form of new-firm start-ups continues to show a positive relation with GDP per capita for the most highly developed countries (Kelley, Bosma and Amorós, 2011), it is also well known that many entries fail. Indeed, the correlation between entry and exit rates tends to be very high (Geroski, 1995). The high entry rates in highly developed economies could imply a more fierce selection process by the market where the surviving entrepreneurs are of a higher quality, compared to a market with a lower number of competitors (Jovanovic, 1982). These higher-quality entrepreneurs, in turn, may run larger firms as their span of control is bigger, consistent with the Lucas model. Hence, a higher number of new firms may go together with a lower number of incumbent firms which are on average larger.

Second, the ICT revolution not only favors small firm activity through lower transaction costs, it also lowers internal communication and coordination costs within larger firms. In terms of industry life cycle models, it may be the case that many new industries that emerged during the ICT revolution are now reaching more mature stages of the industry life cycle where dominant designs are improved upon by way of scale economies and process innovations (Wennekers et al., 2010, p. 222).

A limitation of our study is that average firm size is computed at the aggregate level of countries. Since different sectors may have different optimal firm sizes (minimum efficient scales), different sector structures between countries may explain part of the country differences found in the relation between economic development and average firm size. In particu-

lar, specific specialization patterns in certain countries may have influenced the results.

References

Armour, J. and Cumming, D. (2008), Bankruptcy Law and Entrepreneurship, *American Law and Economics Review* 10(2), 303-350.

Audretsch, D.B., Grilo, I., and Thurik, A.R. (eds.) (2007), *Handbook of Research on Entrepreneurship Policy*, Cheltenham UK: Edward Elgar Publishing.

Carree, M., Van Stel, A., Thurik, R., and Wennekers, S. (2002), Economic Development and Business Ownership: An Analysis Using Data of 23 OECD Countries in the Period 1976-1996, *Small Business Economics* 19(3), 271-290.

Carree, M., Van Stel, A., Thurik, R., and Wennekers, S. (2007), The Relationship Between Economic Development and Business Ownership Revisited, *Entrepreneurship and Regional Development* 19(3), 281-291.

Chandler, A.D. Jr. (1990), *Scale and Scope: The Dynamics of Industrial Capitalism*, Cambridge, MA: Harvard University Press.

Davis, T. (2008), Understanding Entrepreneurship: Developing Indicators for International Comparisons and Assessments, In: Congregado, E. (ed.), *Measuring Entrepreneurship; Building a Statistical System*, New York: Springer, 39-63.

DeJong, D.N., Nankervis, J.C., Savin, N.E., and Whiteman, C.H. (1992), Integration versus Trend Stationarity in Time Series, *Econometrica* 60, 423-433.

Eeckhout, J. and Jovanovic, B. (2010), *Occupational Choice and Development*, NBER Working Paper.

Engle, R. and Granger, C. (1987), Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica* 55, 251-276.

Geroski, P. (1995), What do we Know about Entry?, *International Journal of Industrial Organization* 13, 421-441.

Gregory A.W. and Hansen, B.E. (1996a), Residual-based Tests for Cointegration in Models with Regime Shifts, *Journal of Econometrics* 70, 99-126.

Gregory A.W. and Hansen, B.E. (1996b), Tests for Cointegration in Models with Regime and Trend Shifts, *Oxford Bulletin of Economics and Statistics* 58, 555-560.

Hamermesh, D.S. (1993), *Labour Demand*, Princeton, NJ: Princeton University Press.

Hatemi-J, A. (2008), Tests for Cointegration with Two Unknown Regime Shifts with an Application to Financial Market Integration, *Empirical Economics* 35, 497-505.

Jensen, M.C. (1993), The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems, *Journal of Finance* 48, 831-880.

Johansen, S. (1988), *Statistical Analysis of Cointegration Vectors*, *Journal of Economic Dynamics and Control*, 12, 231-54.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, UK: Oxford University Press.

Jovanovic, B. (1982), Selection and Evolution in Industry, *Econometrica* 50, 649-670.

Kelley, D.J., N. Bosma and J.E. Amorós (2011), *Global Entrepreneurship Monitor, 2010 Global Report*, Wellesley, MA: Babson College and Santiago, Chile: Universidad del Desarrollo.

Koellinger, P. and Thurik, R. (2009), *Entrepreneurship and the Business Cycle*, Tinbergen Institute Discussion Paper TI 2009-032/3, Erasmus University Rotterdam.

Kumar K.B., Rajan, R.G., and Zingales, L. (2001), What Determines Firm Size?, University of Chicago, Graduate School of Business, The Center for Research in Security Prices, Working Paper No. 496.

Kejriwal, M. and Perron, P. (2010), Testing for Multiple Structural Changes in Cointegrated Regression Models, *Journal of Business & Economic Statistics* 28, 503-522.

Loveman, G. and W. Sengenberger (1991), The Re-Emergence of Small-Scale Production; an International Comparison, *Small Business Economics* 3, 1-37.

Lucas, R.E. Jr. (1978), On the Size Distribution of Business Firms, *Bell Journal of Economics* 9, 508-523.

Ng, S. and Perron, P. (1995), Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag, *Journal of the American Statistical Association* 90, 268-281.

Ng, S. and Perron, P. (2001), Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, *Econometrica* 69, 1519-1554.

Nyström, K. (2008). The Institutions of Economic Freedom and Entrepreneurship: Evidence from Panel Data, *Public Choice*, 136(3-4), 269-282.

Okamuro, H., Van Stel, A., and Verheul, I. (2011), Understanding the Drivers of an 'Entrepreneurial' Economy: Lessons from Japan and the Netherlands, EIM Research Report H201102, Zoetermeer, NL: EIM Business and Policy Research.

Parker, S.C. (2004), *The Economics of Self-employment and Entrepreneurship*, Cambridge, UK: Cambridge University Press.

Parker, S.C., Congregado, E., and Golpe, A.A. (2011), Testing for Hysteresis in Entrepreneurship in 23 OECD Countries, *Applied Economics Letters*, forthcoming.

Parker, S.C. and Robson, M.T. (2004), Explaining International Variations in Self-employment: Evidence from a Panel of OECD Countries, *Southern Economic Journal* 71, 287-301.

Perron, P. (1997), Further Evidence on Breaking Trend Functions in Macroeconomic Variables, *Journal of Econometrics*, 80, 355-385.

Perron, P. and Ng, S. (1996), Useful Modifications to some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties, *Review of Economic Studies*, 63, 435-463.

Schwert, G.W. (1989), Tests for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics* 7, 147-159.

Van Praag, M. and A. van Stel (2010), The More Business Owners The Merrier? The Role of Tertiary Education, EIM Research Report H201010, Zoetermeer, NL: EIM Business and Policy Research.

Van Stel, A. (2005), COMPENDIA: Harmonizing Business Ownership Data Across Countries and Over Time, *International Entrepreneurship and Management Journal* 1(1), 105-123.

Van Stel, A., Cieslik, J., and Hartog, C. (2010), Measuring Business Ownership Across Countries and Over Time: Extending the COMPENDIA Data Base, EIM Research Report H201019, Zoetermeer, NL: EIM Business and Policy Research.

Wennekers, S., Van Stel, A., Carree, M., and Thurik, R. (2010), The Relationship Between Entrepreneurship and Economic Development: Is It U-shaped?, *Foundations and Trends in Entrepreneurship* 6(3), 167-237.

Table 1: Summary of linear and non-linear cointegration test results

Linear cointegration (no structural break)	Non-linear cointegration (one break)	Non-linear cointegration (two breaks)
Belgium	The Netherlands	Austria
France	Norway	Denmark
Germany	United States ⁸	Finland
Luxembourg		Greece
Portugal		Ireland
Spain		Italy
Sweden		
United Kingdom		
Iceland		
Switzerland		
Japan		
Canada		
Australia		
New Zealand		

Note: The results in the table refer to the existence of cointegration between GDP per capita and the employment/self-employment ratio.

⁸ For the United States, the Hatemi-J tests indicated two breaks but since they both occurred in the same year (1977), we consider the cointegration relation to have one break.

