

Asociación Española de Profesores Universitarios
de Matemáticas para la Economía y la Empresa



rect@

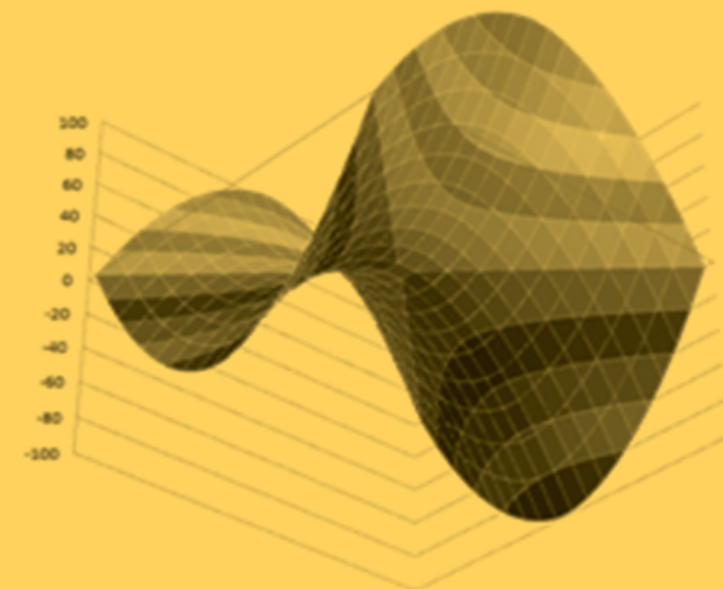
MÉTODOS CUÁNTITATIVOS E INFORMÁTICA

MÉTODOS CUÁNTITATIVOS E INFORMÁTICA

Coordinadores

J. J. Bernal García

S. M. Martínez María-Dolores



ISBN 978-84-9053-481-6



9 788490 534816



Universidad
Politécnica
de Cartagena

f SéNeCa⁽⁺⁾
Agencia de Ciencia y Tecnología
Región de Murcia



SERIE MONOGRAFÍAS
NÚMERO 4



rect@

Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA

SERIE MONOGRAFÍAS • Número 4 • 2013

MÉTODOS CUÁNTICOS E INFORMÁTICA

MÉTODOS CUÁNTICOS E INFORMÁTICA

Coordinadores

J.J. BERNAL GARCÍA

S.M. MARTÍNEZ MARÍA-DOLORES



tirant lo blanch

Valencia, 2013

Copyright ©2013

Todos los derechos reservados. Ni la totalidad ni parte de este libro pueden reproducirse o transmitirse por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética, o cualquier almacenamiento de información y sistema de recuperación sin permiso escrito de los autores y del editor

Rect@ es la Revista Electrónica de ASEPUMA y publica trabajos relacionados con la docencia y la investigación en la matemática aplicada a la economía y la empresa. La revista tiene una periodicidad semestral.

EL CONSEJOR EDITORIAL DE LA REVISTA está compuesto por Dra. Ada Álvarez (Universidad Autónoma de Nuevo León, Monterrey, México) Dr. Juan Jesús Bernal García (Universidad Politécnica de Cartagena y ASEPUMA) Dr. Emilio Cerdá (Universidad Complutense de Madrid y ASEPUMA) Dr. Carlos Coello, (SINVESTAV-IPN, México) Dr. Ángel Corberan Salvador (Universitat de Valencia) Dr. Laureano Escudero Bueno (Universidad Rey Juan Carlos) Dr. José Luis González-Velarde (Instituto Tecnológico de Monterrey, Monterrey, México) Dr. Manuel Laguna (Universidad Pinar del Río, Cuba) Dra. Flor Guerrero (Universidad Pablo de Olavide de Sevilla y ASEPUMA) Dr. José Carlos de Miguel (Universidad de Santiago y ASEPUMA) Dr. Francisco Montes Suay (Universitat de Valencia) y Dr. Carlos Romero (Universidad Politécnica de Madrid).

© J.J. Bernal García
S. M. Martínez María Dolores

© TIRANT LO BLANCH
EDITA: TIRANT LO BLANCH
C/ Artes Gráficas, 14 - 46010 - Valencia
TELF.: 96/361 00 48 - 50
FAX: 96/369 41 51
Email: tlb@tirant.com
<http://www.tirant.com>
Librería virtual: <http://www.tirant.es>
DEPÓSITO LEGAL: V-2208-2013
ISBN: 978-84-9053-481-6
IMPRIME: Guarda Impresores, S.L.
MAQUETA: PMc Media

Si tiene alguna queja o sugerencia, envíenos un mail a: atencioncliente@tirant.com. En caso de no ser atendida su sugerencia, por favor, lea en www.tirant.net/index.php/empresa/politicas-de-empresa nuestro Procedimiento de quejas.

Agradecimientos

Cuando propuse en la Junta Directiva de *ASEPUMA* la realización de un monográfico de *Rect@* compuesto por artículos que integrasen la informática y los métodos cuantitativos, no dudaron en darme su voto por unanimidad, lo cual quiero agradecerles en la figura de su presidente *Ramón Sala* y de *Vicente Liern*, editor responsable de la revista. Cuando finalmente se inició la elaboración del monográfico “Métodos cuantitativos e informática”, desde luego éste no hubiese sido posible sin la contribución que hicieron tanto miembros de esta asociación como otros profesores especialistas en la materia a los que invitamos a participar, enviándonos artículos relacionados con dicha temática, por lo que les estamos muy agradecidos. Hago extensible dicho agradecimiento a aquellos a los que les envié los trabajos para la revisión ciega (*), que a pesar de la premura y las fechas tan poco idóneas, realizaron su labor de forma concienzuda, a veces hasta en varias rondas, antes de validarlos definitivamente para su montaje final, en la que he contado con la valiosa ayuda de la profesora *Soledad M^a Martínez*, y la cuidada edición de la técnica informática *Mariu Alcate* para su impresión posterior, gracias al patrocinio del Vicerrectorado de Investigación de la UPCT, y la *Fundación Séneca de la Agencia de Ciencia y Tecnología*, de la Región de Murcia.

Muchas gracias a todos los que habéis hecho posible este volumen.

JUAN JESÚS BERNAL GARCÍA

Miembro del consejo editorial de la revista *Rect@*

(*) Relación alfabética de los evaluadores:

- *Alfonso González Pareja*
- *Amparo Mármol Conde*
- *Antonio Seijas Macias*
- *Coro Chasco Yrigoyen*
- *Fernando A. López Hernández*
- *Gabriela Fernández Barberis*
- *Joaquín Pacheco Bonrostro*
- *Juan Jesús Bernal García*
- *M Carmen García Centeno*
- *Soledad M^a Martínez M^a Dolores*
- *Matilde Lafuente Lechuga*
- *Rafael Caballero Fernández*
- *Ramón Sala Garrido*

Presentación

Siguiendo la línea abierta por anteriores números monográficos de la revista *Rect@* iniciada en el año 2002, a raíz de la Jornada de *ASEPUMA* realizada en Cartagena con el nombre de “Métodos Cuantitativos e Informática” se ha elaborado el presente libro, que hace el número cuatro de la colección, con el propósito de dejar patente la contribución cada vez más importante de la informática a los métodos cuantitativos, tanto en la docencia como en la investigación aplicada a la economía y la empresa.

El volumen se compone de 16 artículos, 4 de ellos en inglés, en el que han participado 36 autores pertenecientes a 14 Universidades, 2 de ellas extranjeras (Universidad Tecnológica de México y Universidad de Ljubljana de Eslovenia), lo que unido a las que pertenecen los evaluadores, hace un total de 17 Universidades, públicas y privadas, que han contribuido a que este monográfico viese la luz.

Los dieciséis trabajos los hemos ordenados por orden alfabético según la herramienta, o programa informático principal que utilizan, bien en el tratamiento de datos de la investigación realizada, o en las prácticas docentes a que se refieren: *Eviews*, *Excel*, *Frontier Analyst*, *GeoDa*, *Matlab*, *Mathematica*, *R*, *Software propio*, *SPSS* y *STATA*. Aunque cada artículo lleva su correspondiente resumen en dos idiomas, a modo de presentación veamos someramente una descripción de cada uno de ellos:

El primero de los trabajos que emplea el programa **Eviews**, ha sido elaborado por Lourdes Badillo, Rosa Badillo y Rafael Martínez, y realiza un **Estudio econométrico del nivel de satisfacción de los alumnos en la adquisición de conocimientos y competencias con las nuevas metodologías de enseñanza-aprendizaje en el entorno del EEES**, más concretamente de los alumnos de la titulación del Grado en ADE de la Facultad de Ciencias de la Empresa de la UPCT, en relación a diversos procedimientos metodológicos aplicados por los docentes.

Los siguientes tres artículos utilizan como herramienta informática principal **Excel**, complementada en todos los casos con código realizado en **VBA** (Visual Basic for Applications), el primero de ellos y segundo del libro, aborda **El problema del enrutamiento de vehículos. Propuestas para la búsqueda del camino más corto. Aplicación al entorno docente y Pymes**, ha sido realizado por los profesores Juan Jesús Bernal, Eloy Hontoria y Drako Aleksovski, apoyándose además en la utilidad **Solver** (incorporada en Excel) y en **Google Maps**, que gracias a la programación realizada proporciona la matriz de distancias entre ciudades. El siguiente artículo, que aportan Juan Jesús Bernal, Soledad Martínez y Pablo Bernal, bajo el título: **Representación de gráficas y tablas de datos con movimiento en Excel**, consigue también mediante programación en VBA para dotar a las gráficas de movimiento, no solo las funciones matemáticas, sino también emulando los denominados *Motion Chart* de Google Drive. La última de las aplicaciones con esta herramienta es la de

Resolución de ecuaciones diferenciales con Excel, que trata de encontrar aproximaciones, de tipo numérico y gráfico, a las soluciones de problemas de valor inicial (PVI) de ecuaciones diferenciales ordinarias (EDO) mediante el uso de ésta hoja de cálculo, habiendo sido elaborada por Francisco Javier Palencia.

El artículo que ocupa el quinto lugar del monográfico, utiliza el software **Frontier Analyst**, teniendo por objetivo evaluar la producción investigadora de 48 universidades públicas españolas aplicando el Análisis Envolvente de Datos (DEA, con una muestra de datos referidos al año 2010, habiendo sido realizado por María Concepción, Juan Gómez-Gallego, María Gómez-Gallego y Juan Gómez-García, respondiendo al título **Ranking de Productividad en investigación de las Universidades públicas españolas**.

Los dos artículos siguientes tiene en común el empleo del software de estadística espacial **GeoDa**. En el primero de ellos, **Distribución espacial de la actividad económica en la comarca de Cartagena. Un análisis basado en la demografía empresarial usando Arcview y GeoDa**, los profesores Francisco Tomas y Fernando A López, realizan un análisis exploratorio de datos espaciales para describir la estructura económica del campo de Cartagena desde la perspectiva espacial, utilizando la información disponible en la base de datos SABI (sistemas de análisis de balances ibéricos). Empleando el paquete informático **GeoDaSpace** (versión alfa), disponible gratuitamente desde GeoDa Center, la profesora Coro Chasco, nos presenta el artículo **GeoDaSpace: a resource for teaching spatial regression models**, donde nos muestra sus amplias posibilidades tanto por estudiantes como profesores, dada su amplia gama de métodos de estimación.

En el artículo octavo, titulado **Aplicación de la herramienta Fuzzy Logic al proceso de selección de un candidato político**, mediante el programa **Matlab**, y más concretamente su herramienta Fuzzy Logic Toolbox, los profesores M^a Carmen Lozano y Federico Fuentes, diseñan un modelo difuso de clasificación de candidatos a un puesto de responsabilidad política, a partir de la valoración cualitativa de las habilidades sociales y cualidades o atributos psicológicos propios de la llamada “inteligencia emocional”, presentes en cada candidato.

El conocido programa **Mathemática** es la herramienta principal utilizada en los dos siguientes artículos, combinándolo el primero de ellos también con la hoja de cálculo Excel. Éste responde al título **Comparación de diferentes distribuciones trapezoidales, mediante sus momentos y diagramas de cociente de momentos**, de los profesores José Manuel Herrerías y Rafael Herrerías, donde a partir de la distribución trapezoidal y utilizando las tres típicas estimaciones subjetivas del método PERT, tiene por objetivo realizar un análisis comparativo entre esas tres distribuciones. En el siguiente trabajo, **Wiener process in Mathematica**, el profesor José Antonio Seijas, utiliza una construcción primaria del proceso de Wiener (movimiento Browniano), basado en una secuencia adecuada de simples caminos aleatorios simétricos y el programa de cálculo simbólico Mathematica para proporcionar una

implementación de dicho proceso como recursos didáctico que ayuda a comprender dicho fenómeno.

El artículo que ocupa el onceavo lugar, emplea el programa estadístico **R**, de software libre, en el trabajo **Sobre las prácticas de laboratorio en la estadística de un Grado en Turismo**, se muestra como pueden realizarse prácticas de laboratorio utilizando, de modo sencillo, un software aparentemente complicado, mediante un caso práctico relacionado con el futuro desarrollo profesional de los estudiante del Grado de Turismo de la Universitat Jaume I, habiendo sido elaborado por los profesores J. Castelló, G. Galindo, V. Martínez y J. Momparler.

Destacar que no en todos los artículos incluidos en el presente monográfico se ha empleado programas informáticos comerciales estándar, así en los dos siguientes se utilizan programas de desarrollo propio. **GIS'F fractal analysis with pivoting graphic**, se ha realizado en colaboración de profesores de tres Universidades, María Ramos y Oscar Rene, utiliza las herramientas informática **DEFA** (Dinámica economía y financiera actual), y **FEFA** (factrales y la economía financiera actual), desarrollados por el primero de los autores en 2011 en México. Trabajo donde se presentan las estaciones de modelado precios geométricos fractales de Capital Markets en París, Frankfurt, Londres, Tokio, Nueva York y México, con el objetivo es lograr una mayor rentabilidad de las inversiones realizadas en el tiempo Ex Ante y que sirvan de referencia en tiempo iterativos Ex post, mientras que en el siguiente denominado **Iteración fractal de Computo IFS en los mercados financieros**, María Ramos, M^a Jesús Segovia y Marta Miranda, presentan un análisis de precios fractal de las acciones emisoras que cotizan en el mercado de capitales, con la determinación de los cardiodes de Mandelbrot para la fijación de soportes y resistencias en las tendencias del rango de precios, utilizando en este caso la herramienta **Deterministic IFS**, de Richard Bedient del Mathematic Department de Yale University.

El programa estadístico **SPSS** es el principal referenciado en los artículos 14 y 15; en el primero de ellos junto con otros recursos informáticos disponibles en el portal OCW de la Universidad de Murcia: **Enseñanza de la estadística apoyada en las TICS. Una visión desde el OCW de estadística en el Grado en Marketing de la Universidad de Murcia**, donde las profesoras Úrsula Faura y Fuensante Arnaldos, analizan cómo en la actualidad el acceso a la red permite emplear estrategias y herramientas nuevas que mejoran o complementan las utilizadas tradicionalmente, y que ofrecen al estudiante la oportunidad de visualizar conceptos estadísticos que les puedan resultar complejos e interactuar con ellos y aprender de la propia experiencia. Así mismo el siguiente artículo **Influencia de las tecnologías de la información sobre el rendimiento en las pruebas de matemáticas PISA-2009**, estudia la posible influencia de la disponibilidad de Tecnologías de la Información o TIC sobre los resultados del aprendizaje de matemáticas de los alumnos españoles. Para ello se explotan los microdatos del proyecto PISA (Programme for International Student Assessment) de la OCDE correspondientes a 2009, que cuantifica el nivel de conocimientos de alumnos de 15

años en matemáticas, lengua y ciencias, el cual ha sido elaborado por Gonzalo Sanz-Magallón, Manuel M. Molina y M. Cristina Aguirre.

El artículo 16, último del libro, utiliza el programa **STATA**, donde los Profesores Jaime García, Pedro Jesús Hernández y Ángel López Nicolás, bajo el título **An investigation of the relationship between job characteristics and the gender wage gap**, evalúa las diferencias salariales por género teniendo en cuenta no sólo las características de los trabajadores sino las del puesto de trabajo, y donde a raíz de los resultados sugieren que es necesario evitar la sobre-estimación del componente discriminatorio de la brecha salarial entre hombres y mujeres.

...

Esperamos sinceramente que sean de su interés.

Cartagena, julio de 2013

JUAN JESÚS BERNAL GARCÍA
SOLEDAD M^a MARTÍNEZ M^a DOLORES
Coordinadores

Índice Artículos

Estudio econométrico del nivel de satisfacción de los alumnos en la adquisición de conocimientos y competencias con las nuevas metodologías de enseñanza-aprendizaje en el entorno del EEES.

L. Badillo, R. Badillo, R. Martínez..... 1

El problema del enrutamiento de vehículos. Propuestas para la búsqueda del camino más corto. Aplicación al entorno docente y Pymes

J.J. Bernal, E. Hontoria, D. Aleksovski.....25

Representación de gráficas y tablas de datos con movimiento en Excel

J.J. Bernal, S.M. Martínez, P. Bernal 43

Resolución de ecuaciones diferenciales con Excel

F.J. Palencia 57

Ranking de productividad en investigación de las Universidades públicas españolas

M.C. Pérez, J.C. Gómez, M. Gómez, J. Gómez..... 83

Distribución espacial de la actividad económica en la Comarca de Cartagena.

Un análisis basado en la demografía empresarial usando Arcview y GeoDa.

F. Tomás, F.A. López 97

Geodaspace: A Resource For Teaching Spatial Regression Models

C. Chasco..... 119

Aplicación de la herramienta fuzzy logic al proceso de selección de un candidato político

M.C. Lozano, F. Fuentes..... 145

Comparación de diferentes distribuciones trapezoidales, mediante sus momentos y diagramas de cociente de momentos

J.M. Herrerias, R. Herrerias..... 159

Wiener Process In Mathematica

J.A. Seijas..... 177

Sobre las prácticas de laboratorio en la estadística de un grado en turismo

J. Castello, C. Galindo, V. Martínez, J. Momparler 197

Gis'f Fractal Analysis With Pivoting Graphic <i>M. Ramos, O. Rene</i>	209
Iteración fractal de computo IFS en los mercados financieros <i>M. Ramos, M.J. Segovia, M. Miranda</i>	223
Enseñanza de la estadística apoyada en las TICs. una visión desde el OCW de estadística en el grado en marketing de la universidad de Murcia <i>U. Faura, F. Arnaldos</i>	245
Influencia de las tecnologías de la información sobre el rendimiento en las pruebas de matemáticas PISA-2009 <i>G. Sanz-Magallón, M.M. Molina, M.C. Aguirre</i>	265
An investigation of the relationship between job characteristics and the gender wage gap <i>J. García, P.J. Hernández, A.L. Nicolás</i>	281

ESTUDIO ECONOMÉTRICO DEL NIVEL DE SATISFACCIÓN DE LOS ALUMNOS EN LA ADQUISICIÓN DE CONOCIMIENTOS Y COMPETENCIAS CON LAS NUEVAS METODOLOGÍAS DE ENSEÑANZA-APRENDIZAJE EN EL ENTORNO DEL EEES.