Table 2: The estimated values of the parameters

Country	α	β	${}_2R$	α	β	${}_2R$	\hat{T}_b	α	β	${}_2R$	\hat{T}_b	α	β	${}_2R$
Linear relationship (full sample)				Linear estimates for different regimes										
Austria	- 0.040 (0.06 2)	0.012* (0.006)	0 .094	0.862 *** (0.23 7)	- 0.083*** (0.025)	0 .685	1 978	- 0.092 (0.07 4)	0.017 ** (0.00 8)	0 .230	1 996	- 0.702*** (0.04 6)	0.079 *** (0.00 5)	0 .971
Belgium	- 0.099* (0.052)	0.022*** (0.005)	0 .333	0.648 ** (0.168)	- 0.058** (0.018)	0 .721	1 977	- 0.593*** (0.033)	0.072 *** (0.00 3)	0 .967	1 994	1.020 *** (0.129)	- 0.092*** (0.013)	0 .817
Denmark	0.331 *** (0.046)	-0.027*** (0.005)	0 .489	0.762 *** (0.15 8)	- 0.072*** (0.01 7)	0 .674	1 982	- 0.020*** (0.05 0)	0.009 * (0.00 5)	0 .107				
Finland	- 0.249*** (0.027)	0.033*** (0.003)	0 .803	0.090 (0.11 2)	- 0.003 (0.01 2)	0 .008	1 981	- 0.763*** (0.136)	0.087 *** (0.014)	0 .790	1 993	- 0.062* (0.032)	0.015 *** (0.003)	0 .613
France	0.552 *** (0.02 2)	-0.047*** (0.002)	0 .926	0.487 *** (0.036)	0.040*** (0.00 4)	0 .877	1 989	1.732 *** (0.530)	0.168 *** (0.054)	0 .657	1 996	0.186 (0.15 3)	0.010 (0.01 5)	0 .048
Germany	- 0.379*** (0.109)	0.047*** (0.011)	0 .340	0.385 *** (0.054)	0.033 (0.00 6)	0 .738	1 985	- 0.209*** (0.026)	0.029 *** (0.003)	0 .976	1 990	- 0.907*** (0.086)	0.102 *** (0.009)	0 .898
Greece	- 0.162***	0.039*** (0.006)	0 .520	- 0.420***	0.068 ***	0 .670	1 987	- 1.791	0.223 (0.20 .166)	0 .166	1 995	0.077 (0.09	0.013 (0.01	0 .141

	(0.05 8)		(0.112)	(0.013)			(1.82 0)	4)			0)	0)		
Ireland	- 0.161*** (0.023)	0.028*** (0.002)	0 .790	- 0.155** (0.06 8)	0.027 *** (0.00 8)	0 .503	1 985	- 0.449*** (0.06 7)	0.059 *** (0.007)	0 .896	1 995	0.075 * (0.04 1)	0.004 (0.00 4)	0 .078
Italy	- 0.640*** (0.033)	0.086*** (0.003)	0 .949	- 0.325** (0.13 2)	0.052 *** (0.01 4)	0 .582	1 983	- 0.535*** (0.089)	0.076 *** (0.009)	0 .918	1 991	- 0.269*** (0.074)	0.049 *** (0.008)	0 .748
Luxem- bourg	0.428 *** (0.017)	-0.036*** (0.002)	0 .928	0.859 * (0.33 9)	- 0.081* (0.03 6)	0 .558	1 977	0.510 *** (0.02 6)	- 0.044*** (0.003)	0 .960	1 991	0.399 *** (0.041)	- 0.033*** (0.004)	0 .829
Nether- lands	- 0.318*** (0.065)	0.042*** (0.007)	0 .536	0.952 *** (0.108)	- 0.092*** (0.011)	0 .928	1 978	0.138 (0.10 1)	- 0.001 (0.01 1)	0 .036	1 989	- 0.788*** (0.07 6)	0.090 *** (0.008)	0 .894
Portugal	- 0.268*** (0.067)	0.045*** (0.007)	0 .516	0.022 (0.20 7)	0.011 (0.02 4)	0 .044	1 978	- 0.376** (0.15 6)	0.056 *** (0.017)	0 .485	1 991	1.451 *** (0.19 2)	- 0.139*** (0.021)	0 .766
Spain	- 0.213*** (0.031)	0.036*** (0.003)	0 .784	- 0.277*** (0.072)	0.043 *** (0.00 8)	0 .606	1 992	0.045 (0.06 2)	0.009 (0.00 7)	0 .124				
Sweden	- 0.183*** (0.039)	0.027*** (0.004)	0 .567	0.072 (0.09 1)	- 0.000 (0.01 0)	0 .000	1 985	- 0.275 (0.33 6)	0.035 (0.03 5)	0 .340	1 989	- 0.181** (0.07 5)	0.026 *** (0.00 8)	0 .428
U.K.	- 0.455***	0.057*** (0.006)	0 .727	0.685 (0.44)	- 0.065	0 .321	1 977	- 1.198***	0.135 ***	0 .920	1 992	0.146 ***	- 0.004	0 .012

	(0.058)			2)	(0.048)			(0.105)	(0.011)			(0.090)	(0.009)	
Iceland	- 0.303*** (0.074)	0.041*** (0.008)	0 .460	0.550*** (0.092)	- 0.049*** (0.010)	0 .723	1 983	- 0.680*** (0.529)	0.079 (0.054)	0 .260	1 991	0.023 (0.078)	0.009 (0.009)	0 .082
Norway	0.325*** (0.030)	-0.025*** (0.003)	0 .665	0.331*** (0.041)	- 0.026*** (0.004)	0 .729	1 986	0.443*** (0.079)	- 0.037*** (0.008)	0 .728	1 996	1.142*** (0.362)	0.120*** (0.036)	0 .555
Switzerland	- 0.313*** (0.087)	0.039*** (0.009)	0 .360	- 0.126** (0.052)	0.019*** (0.005)	0 .418	1 992	- 2.549* (0.997)	0.264* (0.100)	0 .634	1 998	1.521* (0.649)	- 0.145* (0.065)	0 .416
U.S.	- 0.143** (0.066)	0.024*** (0.0007)	0 .286	0.062 (0.243)	0.002 (0.025)	0 .003	1 975	- 0.844*** (0.192)	0.096*** (0.019)	0 .688	1 988	0.393*** (0.027)	- 0.029*** (0.003)	0 .872
Japan	0.642*** (0.048)	-0.055*** (0.005)	0 .782	0.080 (0.048)	0.005 (0.005)	0 .075	1 985	0.989*** (0.102)	- 0.090*** (0.010)	0 .882	1 997	1.140*** (0.177)	- 0.106*** (0.018)	0 .814
Canada	- 0.836*** (0.090)	0.096*** (0.009)	0 .762	- 0.755*** (0.129)	0.087*** (0.013)	0 .765	1 986	- 0.910 (1.262)	0.104 (0.129)	0 .098	1 994	0.609** (0.209)	- 0.048** (0.021)	0 .325
Australia	0.004 (0.069)	0.016** (0.007)	0 .126	- 2.911*** (0.381)	0.324*** (0.040)	0 .915	1 979	0.308** (0.131)	- 0.016 (0.014)	0 .105	1 992	0.485*** (0.080)	- 0.033*** (0.008)	0 .565
New Zealand	- 0.955*** (0.11)	0.113*** (0.012)	0 .737	- 0.751*** (0.24)	0.090*** (0.026)	0 .465	1 987	- 1.260*** (0.31)	0.146*** (0.033)	0 .708	1 997	0.527* (0.259)	- 0.040 (0.02)	0 .220

0) 5) 5) 7)

Notes: Standard errors are between parentheses. *, ** and *** denotes significance at 10%, 5% and 1% significance level respectively. For full sample estimates, the period covered is 1972-2008. For countries with one structural break the sample periods are 1972- $\hat{T}_{b,1}$ and $\hat{T}_{b,1}$ -2008. For countries with two breaks the sample periods are 1972- $\hat{T}_{b,1}$, $\hat{T}_{b,1}$ - $\hat{T}_{b,2}$, and $\hat{T}_{b,2}$ -2008.

Kerjiwal-Perron tests of multiple structural changes in the long-run relationship between business ownership and GDP per capita, (1972-2007)

$$y_i = \{S_i\} \quad z_i = \{1, \ln GDP\} \quad q=2 \quad p=0 \quad h=0.25 \quad m=2$$

	Austria	Belgium	Denmark	Finland	France	Germany	Greece	Ireland	Italy	Luxemb- rurg	The Neth- erlands	Portugal	Spain	Sweden	United Kingdom	Iceland	Norway	Switzer- land	USA	Japan	Canada	Australia	New Zea- land
<i>UD- max</i>	.790	.418	5.37 ²	2.78 ²	.295	.358	.146	4.05 ²	1.94 ¹	.618	.896	8.96 ²	0.02 ³	.878	.727	1.94 ¹	.525	43.46 ¹	.606	.699	0.65 ¹	1.71 ¹	3.07 ¹
<i>SupF</i>	.466	.697	.078	.339	.714	.443	0.19 ¹	1.66 ¹	.463	.895	2.74 ¹	.859	4.47 ¹	.943	.473	.463	2.50 ¹	2.421	.529	.324	.688	.136	3.07 ¹
<i>SupF</i>	.790	.418	5.37 ²	2.78 ²	.295	.358	0.19 ¹	4.05 ²	1.94 ¹	.618	2.74 ¹	8.96 ²	0.02 ³	.878	.727	1.94 ¹	2.50 ¹	43.46 ¹	.529	.699	0.65 ¹	1.71 ¹	.406
N° of breaks:																							
BIC	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2
LW	2	2	1	2	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	2	1	2
Z																							

SP	0	0	1	1	0	0	0	1	0	0	0	2	1	0	0	0	0	1	0	0	0	0	1
----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Estimated break points:

$\hat{T}_{b,1}$	1 978	1 977	1 982	1 981	1 989	1 985	1 987	1 985	1 983	1 977	1 978	1 978	1 992	1 985	1 977	1 983	1 986	1 992	1 975	1 985	1 986	1 979	1 987
$\hat{T}_{b,2}$	1 996	1 994		1 993	1 996	1 990	1 995	1 995	1 991	1 991	1 989	1 991		1 989	1 992	1 991	1 996	1 998	1 988	1 997	1 994	1 992	1 997

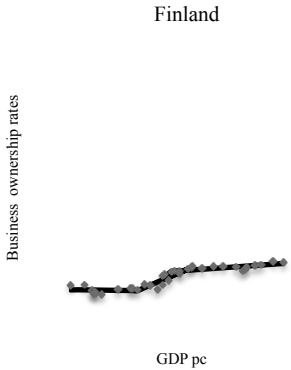
Notes: y, z, q, p, h , and m denote the dependent variable (business ownership rate), the explanatory variable allowed to change (GDP per capita), the number of regressors (2), the number of corrections included in the variance-covariance matrix, the minimum number of observations in each segment, and the maximum number of breaks, respectively. *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

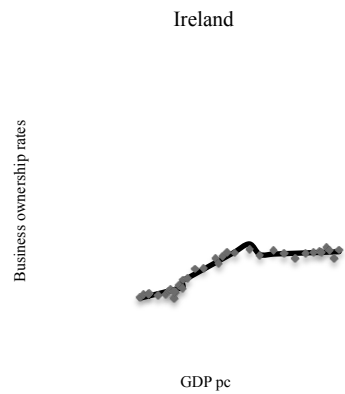
Table 3: Summary of estimation results for parameter β in most recent sub-period for each country

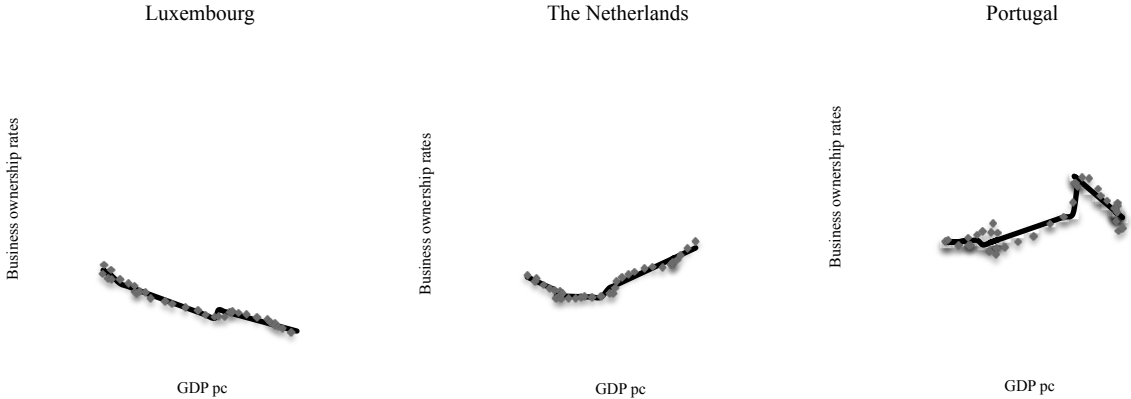
β significantly positive	β not statistically different from zero	β significantly negative
Denmark	Austria	Italy
Finland	Belgium	Portugal
Germany	France	Japan
Greece	Luxembourg	
Ireland	New Zealand	
Netherlands		
Spain		
Sweden		
United Kingdom		
Iceland		
Norway		
Switzerland		
United States		
Canada		
Australia		

Note: The results in the table refer to the estimated relation between GDP per capita and the employment/self-employment ratio.

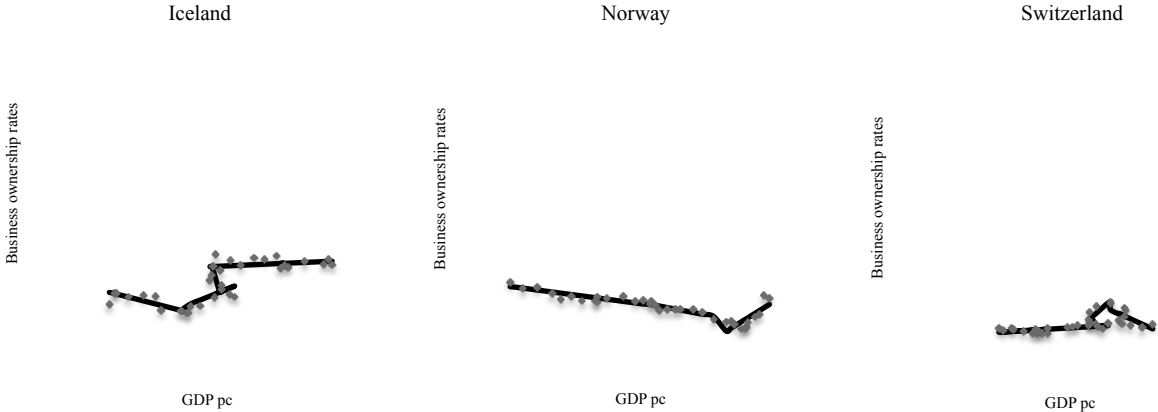


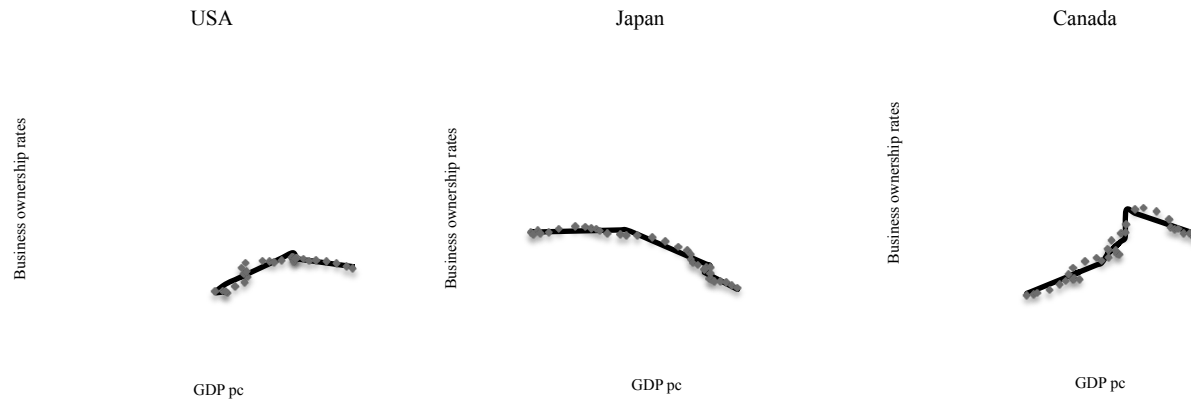


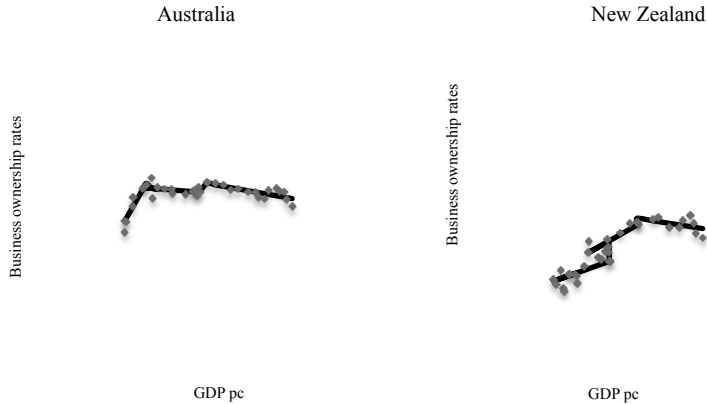












**Parte IV: Aplicaciones a la docencia de
macroeconomía dinámica**

Capítulo 5. Modelización de problemas económicos a través de Sistemas de ecuaciones diferenciales resueltos a través de métodos numéricos

5.1. Introducción

Esta cuarta parte de la tesis presenta dos ensayos que se encuadran en el uso de la programación y de los métodos numéricos en la enseñanza del análisis económico dinámico.

En los dos capítulos que componen esta cuarta parte se presentan sendos ejemplos de modelización de problemas dinámicos mediante sistemas de ecuaciones diferenciales, resueltos a través de métodos numéricos.

5.2. El planteamiento del problema

PROBLEMA: Estudio del comportamiento de una masa bacteriana en un cultivo controlado.

DETALLES: Se va a estudiar la evolución con el tiempo de una masa bacteriana de millones de individuos en un cultivo donde se va a controlar la temperatura, la concentración de nutrientes, la concentración de bacterias y el pH entre otras.

DESARROLLO TEÓRICO:

Para el desarrollo teórico se establecerán ciertas hipótesis basadas en la experiencia y conocimientos previos.

HIPÓTESIS 1: el crecimiento bacteriano puede modelizarse a partir de una ecuación en diferencias finitas. Para comenzar se supone un intervalo infinitesimalmente pequeño de tiempo durante el cual, llamando X a la concentración de bacterias en millones por litro (millones/L):

$$X_{i+1} = X_i + k \cdot \Delta t \cdot X_i$$

Donde el subíndice i se refiere a un estado concreto de la variable para el tiempo t_i , la constante k es la tasa neta de crecimiento por unidad de tiempo e individuo por unidad de volumen, incluyendo la contribución endógena de mortalidad natural, e Δt es el intervalo de tiempo infinitesimal entre ambos estados.