LOURDES BADILLO AMADOR

Lourdes.Badillo@upct.es

*Universidad Politécnica de Cartagena / Departamento de Economía
Facultad de Ciencias de la Empresa. C/Real, 3. 30201 Cartagena (Murcia)*

ROSA BADILLO AMADOR

Rosa.Badillo@upct.es

*Universidad Politécnica de Cartagena / Departamento de Economía
Facultad de Ciencias de la Empresa. C/Real, 3. 30201 Cartagena (Murcia)*

RAFAEL MARTÍNEZ CAPEL

Rafamcapell@hotmail.com

*Universidad Politécnica de Cartagena / Departamento de Economía
Facultad de Ciencias de la Empresa. C/Real, 3. 30201 Cartagena (Murcia)*

RESUMEN: El análisis sobre el nivel de satisfacción de los alumnos, en relación al nivel de conocimientos, competencias y habilidades que perciben los alumnos a través de las diferentes metodologías docentes utilizadas por el docente, resulta de importancia crucial, en el ámbito universitario, en un momento de cambio y adaptación al EEES (Espacio Europeo de Educación Superior). El éxito del mismo depende, en gran parte, del incentivo que tenga el alumno y, en ello, representa un papel fundamental las metodologías docentes utilizadas en su proceso formativo. Unos procedimientos metodológicos que satisfacen al alumno, suelen motivarlo y facilitar su éxito en su paso por la Universidad. En este sentido, el presente trabajo realiza, a través de procedimientos estadísticos y econométricos, un análisis del grado de satisfacción de los alumnos a través de una encuesta a los alumnos de la titulación del Grado en ADE de la Facultad de Ciencias de la Empresa de la UPCT en relación a diversos procedimientos metodológicos aplicados por los docentes.

Palabras clave: Modelos de elección discreta, Espacio Europeo de Educación Superior, Conocimientos, Competencias y Habilidades, Educación Superior.

ABSTRACT: The analysis of the student's satisfaction level in the acquisition of

knowledge, skills and abilities through the new teaching-learning methods is becoming crucial in the University in a key moment of the adaptation of those methods to the guidelines of the European Space for Higher Education. Its success depends, in large part, on the students' incentives, which play a significant role in their training process. Teaching-learning methods that satisfy the students will motivate them and facilitate the success on their way through University. In this sense, the present paper, through statistical and econometric procedures, analyses the degree of satisfaction perceived by university students in the new Degree in Business Administration and Management at the Universidad Politécnica de Cartagena with regard to several methodological procedures applied by University lecturers.

Keywords: Discrete choice models, European Space for Higher Education, Knowledge, Skills and Abilities, Higher Education.

1. Introducción

El 19 de Junio de 1999, veintinueve países europeos suscribieron la declaración de Bolonia. Ésta fija las bases para adaptar el variado sistema universitario existente en cada país, a un nuevo marco de educación común que se conoce como Espacio Europeo de Educación Superior (EEES) conforme a unos principios de calidad, movilidad, diversidad y competitividad. A partir de esta declaración se fija como uno de los principales papeles de la Universidad el formar a estudiantes y futuros profesionales, para que respondan a las demandas sociales y facilitar su incorporación al mercado laboral. En este sentido, se convierte en un reto importante analizar si, efectivamente, las metodologías docentes, que actualmente se vienen utilizando en la Universidad, satisfacen el nivel de conocimientos, competencias y habilidades que los alumnos demandan y desean adquirir, para poder acceder al mundo laboral con unas mínimas garantías de éxito.

Teniendo en cuenta este reto, el objetivo del presente trabajo se centra en analizar de qué factores depende *el grado de satisfacción que perciben los alumnos en relación a la adquisición de conocimientos, competencias y habilidades*, a través de las diferentes metodologías docentes de enseñanza-aprendizaje utilizadas en el Grado de Administración y Dirección de Empresas (GADE) de la Universidad Politécnica de Cartagena (UPCT). Para ello, se ha hecho uso de encuestas dirigidas a los alumnos de GADE a lo largo del curso 2011-2012, y se han utilizado diferentes procedimientos estadísticos y econométricos de tratamientos de datos mediante el programa Eviews.

El trabajo se divide en varias partes. En la Sección 2 se resumen algunas de los procedimientos metodológicos de enseñanza-aprendizaje más conocidos en el marco del EEES. En la Sección 3, se expone de manera resumida, los modelos econométricos que se emplean en el análisis empírico desarrollado en la Sección 4. Finalmente, en la Sección 5, se exponen las conclusiones finales del estudio llevado a cabo en el presente trabajo.

2. Metodologías docentes en el marco teórico del EEES

El nuevo EEES tiene como uno de los objetivos principales conseguir que los alumnos adquieran conocimientos y desarrollen competencias y habilidades a través del aprendizaje autónomo, guiado por la figura del profesor, que asume el rol de gestor del proceso de aprendizaje. La nueva concepción que subyace en las nuevas metodologías docentes aplicadas por el docente es la de “enseñar a aprender”.

La aplicación de los nuevos procedimientos metodológicos de enseñanza-aprendizaje, que surgen en el entorno del EEES, suele presentar una serie de problemas que devienen del espacio, del tiempo, de los recursos económicos, materiales y técnicos para su implantación y de la dificultad de adaptación, por parte de docentes y estudiantes. En este sentido, la coordinación de las actividades docentes, por parte de los profesores, se convierte en un aspecto crucial para que la aplicación de estos procedimientos consiga su finalidad.

Algunos de los procedimientos, más utilizados en el ámbito universitario, se resumen a continuación:

Clase Teórica o Magistral. Es el método tradicional más empleado en los centros docentes. Es el más sencillo, rápido y barato. Actualmente está siendo actualizado y modificado, respecto al modelo clásico de *lección magistral*, a la vez que complementado con otros métodos que implementan el nuevo concepto de enseñanza y aprendizaje, tratando de minimizar las desventajas del método clásico y maximizando sus ventajas. El método clásico de clase magistral ha recibido numerosas críticas por el hecho de que el profesor es el activo o emisor del saber y el alumno es el pasivo o receptor de la información. La transmisión de la información es unidireccional: de profesor a alumno. Este método clásico puede provocar una diferencia en el ritmo del profesor y el de los alumnos, aunque permite estructurar una gran cantidad de materia temática en poco tiempo y mostrarla a un numeroso grupo de estudiantes a la vez, permitiendo un pleno control del tiempo por parte del docente.

Prácticas de Laboratorio y en Aulas de Informática. Tradicionalmente se organizaba prefijando los pasos a seguir, los objetivos y los resultados a obtener, por lo que recibió críticas similares a la clase teórica clásica, ya que el profesor era quien dominaba la situación, la dirigía y controlaba, dejando poco margen a la iniciativa del alumno. Actualmente, y teniendo en cuenta los objetivos que el nuevo EEES establece, se aplica de tal manera que se permite al estudiante una mayor manipulación de materiales, programas, instrumentos, etc. invitándole a investigar por sí mismo y, por ende, incentivando su aprendizaje de manera autónoma.

Tutoría. Se trata de una entrevista del alumno o grupos de alumnos con el profesor para resolver dudas sobre la asignatura y la materia recibida. El aprendizaje, a través de las tutorías, puede convertirse en una especie de entrenamiento para el trabajo autónomo del estudiante y una ocasión para que el profesor evalúe el trabajo del mismo.

Asistencia a Conferencias o Seminarios. Son sesiones de trabajo interactivas entre estudiantes y profesor. Cuando son impartidas por expertos en determinadas materias, permite a los estudiantes adquirir una serie de conocimientos derivados de la experiencia propia del ponente. Cuando son preparados por estudiantes, éstos analizan un tema individualmente o colectivamente y lo presentan a los demás compañeros. En este sentido el seminario se convierte en un procedimiento de [aprendizaje](#) activo, ya que

los participantes no reciben la información ya elaborada, como se hace en la clase magistral, sino que tienen que buscarla por sus propios medios, por lo que promueven un ambiente de colaboración.

Aprendizaje Cooperativo y Colaborativo. Se trata del *Trabajo en grupos o equipos*. Las llamadas Tecnologías del Aprendizaje y del Conocimiento (*TACs*), consisten en crear grupos de alumnos que trabajan en equipo para maximizar el aprendizaje junto con el profesor, siendo evaluados según el resultado del trabajo del grupo. La distinción entre aprendizaje Cooperativo y Colaborativo es difícil, porque ambos procedimientos tienen significados muy parecidos. Algunos expertos no los diferencian y otros los distinguen con pequeños matices. Así, según Roger y Johnson (1987), señalan que el aprendizaje *cooperativo* se basa en el aprendizaje a través de la creación de grupos de alumnos que trabajan juntos para resolver un problema o encontrar una solución, que comparten la información y que se apoyan entre ellos para conseguir una tarea común. El profesor actúa como director, moderador y supervisor del grupo de alumnos, manteniendo en cierto modo su rol tradicional. De manera diferente, en el aprendizaje *colaborativo* tanto alumnos como profesores trabajan conjuntamente para alcanzar los objetivos. Existen multitud de técnicas dentro de estos métodos de aprendizaje. Algunas de las que destacan Barkley *et al.* (2007) son:

- *Rueda de ideas o tormenta de ideas.* A través de ella los integrantes del grupo responden individualmente a alguna pregunta o cuestión planteada por el profesor. Los alumnos responden con una frase corta o pocas palabras.
- *Grupos de conversación.* Se forman equipos para contestar a preguntas sobre un tema.
- *Entrega de ficha.* Los alumnos entregan una ficha cada vez que intervienen, lo que permite que todos puedan participar de igual forma, de manera equitativa.
- *Entrevistas por parejas.* Se entrevistan entre los alumnos y cada uno escribe, de forma concisa, las respuestas del compañero.
- *Críticas y debates.* Se forman grupos donde unos asumen ideas contrapuestas, defendiendo cada uno sus opiniones frente al resto.
- *Toma de apuntes por parejas.* Después de tomar apuntes, se reúnen por parejas y así revisan y mejoran lo que cada uno ha anotado.
- *Celdas de Aprendizaje.* Se trabaja por parejas, donde los alumnos se preguntan entre ellos sobre el tema estudiado. Ellos mismos corrigen sus fallos o errores.
- *Pecera o círculos.* Se hacen dos círculos de alumnos. Uno de los grupos inicia un diálogo o debate sobre algún tema y los otros analizan lo que los primeros hacen.
- *Juegos de Rol.* Los alumnos integrantes de un grupo asumen individualmente diferentes roles, actitudes o personalidades establecidas por el profesor. Actúan desde la perspectiva de esa personalidad, diferente a la que tienen por sí mismos. Se

ponen en situaciones imaginarias, pero que tratan de estar relacionadas con la realidad, lo que les permite aprender mientras intervienen.

- *Rompecabezas.* Cada uno de los alumnos estudia, aprende y domina una parte del tema. Posteriormente, cada uno de ellos expone su parte de forma correlativa o cronológica, dependiendo del producto, asunto o materia que se estudie o analice. Cada alumno se convierte en experto en su parte y es el protagonista de ella. A la hora de la exposición, cada uno debe estar atento a la parte de los demás para comprender y tener la visión del *todo*. Cada estudiante es necesario y esencial para el conjunto.
- *Equipos de exámenes.* Se divide en tres fases. En una primera fase, cada alumno estudia individualmente la materia. En una segunda fase, resuelve un examen también de manera individual. Finalmente, en una tercera fase, los estudiantes dan solución a los exámenes en grupo, lo que les permite beneficiarse de lo que han resuelto los demás.
- *Problemas por pares.* Un alumno intenta resolver un problema y otro escucha la resolución, corrigiéndole en el caso de que el ponente se haya equivocado.
- *Paso de problemas.* Se plantea un problema. Se le da a una pareja para que proponga sus soluciones y lo pase a otra. Esta nueva pareja, también propone sus soluciones, sin tener información sobre las respuestas de la anterior, y lo pasa a otra tercera pareja y así sucesivamente. Al final, la última pareja, observa todas las respuestas de cada una de las anteriores y expone a todo el grupo el resumen de los resultados más acertados, desde su opinión.
- *Resolución de problemas estructurados.* Los estudiantes aprenden a solucionar problemas. Cada uno aporta su opinión y/o el camino a seguir para su resolución. Unos aprenden y parten de las ideas de los otros y, al final, dan al profesor o al resto de la clase una solución consensuada del grupo.
- *Equipos de análisis.* Se le asigna una lectura, una clase, vídeo u otra información a un grupo de alumnos que, posteriormente, la analiza críticamente, la sintetiza, la resume, etc.
- *Investigación en grupo.* Se plantea un tema y los componentes del grupo investigan sobre el mismo, lo analizan y estudian detalladamente y luego lo exponen, creando con el resto de compañeros un debate sobre los resultados de la investigación.
- *Aprendizaje Basado en Equipos (ABE).* En él se establecen unas actividades para que los integrantes del equipo adquieran unos conocimientos y los apliquen. Se les hacen pruebas individualmente y en grupo y luego se les dan ejercicios prácticos para que apliquen los conocimientos. A continuación se exponen en clase los ejercicios y se corrigen en el momento, produciéndose un *feedback* o retroalimentación entre ellos.

- *El Aprendizaje Basado en Problemas (APB)*. Se parte de un problema y/o pregunta que se intenta resolver. El aprendizaje se consigue a través del proceso de resolución del problema. En ese proceso, se adquieren conocimientos, habilidades y competencias para dar solución a problemas reales. Un ejemplo de este tipo de procedimiento es la *revisión de casos*. El profesor se limita a identificar el problema, da las directrices y los guía de forma limitada. Los estudiantes buscan la información e intentan resolverlo. Tienen que encontrar los medios para llegar a la solución identificando las necesidades del aprendizaje, consiguiendo la “construcción” del conocimiento. Es un método más activo que los tradicionales, ya que, partiendo de la teoría, permite aplicarla, investigar y descubrir la solución a algo desconocido. Permite que los estudiantes asuman un papel más activo. Los alumnos son los protagonistas y tienen la responsabilidad de participar activamente en el proceso de aprendizaje. Según McKeachie y Svinicki (2011), “El aprendizaje basado en problemas, junto con el aprendizaje activo, cooperativo y colaborativo y la tecnología, es uno de los desarrollos más importantes en la educación superior contemporánea”.
- *Metodologías derivadas de las Tecnologías de la Información y la Comunicación (TICs)*. El uso de la tecnología, los avances técnicos e informáticos de que disponemos proporcionan nuevas formas de aprender, permitiendo compartir la información sin necesidad de coincidir, profesor y alumnos, en un mismo espacio físico ni temporal. Facilitan la comunicación rápida, eficaz, el intercambio de información, de trabajos, artículos, etc. y permiten poner en común ideas y actividades. En el ámbito universitario, los procedimientos más utilizados para el proceso formativo son:
 - *E-mail o correo electrónico*. Consiste en enviar y recibir mensajes y archivos a través de Internet, de forma rápida.
 - *Chats*. Se basan en la comunicación escrita entre dos o más personas a través de Internet ya sea de forma pública o privada. Se utilizan para ello diversos programas o *software*.
 - *Foros*. Son lugares o espacios en la Red en los que se promueven discusiones o se permite expresar opiniones en línea. Los usuarios o participantes en ellos pueden incorporar sus ideas, leer las de los otros participantes, comentarlas, responder, replicarlas, etc.
 - *Video-conferencias o video-llamadas*. Permiten la comunicación entre dos o más personas, a través de la utilización de audio y vídeo. Proporcionan muchas ventajas, por la facilidad de conexión entre personas o grupos distantes geográficamente, ya sea con fines privados, empresariales, de colaboración, estudio, etc.
- *Aula Virtual*. Facilitan la interrelación alumno-profesor-Universidad o centro docente en un espacio virtual bien a través de un registro o validación o bien con acceso totalmente libre a ciertos servicios. El alumno tiene total libertad de espacio y horario, evitando desplazamientos. A través de ella se ofrecen cursos, recursos,

aprendizaje, etc. mediante el uso de foros, correos, chats o wikis, etc.

- *Educlick*. Es un sistema interactivo de enseñanza y aprendizaje con mandos a distancia que permite hacer tests anónimos a los estudiantes con diferentes posibilidades de evaluación de manera individual o conjunta, obteniendo los resultados en el momento. A través de este procedimiento, se puede evaluar y controlar de forma rápida y continua el aprendizaje de los alumnos y, al mismo tiempo, establecer un *feedback* entre profesor y alumno.
- *Portafolio electrónico*. Consistente en la incorporación de documentos, actividades, trabajos, artículos por parte del alumno y de manera individual en un espacio en la Red. A través del portafolio, se puede ver el trabajo, la evolución del mismo, su actividad, su capacidad de aprendizaje, etc. Se realiza de forma continua durante un período de tiempo fijado, incorporando material de forma regular y periódica.
- *Wikis*. Se basan en colaborar y compartir la construcción del conocimiento, generalmente de forma anónima, entre iguales. La más conocida es la Wikipedia.
- *Weblogs o Blogs*. Son sitios en Internet donde se publican anotaciones, artículos, historias e incluso *links* a otras páginas. Pueden ser creados por personas en particular, empresas u organismos y la información se va añadiendo por orden cronológico, permitiendo la participación de los lectores.

3. Modelos de elección discreta

La naturaleza cualitativa y no continua de la variable objeto de estudio en el presente trabajo, como es el *grado de satisfacción de los alumnos, en relación a las metodologías docentes utilizadas en el Grado en ADE*, lleva a que su proceso generador se modelice a través de los denominados *modelos de elección discreta*. Estos modelos pertenecen al grupo de *modelos de variable dependiente limitada*. Su objetivo no es tanto la determinación del comportamiento medio de la población sino obtener el valor de la probabilidad de que el individuo promedio adopte una elección determinada dados unos valores de las variables explicativas.

En el presente estudio, nos centramos en los *Modelos de Elección Discreta categóricos ordenados*, debido a que la variable dependiente posee un rango natural de posibles alternativas. La estimación de un modelo lineal de probabilidad, es decir, la estimación de un modelo donde la variable dependiente posee alternativas mutuamente excluyentes a través del procedimiento de Mínimos Cuadrados Ordinarios (MCO), no sería adecuada debido a que ajusta un modelo lineal a la nube de puntos como la que aparece en la Figura 1.

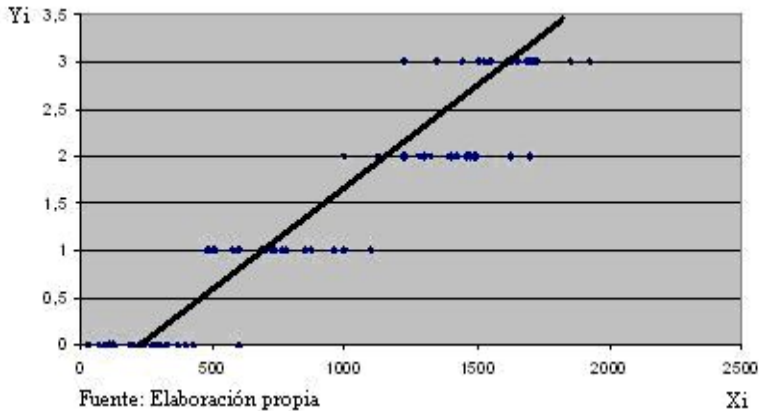


Figura 1: Recta de regresión ajustada a una variable de respuesta múltiple.

Este procedimiento de estimación presenta limitaciones considerables, debido a la naturaleza discreta y no continua de la variable dependiente (Y_i):

- El valor estimado del regresando puede estar fuera del rango de valores de la variable dependiente, pudiendo ser, incluso, negativo.
- La perturbación aleatoria no sigue una distribución normal y su varianza no es constante.
- Los estimadores de los coeficientes de regresión presentan sesgo.
- El coeficiente de determinación (R^2) está subestimado.

Ello obliga a estimar estos modelos por métodos alternativos a MCO, por ejemplo a través de máxima verosimilitud. Para ello, haremos uso de funciones no lineales que aseguren la restricción de las probabilidades, como son la función normal (modelo probit) y la función logística (modelo logit). Estos modelos son modelos de elección discreta en los que la variable dependiente Y_i toma los valores de 0 a J y permiten una relación no lineal con los regresores.

El modelo de respuesta múltiple ordenado relaciona la variable dependiente Y con las distintas variables explicativas X a través de la siguiente ecuación:

$$Y^* = \Xi(\beta'X) + u \quad (1)$$

donde:

Y^* : es una variable latente, no observada y no limitada en su rango, que cuantifica las distintas categorías que puede tomar nuestra variable dependiente.

$\Xi(\square)$: es una función no lineal de tipo logística (Λ), en el modelo logit, y de tipo normal, en el modelo probit (Φ).

$\beta' X$: es una combinación lineal de las variables o características a la que se le denomina 'índice del modelo', siendo β el vector de parámetros y X la matriz de regresores.

u : es el término de perturbación aleatoria.

Como la variable latente Y^* es inobservable, se utiliza, en su lugar, una variable aproximada que sí se puede observar a la que denotamos como Y , y que se define de la siguiente manera:

$$Y = \begin{cases} 0 & \text{si } Y^* \leq c_1 \\ 1 & \text{si } c_1 \leq Y^* \leq c_2 \\ \vdots & \dots \\ J & \text{si } c_{J-1} \leq Y^* \end{cases} \quad (2)$$

donde:

c_1, c_2, \dots, c_{J-1} son los puntos de corte o valores de los umbrales o barreras desconocidos (*threshold parameters*), que deben cumplir la siguiente restricción: $0 < c_1 < c_2 < \dots < c_{J-1}$.

Las funciones de distribución que permiten relacionar a la variable dependiente Y , con las variables explicativas X , son las siguientes:

Modelo Logit:

$$Y = \frac{e^{(\beta' X)}}{1 + e^{(\beta' X)}} + u = \Lambda(\beta' X) + u \quad (3)$$

Modelo Probit:

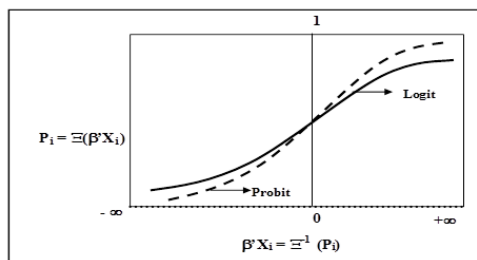
$$Y = \int_{-\infty}^{\beta' X_0} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{(\beta' X)^2}{2}\right)} d(\beta' X) + u = \Phi(\beta' X) + u \quad (4)$$

El criterio de selección entre las diferentes opciones o categorías en un modelo de elección binaria o multinomial, depende de las probabilidades asociadas a cada una de las posibles alternativas que tiene el sujeto.

La distribución condicional de Y dada la matriz X de regresores, es la siguiente:

$$\begin{aligned}
 \text{Prob}(Y = 0/\beta' X) &= \Xi(c_1 - \beta' X) \\
 \text{Prob}(Y = 1/\beta' X) &= \Xi(c_2 - \beta' X) - \Xi(c_1 - \beta' X) \\
 \text{Prob}(Y = 2/\beta' X) &= \Xi(c_3 - \beta' X) - \Xi(c_2 - \beta' X) \\
 &\dots \\
 &\dots \\
 \text{Prob}(Y = J/\beta' X) &= 1 - \Xi(c_J - \beta' X)
 \end{aligned} \tag{5}$$

La función de distribución logística es muy parecida a la normal, excepto por las colas. Tal y como se puede ver en la Figura 2, las dos distribuciones dan unos valores de probabilidad parecidos en los valores medios de $\beta' X$. La distribución logística (del modelo logit) proporciona probabilidades más altas que la normal (del modelo probit), para $Y=J$, cuando $\beta' X$ es muy pequeño y, al contrario, menores que la distribución normal cuando $\beta' X$ es muy grande. Para poder resolver esta cuestión, habría que conocer el valor de β . Puede haber razones para optar por una u otra, como su facilidad de cálculo o la facilidad de operar con ella, pero, según Greene y Hensher (2010) “elegir entre un modelo u otro, continúa siendo una cuestión teórica sin resolver y difícil de justificar”.



Fuente: Elaboración propia.

Figura 2. Funciones de distribución acumuladas logística y normal

La interpretación de los coeficientes en un modelo de elección discreta de respuesta múltiple es más complicada que en otros tipos de modelos, véase Daykin y Moffatt (2002), ya que no se corresponden con los efectos marginales, es decir, no indican la magnitud de la variación en la probabilidad de respuesta sobre Y de variaciones marginales en X . La variable dependiente (Y) no tiene una media natural en el modelo que se pueda analizar, ya que no existe una función media condicional natural. Los valores que toma la variable explicada son simplemente etiquetas o categorías. Por lo tanto, según Greene (2008), no se podría analizar la función como el valor esperado promedio de Y , condicionado al valor de los regresores. Ni el signo, ni la magnitud de los coeficientes, en los modelos de elección múltiple, nos informan del efecto marginal, por lo que la interpretación de los coeficientes es ambigua, véase

Greene y Hensher (2010). Sin embargo, en el caso de los modelos de respuesta múltiple de variables dependientes ordenadas, el signo de los coeficientes sí indica el sentido del cambio que experimenta la probabilidad de que se dé el suceso J . Es decir, si el coeficiente de la variable explicativa es positivo, incrementos marginales en la variable explicativa, aumentan la probabilidad de Y en la categoría J más alta y disminuye en la más baja (0). Es por ello por lo que el signo del efecto marginal es diferente en la categoría más baja (0) respecto a la más alta (J), coincidiendo con el signo del efecto marginal en la categoría más elevada. A pesar de ello, a partir de los coeficientes estimados no se conoce la magnitud de dicho efecto marginal, ni el signo para categorías inferiores a J .

En muchos estudios, es de gran interés conocer la magnitud y signo de los efectos de variaciones de X en las probabilidades de respuesta: $P(Y = j | \beta' X)$, para $j = 1, 2, \dots, J$, o *efecto marginal*. Sin embargo, no se calcula de la misma manera si las variaciones de X vienen provocadas por variables de naturaleza cuantitativa o cualitativa.

Así, el efecto marginal, en los modelos de elección múltiple ordenados, para el caso de las *variables explicativas continuas o cuantitativas* (X_i), es la variación o derivada en la probabilidad de que la variable dependiente Y adopte el valor de cada una de las categorías ($j = 0, 1, 2, \dots, J$) ante un cambio unitario de alguna de las variables explicativas (X_i), manteniendo el resto constantes.

De forma esquemática, su cálculo sería el siguiente:

$$\begin{aligned} \frac{\partial \text{Prob}(Y = 0)}{\partial X_i} &= -\xi(\beta' X)\beta_i \\ \frac{\partial \text{Prob}(Y = 1)}{\partial X_i} &= -\xi(c_1 - \beta' X)\beta_i + \xi(\beta' X)\beta_i \\ &\dots \\ \frac{\partial \text{Prob}(Y = J)}{\partial X_i} &= \xi(c_{J-1} - \beta' X)\beta_i, \end{aligned} \quad (6)$$

siendo $i = 1, 2, \dots, k$, k el total de regresores de naturaleza cuantitativa y continuos del modelo y ξ la función de densidad del modelo.

Como contrapartida, cuando se trata de *variables explicativas discretas* (X_l) que toman valor 0 o 1, el efecto marginal se calcula como diferencias de probabilidad de la función entre sus valores, cuando X_k pasa de 0 a 1, considerando que el resto de valores de las restantes variables explicativas permanecen fijos.

De forma esquemática, su cálculo sería el siguiente:

$$\begin{aligned} \Delta \text{Prob}(Y = 0, \beta' X) &= \Xi(c_1 - \beta' X + \gamma) - \Xi(c_1 - \beta' X) \\ \Delta \text{Prob}(Y = 1, \beta' X) &= [\Xi(c_2 - \beta' X + \gamma) - \Xi(c_1 - \beta' X + \gamma)] - [\Xi(c_2 - \beta' X) - \Xi(c_1 - \beta' X)] \\ \Delta \text{Prob}(Y = 2, \beta' X) &= [1 - \Xi(c_2 - \beta' X + \gamma)] - [1 - \Xi(c_2 - \beta' X)], \end{aligned} \quad (7)$$

siendo \square el coeficiente de la variable explicativa dicotómica para la que se calcula el efecto marginal (X_l),

$l=k+1, \dots, K$ es el total de variables explicativas dicotómicas del modelo y Ξ la función de distribución del modelo.

4. Datos y resultados del análisis empírico

4.1. Descripción de las variables:

Para llevar a cabo el estudio empírico, se seleccionan diez variables dependientes, que corresponden a las cuestiones sobre los diez métodos de enseñanza y aprendizaje más utilizados (Y1 a Y10)¹ en el Grado en ADE de la Facultad de Ciencias de la Empresa de la UPCT (ver Tabla 1).

Tabla.1. Variables Dependientes utilizadas

Y1 - Exposición Teórico-Práctica del Profesor
Y2 - Prácticas en el aula de informática
Y3 - Exposición en clase de Trabajos por Grupos
Y4 - Revisión y Análisis de CASOS reales de Empresas
Y5 - Portafolio : Recopilación y Análisis de datos e Información Económica/Empresarial
Y6 - Foros en Aula Virtual
Y7 - Tests / Cuestionarios en el Aula Virtual
Y8 - Tests / Cuestionarios Anónimos con Educlick
Y9 - Asistencia a Conferencias/Seminarios
Y10- Tutorías

Para medir “el grado de satisfacción en relación a los conocimientos, competencias y habilidades adquiridas en el proceso de aprendizaje de los alumnos” (variable latente e inobservable Y^*), se ha utilizado como variable aproximada (Y) “el grado de conocimientos, competencias y habilidades que el alumno considera que ha adquirido” cuando el docente utiliza las diferentes metodologías formativas expuestas

¹ Se escogen tras una encuesta realizada a profesores y alumnos.

en la Tabla 1. Se permite que esta variable tome 5 categorías o posibilidades: “*Muy bajo*”, “*Bajo*”, “*Medio*”, “*Alto*” y “*Muy alto*”, las cuales se codifican mediante 5 dígitos o categorías que van de 0 a 4, respectivamente.

Como variables explicativas del comportamiento de la variable dependiente, se consideran algunas de naturaleza cualitativa y otras cualitativas (ver Tabla 2). Las variables explicativas cualitativas escogidas toman valor 1 o 0 exclusivamente, indicando si el entrevistado pertenece o no a la categoría. Las variables explicativas cuantitativas son continuas, aunque su valor se halla acotado, dentro de lo razonable. Así, la variable *Edad*, toma en nuestra encuesta valores entre 18 y 50. Las variables *Notas* medias de cada alumno y *Horas* diarias de estudio, toman valores comprendidos entre 0 y 10.

Tabla.2. Variables independientes o explicativas utilizadas.

Curso1	(Variable Dicotómica: 1 Pertenece al curso ; 0 Resto de Cursos)
Curso2	(Variable Dicotómica: 1 Pertenece al curso ; 0 Resto de Cursos)
Curso3	(Variable Dicotómica: 1 Pertenece al curso ; 0 Resto de Cursos)
Cursopas	(Variable Dicotómica: 1 Pertenece al curso ; 0 Resto de Cursos)
Sexo	(Variable Dicotómica: 1 Hombre ; 0 Mujer)
Edad	(Variable Cuantitativa: (De 18 a 50).
Trabaja	(Variable Dicotómica: 1 Trabaja ; 0 No Trabaja)
Hijos	(Variable Dicotómica: 1 Tiene hijos ; 0 No Tiene)
Español	(Variable Dicotómica: 1 Español ; 0 Resto de Nacionalidades)
AmericHisp	(Variable Dicotómica: 1 HispanoAmericano ; 0 Resto de Nacionalidades)
AmericnoHisp	(Variable Dicotómica: 1 Americano No hispanoparlante ; 0 Resto de Nacionalidades)
África	(Variable Dicotómica: 1 Africano ; 0 Resto de Nacionalidades)
Asiático	(Variable Dicotómica: 1 Asiático ; 0 Resto de Nacionalidades)
Europeo	(Variable Dicotómica: 1 Europeo No Español ; 0 Resto de Nacionalidades)
Notas	(Variable Cuantitativa: Nota media de las asignaturas aprobadas hasta el momento 0 a 10).
Horas	(Variable Cuantitativa: Horas diarias que dedica el alumno al estudio (0 a 10).

4.2. Datos.

El modelo de la encuesta que se pasó a los alumnos del Grado en ADE de la UPCT en el curso 2011-2012, fue el siguiente:

Tabla.3. Modelo de Encuesta realizada sobre las metodologías en el grado de ADE de la UPCT.

Encuesta sobre las Metodologías de enseñanza y aprendizaje en el Grado de A.D.E. de la UP.C.T.					
<p>Por favor, dedica unos 2 minutos a completar este pequeño cuestionario. La información que nos proporciones será utilizada para conocer el Grado de Adquisición de conocimientos/habilidades/competencias de los alumnos/as en relación a las metodologías de enseñanza y aprendizaje aplicadas en el Grado en Administración y Dirección de Empresas (A.D.E.) de la UPCT.</p> <p>Tus respuestas serán tratadas de forma confidencial Señale con un aspa la opción seleccionada:</p> <p>1. Curso en el que está matriculado principalmente : <input type="checkbox"/> 1º <input type="checkbox"/> 2º <input type="checkbox"/> 3º <input type="checkbox"/> Pasarela a GADE</p> <p>2. Sexo : H <input type="checkbox"/> M <input type="checkbox"/></p> <p>3. Edad : _____ Años</p> <p>4. ¿Trabaja? Si <input type="checkbox"/> No <input type="checkbox"/></p> <p>5. ¿Tiene hijos? Si <input type="checkbox"/> No <input type="checkbox"/></p> <p>6. ¿Español? Si <input type="checkbox"/> No <input type="checkbox"/></p> <p>7. En caso de ser extranjero señale si es: <input type="checkbox"/> Americano hispanoparlante <input type="checkbox"/> Americano no hispanoparlante <input type="checkbox"/> Asiático <input type="checkbox"/> Africano <input type="checkbox"/> Europeo</p> <p>8. Nota Media (aproximada) hasta el momento en la carrera (Escala entre 0 y 10) : _____</p> <p>9. Número de horas diarias, por término medio, dedicadas al estudio en casa/biblioteca : _____</p> <p>10. Determine el Grado de Adquisición de Conocimientos, competencias y habilidades que considera que ha adquirido al aplicar, en las asignaturas cursadas, las siguientes actividades docentes: (Nota: Si no se le ha aplicado a lo largo de la carrera alguna de esas metodologías, por favor, deje los cuadros en blanco relativos a esa actividad):</p>					
	Muy Bajo	Bajo	Medio	Alto	Muy Alto
Exposición Teórico-Práctica del Profesor	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Prácticas en el aula de informática	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Exposición en clase de Trabajos por Grupos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Revisión y Análisis de CASOS reales de Empresas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Portafolio: Recopilación y Análisis de datos e Información Económica/Empresarial	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Foros en Aula Virtual	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tests / Cuestionarios en el Aula Virtual	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tests /Cuestionarios Anónimos con Educlick	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Asistencia a Conferencias/Seminarios	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tutorías	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Se recogieron 243 encuestas válidas, correspondientes a respuestas de 142 son mujeres y 101 varones, 105 alumnos pertenecen al primer curso, 75 al segundo y 60 al curso tercero y 3 al curso pasarela². Del total de estudiantes, 48 alumnos trabajan y 8 tienen hijos. En cuanto a su nacionalidad, 216 son españoles, 13 Americanos Hispanoparlantes, 9 Europeos no españoles y 5 africanos. La media de edad es de 21,29 años, la de las notas es de 5,41 puntos sobre diez y las horas medias de estudio son 2,47. Las frecuencias relativas y absolutas se muestran en la Tabla.4.

Tabla.4. Frecuencias de las variables explicativas.

Variables explicativas	Frecuencias relativas	Frecuencias absolutas
Alumnos		243
Curso 1	43,21 %	105
Curso 2	30,86 %	75

² Este grupo está formado por alumnos que pertenecen a otras titulaciones y que aprobando el curso *pasarela* obtienen el Grado en Administración y Dirección de Empresas.

Curso 3	24,70 %	60
Curso Pasarela	1,23 %	3
Hombres	41,56 %	101
Mujeres	58,44 %	142
Media de Edad		21,29
Trabajan	19,75 %	48
Con Hijos	3,29 %	8
Españoles	88,89 %	216
Amer.Hisp.Parlantes	5,35 %	13
Europeo-no español	3,70 %	9
Africanos	2,06 %	5
Media de Notas		5,41
Media de Horas		2,47

El tamaño de la muestra (n) es apropiado para poder extrapolar el estudio a la población (N) formada por 507 alumnos matriculados en el Grado en ADE. Teniendo en cuenta que la población se considera finita (al estar formada por menos de 100.000 observaciones), considerando un margen de error (e) cuando extrapolemos los resultados a la población del 4,54%, un nivel de confianza estándar del 95% (o nivel de significación del 5%), al que le corresponde un valor $z=1,96$, y suponiendo un valor de la varianza máximo $pq=0,25^3$, el tamaño muestral (n) al aplicar la fórmula siguiente:

$$n = \frac{N}{1 + \frac{e^2(N-1)}{z^2 pq}} \quad (8)$$

es de 243. Cabe señalar que si se considerara un nivel de significación superior, el tamaño muestral sería inferior para los mismos valores de e , pq y N .

4.3. Estimación de los modelos logit y probit

Para analizar la probabilidad de que un alumno elija alguna opción de Y dados unos valores de los regresores, se procede a estimar diez modelos logit y otros diez probit, para cada una de las variables dependientes indicadas en la Tabla 1. Los criterios para seleccionar las variables explicativas de cada modelo se basan en la información obtenida a partir de diversos criterios de información, medidas de bondad de ajuste y contrastes de significatividad, que se exponen en la subsección 4.4.

En las tablas 5 y 6, se muestran los resultados de las estimaciones de los modelos logit y probit, respectivamente:

³ Morales, P. (2012) argumenta que este dato de la varianza es válido para variables dicotómicas, pero se puede utilizar para el caso de que no sean dicotómicas.

Tabla 5. Coeficientes estimados de los modelos Logit.

LOGIT										
Variables Dependientes	Variables explicativas									
	CURS O1	CURS O2	CURS O3	SEXO	EDAD	TRABA JA	HIJOS	ESPAÑ OL	NOTA S	HORA S
Y1	1,1492* **	-	-	-0,2848	0,0625	-	-	-	0,1489	-0,9997
Y2	0,7838	1,7027	1,79543	0,4845*	-0,0012	0,9462***	0,2628	0,0935	0,0271	-0,1171
Y3	1,0039* **	-	-	-	0,0355	0,3962	-1,1016	-	0,1699	0,1997* *
Y4	-0,4524	-0,4355	-	0,2175	0,1008* **	0,7557**	0,1574	-0,0493	0,1307	-0,1275
Y5	0,4288	-	-	-	-0,0254	0,1833	-	-	0,3382* **	-0,0537
Y6	0,8306* *	-	-	0,2035	-0,0418	-	1,2713	-	0,0124	-
Y7	-	-	-	0,6521* **	-	-	-	-	-	-
Y8	0,5307	0,4300	-	1,0518* **	-0,0579	-0,5318	-1,0569	-0,4966	0,2064*	0,14745 *
Y9	-1,4262*** 0,7036*	-	-	-0,1519	-0,0486	-	1,3801 *	-	0,1156	0,0485
Y10	-	-	-	-0,2811	0,0118	-0,0545	-1,2013	-	0,3699* **	0,0644
*	Variable significativa al 10 %									
**	Variable significativa al 5 %									
***	Variable significativa al 1 %									

Tabla 6. Coeficientes estimados de los modelos Probit.