Así pues la ecuación en diferencias quedaría de la siguiente forma:

$$\frac{\Delta X}{\Delta t} = k \cdot X_i$$

Y haciendo límites cuando $\Delta t \rightarrow 0$:

$$\frac{dX}{dt} = k X$$

Este resultado es ya una ecuación diferencial, fácilmente resoluble por el método de variables separadas, resultando:

$$X(t) = X_0 e^{kt}$$

De lo cual se deduce un crecimiento exponencial en el número de individuos funcionales que puede encontrarse por volumen.

HIPÓTESIS 2: la tasa de crecimiento puede considerarse constante bajo ciertas circunstancias controladas durante un cierto tiempo, pero se sabe que ésta depende de factores como la disponibilidad de nutrientes en el sustrato, la temperatura o el pH del medio. Considerando constantes y no influyentes la temperatura y el pH, se va a suponer un comportamiento entre la tasa de crecimiento y la concentración de sustrato medida en miligramos de nutrientes por volumen (mg/L) que sigue la siguiente correlación:

$$k(M) = A M^2 + B M$$

Estudiando la correlación se observa que es una relación cuadrática entre ambas variables. Para encontrar los parámetros A y B se va a suponer que para un cierto valor de M (denominado M_m) se alcanza un máximo para k de valor k_0 .

Se supone también que una vez alcanzado el valor máximo k_0 éste se mantiene constante aunque la concentración de nutrientes en el sustrato sea aún mayor. Por tanto, la correlación quedaría de la siguiente forma:

$$k(M) = \begin{cases} k_0 & \text{si } M > M_m \\ \frac{k_0 M}{M_m^2} \cdot (2M_m - M) & \text{si } 0 \leq M \leq M_m \end{cases}$$

HIPÓTESIS 3: la concentración de nutrientes en el sustrato también variará debido al consumo por el crecimiento bacteriano. Considerando una tasa de consumo Q medido en concentración de nutrientes consumida por unidad de tiempo y concentración de bacterias, la ecuación en diferencias sería:

$$M_{i+1} = M_i - Q \cdot \Delta t \cdot X_i$$

Si además se añade una contribución extra de nutrientes con un flujo constante P_M :

$$M_{i+1} = M_i + P_M \cdot \Delta t - Q \cdot \Delta t \cdot X_i$$

Cuya ecuación en diferencias quedaría de la siguiente forma:

$$\frac{\Delta M}{\Delta t} = P_M - Q \cdot X_i$$

Y como ecuación diferencial:

$$\frac{dM}{dt} = P_M - Q X$$

HIPÓTESIS 4: dado que un aumento exponencial de la concentración de bacterias provocaría una competencia de las mismas por el espacio y los nutrientes, esto producirá una mortalidad añadida a la natural, provocada por las condiciones del estudio que va a suponerse se la siguiente forma: dada una tasa de mortalidad β para la ausencia total de nutrientes, y supuesto un margen crítico C de disponibilidad de nutrientes que se va a considerar con el cociente M/X se puede deducir que la tasa neta de mortalidad R se puede calcular como:

$$R(X, M) = \beta \cdot \frac{C - M/X}{C}$$

Y añadiendo la restricción que para valores del cociente M/X superiores a C el valor de R es nulo, se concluye con:

$$R(X, M) = \begin{cases} 0 & \text{si } \frac{M}{X} > C \\ \beta \cdot \frac{C - M/X}{C} & \text{si } 0 \leq \frac{M}{X} \leq C \end{cases}$$

PLANTEAMIENTO: dadas las hipótesis planteadas, se consideran dos variables principales de control, la concentración de bacterias X y la concentración de nutrientes M , dependientes ambas entre sí y del tiempo, por lo

que el problema se convierte en un sistema de ecuaciones diferenciales en derivadas parciales (SEDP).

Las ecuaciones diferenciales en derivadas parciales a tener en cuenta son:

$$\begin{aligned}\frac{\partial X}{\partial t} &= k X - R X \\ \frac{\partial M}{\partial t} &= P_M - Q X\end{aligned}$$

Los coeficientes k y R son variables y condicionados, y sus expresiones corresponden a:

$$k(M) = \begin{cases} k_0 & \text{si } M > M_m \\ \frac{k_0 M}{M_m^2} \cdot (2M_m - M) & \text{si } 0 \leq M \leq M_m \end{cases}$$

$$R(X, M) = \begin{cases} 0 & \text{si } \frac{M}{X} > C \\ \beta \cdot \frac{C - M/X}{C} & \text{si } 0 \leq \frac{M}{X} \leq C \end{cases}$$

5.3. Resolución del problema

Buscar una solución para este tipo de problemas de forma analítica, es muy difícil, y sólo será posible en casos muy concretos con suposiciones muy extremas. Por tanto, la mejor opción es utilizar algún método numérico para resolver este tipo de SEDP.

El método de Runge-Kutta es posible utilizarlo, adaptándolo al problema de coeficientes variables, lo cual resulta bastante obvio y sencillo.

- MÉTODO RUNGE-KUTTA DE CUARTO ORDEN

El método generalizado para 3 variables puede definirse aplicado a este problema de la siguiente forma:

$$\begin{aligned}t_{i+1} &= t_i + h \\ X_{i+1} &= X_i + h \cdot \varphi_X(t_i, X_i, M_i)\end{aligned}$$

$$M_{i+1} = M_i + h \cdot \varphi_M(t_i, X_i, M_i)$$

Sabiendo que h es el paso de iteración, en este caso el lapso de tiempo mínimo considerado. Hay que tener en cuenta que cuanto más cercano a 0 este valor, más fidedignos son los resultados obtenidos.

En todo problema diferencial, hay que tener en cuenta los valores en la frontera, que para este caso son valores iniciales. Para las tres variables se consideran:

$$t = 0$$

$$X = X_0$$

$$M = M_0$$

En este problema las funciones incremento φ vendrán definidas como:

$$\varphi_X = \frac{x_1 + 2x_2 + 2x_3 + x_4}{6}$$

$$\varphi_M = \frac{m_1 + 2m_2 + 2m_3 + m_4}{6}$$

Y los parámetros intermedios x y m se definen como:

$$x_1 = f(t_i, X_i, M_i)$$

$$x_2 = f\left(t_i + \frac{h}{2}, X_i + \frac{h}{2} \cdot x_1, M_i + \frac{h}{2} \cdot m_1\right)$$

$$x_3 = f\left(t_i + \frac{h}{2}, X_i + \frac{h}{2} \cdot x_2, M_i + \frac{h}{2} \cdot m_2\right)$$

$$x_4 = f(t_i + h, X_i + h \cdot x_3, M_i + h \cdot m_3)$$

$$m_1 = g(t_i, X_i, M_i)$$

$$m_2 = g\left(t_i + \frac{h}{2}, X_i + \frac{h}{2} \cdot x_1, M_i + \frac{h}{2} \cdot m_1\right)$$

$$m_3 = g\left(t_i + \frac{h}{2}, X_i + \frac{h}{2} \cdot x_2, M_i + \frac{h}{2} \cdot m_2\right)$$

$$m_4 = g(t_i + h, X_i + h \cdot x_3, M_i + h \cdot m_3)$$

Y las funciones f y g se deducen de las ecuaciones diferenciales como:

$$f(t_i, X_i, M_i) = k(M_i) \cdot X_i - R(X_i, M_i) \cdot X_i$$

$$g(t_i, X_i, M_i) = P_M - Q \cdot X_i$$

Donde, además:

$$k(M_i) = \begin{cases} k_0 & \text{si } M_i > M_m \\ k_0 \cdot \frac{M_i}{M_m^2} \cdot (2M_m - M_i) & \text{si } 0 \leq M_i \leq M_m \end{cases}$$

$$R(X_i, M_i) = \begin{cases} 0 & \text{si } \frac{M_i}{X_i} > C \\ \beta \cdot \frac{C - M_i/X_i}{C} & \text{si } 0 \leq \frac{M_i}{X_i} \leq C \end{cases}$$

5.4. Implementación y programación en hoja de cálculo

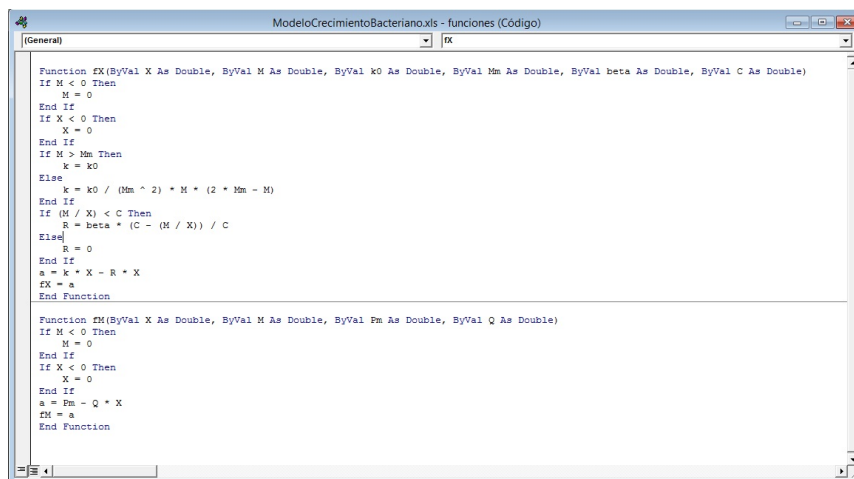
Es posible implementar el proceso de resolución y estudio de un sistema de este tipo, y variando los parámetros comprobar los distintos tipos de soluciones que se ofrecen para el mismo.