PROBIT										
Variables Dependientes	Variables explicativas									
	CURS O1	CURS O2	CURS O3	SEXO	EDAD	TRABA JA	HIJOS	ESPAÑ OL	NOTA S	HORA S
Y1	0,6062* **	-	-	-0,2015	0,0323	-	-	-	0,0684	-0,0495
Y2	0,5368	1,0372	1,0785	-0,2870*	0,0004	0,5473***	0,1112	0,0795	0,0134	-0,0571
Y3	0,5028* **	-	-	-	0,0191	0,1650	0,5697	-	0,0899	0,1123* *
Y4	-0,2872	-0,3482	-	0,1173	0,0638* **	0,4458**	0,1268	-0,0297	0,0861	-0,0750
Y5	0,2346	-	-	-	-0,0167	0,1128	-	-	0,2047* **	-0,0451
Y6	0,4930*	-	-	0,1004	-0,0268	-	0,8007	-	0,0040	-

	*									
Y7	-	-	-	0,3876*	-	-	-	-	-	-
Y8	0,3285	0,2017	-	0,5627*	-0,0283	-0,3329*	0,7252	-0,3036	0,1304*	0,0895*
Y9	0,7928*	-	-	-0,0971	-0,0265		0,7766	-	0,0805	0,0265
Y10	0,4133*	-	-	-0,1948	0,0027	-0,0736	0,6616	-	0,2169*	0,0248
	Variable significativa al 10 %									
*	Variable significativa al 5 %									
**	Variable significativa al 1 %									

Tal y como se puede observar en las tablas anteriores, del total de regresores considerados en la encuesta y en los modelos de regresión, sólo un número no muy amplio de ellos han sido significativos, individualmente considerados, para niveles de significación inferiores al 10%.

Aunque el signo de los coeficientes en los modelos logit y probit ordenados coincide con el del efecto marginal en la categoría más elevada de la variable dependiente, se procede al cálculo de su magnitud o impacto en la Subsección 4.5. Debido a que las diferencias entre los resultados obtenidos de la estimación de los modelos logit y probit no difieren mucho y con el fin de facilitar la interpretación de los resultados, se selecciona el modelo logit para el cálculo de los efectos marginales, ya que presenta un comportamiento marginalmente mejor, teniendo en cuenta los criterios de información, medidas de bondad de ajuste y contrastes de hipótesis expuestos en la Subsección 4.4.

4.4. Criterios de información, medidas de bondad de ajuste y contrastes de hipótesis

En las tablas 7 y 8, se expone la información siguiente:

Estadísticos de selección de modelos:

- Criterio de Información de Akaike (*AIC*)
- Criterio de Información de Schwarz (*SC*)
- Criterio de Información de Hannan-Quinn (*HQ*)
- Logaritmo de la Función de Verosimilitud (*Log likelihood*)

Contrastes de hipótesis:

- Nivel de significación exacto del estadístico de razón de verosimilitud para contrastar la significatividad conjunta del modelo (*p-valor. LR Statistic*).

Estadísticos de bondad de ajuste:

- Porcentaje de predicciones correctas (*Evaluac. de Predicción. % Correct*)
- Pseudo R² (*McFadden*)

Tabla 7. Criterios de Información, medidas de bondad de ajuste y contraste de significatividad conjunta.
Modelo Logit

Variables dependientes	Selección de modelos Logit					Pseudo R2 (McFadden)	Evaluac. de Predicción
	p-valor (LR Statistic)	AIC	SC	HQ	Log Likelihood		% Correct
Y1	0,0233	2,0083	2,161628	2,0702	-212,9274	0,0332	59,4590
Y2	0,0023	2,6174	2,831999	2,7040	-276,5332	0,0470	39,1890
Y3	0,0143	2,6002	2,753037	2,6619	-279,9278	0,0275	52,0180
Y4	0,0274	2,8555	3,064882	2,9402	-282,5526	0,0321	39,1300
Y5	0,0681	2,5225	2,673110	2,5835	-238,2133	0,0210	44,8980
Y6	0,0600	2,7821	2,929586	2,8418	-272,0010	0,0191	40,5940
Y7	0,0088	2,7740	2,850423	2,8048	-304,3042	0,0111	38,1170
Y8	7,41E-06	2,7982	3,002077	2,8806	-287,8142	0,0650	45,5810
Y9	0,0000	2,6839	2,847125	2,7499	-262,4172	0,0587	38,9160
Y10	0,0395	2,8366	3,009101	2,9063	-293,9399	0,0244	41,8600

Tabla 8. Criterios de Información, medidas de bondad de ajuste y contraste de significatividad conjunta.
Modelo Probit

Variables dependientes	Selección de modelos Probit					Pseudo R2 (McFadden)	Evaluac. de Predicción
	p-valor (LR Statistic)	AIC	SC	HQ	Log Likelihood		% Correct
Y1	0,0398	2,0147	2,1679	2,0766	-213,6344	0,0299	59,4590
Y2	0,0040	2,6243	2,8389	2,7110	-277,3081	0,0444	37,8380
Y3	0,0427	2,6131	2,7659	2,6748	-281,3632	0,0226	48,8790
Y4	0,0189	2,8503	3,0596	2,9349	-282,0094	0,0339	40,5940
Y5	0,0391	2,5152	2,6657	2,5761	-237,4932	0,0240	44,3880
Y6	0,0662	2,7834	2,9308	2,8431	-272,1300	0,0186	40,5940
Y7	0,0073	2,7725	2,8504	2,8048	-304,3042	0,0116	38,1170
Y8	2,12E-05	2,8100	3,0138	2,8924	-289,0817	0,0609	46,0470
Y9	0,0000	2,6801	2,8433	2,7461	-262,0363	0,6007	38,9160
Y10	0,0350	2,8350	3,0075	2,9047	-293,7722	0,0250	42,3260

Tal y como se puede ver en las tablas 7 y 8, los estadísticos que permiten comparar modelos (*AIC*, *SC*, *HQ*, *Log likelihood*) y medidas de bondad de ajuste (*Pseudo R²* y *% correct*), presentan valores muy parecidos entre los modelos logit y probit, aunque parecen sensiblemente mejores en el logit. Es por ello, por lo que se procede al cálculo de los efectos marginales para este modelo y para las variables explicativas que son individualmente significativas, tal y como se ha expuesto anteriormente.

4.5. Efectos Marginales

El efecto marginal, en los modelos de elección múltiple ordenados mide la variación en la probabilidad de que la variable dependiente *Y* adopte el valor de cada una de las 5 categorías (en nuestro estudio de 0 hasta 4), ante un cambio de una unidad de medida de las variables explicativas de naturaleza continua o cuando cambia del valor 0 a 1, en las dicotómicas.

Para los *regresores dicotómicos o cualitativos*, se generan dos índices del modelo ($\beta' X$). Uno considera que el regresor toma valor 1 y otro que toma valor 0. Estos índices se calculan en función de los coeficientes estimados, y se calculan fijando los valores del resto de variables explicativas, considerando el individuo de referencia, para el caso de las variables dicotómicas, y el valor medio, para las variables de naturaleza cuantitativa. El efecto marginal se obtiene como la diferencia en la probabilidad de que la variable dependiente *Y* tome el valor de cada una de las 5 categorías posibles, ante un cambio de 0 a 1 en las dicotómicas a las que se refiere y considerando constantes el resto de valores del resto de regresores.

Para los *regresores continuos o cuantitativos*, también se genera el índice del modelo, para el individuo de referencia y resto de variables explicativas calculadas en el punto medio. Se calcula como la variación o derivada en la probabilidad de que la variable dependiente *Y* adopte el valor de cada una de las categorías (en nuestro estudio de 0 hasta 4) ante un cambio unitario de alguna de las variables explicativas, manteniendo constantes el resto de regresores.

En la Tabla 9 se muestran los resultados obtenidos del cálculo de los Efectos Marginales de cada una de las regresiones. Como se puede ver en la misma, la suma de los efectos marginales para cada variable explicativa es cero, lo que deviene por el hecho de que la suma de las probabilidades tienen que ser uno y también el signo de los efectos marginales es opuesto cuando se calcula el efecto parcial para la categoría más baja ($Y=0$) con respecto a la más alta ($Y=4$), coincidiendo el signo de esta última categoría con el del coeficiente de regresión asociado a cada variable explicativa.

La *interpretación de los efectos marginales* expuestos en la Tabla 9, la clasificamos en función de la variable dependiente, para el caso de regresores individualmente significativos y considerando que el resto de variables explicativas se mantienen constantes, como se expone a continuación:

- *Exposición Teórico-Práctica (Y1):*

La probabilidad de que, como respuesta a la pregunta planteada, el individuo elija la opción 4 (grado de satisfacción “muy alto”) aumenta en un 2,91% cuando el valor del regresor (al ser dicotómico) cambia de 0 a 1, es decir cuando se pasa de no pertenece al primer curso del Grado en ADE respecto a sí pertenecer al mismo. Por el contrario, la probabilidad de que los alumnos con hijos, frente a los que no tienen, tomen la opción “muy alto” disminuye en un 1,23%.

- *Prácticas en el aula de informática (Y2):*

La probabilidad de que los varones elijan la opción “muy alto”, respecto a las mujeres, disminuye en un 0,17%. También se reduce esta probabilidad si los alumnos trabajan, frente a los que no lo hacen. En particular, la reducción es del 0,27%.

- *Exposición en clase de trabajos por grupos (Y3):*

La probabilidad de que los alumnos de primer curso elijan la opción “muy alto” aumenta en un 5,49%, respecto a los alumnos del resto de cursos. Asimismo, se incrementa, pero sólo en un 0,68%, la probabilidad elegir la opción 4 por cada hora de estudio diaria que dediquen más los alumnos.

Tabla 9. Efectos marginales

Variable Dependiente	Regresores	Prob. Y=0	Prob. Y=1	Prob. Y=2	Prob. Y=3	Prob. Y=4
Y1	curso1	-0,017777590	-0,037627922	-0,203368955	0,229640730	0,029133737
	D sexo	0,008580667	0,016610695	0,027079466	-0,048592162	-0,003678667
	C edad	-0,001628369	-0,003249031	-0,007278554	0,011253803	0,000902151
	D hijos	0,113627646	0,157330551	-0,062311858	-0,196291060	-0,012355279
	C notas	-0,003875432	-0,007732523	-0,017322576	0,026783459	0,002147072
	C horas	0,002601085	0,005189860	0,011626444	-0,017976332	-0,001441056
	Y2	curso1	-0,051584826	-0,115632057	0,074544426	0,088359650
D curso2		-0,086974353	-0,278352264	0,079455667	0,268456902	0,017414048
D curso3		-0,089347119	-0,293539201	0,071920350	0,291231303	0,019734668
D sexo		0,057334443	0,051828932	-0,066165701	-0,041281146	-0,001716529
C edad		0,000138957	0,000167710	-0,000177642	-0,000123750	-0,000005275
D trabaja		0,130773479	0,064019781	-0,124378445	-0,067682267	-0,002732548
D hijos		-0,027226797	-0,049047745	0,038995620	0,035653992	0,001624930
D espaol		-0,008491661	-0,013047832	0,011710838	0,009413287	0,000415369
C notas		-0,002911884	-0,003514399	0,003722532	0,002593207	0,000110545
C horas		0,002601085	0,005189860	0,011626444	-0,017976332	-0,001441056
Y3	curso1	-0,038740168	-0,083422754	-0,117194685	0,184436971	0,054920636
	D trabaja	-0,019438025	-0,038689652	-0,031170184	0,073305073	0,015992788
	C edad	-0,002107341	-0,003873710	-0,001652294	0,006427135	0,001206209
	D hijos	0,105979276	0,131268787	-0,054653748	-0,159196499	-0,023397816
	C notas	-0,010084106	-0,018536588	-0,007906602	0,030755309	0,005771988
	C horas	-0,011853595	-0,021789260	-0,009293998	0,036152037	0,006784816
	Y4	curso1	0,052268851	0,053698686	-0,046701475	-0,049801652
D curso2		0,053753210	0,054798922	-0,048039819	-0,050862757	-0,009649557
D sexo		-0,021702236	-0,033001365	0,016286318	0,031612635	0,006804648
C edad		0,010466450	0,014380732	-0,008774852	-0,013346654	-0,002726676
D trabaja		-0,058847388	-0,110130663	0,023929445	0,1158774416	0,029174190
D hijos		-0,005811771	-0,008110005	0,004740651	0,007606458	0,001574667
D espaol		0,005309496	0,006981835	-0,004491020	-0,006486650	-0,001313660
C notas		-0,020137354	-0,011704643	0,017709174	0,012081245	0,002051578
C horas		0,019651608	0,011422308	-0,017282001	-0,011789826	-0,002002091

Y5	curso1	-0,031379926	-0,073312154	0,048158420	0,054449088	0,002084571
C	edad	0,002151189	0,004123752	-0,003496877	-0,002682107	-0,000095957
D	trabaja	-0,014221464	-0,029664132	0,022919855	0,020223838	-0,000741902
C	notas	-0,026638872	-0,055796411	0,043061792	0,037983628	0,001389862
C	horas	0,004229682	0,008859275	-0,006837291	-0,006030987	-0,000220680
Y6	curso1	0,188441798	-0,010870375	-0,108103613	-0,060295453	-0,009172356
D	sexo	-0,039633319	-0,013185795	0,026351330	0,022626257	0,003841526
C	edad	0,008101826	0,002008142	-0,005309073	-0,004120862	-0,000680033
D	hijos	-0,178007403	-0,138866547	0,082694165	0,190890417	0,043289368
C	notas	-0,002410142	-0,000597385	0,001579350	0,001225879	0,000202297
Y7	sexo	0,093863277	0,064897424	-0,079227878	-0,066342825	-0,013189998
Y8	curso1	-0,023026947	-0,068949400	-0,039976868	0,100728680	0,031224534
D	curso2	-0,018858188	-0,055153724	-0,027664427	0,078734195	0,022942144
D	sexo	0,088996742	0,151713134	-0,051009411	-0,160122171	-0,029578294
C	edad	0,003064706	0,008129746	0,001952334	-0,010579332	-0,002567454
D	trabaja	0,036012941	0,078945834	-0,005759181	-0,090583971	-0,018615623
D	hijos	0,098145363	0,160489465	-0,059639932	-0,168270225	-0,030724671
D	espaol	0,032985656	0,073405089	-0,003896059	-0,084885183	-0,017609503
C	notas	-0,010923183	-0,028975932	-0,006958482	0,037706711	0,009150885
C	horas	-0,007805067	-0,020704503	-0,004972124	0,026943007	0,006538687
Y9	curso1	0,185518534	0,156449302	-0,181508217	-0,140355394	-0,020104225
C	edad	0,003576577	0,006964232	-0,002024062	-0,007254524	-0,001262222
D	sexo	0,011901482	0,021876619	-0,008302783	-0,021806285	-0,003669034
D	hijos	-0,058474006	-0,153956283	-0,098776237	0,239733490	0,071473037
C	notas	-0,008492219	-0,016535862	0,004805930	0,017225132	0,002997019
C	horas	-0,003565972	-0,006943582	0,002018060	0,007233014	0,001258480
Y10	curso1	-0,051774371	-0,086207908	0,002434699	0,105391882	0,030155698
D	sexo	0,031102953	0,035835667	-0,023221005	-0,035787245	-0,007930371
C	edad	-0,001179564	-0,001555645	0,000735820	0,001620157	0,000379232
D	trabaja	0,004000111	0,005186992	-0,002561126	-0,005376288	-0,001249690
D	hijos	0,157482184	0,104546528	-0,132294579	-0,108115495	-0,021618639
C	notas	-0,036830785	-0,048573569	0,022975286	0,050587895	0,011841173
C	horas	-0,006413062	-0,008457743	0,004000510	0,008808482	0,002061813
C - Var. Cuantitativas --- D - Var. Dummy, Dicotómicas.						
Nota: En negrita aparecen los efectos marginales de las variables explicativas individualmente significativas.						

- *Revisión y análisis de casos reales de empresas (Y4):*

Un incremento en la edad del alumno de un año, reduce la probabilidad de escoger el valor “muy alto” en un 0,27%, mientras que, por el contrario, aumenta la probabilidad en un 2,92% si el individuo trabaja, frente a los que no lo hacen, manteniendo constantes el resto de regresores.

- *Portafolio: Recopilación y análisis de datos e información (Y5):*

La probabilidad de que el alumno escoja la opción 4, mayor grado de satisfacción, se incrementa en un 0,14%, por cada punto de incremento en sus *notas* medias.

- *Foros en Aula Virtual (Y6):*

La probabilidad de tener un grado de satisfacción “muy alto” disminuye en un 0,917% para los alumnos de primer curso respecto a los demás.

- *Tests/Cuestionarios en Aula Virtual (Y7):*

La probabilidad de que los alumnos escojan la opción 4 disminuye en un 1,32% para los varones, frente a las mujeres.

- *Tests/Cuestionarios anónimos con Educlick (Y8):*

La probabilidad de escoger la opción “más alto” disminuye en los varones en un 2,96%, respecto a las mujeres.

Con respecto a los regresores de naturaleza continua, *notas* y *horas*, se obtiene que las probabilidades de escoger la opción 4 aumentan, respectivamente, en un 0,91% y 0,65%, cuando se incrementa en una unidad la media de las notas o las horas de estudio.

- *Asistencia a conferencias y seminarios (Y9):*

La probabilidad de elegir la categoría más alta o el nivel de satisfacción más elevado, se reduce en un 2,01% cuando el alumno es de primer curso, respecto al resto. Por el contrario, esta probabilidad se incrementa en un 7,15% si los alumnos tienen hijos, frente a los que no los tienen.

- *Tutorías (Y10):*

La probabilidad de escoger la opción “muy alto” se incrementa si el alumno pertenece a primer curso en un 3,01% respecto a si es de otro curso y aumenta en un 1,18% si las notas medias del alumno crecen en un punto sobre diez.

5. Conclusiones

El análisis sobre el nivel de satisfacción de los alumnos, en relación al nivel de conocimientos, competencias y habilidades que ellos mismos perciben a través de las diferentes metodologías docentes utilizadas por el docente, resulta de importancia crucial, en el ámbito universitario, en un momento de cambio y adaptación al EEES. El éxito del proceso de aprendizaje depende, en gran parte, del incentivo que tenga el alumno y en ello representa un papel fundamental las metodologías docentes utilizadas en su proceso formativo. Unos procedimientos metodológicos que satisfacen al alumno, suelen motivarlo y facilitar su éxito en su paso por la Universidad.

En este sentido, a partir del presente estudio, aplicado a los alumnos de la titulación del Grado en ADE de la Facultad de Ciencias de la Empresa de la UPCT, se obtiene que existe una diferencia importante entre el grado de satisfacción, en relación al nivel de conocimientos, competencias y habilidades que los alumnos perciben cuando el docente aplica diferentes metodologías en su proceso de enseñanza-aprendizaje, entre los alumnos de primer curso respecto al resto de estudiantes. En particular, los alumnos de primer curso valoran más, que el resto, procedimientos metodológicos más pasivos para ellos y en los que el profesor actúa como gestor de su proceso formativo. Este resultado deja patente que este colectivo de estudiantes se siente más *cómodo* con procedimientos que resultan más acordes con el nivel, todavía bajo, de conocimientos

sobre la carrera y con el de competencias y habilidades adquiridas dado su escaso recorrido en su paso por la Universidad. Este resultado, así como el hecho de que alumnos de cursos más elevados se sientan más satisfechos con procedimientos que implican un mayor trabajo autónomo del alumno, ponen de manifiesto que la adaptación de los estudiantes a los procedimientos metodológicos docentes que impliquen un mayor grado de razonamiento, reflexión, capacidad de síntesis, exposición, etc. del alumno, no es un aspecto banal y debe evolucionar a la par con el nivel de conocimientos, competencias y habilidades que con el paso del tiempo va adquiriendo el estudiante. Este proceder llevaría a un incremento en su grado de satisfacción y, por ende, una mayor motivación y mayor éxito en sus resultados.