Un esquema básico para este tipo de entornos de resolución podría ser el siguiente:

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Datos iniciales			Parámetros		Runge-Kutta4		Paso		t	X	M	x1	x2	x3	x4	m1	m2	m3	m4	
2	X0	S	K0	0.1	h	0.2				0	0	5	2000								
3	MO	2000	MM	100						1	0.2	5.1010	2074.7483	0.5000	0.5050	0.5051	0.5101	375.0000	373.7500	373.7375	372.4748
4			beta	0.2						2	0.4	5.2041	2148.9865	0.5101	0.5152	0.5153	0.5204	372.4748	371.1996	371.1868	369.8986
5			C	200						3	0.6	5.3092	2222.7043	0.5204	0.5256	0.5257	0.5309	369.8987	368.5976	368.5846	367.2703
6			PM	500						4	0.8	5.4164	2295.8912	0.5309	0.5362	0.5363	0.5416	367.2704	365.9431	365.9299	364.5890
7			Q	25						5	1	5.5259	2368.5364	0.5416	0.5471	0.5471	0.5526	364.5891	363.2350	363.2215	361.8535
8										6	1.2	5.6375	2440.6289	0.5526	0.5581	0.5582	0.5637	361.8536	360.4722	360.4584	359.0628
9										7	1.4	5.7514	2512.1578	0.5637	0.5694	0.5694	0.5751	359.0629	357.6535	357.6394	356.2157
10										8	1.6	5.8676	2583.1114	0.5751	0.5809	0.5809	0.5868	356.2158	354.7779	354.7636	353.3110
11										9	1.8	5.9861	2653.4783	0.5868	0.5926	0.5927	0.5986	353.3111	351.8443	351.8296	350.3477
12										10	2	6.1070	2723.2466	0.5986	0.6046	0.6047	0.6107	350.3478	348.8513	348.8363	347.2246
13										11	2.2	6.2304	2792.4041	0.6107	0.6168	0.6169	0.6230	347.3247	345.7979	345.7826	344.2403
14										12	2.4	6.3562	2860.9386	0.6230	0.6293	0.6293	0.6356	344.2404	342.6828	342.6672	341.0938
15										13	2.6	6.4847	2928.8374	0.6356	0.6420	0.6420	0.6485	341.0939	339.5048	339.4889	337.8836
16										14	2.8	6.6156	2996.0877	0.6485	0.6549	0.6550	0.6616	337.8837	336.2626	336.2464	334.6087
17										15	3	6.7493	3062.6783	0.6616	0.6682	0.6682	0.6749	334.6088	332.9549	332.9383	331.2675
18										16	3.2	6.8856	3128.5903	0.6749	0.6817	0.6817	0.6886	331.2676	329.5803	329.5633	327.8589
19										17	3.4	7.0247	3193.8155	0.6886	0.6954	0.6955	0.7025	327.8590	326.1376	326.1204	324.3814
20										18	3.6	7.1666	3258.3382	0.7025	0.7095	0.7096	0.7167	324.3816	322.6254	322.6078	320.8337
21										19	3.8	7.3114	3322.1443	0.7167	0.7238	0.7239	0.7311	320.8338	319.0422	319.0242	317.2143
22										20	4	7.4591	3385.2191	0.7311	0.7385	0.7385	0.7459	317.2144	315.3886	315.3683	313.5218
23										21	4.2	7.6096	3447.5481	0.7459	0.7534	0.7534	0.7610	313.5219	311.6571	311.6353	309.7547
24										22	4.4	7.7635	3509.1160	0.7610	0.7686	0.7687	0.7764	309.7548	307.8524	307.8333	305.9115
25										23	4.6	7.9204	3569.9075	0.7764	0.7841	0.7842	0.7920	305.9116	303.9707	303.9513	301.9906
26										24	4.8	8.0804	3629.9070	0.7920	0.8000	0.8000	0.8080	301.9908	300.0107	299.9909	297.9906
27										25	5	8.2436	3689.0984	0.8080	0.8161	0.8162	0.8244	297.9907	295.9706	295.9504	293.9097
28										26	5.2	8.4101	3747.4654	0.8244	0.8326	0.8327	0.8410	295.9708	293.9489	293.9283	291.7444
29										27	5.4	8.5800	3804.9914	0.8410	0.8494	0.8495	0.8580	293.9490	291.9148	291.8938	289.5990
30										28	5.6	8.7534	3861.6594	0.8580	0.8666	0.8667	0.8753	291.9149	289.8683	289.8468	287.4990
31										29	5.8	8.9302	3917.4520	0.8753	0.8841	0.8842	0.8930	289.8684	287.8116	287.7897	285.3690
32										30	6	9.1106	3972.3515	0.8930	0.9019	0.9020	0.9111	287.8117	285.7426	285.7203	283.2390
33										31	6.2	9.2946	4026.3599	0.9111	0.9202	0.9203	0.9295	285.7427	283.6615	283.6388	281.1018
34										32	6.4	9.4824	4079.3989	0.9295	0.9388	0.9389	0.9482	283.6616	281.5703	281.5472	278.9317
35										33	6.6	9.6740	4131.5096	0.9482	0.9577	0.9578	0.9674	281.5704	279.4688	279.4453	276.7517
36										34	6.8	9.8694	4182.6528	0.9674	0.9771	0.9772	0.9869	279.4689	277.3573	277.3334	274.5817
37										35	7	10.0688	4232.8091	0.9869	0.9969	0.9969	1.0069	277.3574	275.2363	275.2119	272.4017
38										36	7.2	10.2722	4281.9555	1.0069	1.0169	1.0170	1.0272	275.2364	273.1057	273.0809	270.2117
39										37	7.4	10.4797	4330.0806	1.0272	1.0375	1.0376	1.0480	273.1058	270.9652	270.9400	267.9117
40										38	7.6	10.6914	4377.1847	1.0480	1.0584	1.0586	1.0691	270.9653	268.8147	268.7892	265.6817

Además, el programa EXCEL permite diseñar funciones definidas por el

usuario a través de un lenguaje muy sencillo llamado Visual Basic for Applications (VBA) donde podemos definir esas funciones con condiciones, de tal forma que quedarían programadas así:



```
Function fX(ByVal X As Double, ByVal M As Double, ByVal k0 As Double, ByVal Mm As Double, ByVal beta As Double, ByVal C As Double)
    If M < 0 Then
        M = 0
    End If
    If X < 0 Then
        X = 0
    End If
    If M > Mm Then
        k = k0
    Else
        k = k0 / (Mm ^ 2) * M * (2 * Mm - M)
    End If
    If (M / X) < C Then
        R = beta * (C - (M / X)) / C
    Else
        R = 0
    End If
    a = k * X - R * X
    fX = a
End Function

Function fM(ByVal X As Double, ByVal M As Double, ByVal Pm As Double, ByVal Q As Double)
    If M < 0 Then
        M = 0
    End If
    If X < 0 Then
        X = 0
    End If
    a = Pm - Q * X
    fM = a
End Function
```

5.5. Análisis

ESTUDIO DE LAS POSIBLES SOLUCIONES:

Una vez planteado el problema físico, elaborado una teoría de resolución y diseñado la herramienta informática para manipular los datos, se pueden jugar con los distintos parámetros, o probar, si se dispone, datos reales y comprobar las soluciones encontradas, razonando si son coherentes los resultados con las conclusiones que se puedan deducir de las hipótesis de partida.

Se van a suponer unos valores iniciales para todos los casos de:

	valor	unidades
X_0	5	millones/L
M_0	2000	mg/L

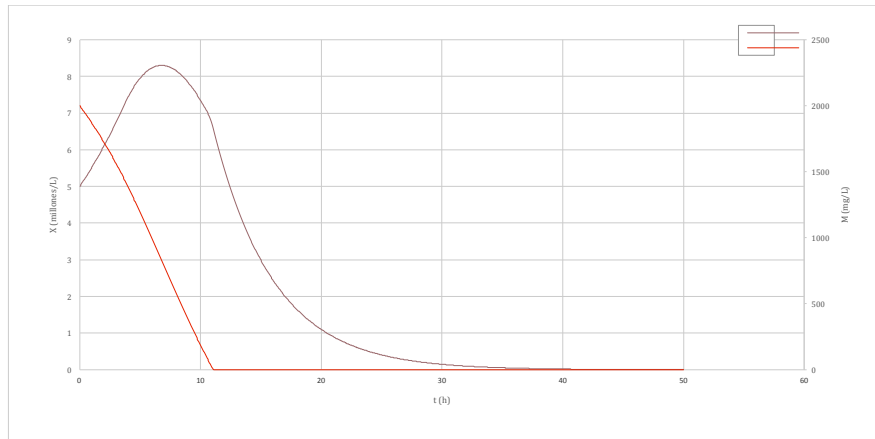
- ESTUDIO DEL SISTEMA SIN CONTRIBUCIÓN DE NUTRIENTES

Para este caso el valor de $P_M = 0$. Es decir, no hay contribución de nutrientes por lo que el disponible es el que existe al principio del estudio. Por tanto, lo previsible es esperar que el sistema acabará agotándose por completo y la masa bacteriana muriendo por completo al cabo de un tiempo.