La coordinación transversal, entre profesores de un mismo curso, no sólo de las materias a impartir en cada uno de ellos, sino también longitudinal, entre profesores de diferentes cursos, orientada, entre otros aspectos, a la implantación gradual de procedimientos que impliquen un mayor grado de trabajo autónomo del alumno a medida que éste pase a cursos más avanzados, se convierte en un reto importante y fundamental en el proceso de enseñanza-aprendizaje. Ello puede ayudar a la consecución de los objetivos de la Universidad y el espíritu del Proceso de Bolonia: *formar a los alumnos en los conocimientos de las diferentes materias y proporcionarles los recursos metodológicos que les hagan posible desarrollarse integralmente para el desarrollo con éxito en su vida profesional y personal.*

6. Referencias Bibliográficas

Johnson, D. H. y Johnson, R.T., *Learning together and alone: Cooperative, competitive and individualistic learning*. (5ª ed. Prentice Hall, NJ, 1998).

Barkley, E., Cross, K.P. y Major, C.H., *Técnicas de aprendizaje colaborativo* (Morata: Madrid, 2007).

Greene, W. H: *Econometric Analysis*. (7ª ed. Prentice Hall, NY., 2008)

Greene, W. H. y Hensher, D.A: *Modeling Ordered Choices: A Primer and Recent Developments*, Cambridge University Press, Cambridge, (2010).

Daykin, A. R. y Moffatt, P.G. *Analysing Ordered Responses: A Review of the Ordered Probit Model*. *Understanding Statistics*, vol. 1, (2002), pp. 157-166.

McKeachie, W.J. y Svinicki M. D: *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers*, 13th Edition, Univ. Of Michigan, (2011).

Morales, P: *Tamaño necesario de la muestra: ¿Cuántos sujetos necesitamos?*, Universidad Pontificia Comillas (2012), en: <http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oMuestra.pdf>

**EL PROBLEMA DEL ENRUTAMIENTO DE VEHÍCULOS.
PROPUESTAS PARA LA BUSQUEDA DEL CAMINO MAS CORTO.
APLICACIÓN AL ENTORNO DOCENTE Y PYMES**

JUAN JESÚS BERNAL GARCÍA

Juanjesus.bernal@upct.es

*Universidad Politécnica de Cartagena (UPCT) / Métodos Cuantitativos e Informáticos
C/Real3. 30201 Cartagena*

ELOY HONTORIA HERNÁNDEZ

eloy.hontoria@upct.es

*Universidad Politécnica de Cartagena / Organización de Empresas.
Escuela Técnica Superior de Ingenieros Industriales. 30202 Cartagena*

DARKO ALEKSOVSKI

darko.aleksovski@ijs.si

*Jožef Stefan Institute / Department of Knowledge Technologies.
Ljubljana (Eslovenia)*

RESUMEN: El enrutamiento de vehículos o VRP (Vehicle Routing Problem) es un problema típicamente planteado para la optimización de las rutas de transporte y vehículos asociados a ellas. Para su planteamiento y búsqueda de soluciones óptimas se recurre en investigación operativa a modelos matemáticos como *Branch and Bound* o programación dinámica entre otros. Estos modelos proporcionan soluciones exactas pero presentan una gran complejidad para su uso práctico en el entorno docente y de las PYMES. El objetivo del presente trabajo es la presentación de la utilidad informática Solver, que incorporada en Excel y debidamente programada con VBA (*Visual Basic for Applications*) permite la obtención de soluciones de manera sencilla para su uso docente y para la pequeña y mediana empresa.

Como aplicación práctica para mostrar la metodología elaborada y su operatividad, se ha recurrido a un caso práctico consistente en la determinación de la ruta más corta entre ciudades españolas. Para contrastar la validez de los resultados obtenidos, se han comparado éstos con los proporcionados con un algoritmo de búsqueda exhaustiva mediante fuerza bruta (*BSF*) diseñado “ad hoc”. Este algoritmo que explota la aplicación de distancias de rutas de *Google Maps*, proporciona mejoras con respecto al modelo anterior basado en Solver [1], suministrando información más actualizada y completa sobre el problema y, pudiendo aportar en un futuro una visión de costes al problema de enrutamiento de vehículos.

Palabras clave: Excel, Solver, Google Maps, Optimización de Rutas, Algoritmo de Búsqueda por Fuerza Bruta.

ABSTRACT: The Vehicle Routing Problem (VRP) is a common problem presented to optimize the transport routes and vehicles involved. For its approach and search of optimal solutions in operational research, mathematical models like Branch and Bound

or dynamic programming among other options are used. These tools provide exact solutions, but the disadvantage is it's difficult practical application for teaching or SME's. The purpose of this researching work is to present Solver, a deployed application in Excel which properly programmed in BVA (Visual Basic for Applications) allows us to obtain easily solutions for teaching purpose or to be applied in SME's.

As a practical application to show the designed methodology and its operability, a practical case consisting in the calculation of the shortest path between several Spanish cities will be presented. To test the validity of the results obtained through Solver's application, they will be contrasted with others obtained through an algorithm based in the Brute Force Search (BFS) specifically designed for this purpose. By using this algorithm, which uses Google Maps's application for route distances, some improvements are reached compared with the previous model based in Solver, providing more current and complete information about the problem and a future cost perspective to the Vehicle Routing Problem.

Keywords: Excel, Solver, Google Maps, Routes Optimization, Brute Force Search Algorithm.

1. Introducción

El problema del enrutamiento de vehículos o VRP (Vehicle Routing Problem) es un nombre genérico dado a toda aquella clase de problemas que implica la visita a "clientes" mediante vehículos [2]. El VRP ha sido profundamente estudiado por diversos autores como [3], [4], [5] y [6] y en su situación inicial se parte de un conjunto de clientes con localizaciones y requerimientos conocidos, que tienen que ser servidos desde uno o varios depósitos por una flota de vehículos. El problema del diseño de las rutas para esos vehículos que satisfagan los requerimientos de los clientes y al menor coste de distribución posible, es lo que se conoce con el nombre de problema del enrutamiento de vehículos.

Las variantes del VRP son muy diversas y vienen dadas por los requerimientos y restricciones operativas impuestas por el diseño de rutas en situaciones prácticas [7]. Entre estas variantes se puede destacar que los clientes no sean conocidos a priori en el caso del DVRP (Dynamic Vehicle Routing Problem) estudiado por [8] o que las entregas solo puedan ser realizadas en una determinada franja horaria (Time Window), tratándose del VRPTW [9] o que la capacidad del vehículo sea una restricción en el Capacitated Vehicle Routing Problem (CVRP) analizado por [7] y una infinidad de variantes (entrega y recogida simultánea al cliente, demandas deterministas o aleatorias, etc.). En determinadas ocasiones se ha recurrido a técnicas de simulación para la composición de una flota de vehículos heterogénea (Heterogeneous fleet VRP, HVRP), con demanda determinista atendiendo a criterios económicos y restricciones de tiempo máximo para realizar el reparto diario [10].

VRP es en definitiva una extensión del problema del Viajante de Comercio (Traveling Salesman Problem, TSP) donde el vendedor debe visitar un conjunto de clientes o

ciudades una sola vez, para después volver a la ruta de partida, construyéndose un grafo donde figuran los clientes (vértices) y las rutas posibles entre un cliente y otro (aristas). El TSP empezó a ser contemplado mediante técnicas informáticas con el trabajo de [11] en el que se diseñó un algoritmo para el cálculo óptimo de una ruta a través de 49 ciudades. El VRP según [12] es considerablemente más difícil de resolver que un TSP para un problema del mismo tamaño. Un problema de TSP implicando cientos o incluso miles de vértices puede ser resuelto mediante algoritmos avanzados de branch-and-cut-and-price [12] y en contraste [14] creó uno de los más sofisticados algoritmos exactos donde resolvía con una tasa de éxito variable un máximo de 100 clientes.

Otro capítulo a analizar es el método empleado para la resolución del problema planteado. Las técnicas de optimización se enfrentan generalmente a objetivos contrapuestos como son encontrar la solución óptima y además hacerlo en el menor tiempo posible o realizarlo mediante un modelo matemático que pueda imitar fielmente el experimento o conseguirlo a un coste razonable, etc. En determinadas ocasiones se valora más la obtención de soluciones aceptables aunque no sean las óptimas si éstas han sido obtenidas más rápidamente. En este aspecto se debe destacar tres propuestas de resolución: Métodos exactos, métodos basados en el TSP y Heurísticos. Los heurísticos son métodos aproximados cuya versión más compleja se alcanza con los Metaheurísticos. Referido a la primera opción de métodos exactos y para la optimización de las rutas de transporte y vehículos asociados a ellas, se recurre en investigación operativa a modelos matemáticos o a programación dinámica entre otras opciones. Estas herramientas proporcionan soluciones exactas, pero el diseño de un modelo matemático que refleje la realidad con precisión es complicado y además requiere mucho tiempo para la búsqueda de soluciones.

Para la resolución de estos modelos matemáticos en docencia no especializada, se suele recurrir a programas específicos como *WinQSB* [15], *LINDO Systems* [16], etc., pero en este trabajo se ha preferido la utilidad informática Solver, que incorporada en Excel y a pesar de su metodología limitada, permite determinar la ruta más corta de forma sencilla y por tanto muy útil para su utilización.

En el presente trabajo se pretende contrastar la validez de los resultados aportados por Solver para el cálculo del camino óptimo, mediante su comparación con los obtenidos con un algoritmo diseñado para este propósito, que al utilizarlo conjuntamente con *Google Maps* permite además facilitar la toma de distancias, y aportar información adicional a la solución obtenida con la hoja de cálculo.

2. Camino más corto entre varios puntos. Caso práctico con de la herramienta Solver para la obtención de resultados.

El TSP nombrado anteriormente, es uno de los problemas más conocidos de la optimización combinatoria donde la dificultad de buscar soluciones crece de manera exponencial al aumentar el número de ciudades a visitar. A modo de ejemplo y para un planteamiento de 12 ciudades, el viajante deberá decidirse entre las 479.001.600 combinaciones posibles.

El problema del camino más corto planteado en este trabajo es una variación del TSP,

consistente en que en este caso se debe buscar el camino más corto entre un origen y un destino, sin la obligación de visitar o pasar por todos los puntos (ciudades) posibles. En la resolución mediante *Solver* el planteamiento del problema es sencillo y las soluciones son encontradas sin un elevado coste computacional cuando el número de ciudades es relativamente reducido. Estas ventajas en cuanto a facilidad, economía y rapidez hacen de este complemento de Excel una herramienta de gran utilidad para introducir tanto a los alumnos, cómo a los técnicos de las pequeñas empresas en el campo de métodos cuantitativos, al mismo tiempo que se familiarizan con el manejo de hojas de cálculo. Sin embargo, los resultados obtenidos deben ser contrastados convenientemente.

Para la citada validación se recurre al planteamiento de un caso práctico consistente en encontrar el camino más corto entre dos ciudades, una origen y otra destino, sabiendo que existen diversas rutas alternativas dependiendo de las 5 posibles ciudades intermedias por las que se pase. Por tanto se trata de un problema con cinco vértices y 2 aristas, incluidos el origen y el destino, con 120 combinaciones posibles (*Figura 1*), del que se conoce la distancia entre ellos; concretamente se trata de considerar las ciudades españolas: Cartagena (CT-Origen), Albacete (AB), Cuenca (CU), Toledo (TO), Segovia (SG), Soria (SO) y Burgos (BU-Destino).

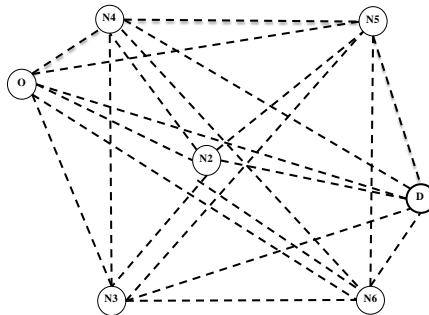


Figura 1: Siete nodos y sus conexiones posible

Cómo información de partida se dispone de las distancias entre las distintas ciudades, que pueden ser introducidas manualmente, tecleándolos en la celda correspondiente, o tomarlas directamente de una tabla de distancias entre ciudades importada a Excel desde una página Web (*Tabla 1*).

Tabla 1. Matriz de distancias entre las ciudades del problema

	CT	AB	CU	TO	SG	SO	BU
CT		192	352	438	541	649	677
AB	192		164	249	353	461	489
CU	352	164		179	260	272	398
TO	438	249	179		158	299	318
SG	541	353	260	158		191	200
SO	649	461	272	299	191		143
BU	677	489	398	318	200	143	

La situación real suele implicar el planteamiento inicial de una serie de rutas o conexiones convenientes, de entre todas las posibles (*Figura 1*). En este caso, el grafo con las posibles rutas, con líneas de trazo discontinuo, quedaría de la forma que se muestra en la *Figura 2*, tanto con aristas, de forma general, como con las ciudades del caso práctico. La programación mediante macros en VBA [17] realizada para dibujar las conexiones se adjunta en el *Anexo B*.

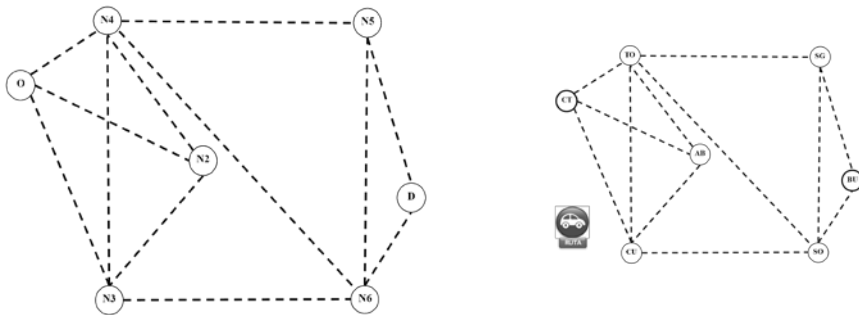


Figura 2: Conexiones deseadas entre nudos y entre ciudades

La forma práctica de selección de caminos entre ciudades en el modelo elaborado en Excel, es en la tabla de doble entrada, sustituir el valor “0” por un “1” en la celda intersección de ambas ciudades, lo que establece la existencia de conexión deseada entre ellas, siendo necesario hacerlo solamente en la parte superior de la misma, ya que al tratarse de una matriz transpuesta, la inferior se rellena automáticamente (*Tabla 2*). Los caminos posibles que sean señalados determinarán la matriz de distancias a utilizar entre las ciudades conectadas, haciéndose cero el resto (*Tabla 3*). A modo de ejemplo, al teclear un 1 en el elemento intersección CT-AB, aparece la distancia de 192 km., que es la existente entre ambas ciudades.

Tabla 2. Matriz (transpuesta) de “unos y fondo gris” para indicar las conexiones posibles

	1						0
	CT	AB	CU	TO	SG	SO	BU
CT	0	1	1	1	0	0	0
AB	1	0	1	1	0	0	0
CU	1	1	0	1	0	1	0
TO	1	1	1	0	1	1	0
SG	0	0	0	1	0	1	1
SO	0	0	1	1	1	0	1
BU	0	0	0	0	1	1	0

Seguidamente al invocar la macro, o pulsar el icono elaborado para ello con forma de coche (*Figura 2*), se ejecuta el Solver programado con la ecuación a minimizar y distintas restricciones (*Anexo A*) [6], obteniéndose la solución de ruta óptima encontrada en forma de tabla (*Tabla 4*). La solución encontrada incluye las ciudades intermedias de paso, en este caso: N3 (CU) y N6 (SO), y la distancia total de dicha ruta óptima: $352 \text{ (CT-CU)} + 272 \text{ (CU-SO)} + 143 \text{ (SO-BU)} = 767 \text{ kms.}$

Tabla 3. Matriz de distancias entre ciudades conectadas

	CT	AB	CU	TO	SG	SO	BU
CT	0	192	352	438	0	0	0
AB	192	0	164	249	0	0	0
CU	352	164	0	179	0	272	0
TO	438	249	179	0	158	299	0
SG	0	0	0	158	0	191	200
SO	0	0	272	299	191	0	143
BU	0	0	0	0	200	143	0

Tabla 4. Matriz con las distancias de la ruta óptima

	CT	AB	CU	TO	SG	SO	BU
CT	0	0	1	0	0	0	0
AB	0	0	0	0	0	0	0
CU	0	0	0	0	0	1	0
TO	0	0	0	0	0	0	0
SG	0	0	0	0	0	0	0
SO	0	0	0	0	0	0	1
BU	0	0	0	0	0	0	0

Como aportación a otros modelos análogos existentes en Excel, se ha programado también la solución en forma gráfica, ya que de manera automática se muestra la ruta más corta mediante conexiones de trazo continuo (en color verde en el programa de hoja de cálculo) (*Figura 5*). Macro en *Anexo B*.

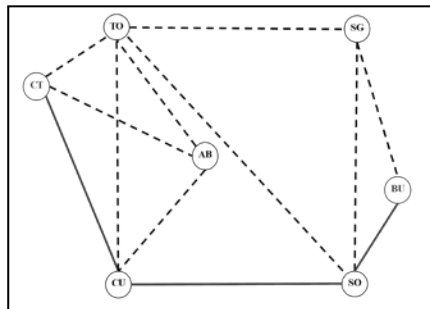




Figura 5: Solución gráfica y con iconos de la ruta más corta: CT-BU.

Aunque se ha hablado inicialmente de vértices para generalizar el uso del modelo, como el ejemplo tratado hace referencia a siete ciudades concretas, se ha pensado representar sobre una imagen con mapa de España, o un mapa de Google, dichas ciudades así como la solución, resultando de esta forma más evidente la lógica de la misma (Figura 6). Se debe comentar en este punto que Excel no cuenta en las versiones actuales con la inclusión de *Microsoft Maps*, que permitía incluir directamente mapas en la hoja de cálculo e insertar datos en ellos.



Figura 6: Solución del problema sobre mapa con siete ciudades

Una vez encontrada la solución óptima con Solver, se estimó conveniente someter al modelo a diversas pruebas para comprobar su correcto funcionamiento. En primer lugar se quiso analizar su comportamiento ante variaciones en las conexiones existentes entre ciudades, y en segundo, comprobar la reversibilidad de la solución obtenida. Para la primera prueba, y dado que el establecimiento de estas conexiones son fundamentales a la hora de plantear y resolver el problema, se ha probado a marcar como posible la conexión que une directamente el origen “O” (CT) con el destino “D” (BU), marcando con 1 el valor correspondiente en la matriz de conexiones (Tabla 5). Se comprueba que efectivamente la nueva solución óptima encontrada es ésta, ya que implicaría menos kilómetros que la anterior, con varias ciudades, de 767 km., y por tanto supondría un ahorro de 90 kms. (Figura 7).

Tabla 5. Conexión directa entre origen y destino

	O	N2	N3	N4	N5	N6	D
O	0	1	1	1	0	0	0
N2	1	0	1	1	0	0	1
N3	1	1	0	1	0	1	0
N4	1	1	1	0	1	1	0
N5	0	0	0	1	0	1	1
N6	0	0	1	1	1	0	1
D	0	1	0	0	1	1	0

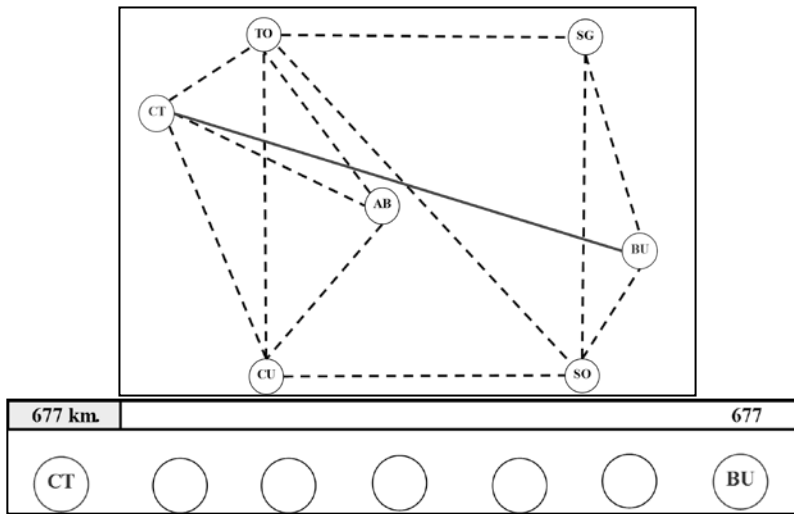


Figura 7: Nueva solución tras añadir una nueva conexión CT-BU.