Para los siguientes valores supuestos:

	valor	unidades
k_0	0.1	h-1
M_m	100	mg/L
β	0.2	h-1
C	200	mg/millones
P_M	0	mg/(h·L)
Q	25	mg/(millones·h)

El resultado obtenido es el siguiente:



Donde puede observarse que hay un crecimiento exponencial bacteriano, como era previsible, que se detiene debido a la mortalidad provocada por la disminución drástica de nutrientes, la cual experimenta una caída exponencial al principio que más adelante se armoniza con la caída de bacterias. Finalmente se observa como al agotarse todos los nutrientes del sustrato, se produce una caída exponencial del cultivo debido a la mortalidad natural por ausencia de nutrientes.

- ESTUDIO DEL SISTEMA CON CONTRIBUCIÓN DE NUTRIENTES

En este caso habrá una contribución de nutrientes que se contrarrestará con un aumento de bacterias. Este caso podría plantear tres escenarios, previstos para otro tipo de modelos de control mutuo entre dos variables y el tiempo, basados en el amortiguamiento. El cual, podría definirse someramente como la capacidad que tiene un sistema de absorber el cambio de una de sus variables y contrarrestar los efectos provocados por dicho cambio para volver a un estado de equilibrio.

En este caso, después de realizar numerosas simulaciones vamos a considerar los 3 casos:

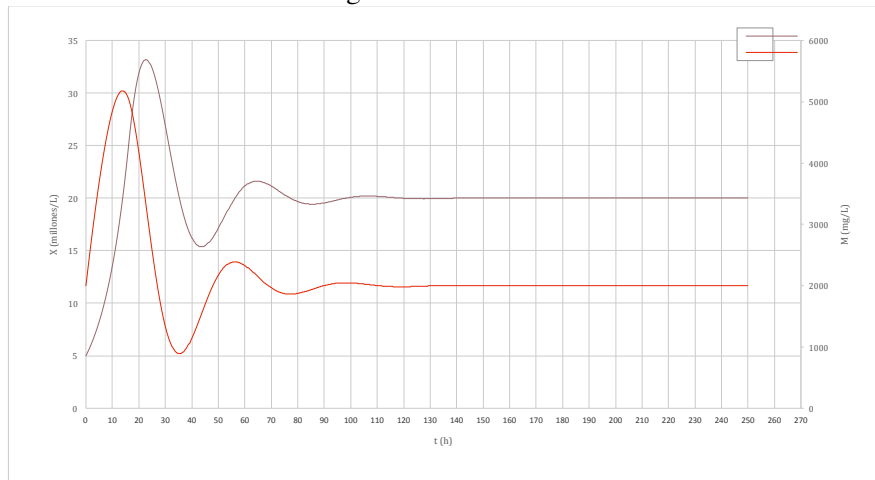
Sistema subamortiguado:

En estos casos, el exceso de nutrientes provoca un crecimiento masivo de bacterias que llegado un punto hace que disminuya drásticamente la concentración de nutrientes, lo cual a su vez hace que vuelva a disminuir las bacterias y este ciclo se repite cada vez con menor intensidad hasta que se establece un equilibrio entre la aportación de nutrientes, la mortalidad de bacterias y el crecimiento bacteriano.

De todas las simulaciones realizadas se expone un ejemplo para el caso:

	valor	unidades
k_0	0.1	h-1
M_m	100	mg/L
β	0.2	h-1
C	200	mg/millones
P_M	500	mg/(h·L)
Q	25	mg/(millones·h)

El resultado obtenido es el siguiente:



Sistema sobreamortiguado:

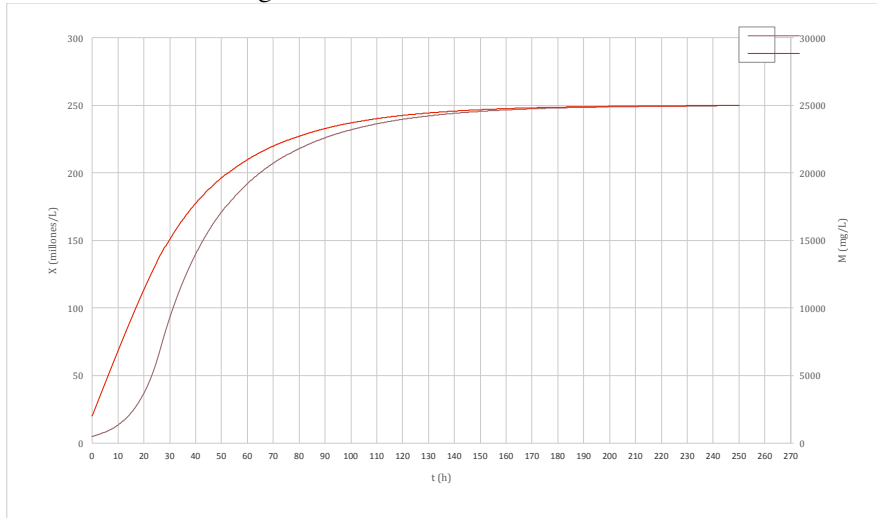
El sobreamortiguamiento se da cuando la contribución de nutrientes al medio está muy por encima de lo que la masa bacteriana, incluido su crecimiento exponencial, es capaz de asimilar, produciendo un aumento de la concentración en el tiempo, dando lugar a un largo período transitorio hasta llegar al estado estacionario.

De todas las simulaciones realizadas se expone un ejemplo para el caso:

	valor	unidades
k_0	0.1	h-1
M_m	100	mg/L
β	0.2	h-1

C	200	mg/millones
P_M	500	mg/(h·L)
Q	2	mg/(millones·h)

Y el resultado es el siguiente:



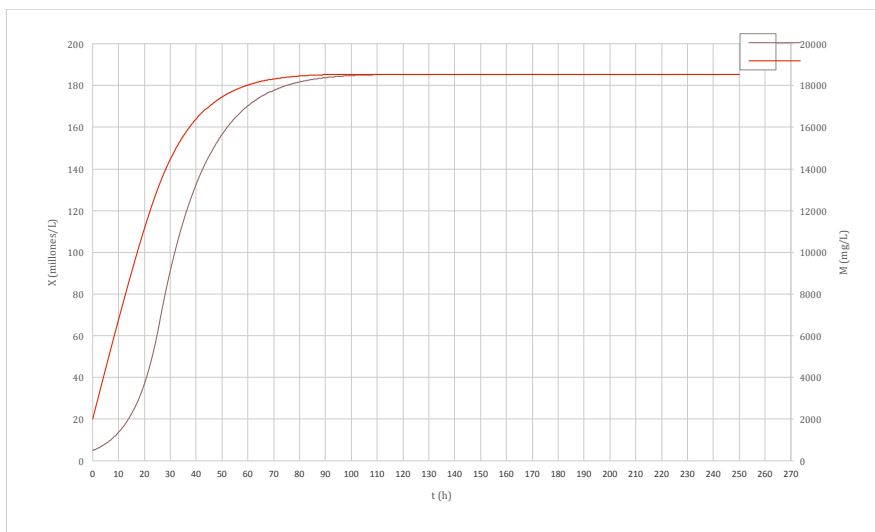
Sistema críticamente amortiguado:

El amortiguamiento crítico es ese estado de metaequilibrio donde el sistema alcanza el equilibrio de forma estable más rápidamente. Evita las etapas de subidas y bajadas que provoca el subamortiguamiento, pero también optimiza el tiempo que tarda el sistema en alcanzar el estado estacionario.

De todas las simulaciones realizadas se expone un ejemplo para el caso:

	valor	unidades
k_0	0.1	h-1
M_m	100	mg/L
β	0.2	h-1
C	200	mg/millones
P_M	500	mg/(h·L)
Q	2.7	mg/(millones·h)

Y el resultado es el siguiente:



Capítulo 6. Un simulador de un modelo dinámico para un curso de Macroeconomía Intermedia

Resumen

Esta comunicación presenta dos aplicaciones de un simulador de una versión básica del modelo de oferta y demanda agregada dinámica –del conocido modelo de síntesis neoclásica– programada en Excel que ha de permitir que un alumno de un curso de Macroeconomía Intermedia conozca como se formulan, resuelven y los resultados que proporcionan los modelos dinámicos frente a los tradicionales ejercicios de estática comparativa que suelen componer el grueso de las programaciones de un curso estándar de Macroeconomía de este nivel. Así se pretende que estas aplicaciones sirvan al alumno para apreciar la potencia de los modelos dinámicos y la capacidad de éstos de reproducir ajustes cíclicos en el que el ajuste de las variables a sus valores de largo plazo dista de ser un ajuste gradual y lineal.

Palabras clave: Modelo de síntesis neoclásica; resolución numérica; OAD-DAD; macroeconomía dinámica aplicada.

6.1. Introducción

Uno de los problemas asociados al aprendizaje de la macroeconomía en cursos introductorios o intermedios, es que los continuos ajustes en los planes de estudios se han saldado con la práctica desaparición de la dinámica en la programación de matemáticas. De esta forma, la enseñanza de los comportamientos dinámicos en general y en Macroeconomía en particular han de ser suplidos con estrategias imaginativas que permitan al alumno adquirir ciertas competencias de formas alternativas apreciando la potencia del análisis dinámico, de la computación y de la calibración, así como de su capacidad para ayudarnos a entender los procesos de ajuste cíclico.

6.2. Objetivos

En este contexto la experiencia que presentamos plantea dos estrategias complementarias de enseñanza de un modelo dinámico (una estrategia gráfica y la construcción y uso de un simulador), que debe ayudar al alumno a:

- Apreciar la aplicabilidad de lo aprendido.
- Iniciarse en la importancia de la programación los modelos, de la computación.
- Entender la efectividad y/o ineffectividad de las políticas en función de la importancia de los valores de los parámetros.
- Realizar una primera aproximación a la macroeconomía dinámica.

6.3. El modelo y su solución gráfica

Partimos de una expresión analítica de la oferta agregada en la que hemos incorporado, por simplicidad el supuesto de expectativas ingenuas,

$$OAD_t \equiv \pi_t = \pi_{t-1} + \lambda(Y_t - Y_{pe})$$

Esta curva de oferta tiene carácter dinámico, al incorporar una variable y su retardo, y el estudiante ha de comprobar este carácter dado que esta relación se irá desplazando siempre que la producción no haya alcanzado su nivel de pleno empleo. Por su parte, la demanda agregada dinámica se obtiene sin más que diferenciar una curva de demanda agregada estática obtenida del equilibrio IS-LM. Así diferenciando una demanda agregada como ésta:

$$Y = \gamma A_0 + \beta \frac{M}{P}$$

y suponiendo que los multiplicadores de política monetaria y fiscal se mantienen constantes, tenemos que:

$$dY = \gamma dA_0 + \beta d\left(\frac{M}{P}\right)$$

expresión que podemos describir como :

$$\underbrace{\frac{dY}{Y_t - Y_{t-1}}}_{\pi_t} = \gamma dA_0 + \beta \frac{M}{P} \left(\underbrace{\frac{dM}{M}}_m - \underbrace{\frac{dP}{P}}_{\pi_t} \right)$$

o alternativamente como:

$$DAD_t \equiv Y_t = Y_{t-1} + \gamma dA_0 + \beta \frac{M}{P} (m - \pi_t)$$

Analicemos la dinámica del output y de la inflación en ausencia de perturbaciones. Para ello, volvamos a fijarnos en las expresiones de la oferta y de la demanda agregada dinámicas:

$$DAD_t \equiv Y_t = Y_{t-1} + \underbrace{\gamma dA_0}_{=0} + \beta \frac{M}{P} (m - \pi_t)$$

$$OAD_t \equiv \pi_t = \pi_{t-1} + \lambda(Y_t - Y_{pe})$$

Analicemos los diferentes equilibrios del modelo suponiendo que se parte de una situación inicial de desempleo e inflación superior al nivel de inflación de largo plazo, tal y como indica la figura.

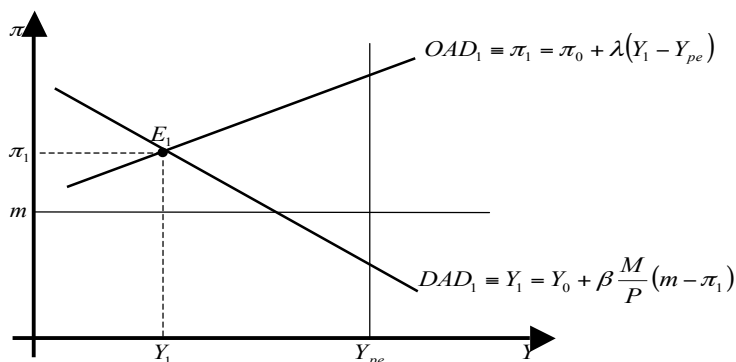


Figura 1: Modelo de OAD-DAD. Situación inicial de desempleo y tasa de inflación superior a la tasa de crecimiento monetario

El punto de equilibrio del periodo 1 E_1 no es estable, puesto que ni la producción coincide con su nivel de pleno empleo, ni la tasa de inflación coincide con la de largo plazo. Siempre que la producción no coincida con la de pleno empleo, el nivel de inflación variará y, por tanto, la curva de oferta agregada dinámica se desplazará. Por su parte, siempre que la tasa de inflación no coincida con la tasa de crecimiento monetario, el nivel de

output variará, lo que lleva consigo un desplazamiento de la demanda agregada dinámica. En el caso que estamos analizando, por tanto, se desplazarán tanto la oferta agregada dinámica, como la demanda agregada dinámica. Las ecuaciones que describen el estado de la economía en el periodo 2 vienen dadas por las siguientes expresiones:

$$OAD_2 \equiv \pi_2 = \pi_1 + \lambda(Y_2 - Y_{pe})$$

$$DAD_2 \equiv Y_2 = Y_1 + \beta \frac{M}{P} (m - \pi_2)$$

Si queremos saber por dónde han de pasar exactamente estas nuevas curvas, podemos hacer los siguientes razonamientos. Si la producción del periodo 2 coincidiese con el nivel de pleno empleo, $Y_1 = Y_{pe}$, la curva de oferta agregada dinámica del periodo 2 nos dice que la tasa de inflación debería ser igual a la del periodo anterior, $\pi_2 = \pi_1$. Por tanto, la curva de oferta agregada dinámica del periodo 2 pasa por el punto blanco denotado en la figura 2 como A. Por su parte, si la tasa de inflación del periodo 2 coincide con la tasa de crecimiento monetario, $\pi_2 = m$, según la función de demanda agregada dinámica y en ausencia de perturbaciones, el nivel de producción debe coincidir con el del periodo 1, $Y_2 = Y_1$. Esto quiere decir que la demanda agregada dinámica del periodo 2 debe pasar por el cuadrado blanco denotado en la figura 2 como a. Las nuevas curvas de oferta agregada y demanda agregada dinámicas del período 2 determinan un nuevo equilibrio para la economía, representado por el punto E_2 .

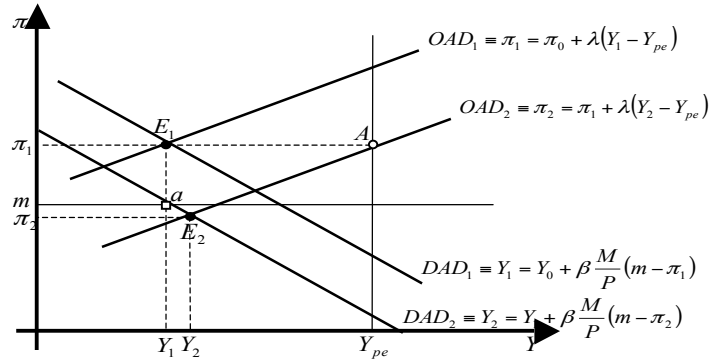


Figura 2: La dinámica de transición al estado estacionario. Periodo 2

Si repitiésemos este ejercicio obtendríamos unas trayectorias temporales como las de las figuras 3 y 4 es decir, el proceso continuaría hasta alcanzar el estado estacionario de la economía, determinado por la producción de pleno empleo y la tasa de crecimiento monetario.