Para la segunda comprobación sobre si el modelo responde correctamente a analizar el camino en sentido inverso se programó la posibilidad de permutar origen y destino. Para ello se cambió simplemente el “1” por “0” sobre la O de origen (CT), y se comprueba que la solución implica los mismos nodos N6 (SO) y N3 (CU), y la misma distancia suma (767 Kms.) (Tabla 6), lo cual demuestra nuevamente el correcto funcionamiento del modelo planteado y la herramienta Solver.

Tabla 6: Marcar rutas en sentido inverso (permutar origen y destino). Ruta BU-CT.

	0						1
	CT	AB	CU	TO	SG	SO	BU
CT	0	1	1	1	0	0	0
AB	1	0	1	1	0	0	0
CU	1	1	0	1	0	1	0
TO	1	1	1	0	1	1	0
SG	0	0	0	1	0	1	1
SO	0	0	1	1	1	0	1
BU	0	0	0	0	1	1	0

3. Propuesta de solución alternativa. Algoritmo de Búsqueda mediante Fuerza Bruta y Google Map.

Como se ha podido apreciar en el ejemplo anterior, Solver es una herramienta válida para conseguir soluciones de manera rápida y económica. Sin embargo, el administrador del sistema puede requerir de soluciones más restrictivas en cuanto a la exactitud,

fundamentalmente para distancias pequeñas como puede ocurrir en rutas urbanas, y/o que la actualización de las mismas se efectuó de manera automática sin requerir introducir nuevos valores en la hoja Excel. Por todo ello, se planteó la posibilidad de buscar la solución óptima por otros métodos alternativos al Solver, que a su vez sirviesen de contraste de la solución obtenida anteriormente para el caso práctico planteado.

El método elegido está basado en el Algoritmo de Búsqueda mediante Fuerza Bruta (en inglés *Brute Search Force*, *BSF*), aplicado al problema del camino más corto. La búsqueda por Fuerza Bruta también es conocida como búsqueda exhaustiva y consistiría en generar todas las soluciones factibles, calcular para todas ellas el coste asociado (distancia, tiempo, unidades monetarias) y posteriormente elegir el menor de estos costes.

De una manera más formal sería equivalente a definir una función de coste $f: S \rightarrow \mathbb{R}$, siendo S el espacio de soluciones factibles y pudiéndose calcular $f(s) \forall s \in S$ y dar como solución final el óptimo, que para el problema que se intenta resolver se correspondería con el camino más corto.

La generación de una secuencia con todas las soluciones factibles no es labor fácil, pero para el ejemplo considerado con un número reducido de nodos, es posible encontrar todo el espacio de soluciones mediante el modelo diseñado y además elegir el óptimo dentro de él.

Cuando el número de ciudades intermedias aumenta de manera considerable, este procedimiento no es viable debido a la magnitud del espacio de soluciones S , desvaneciéndose el objetivo de encontrar el óptimo. En estos casos adquiere mayor importancia la obtención de buenas soluciones aunque no sean las óptimas, recurriéndose al empleo de heurísticos o metaheurísticos en su versión más sofisticada.

Abordaremos el mismo ejemplo anterior de siete ciudades combinando el uso del servidor de aplicaciones de mapas en la web *Google Maps* y *BSF*. La ventaja es que la toma de distancias entre ciudades se hace de manera automática y exacta, al ser obtenida mediante la posibilidad existente en *Google Maps* de disponer de la matriz de distancias. La aplicación Web *Google Maps* permite al usuario mediante su *API*¹ generar una matriz de distancias que será la base para la resolución del problema del camino más corto entre dos ciudades. Para ello todas las direcciones o coordenadas de las ciudades son introducidas en esta *API* de *Google Maps* y la mencionada aplicación devuelve unos resultados correspondientes a las distancias entre cada una de los nodos pero en un formato no demasiado operativo para el ámbito docente y profesional. Para obtener las distancias en formato de hoja de cálculo Excel, se precisa de un “script”² desarrollado “ad hoc” por los autores (*Figura 8*).

¹ Application Programming Interface (Interfaz de Programación de Aplicaciones)

² Programa que proporciona un formato estándar.

```

C:\Ruby193\bin\ruby.exe
0 Cartagena, Spain
1 Albacete, Spain
2 Cuenca, Spain
3 Toledo, Spain
4 Soria, Spain
5 Segovia, Spain
6 Burgos, Spain
URL for testing in browser: maps.googleapis.com/maps/api/distancematrix/xml?sensor=false&origins=Cartagena,Spain!Albacete,Spain!Cuenca,Spain!Toledo,Spain!Soria,Spain!Segovia,Spain!Burgos,Spain&destinations=Cartagena,Spain!Albacete,Spain!Cuenca,Spain!Toledo,Spain!Soria,Spain!Segovia,Spain!Burgos,Spain

```

Figura 8: Script para la obtención de las matrices de tiempos y distancias en formato Excel

El resultado devuelto mediante la aplicación de este script a los valores proporcionados por *Google Maps* es una Matriz de distancias (Distance Matrix) en formato Excel con las distancias en Kms. entre cada una de las 7 ciudades tal y como muestra la *Tabla 7*. Se debe aclarar que el resultado puede depender de cómo sean medidas las distancias entre ciudades, siendo el centro de la ciudad el punto de referencia desde el cual toma medidas Google Maps.

Tabla 7: Matriz de distancias (Google Maps)

DISTANCE MATRIX: (In Kilometers)							
	Cartagena	Albacete	Cuenca	Toledo	Soria	Segovia	Burgos
Cartagena	0	192	352	438	649	541	677
Albacete	190	0	164	249	461	353	489
Cuenca	351	163	0	179	272	260	398
Toledo	439	250	182	0	299	158	318
Soria	651	462	274	300	0	191	143
Segovia	541	353	260	165	190	0	200
Burgos	677	488	400	321	144	200	0

Como se puede apreciar en la anterior tabla, la matriz devuelta no es simétrica, lo cual tiene su lógica debido a que los arcos en el recorrido de ida y vuelta no utilizan en todo momento el mismo camino. Aunque para el problema en cuestión, estas diferentes distancias en ambos sentidos no tiene demasiada importancia, sí que es un factor a considerar en la logística urbana, donde algunas calles son utilizadas plenamente en un sentido y usándose otras en sentido contrario. La asimetría en la matriz de distancias afecta a los cálculos de distancias, tiempos y costes de las empresas de transporte exclusivamente urbano. Para la depuración de la matriz de distancias convirtiéndola en simétrica para una mejor operatividad, se ha optado por el criterio de tomar la menor de las dos distancias como se aprecia en la *Tabla 8*.

Tabla 8: Matriz de distancias simétrica (Google Maps)

DISTANCE MATRIX: (In Kilometers)							
	Cartagena	Albacete	Cuenca	Toledo	Soria	Segovia	Burgos
Cartagena	0	190	351	438	649	541	677
Albacete	190	0	163	249	461	353	488
Cuenca	351	163	0	179	272	260	398
Toledo	438	249	179	0	299	165	321
Soria	649	461	272	299	0	190	143
Segovia	541	353	260	165	190	0	200
Burgos	677	488	398	321	143	200	0

En el ejemplo a resolver y al tratarse de un número reducido de 7 ciudades, de las cuales dos son inalterables en el orden de visita (Origen y destino) se buscará la solución óptima y para ello se recurrirá al algoritmo BSF. Éste buscará todas y cada una de las combinaciones posibles entre todas las ciudades con la única restricción de no repetir el recorrido a través de alguna de ellas. Una vez obtenido el espacio de soluciones, éstas se filtrarán, quedando sólo aquellas que tengan su origen en Cartagena y terminen en Burgos. Con las soluciones resultantes se procederá a cuantificarlas kilométricamente y para ello y mediante una aplicación informática desarrollada por los autores de este trabajo, se establecerá una conexión entre ellas y la matriz de distancias convirtiéndose las soluciones en distancias numéricas. El siguiente paso es la selección de la menor de ellas, que será el camino óptimo más corto.

#1|Cartagena,Spain

#2|Albacete,Spain

#3|Cuenca,Spain

#4|Toledo,Spain

#5|Soria,Spain

#6|Segovia,Spain

#7|Burgos,Spain

CT, AB, CU, TO, SO, SG, BU

La solución óptima encontrada para el cálculo del camino más corto mediante el método propuesto es:

CT - CU - SO - BU = 351 + 272 + 143 = **766 Kms**, ruta coincidente con la obtenida mediante Solver (Figura 9).



Figura 9: Camino más corto entre las 7 ciudades. Ruta CT-BU (Mapa Google)

Como aportación y mejora del modelo inicial, destacar que junto a la matriz de distancias se obtiene la “Matriz de Tiempos” (Time Matrix) que permite conocer los tiempos que emplearía un vehículo en recorrer esas distancias (Tabla 9).

Tabla 9: Matriz de tiempos de recorrido de rutas (Google Map)

Time Matrix: (in minutes)	Cartagena	Albacete	Cuenca	Toledo	Soria	Segovia	Burgos
Cartagena	0	115	206	243	363	292	365
Albacete	115	0	106	143	261	191	266
Cuenca	206	106	0	110	190	151	216
Toledo	243	143	110	0	179	97	183
Soria	363	261	190	179	0	142	107
Segovia	292	191	151	97	142	0	127
Burgos	365	266	216	183	107	127	0

$$CT - CU - SO - BU = 206 + 190 + 107 = 503 \text{ Minutos} = 8,38 \text{ HORAS}$$

Esta posibilidad de obtener los tiempos necesarios de recorrido, permite un cálculo más completo de los costes totales de transporte al posibilitar la incorporación el estudio de los costes laborales u otros.

4. Conclusiones

La herramienta Solver incorporada en Excel ha demostrado su utilidad tanto en su aspecto docente para introducir a los alumnos en el campo de los métodos cuantitativos como para su uso empresarial a niveles no demasiado complejos, como los requeridos con frecuencia en PYMES, al mostrar los resultados de forma sencilla, analítica y gráficamente.

A su vez y con el propósito de testar los resultados obtenidos mediante Solver se ha desarrollado un algoritmo que ha servido para comprobar la fiabilidad de ambos

métodos. El algoritmo basado en la Búsqueda por Fuerza Bruta invoca a la aplicación de distancias de Google Maps obteniendo de manera rápida y automática las distancias entre los vértices. Este punto es interesante ante cualquier variación kilométrica acaecida por apertura de nuevas rutas, evitando la incorporación manual necesaria para el desarrollo mediante Solver. Otra de las aportaciones del algoritmo diseñado es la búsqueda de todas las combinaciones posibles y la selección posterior de la más idónea, es decir, el camino más corto buscado.

Junto con las distancias obtenidas mediante este segundo método, se ha conseguido una información adicional de gran utilidad para el ámbito del enrutamiento de vehículos como es el tiempo necesario de recorrido de un vehículo entre los diferentes puntos, así como el de la solución final. Esta información representada en la matriz de tiempos, permite cálculos de disponibilidad de flotas, así como costes relacionados con ellas, lo que permitirá ampliar los resultados y la aplicabilidad del modelo diseñado a las compañías de transporte.

5. Referencias Bibliográficas

- [1] Bernal García, J.J. (2000). “*Los modelos de red y la minimización de costes. Monta S.A., desea conocer la mejor forma de organizar el transporte de abastecimiento de sus fábricas, así como de elegir la ruta más corta para hacerlo a su nueva planta*”. Estrategia Financiera (CISS-Especial Directivos). Volumen N°. 160, pp 22-28. Madrid.
- [2] Christofides, Nico (1976). “*The Vehicle Routing Problem*”. *Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle*, 10 (2), pp. 55-70
- [3] Laporte, G. (1992). “*The Vehicle Routing Problem: An overview of exact and approximate algorithms*”. *European Journal of Operational Research*, 59 (3), pp. 345-358.
- [4] Laporte, G., Gendreau, M., Potvin, J.Y., Semet, F. (2000). “*Classical and Modern Heuristics for the Vehicle Routing Problem*”. *International Transactions in Operational Research* 7, pp 285-300.
- [5] Fisher, Marshall L. (1994). “*Optimal solution of vehicle routing problems using minimum K-trees*”. *Operations Research*, 42 (4), pp. 626-642.
- [6] Bachelet, B., Yon, L. (2007). “*Model enhancement: Improving theoretical optimization with simulation*”. *Simulation Modelling Practice and Theory* 15 (6), pp 703-715.
- [7] Toth, P., Vigo, D. (2002). “*Models, relaxations and exact approaches for the capacitated vehicle routing problem*”. *Discrete Applied Mathematics*, 123 (1-3), pp. 487-512.
- [8] Montemanni, R., Gambardella, L.M., Rizzoli, A.E., Donati, A.V. (2005). “*Ant*

Colony System

for a *Dynamic Vehicle Routing Problem*". *Journal of Combinatorial Optimization* 10 (4) pp 327-343.

[9] Solomon, Marius M. (1987). "*Algorithms for the Vehicle Routing and scheduling problems with time window constraints*". *Operations Research*, 35 (2), pp. 254-265.

[10] De la Fuente-Aragon, MV., Hontoria, E., Ros-McDonell, L. (2012). "*A Simulation-Based Solution for Optimal Logistics of Heavy and Variable-Size Items*". *Industrial Engineering: Innovative Networks*. Springer Link pp 291-298

[11] Dantzig, G., Fulkerson, R., Johnson, S. (1954). "*Solution of a large scale traveling salesman problem*". *Oper. Res.* 2, pp 393-410.

[12] Laporte, G. (2007). "*What you should know about the vehicle routing problem*". *Naval Research Logistics* 54 (8), pp. 811–819.

[13] Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J. (2007). "*The traveling salesman problem, A computational study*". Princeton. University Press, Princeton, NJ.

[14] Baldacci, R., Christofides, N., Mingozzi, A. (2008). "*An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts*". *Mathematical Programming* 115 (2), pp 351–385.

[15] <http://www.wiley.com/college/tech/winqsb.htm> (03/04/2013).

[16] <http://www.lindo.com/> (03/04/2013).

[17] Walkenbach, J. (2011). "*Excel 2010. Programación con VBA*". Anaya Multimedia.-Wiley. Madrid.

Anexo A Programación del Solver

	A	B	C	D	E	F	G	H	I
33		1						0	
34		CT	AB	CU	TO	SG	SO	BU	Total
35	CT	0	0	1	0	0	0	0	1
36	AB	0	0	0	0	0	0	0	0
37	CU	0	0	0	0	0	1	0	1
38	TO	0	0	0	0	0	0	0	0
39	SG	0	0	0	0	0	0	0	0
40	SO	0	0	0	0	0	0	1	1
41	BU	0	0	0	0	0	0	0	0
42	Total	0	0	1	0	0	1	1	
43		1	0	1	0	0	1	0	
44	Dif	1	0	0	0	0	0	-1	0
45	Neto	1	0	0	0	0	0	-1	0

Figura 10: Celdas del problema

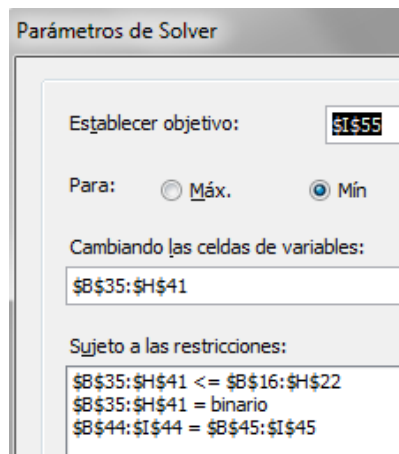


Figura 11: Celda objetivo y restricciones del Solver

Anexo B

Sub RUTA()

' RUTA Macro

```

SolverOk      SetCell:="$I$55",      MaxMinVal:=2,      ValueOf:=0,
ByChange:="$B$35:$H$41", _
Engine:=2, EngineDesc:="Simplex LP"
SolverOk      SetCell:="$I$55",      MaxMinVal:=2,      ValueOf:=0,
ByChange:="$B$35:$H$41", _
Engine:=2, EngineDesc:="Simplex LP"
SolverSolve UserFinish:=True
Linea = 1
    
```

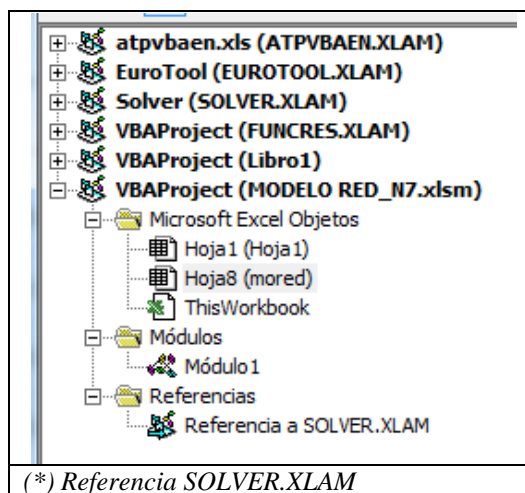
```
FilaIni = 35
ColIni = 3
Inij = 0
Application.ScreenUpdating = False
For i = 0 To 6
  For j = Inij To 5
    If Cells(FilaIni + i, ColIni + j) <> 0 Then
      ActiveSheet.Shapes.Range(Array("linea" & Linea)).Select
      Selection.ShapeRange.Line.ForeColor.RGB = RGB(255, 0, 0)
      Selection.ShapeRange.Line.DashStyle = msoLineSolid
      Selection.ShapeRange.Line.Weight = 2.25
    Else
      ActiveSheet.Shapes.Range(Array("linea" & Linea)).Select
      Selection.ShapeRange.Line.ForeColor.RGB = RGB(0, 0, 0)
      Selection.ShapeRange.Line.DashStyle = msoLineDash
      Selection.ShapeRange.Line.Weight = 2#
    End If
    Linea = Linea + 1
  Next j
  Inij = Inij + 1
Next i
ActiveSheet.Shapes.Range(Array("Botón 1")).Select
Application.ScreenUpdating = True
End Sub
```

```
Private Sub Worksheet_Change(ByVal Target As Range)
If InRange(Target, Range("B16:H22")) Then
Linea = 1
FilaIni = 16
ColIni = 3
Inij = 0
Application.ScreenUpdating = False
For i = 0 To 6
  For j = Inij To 5
    If Cells(FilaIni + i, ColIni + j) <> 0 Then
      ActiveSheet.Shapes.Range(Array("linea" & Linea)).Select
      Selection.ShapeRange.Line.Visible = msoTrue
    Else
      ActiveSheet.Shapes.Range(Array("linea" & Linea)).Select
      Selection.ShapeRange.Line.Visible = msoFalse
    End If
    Linea = Linea + 1
  Next j
  Inij = Inij + 1
Next i
Application.ScreenUpdating = True
Target.Select
```

```
End If  
End Sub
```

```
Function InRange(Range1 As Range, Range2 As Range) As Boolean  
' returns True if Range1 is within Range2  
Dim InterSectRange As Range  
    Set InterSectRange = Application.Intersect(Range1, Range2)  
    InRange = Not InterSectRange Is Nothing  
    Set InterSectRange = Nothing  
End Function
```

Nota: Para programar Solver es preciso activar la referencia correspondiente en VB (*)



REPRESENTACIÓN DE GRÁFICAS Y TABLAS DE DATOS CON MOVIMIENTO EN EXCEL

JUAN JESÚS BERNAL GARCÍA

juanjesus.bernal@upct.es

*Universidad Politécnica de Cartagena (UPT)/Métodos Cuantitativos e Informáticos
C/Real,3. 30201 Cartagena*

SOLEDAD M^a. MARTÍNEZ M^a DOLORES

soledad.martinez@upct.es

*Universidad Politécnica de Cartagena/Métodos Cuantitativos e Informáticos
C/Real,3. 30201 Cartagena*

PABLO BERNAL SOTO

pablo.bernal@gmail.com

*UPCT /Grupo de Investigación de "Métodos cuantitativos e informáticos para la economía y la empresa
C/Real,3. 30201 Cartagena*

RESUMEN: Continuando con anteriores trabajos sobre representación de funciones en Excel, se trata de recurrir a opciones especiales y a la programación en VBA para dotar a dichas gráficas de automatismos que nos permitan darles movimiento, tanto de rotación como de avance lineal, pudiendo generar una vista en cualquier posición. Las gráficas así elaboradas sirven tanto para visionar dinámicamente, no solo las funciones matemáticas, sino también la representación temporal de datos económico-financieros, llegando incluso a emular los gráficos de movimiento (denominados Motion Chart) de Google Drive.

Palabras clave: Excel, representación de funciones, gráficos en movimiento.

ABSTRACT: Following previous works on graphical representation of functions, this article introduces special options and VBA programming methods to provide motion capabilities to these graphs, such as spinning or lineal advance, which allow to set a the chart view from any position. These charts serve to display not only mathematical functions but also timeline representation of economic and financial data, even to emulate the Motion Charts of Google Drive.

Keywords: Excel, graphical representation of functions, motion charts

1. Introducción

Excel, a semejanza de otras hojas de cálculo, cuenta con numerosas funciones de tipo matemático que permiten representaciones gráficas de calidad, así como es posible mostrar de forma visual Tablas de datos, no obstante estas son de tipo estático. En un artículo anterior se expuso cómo representar una función en Excel, simplemente escribiendo su expresión en una celda [1]. En el presente trabajo abordaremos cómo lograr que dichas gráficas puedan ser modificadas a partir de unos controles, para que sea posible mostrar distintas visiones espaciales de las mismas, mediante rotaciones y avances lineales. Los efectos conseguidos son aplicables tanto para funciones de tipo matemático como económico-empresarial, lo cual las hace interesante para la investigación y para la docencia.

1.1. Antecedentes

Ciertamente existen programas en el mercado para la representación móvil de gráficas, normalmente de tipo matemático; algunos son de software libre en Internet, y de entre ellos el más conocido es el GeoGebra [2], que en su versión 4 incluye una versión Web, tratándose fundamentalmente de representaciones geométricas, donde la movilidad se consigue variando la posición de un punto en tiempo real. Existen no obstante, páginas en Internet donde se pueden encontrar ejemplos de gráficas en 3D, también a partir de esta aplicación [3] (*Figura 1*). Además existen complementos para dotar a Excel de mayor potencia a la hora de representar funciones, como la aplicación XLFit [4], pero dichas soluciones, lógicamente, no son gratuitas.

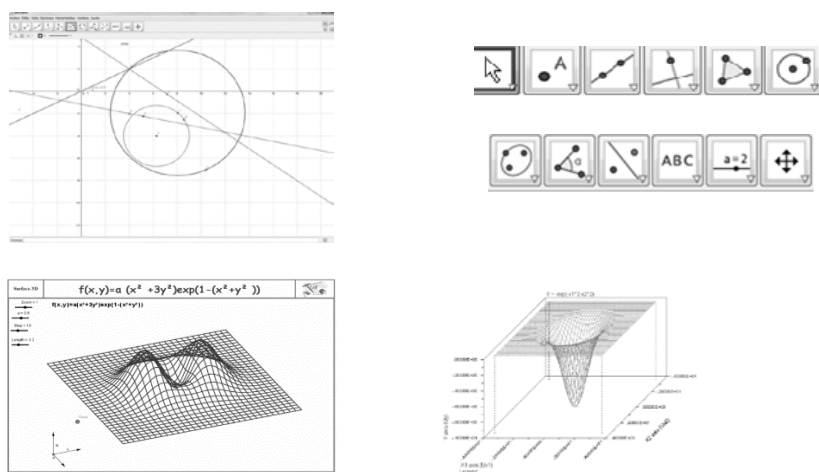


Figura 1: Ejemplos en GeoGebra Web y en 3D y en XLFit

2. Gráficos en movimiento

Veamos a continuación, distintas propuestas para conseguir gráficos en movimiento:

2.1. Gráficas sin refresco de pantalla

Cuando se varían los datos de una Tabla, el gráfico correspondiente lo hace de forma global, es decir, se muestra en la hoja modificada, pero puede resultar interesante, sobre todo si se desea visualizar un proceso temporal, ir mostrando la evolución instantánea del mismo. La actualización continua del gráfico proporciona una secuencia tipo vídeo de dicha evolución, por esta razón le dimos el nombre de “osciloscopio”. Así, en la *Figura 2*, se muestra la simulación del número pi, mediante un algoritmo de simulación, para lo cual se elaboró una macro en VBA (*Visual Basic for Applications*) [5], [6] y [7], que realiza trescientas tiradas, almacenando en una Tabla el resultado obtenido para cada una de ellas; se añade de esta forma una línea en la misma y un punto en el gráfico correspondiente, lo que proporciona una actualización o moviendo en tiempo real, que se consigue desactivando el “refresco de pantalla” de la macro, mediante la instrucción *Application.ScreenUpdating = False* [8]

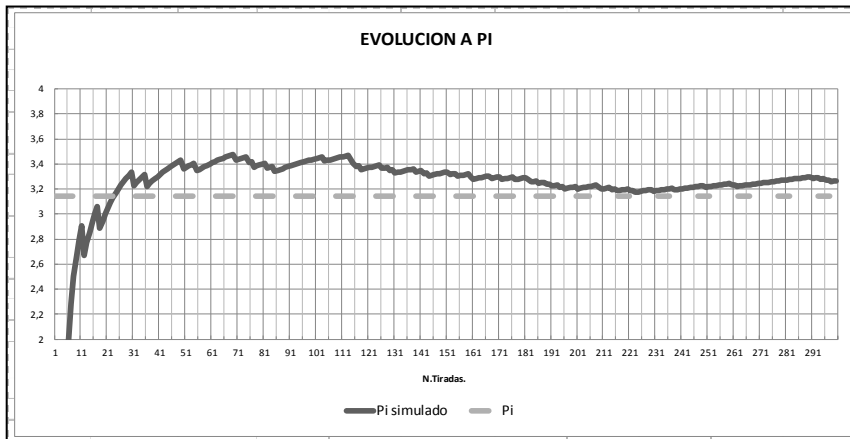


Figura 2: Representación en tiempo real de una gráfica

2.2 Gráficas automáticas con barra de desplazamiento

En otros trabajos explicamos cómo conseguir variaciones automáticas en los gráficos gracias al uso de botones de formulario y/o “controlesX” [8], y los hemos añadido a la representación automática citada [1], con lo cual hemos conseguido contar con una hoja que representa de forma automática una ecuación y permite cambiar sus parámetros.

A modo de ejemplo se ha representado las curvas de ingresos y costes totales, por cuya diferencia determinaremos el beneficio. Para unos costes fijos y variables prefijados, y un precio de venta dado, las rectas dependerán del número de unidades vendidas

(Figura 3). Partiendo de la gráfica superior, podemos, mediante una “barra de desplazamiento”, dotar de movimiento al gráfico al variar el número de unidades vendidas, consiguiendo que para el caso de 8 unidades vendidas, encontremos el beneficio cero o umbral de rentabilidad.

Se podría utilizar un programa de captura de pantalla como *Camtasia Studio* [9], para grabar la evolución del proceso.

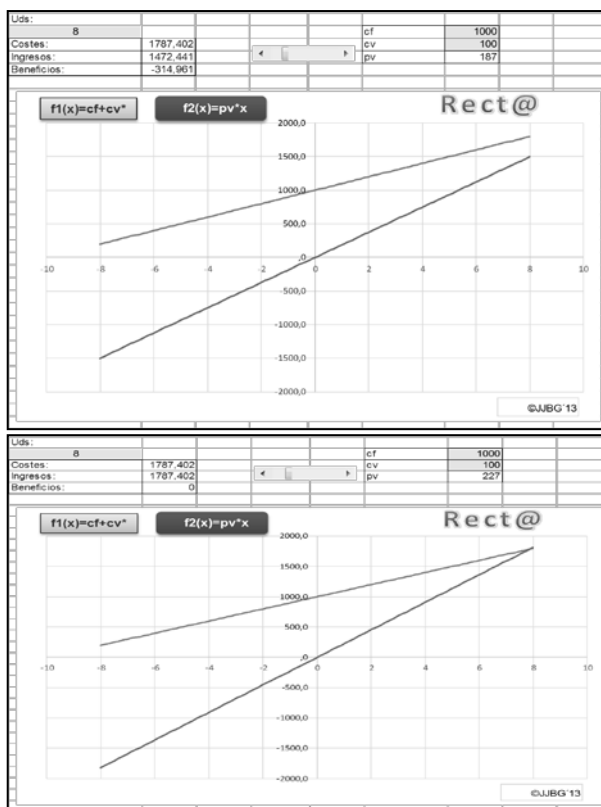


Figura 3: Líneas de costes totales e ingresos

2.3 Movimiento con VBA

La alternativa a los controles anteriores es la programación mediante código en VBA, es posible realizar un ciclo con variaciones incrementales de una función dada; así en el ejemplo presentado SENO(X), la Figura 4 muestra una instantánea de dicha gráfica en movimiento. Con la finalidad de poder fijar un momento dado en la representación, disponiendo así de una “foto fija”, se ha añadido en la macro realizada, y que se adjunta en el Anexo A, unas líneas de código que detienen el movimiento al volver a pulsar el

botón de “mover/parar”.

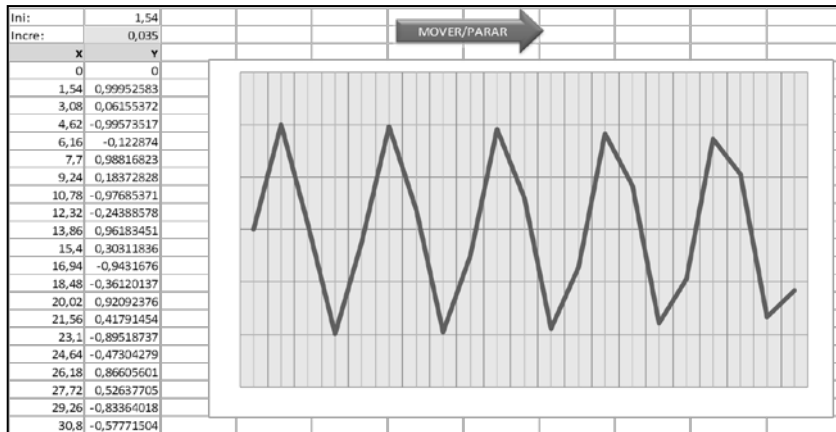


Figura 4: Gráfica SENO(X) en movimiento

2.4 Movimientos espaciales

Si representamos una función en 3D, lo que es posible en Excel mediante la elaboración de una Tabla de dos dimensiones, podremos programar macros para conseguir rotar en sentido horizontal, voltear verticalmente, o variar la perspectiva de dicha gráfica. A modo de ejemplo, en la Figura 5 se ha representado la parábola hiperbólica: $z=x^2-y^2$, y en la Figura 6: $z=x^2+y^2$; además se han añadido “formas” o “botones” [10] que accionan los macros citados (*Anexo B*). Se ha incluido asimismo en todos los casos, la posibilidad de parada, con lo cual podemos obtener distintas vistas en 3D. También es posible, en el visor de VBA pulsar {F5} y realizar el proceso paso a paso, lo que permite visualizarlo a voluntad.

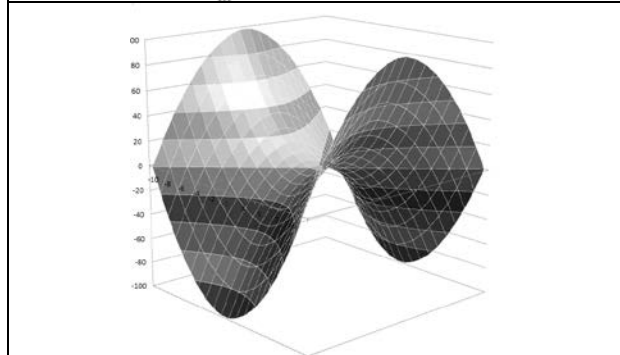
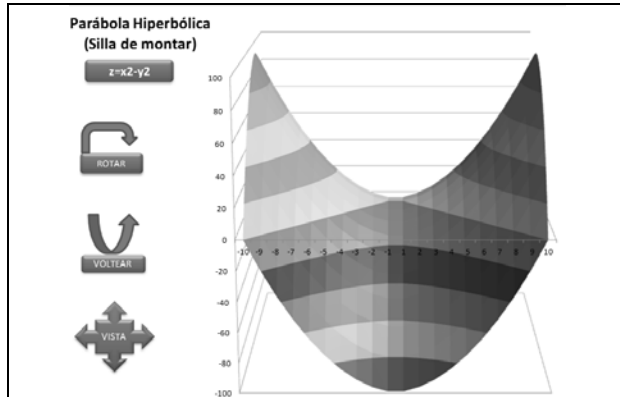


Figura 5: Gráfica $z=x^2-y^2$ en 3D con distintas vistas

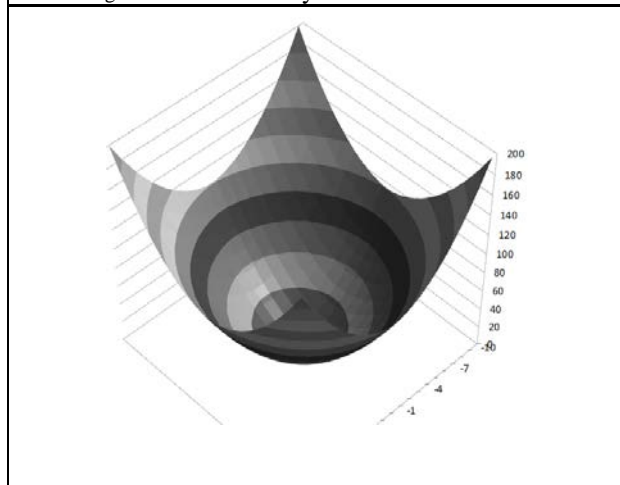


Figura 6: Gráfica $z=x^2+y^2$ en 3D

2.5. Gráficos con datos temporales

Si algún tipo de datos se presta al movimiento, son aquellos de tipo económico-empresarial, o series temporales. En el supuesto que aquí presentamos, se dispone de una larga lista de ventas realizadas comparadas con las previsiones de demanda, ordenadas por fechas; en concreto hablamos de mil entradas desde el 2009 al 2012.

El programa realizado permite elegir el día de inicio, el número de observaciones a mostrar a la vez, y el incremento o día de desplazamiento. La *Figura 7* presenta parte de la Tabla de datos, la entrada de parámetros citados, una barra de desplazamiento para variar el incremento, más una instantánea de la gráfica, ya que el parar el macro nos permite también mostrar los 20 días seleccionados. Se ha incluido la programación en VBA en el *Anexo C*.

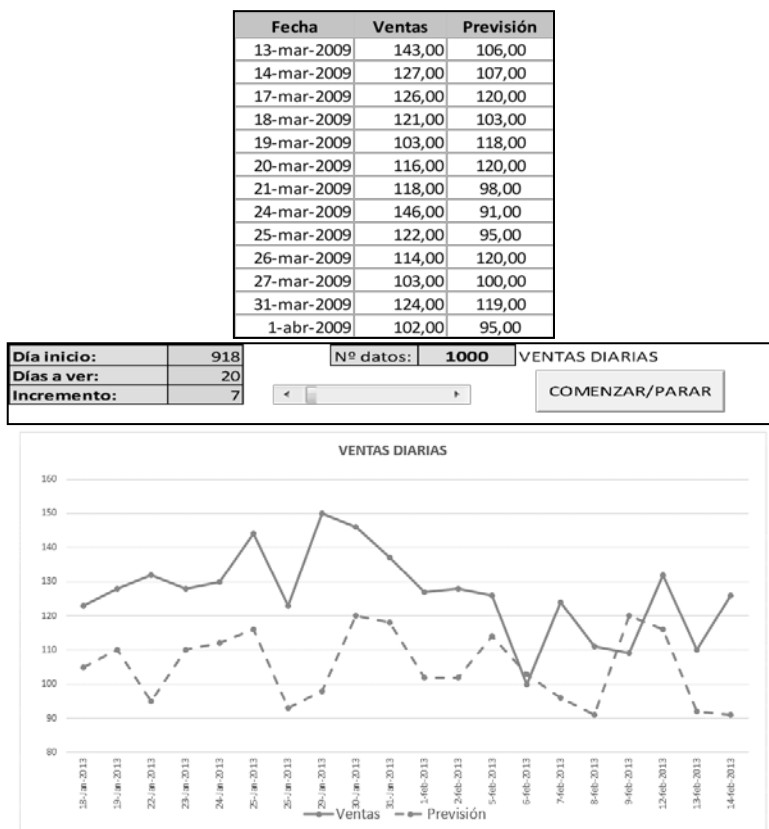


Figura 7 Gráfica en movimiento de ventas reales y esperadas

3. Gráfico de movimiento (Motion Chart) en Excel con VBA

Continuando con la anterior idea de dotar a los datos empresariales de movimiento, señalar que *Google Drive* (antes *Google Doc*) [11], ya cuenta con la posibilidad de mostrar gráficos de burbuja en movimiento (*Figura 8*).

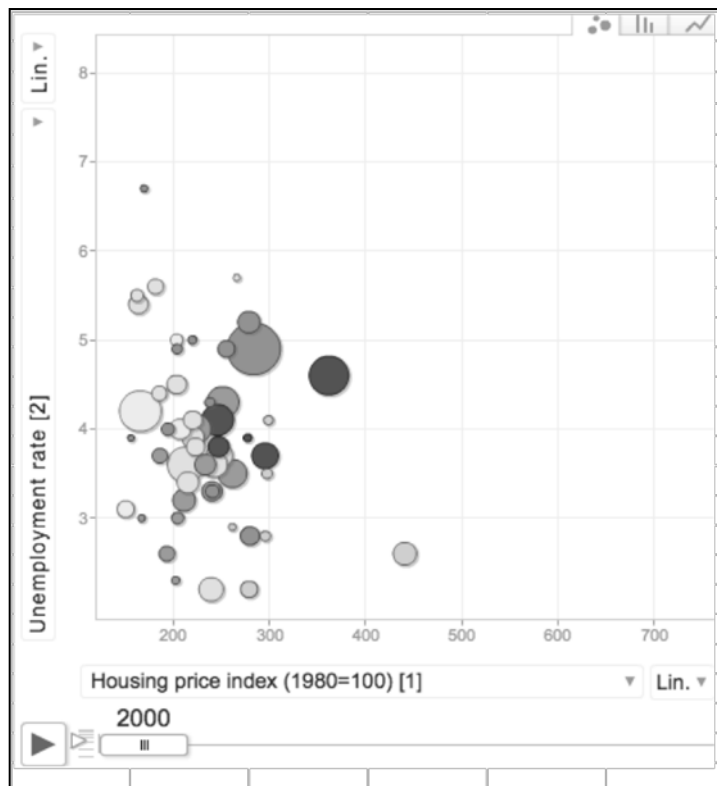


Figura 8 Motion Graf de Google drive

Nos propusimos emular con Excel ese tipo de gráficos, para lo cual recurrimos a los de burbuja 3D, que contienen una doble información, según su posición en el plano, y su tamaño. Para ello se parte de una Tabla con los valores de las ventas y el beneficio neto (BN) obtenido, en un listado por años y meses (*Figura 9*).