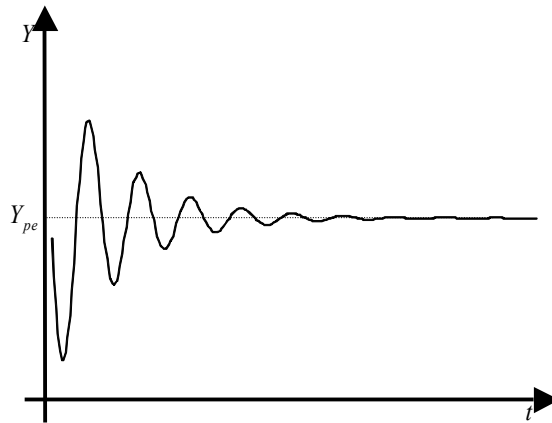


Figura 3: Evolución temporal del output

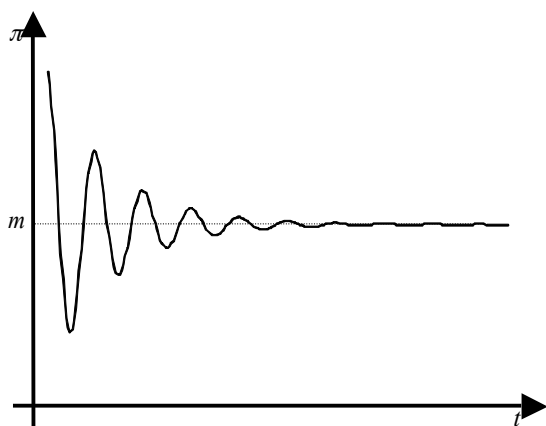


Figura 4: Evolución temporal de la tasa de inflación

6.4. El simulador

Una vez logrado el que el alumno resuelva de forma intuitiva a través de una aproximación gráfica un sistema de ecuaciones en diferencias como el planteado en la sección anterior, el siguiente paso es el de su resolución a través de procedimientos numéricos en una hoja de cálculo. Para ello se presentan dos aplicaciones realizadas con Excel en las que a partir de un conjunto paramétrico inicial, sobre los que el alumno puede introducir cambios, la aplicación permite resolver el modelo y reproducir incluso de forma gráfica las trayectorias temporales del output y la inflación. Se trata pues de un simulador, en el que el alumno puede apreciar los efectos en el largo y en el corto plazo de shocks de oferta y de demanda, apreciar el concepto de estado estacionario/largo plazo, así como realizar calibracio-

nes y analizar la sensibilidad de la inflación y del output a los valores de los parámetros de la economía.

El simulador ha de permitir al alumno analizar los efectos de diferentes tipos de shocks de oferta y de demanda sobre las trayectorias temporales de la inflación y el desempleo replicando las soluciones del modelo dinámico de oferta y demanda en una economía cerrada, dado un determinado espacio paramétrico.

6.5. Referencias

- BAJO, O. y MONÉS, M.A. (2000): *Curso de Macroeconomía*, Antoni Bosch, Barcelona (2ª edición).
- BARREIRO, F., LABEAGA, J.M. y MOCHÓN, F. (1998): *Macroeconomía Intermedia*, McGraw-Hill, Madrid.
- BARRERAS, M. (2005). *Matemáticas con Excel*. Ed. Rama.
- BLANCHARD, O. (2009): *Macroeconomía con soporte interactivo Moodle*, Madrid (4ª edición).
- CARRASCAL, U. (2011). *Estadística Descriptiva con Microsoft Excel 2010*. Ed. Rama.
- CONGREGADO, E, ROMÁN, C. (2014): *Macroeconomía Intermedia*, Pearson, Madrid.
- DORNBUSCH, R., FISCHER, S. y STARTZ, R. (2004): *Macroeconomía*, McGraw-Hill, Madrid (9ª edición).
- JONES, CHARLES I. (2009): *Macroeconomía*, Antoni Bosh, Barcelona.
- JONES, CHARLES I. (2010): *Macroeconomía*, Suplemento sobre la actualidad económica, Antoni Bosh, Barcelona.

SHONE, R. (2002): *Economic Dynamics*. Cambridge University Press

Premio Recibido

e -pública

Revista electrónica sobre la enseñanza de la Economía Pública

Estimados Juan Luis, Emilio y Antonio:

Queremos felicitaros muy sinceramente por el Primer Premio *e-pública*, *Revista electrónica sobre la enseñanza de la Economía Pública*, concedido desde el Consejo Editor de la revista, dentro de las VII Jornadas de Economía Pública celebradas recientemente en Palma de Mallorca, a vuestro excelente artículo “El uso de métodos numéricos y su aplicación en Excel para el modelo de Oferta y Demanda agregada dinámica”.

Entre otros méritos destacables, desde el Consejo Editor hemos valorado muy positivamente la gran utilidad para la docencia en los grados en Economía y ADE de la herramienta que ofrecéis en vuestro trabajo, así como la claridad expositiva del texto, lo que permite un acceso sencillo para los usuarios de la misma.

De cara a la próxima publicación de vuestro artículo, hemos revisado el texto en su versión actual, así como su adecuación a los criterios de edición establecidos en nuestras “normas para los autores”. En este sentido, hemos comprobado que el manuscrito respeta convenientemente el formato de edición, aunque hemos encontrado una serie de erratas tipográficas,

especialmente palabras que no aparecen separadas de espacios o algún uso indebido de mayúsculas/minúsculas. Asimismo, os recuerdo que esta nueva versión debe incluir, conforme a nuestras reglas de edición, un resumen (*abstract*) en inglés, así como las correspondientes palabras clave también en inglés (*keywords*).

Por ello, os animo a que, de acuerdo con lo expresado en el párrafo anterior, reviséis el original, remitiéndonos la versión final en un plazo que no debería superar un mes. Nuestra idea es que vuestro trabajo sea incluido en el sumario del número 17, que publicaremos en el próximo mes de septiembre.

Recibid un cordial saludo,

Jorge Onrubia
Editor Ejecutivo

Parte V: Conclusiones

Summarise

This thesis is a compendium of works in three different lines. The first one dedicated to exploring both theoretically and empirically the use of graphs of visibility in the analysis of temporal series. Thus, in the face of the traditional analysis techniques used to date macroeconomic time series with the objective of analyzing summits and valleys and analyzing the depth and amplitude of the different cyclic phases, this thesis develops applications of the theory of Graphs to series analysis following the approach opened by Lacasa et al. At the same time as making a theoretical contribution on one of the indexes most commonly used in the measurement of resilience.

Thus, this contribution arises from the need to arbitrate some mechanism to avoid the lack of sensitivity of the index when the time-ral series has a very high number of observations.

All these ideas and algorithms are applied to the analysis of two types of economic series for the North American economy: those of the gross do-

mestic product and the series of energy consumption.

The choice of these series is not arbitrary insofar as it is a question of making an alternative approach, so that its value must be made in a key way compared with the techniques traditionally used for these issues. For this reason, we opted to approach a classical series, which is dated by the NBER as well as being the object of numerous studies and analyzes, as well as the North American energy consumption series that have been the object of a profuse analysis in the Field of the Energy Economics.

The visibility graph (VGA) algorithm allows the transformation of a time series into a complex network in a very simple way. The study of the topological properties of the network contributes knowledge about the behavior of the series, as the graph inherits many of its properties. In Chapters 2 and 3 we proceed to transform the temporal series corresponding to the total energy consumption and by type of source and the quarterly rate of change of the GDP of the USA in a graph and apply the theory of complex networks to study it. The resulting approach, when applied to different metrics, shows results that are quite in line with those obtained with traditional algorithms, so that it can be concluded that these applications provide quite useful representations for forecasting and tracking.

This part closes with a question related to the analysis of the time series, and more specifically with the capacity of absorption of a shock by the series. In particular, it is analyzed as one of the most commonly used indices to measure resilience, shows a certain insensitivity when the sample size is large. To address this problem, a digression and theoretical input is made that can correct this bias.

The third part provided empirical evidence supporting a stylized fact observed in some economies with regard to the business ownership rates development: after a long-decline in the rate of self-employment and since 1978, self-employment rates appeared to increase in many industrialized countries in such a way that the trend of self-employment rates seemed to show a structural shift in terms of a revival (U-shape) or at least a stabilization (L-shape).

To this end, they developed an empirical model in which estimates of different functional forms of the relationship between the 'equilibrium' self-employment rate and the GDP per capita allowed them to infer the shape of this relationship for 23 OECD countries. In their empirical model, the equilibrium self-employment rate was obtained using some assumptions about the relationship between self-employment, unemployment, labour incomes and some lags structures. Opposite to this approach, we propose to test the U/L-shape hypothesis using statistical methods: i) first, decomposing the self-employment rate into their two components –i.e. the cyclical and natural components-, and, ii) using a recent econometric approach for detecting the presence of structural breaks (Kejriwal and Perron, 2010). In this way, the U-shape hypothesis is tested for 23 OECD countries, using data on GDP per capita and the natural self-employment rate component (a proxy for the 'equilibrium' self-employment rate) over the period 1972-2008. Our results only provide a partial support for the U-shape hypothesis: for 15 out of 23 countries we find a significantly positive relation between GDP per capita and the natural self-employment rate. Our results suggest that, notwithstanding the rise of self-employment observed in many countries over the last few decades, economies of scale and scope continue to play an important role in many advanced economies.

The fourth part of the thesis addresses issues related to the use of numerical methods and the development of applications that can be used in teaching the modeling of economic dynamics. In particular, two trials are presented. The first one is an essay of modeling and programming of dynamic problems through systems of differential equations solved through numerical methods. In particular, the study of the behavior of a bacterial mass in a controlled crop that presents analogies to the evolution and business survival is presented.

The second presents two applications of a simulator of a basic version of the dynamic aggregate supply and demand model - the well-known neo-classical synthesis model - programmed in Excel that will allow a student of an Intermediate Macroeconomics course to know how they are formulated, solved And the results provided by the dynamical models versus the traditional benchmarking exercises that usually compose the bulk of the programming of a standard course of Macroeconomics of this level. It is intended that these applications serve the student to appreciate the power of dynamic models and the ability of these to reproduce cyclical adjustments in which the adjustment of the variables to their long-term values is far from a Gradual and linear adjustment.