Meses	Año	número mes	Ventas	BN
enero	2010	1	29.155,10 €	380.536,71 €
febrero	2010	2	29.155,02 €	380.536,71 €
marzo	2010	3	44.264,54 €	380.536,71 €
abril	2010	4	27.279,94 €	380.536,71 €
mayo	2010	5	33.945,16 €	380.536,71 €
junio	2010	6	36.168,90 €	380.536,71 €
julio	2010	7	34.516,12 €	380.536,71 €
agosto	2010	8	15.878,73 €	380.536,71 €
septiembre	2010	9	27.886,96 €	380.536,71 €
octubre	2010	10	29.173,12 €	380.536,71 €
noviembre	2010	11	40.826,75 €	380.536,71 €
diciembre	2010	12	32.286,37 €	380.536,71 €

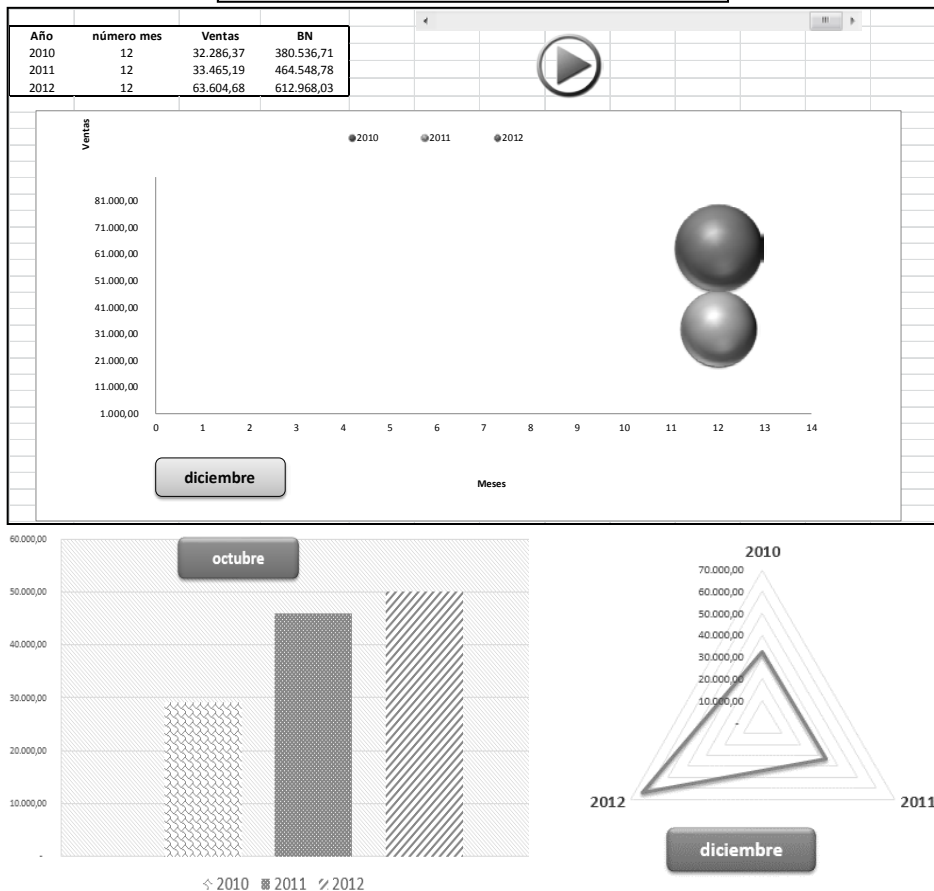


Figura 9: Gráfico en movimiento tipo Motion Char de bolas, barras y radial

Tras preparar una Tabla intermedia, que va extrayendo los datos de tres en tres (por año), podemos representar el gráfico de burbujas, al que posteriormente hemos dotado de una barra de desplazamiento. Si se desea, podemos variar los datos mostrados, o pulsar el icono de avance (>), que activa la macro correspondiente. Como mejora hemos

añadido la forma de texto variable con el mes [8], y en el código la instrucción (Sleep n), nos permite ralentizar el movimiento todo lo deseado para apreciar con detalle la evolución del mismo (*Anexo D*). Se ha probado, con esta técnica, otros tipos de gráficos en movimiento, resultando los más interesantes los de tipo barras y radial.

4. Futuras mejoras

Seguimos trabajando para obtener mejoras en esta línea de investigación, siendo la más inmediata la de unir las dos técnicas presentadas en Rect@, la de representación automática de gráficas en Excel [1] con la de dotarlas a su vez de movimiento, del presente trabajo.

5. Conclusiones

Volvemos a recalcar que la utilización avanzada de la hoja de cálculo Excel, permite conseguir mayores prestaciones, en un entorno muy versátil. Lo que lo hace asequible para la docencia, además de para la investigación y el mundo empresarial.

Referencias Bibliográficas (Ejemplos)

1. Bernal García, J.J, Martínez M^a Dolores, S.M. y Bernal Soto, P. Representación automática de funciones en Excel y su aplicación docente, Revista Rect@. Vol. **12**. Pág. 141-157. 20112-412. 2012.
2. <http://www.geogebra.org/cms/es/download/> (13/05/13)
3. <http://dmentrard.free.fr/GEOGEBRA/Maths/Surface3D/surfexpoMD.html> (13/05/13)
4. <http://idbs-xlfit5.software.informer.com/> (13/05/13)
5. Walkenbach,, J. Excel 2010.Programación con VBA, *Anaya Multimedia.-Wiley* (Madrid, 2011)
6. Walkenbach, J. , Excel 2010. Los mejores trucos, *Anaya multimedia*, (Madrid, 2011)
7. Jelen, B. y Syrstad,T. Excel. Macros y VBA. Trucos esenciales, *Ed Anaya*, (Madrid, 2008)
8. Bernal García, J.J. Aportaciones para la mejora de la presentación grafica de datos cuantitativos en Excel, Revista Rect@. Vol. **16**: 41242-412, XVI Congreso ASEPUMA y IV Encuentro Internacional. (Cartagena, 2008).
9. <http://www.techsmith.com/camtasia.html> (13/05/13)

10. Bernal García, J.J. Nuevas contribuciones a la mejora de la representación gráfica en Excel, *Anales ASEPUMA. N° 18.2010*, XVIII Congreso ASEPUMA y VI Encuentro Internacional (Santiago de Compostela, 2010)
11. <http://support.google.com/drive/bin/answer.py?hl=en&answer=1047434> (13/05/13)

Anexo A

Código VBA para función SENO() en movimiento

```

Option Explicit
Public MovtriIsRunning As Boolean
Sub movtri()
If MovtriIsRunning Then
    MovtriIsRunning = False
    End
End If
MovtriIsRunning = True
Dim i As Long, incre As Double
Range("ini") = 0
incre = Range("incre")
For i = 1 To 150
    DoEvents
    Range("ini") = Range("ini") + incre
    DoEvents
StopIt:
    Next i
    Rem Range("ini") = 0
End Sub

```

Anexo B

Código VBA para ROTAR, VOLTEAR y PERSPECTIVA

```

Sub Rotar()
Dim i As Long
If SilladuraIsRunning Then
    SilladuraIsRunning = False
    End
End If

```

```
SilladuraIsRunning = True
With ThisWorkbook.Sheets("Silladura").ChartObjects(1).Chart
    For i = 0 To 360 Step 8
        .Rotation = i
        '.Elevation = i - 90
        DoEvents
    Next i
End With
SilladuraIsRunning = False
End
```

```
SubSub voltrear()
    Dim i As Long
    ' Elevation range: -90 to +90
    If SilladuraIsRunning Then
        SilladuraIsRunning = False
        End
    End If
    SilladuraIsRunning = True
    With ThisWorkbook.Sheets("Silladura").ChartObjects(1).Chart
        For i = -90 To 90 Step 2
            .Elevation = i
            Application.Wait (Now + 0.000002)
            DoEvents
        Next i
        .Elevation = 15
    End With
    SilladuraIsRunning = False
End Sub
```

```
Sub perspectiva()
    Dim i As Long
    ' Perspective range = 0 to 100
    If SilladuraIsRunning Then
        SilladuraIsRunning = False
        End
    End If
    SilladuraIsRunning = True
```

```
With ThisWorkbook.Sheets("Silladura").ChartObjects(1).Chart
  For i = 0 To 100 Step 2
    .Perspective = i
    DoEvents
  Next i
  .Perspective = 30
End With
SilladuraIsRunning = False
End Sub
```

Anexo C

Public AnimationInProgress As Boolean

```
Sub AnimateChart()
  Dim Dstar As Long, r As Long
  If AnimationInProgress Then
    AnimationInProgress = False
  End
End If
AnimationInProgress = True
Dstar = Range("diainicio")
For r = Dstar To Range("ndatos") - Range("ndiasven") Step
Range("Ince")
  Range("diainicio") = r
  DoEvents
Next r
  AnimationInProgress = False
End Sub
```

Anexo D

Código VBA para Motion Graf para Rect@

```
Declare Sub Sleep Lib "kernel32" (ByVal dwMilliseconds As Long)
Private Sub Play()
  Dim i As Integer
  For i = 1 To 12:
    If grafmovi_rectaIsRunning Then
```

```
    grafmovi_rectaIsRunning = False
    End
End If
    Range("d5").Value = i
    Application.Calculate
    Sleep (500)
Next
End Sub
```

RESOLUCIÓN DE ECUACIONES DIFERENCIALES CON EXCEL

F^o JAVIER PALENCIA GONZÁLEZ

jpalencia@cee.uned.es

UNED – Departamento Economía Aplicada Cuantitativa I
P^o Senda del Rey, 11. 28040 – Madrid

RESUMEN: En este artículo se va a tratar de encontrar aproximaciones, de tipo numérico y gráfico, a las soluciones de problemas de valor inicial (PVI) de ecuaciones diferenciales ordinarias (EDO) mediante el uso de la hoja de cálculo Excel, la cual va a permitir realizarlo con cierta facilidad.

La aproximación a la solución se realiza de dos formas distintas: a) mediante la confección de una plantilla en Excel de un método numérico; b) mediante la programación en VBA de un procedimiento que calcule el método numérico en cuestión. En el primer caso se van a ir creando una serie de hojas para cada uno de los métodos en estudio, en las cuales se introducirán las fórmulas de los métodos numéricos: Euler, Taylor, Runge-Kutta, etc. En el segundo caso se creará un programa en VBA que a partir de unos datos iniciales genere de forma automática la tabla de datos aproximados a la solución aplicando uno u otro método. En cualquiera de los dos casos, una vez obtenida la tabla de datos y mediante la inserción de gráficos XY, se podrá tener una aproximación gráfica a la solución.

Palabras claves: Excel, ecuaciones diferenciales, métodos numéricos, gráficos, Euler, Taylor, Heun, Runge-Kutta.

ABSTRACT: This article will try to find approximations, both of numerical and graphical type, to the solutions of initial value problems (IVP) of ordinary differential equations (ODE) by using the Excel spreadsheet, which will allow to do it with some ease.

The approximation to the solution is made in two ways: a) making a template of a numerical method in Excel, and b) programming in VBA the method in question. In the first case we will create a series of sheets for each of the methods under study, in which the formulas of the numerical methods will be introduced: Euler, Taylor, Runge-Kutta, etc. In the second case, it will be created a VBA program to automatically generate the data table solution from initial data using any method. In both cases, once the data table is generated you can obtain a graphical approach to the solution by the insertion of a XY graph.

Keywords: Excel, differential equations, numerical methods, graphics, Euler, Taylor, Heun, Runge-Kutta.

1. Introducción

Las ecuaciones diferenciales se utilizan para describir, representar o modelizar distintos tipos de sistemas: físicos, químicos, biológicos, económicos, etc... Se puede por tanto inferir, que alrededor nuestro existen multitud de ecuaciones diferenciales que nos pueden explicar el funcionamiento de estos sistemas.

Se recuerdan en primer lugar algunos conceptos respecto de las ecuaciones diferenciales.

Una ecuación diferencial, ED, es una ecuación que contiene las derivadas de una o más variables dependientes, con respecto a una o más variables independientes.

$$F(x, y, \dots, f(x, y, \dots), \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \dots, \frac{\partial^2 f}{\partial x^2}, \dots) = 0 \quad (0.1)$$

dónde f es la variable dependiente que depende de x, y, \dots que son las variables independientes.

Si una ecuación contiene derivadas de una o más variables dependientes respecto a una sola variable independiente se dice que es una ecuación diferencial ordinaria, EDO, y en este caso se utiliza x como variable independiente e y como variable dependiente. Si una ecuación contiene las derivadas parciales de una o más variables dependientes respecto a dos o más variables independientes se dice que es una ecuación en derivadas parciales, EDP.

Se llama orden de una ecuación diferencial al orden de la derivada mayor que aparece en la ecuación. Luego si la derivada mayor que aparece en la ecuación es y' , se dice que la ecuación diferencial es de primer orden, si es y'' , se dice que es de segundo orden, y así sucesivamente. En este artículo se estudia la obtención de soluciones de ecuaciones diferenciales ordinarias de primer orden.

Se dice que una ecuación diferencial ordinaria está expresada en forma general o implícita cuando es de la forma $F(x, y, y', y'', \dots, y^{(n)}) = 0$, y se dice que una ecuación diferencial ordinaria está expresada en forma normal o explícita cuando es de la forma $y^{(n)} = f(x, y, y', \dots, y^{(n-1)})$.

Una ecuación diferencial en forma normal o explícita es lineal cuando f es una función lineal de $y, y', \dots, y^{(n-1)}$, esto significa que una ecuación diferencial es lineal si se puede escribir de la forma:

$$a_n(x) \frac{d^n y}{dx^n} + a_{n-1}(x) \frac{d^{n-1} y}{dx^{n-1}} + \dots + a_1(x) \frac{dy}{dx} + a_0(x) y = g(x) \quad (0.2)$$

con las propiedades siguientes:

- La variable dependiente y y todas sus derivadas $y', \dots, y^{(n)}$ son de primer grado, es decir, la potencia de todo término donde aparece y es 1.
- Los coeficientes a_0, a_1, \dots, a_n correspondientes a $y, y', \dots, y^{(n)}$ dependen únicamente de la variable independiente, x .

Si en (0.2), $g(x) = 0$, entonces se dice que la ecuación diferencial es homogénea.

En determinadas ocasiones se pretende resolver ecuaciones diferenciales que satisfacen unas condiciones dadas, en particular se llama problema de valor inicial, PVI, a:

$$\begin{cases} y^{(n)} = f(x, y, y', y'', \dots, y^{(n-1)}) \\ y(x_0) = y_0, y'(x_0) = y_1, \dots, y^{(n-1)}(x_0) = y_{n-1} \end{cases} \quad (0.3)$$

Los valores de y , y sus $n-1$ derivadas en el punto x_0 , se llaman condiciones o valores iniciales.

2. Soluciones de ecuaciones diferenciales

Dada una función $y(x)$, que está definida en un determinado intervalo, que tiene al menos n derivadas continuas en el intervalo dado, y que al sustituirse en la ecuación diferencial ordinaria de orden n que se pretende resolver, reduce la ecuación a una identidad, se considera solución (o integral) de la ecuación diferencial en el intervalo.

La gráfica de la solución $y(x)$ de una EDO se llama curva solución o curva integral.

Si la solución $y(x)$ está expresada solamente en términos de la variable independiente x , y/o en términos de constantes, entonces se dice que la solución es explícita. Un caso particular es cuando la solución dada es idéntica a 0 en un determinado intervalo, entonces se llama solución trivial.

En un PVI, si $n=1$, se está buscando una solución de la ecuación diferencial en el intervalo de manera que la curva solución pase por el punto (x_0, y_0) . Si $n=2$, se está buscando una solución de la ecuación diferencial en el intervalo de forma que la curva solución pase por el punto (x_0, y_0) , y que la pendiente de la curva solución en ese punto sea y_1 .

Encontrar soluciones a las ecuaciones diferenciales ordinarias no es tarea fácil en según qué casos, convirtiéndose en tarea compleja en algunos de ellos.

Si se tiene en cuenta su clasificación, se sabe que:

- para las ecuaciones diferenciales ordinarias lineales de primer orden, siempre puede encontrarse su solución exacta,
- para las ecuaciones diferenciales ordinarias no lineales de primer orden, únicamente pueden resolverse algunos casos especiales (ecuación de Bernoulli, ecuaciones exactas, ecuaciones separables, ecuaciones homogéneas, etc...),
- para las ecuaciones diferenciales ordinarias lineales de orden 2 o superior, se puede hallar una solución, bien exacta o bien mediante serie de potencias,
- para las ecuaciones diferenciales ordinarias no lineales de orden 2 o superior no existen métodos generales para llegar a una solución.

La opción resultante en los casos en que no se llega a obtener la solución exacta es analizar la ecuación bien de forma cualitativa, a través de los campos de direcciones, los puntos críticos y los diagramas de fase, o bien obtener soluciones cuantitativas, mediante la aplicación de métodos numéricos: Euler, Taylor, Runge-Kutta, etc...

La aplicación de hoja de cálculo Excel va a permitir obtener soluciones aproximadas a la solución real de forma cuantitativa, es decir, obtener la solución en determinados puntos de la ecuación diferencial de forma relativamente sencilla mediante el uso de diversos métodos numéricos, tal y como se expone a continuación.

La aplicación Excel va a permitir igualmente obtener la curva solución aproximada de la ecuación diferencial ordinaria que se esté estudiando.

3. Métodos numéricos

Se quiere hallar una aproximación a la solución de un problema de valor inicial de primer orden

$$\begin{cases} y'(x) = f(x, y(x)) \\ y(x_0) = y_0 \end{cases} \quad (2.1)$$

y se calculará por varios métodos numéricos, en todos ellos la idea que subyace es ir hallando valores $y_0, y_1, y_2, \dots, y_k$, cercanos a los de la solución $y(x)$ en una serie de puntos $x_0 < x_1 < x_2 < \dots < x_k$ separados entre sí una distancia h fija, también llamado paso, es decir, $x_1 = x_0 + h$, $x_2 = x_0 + 2h$, \dots , $x_k = x_0 + kh$.

3.1. Método de Euler

El más sencillo de los métodos es el de Euler, que consiste en aproximar la solución desconocida por su tangente conocida.

Como se conocen x_0 e y_0 , se conoce también el valor de la pendiente de la tangente de la curva solución en x_0 , $y'(x_0) = f(x_0, y_0)$ y se puede dibujar la recta tangente a la solución que pasa por el punto x_0 .

Si el paso h es pequeño es de esperar que el valor de la solución en el punto x_1 , $y(x_1) = y(x_0 + h)$, sea próximo al valor de la recta tangente en ese mismo punto:

$$y_1 = y_0 + hf(x_0, y_0).$$

Como el par (x_1, y_1) se parece al desconocido $(x_1, y(x_1))$ se puede aproximar la solución $y(x_2)$ por el valor y_2 que se obtiene a partir del punto (x_1, y_1) de la misma forma que se obtuvo la solución y_1 a partir del punto (x_0, y_0) inicial. Prosiguiendo así se van obteniendo las soluciones y_k aproximadas (más inexactas según se alejan de x_0), las cuales vienen dadas por:

$$y_{k+1} = y_k + h f(x_k, y_k) \quad (2.2)$$

Puesto que el PVI está planteado para ser resuelto en un cierto intervalo $[a, b] = [x_0, x_k]$, se divide el intervalo en m subintervalos de forma que $h = \frac{b-a}{m}$, y por tanto $x_n = a + nh$, para $n=0, 1, \dots, m$.

Se muestra a continuación una aplicación práctica de este método.

Ejemplo 1.- Dado el $\begin{cases} y'(x) = 0.1\sqrt{y} + 0.4x^2 \\ y(2) = 4 \end{cases}$ PVI: Resolver y dibujar la solución para $2 \leq x \leq 3$.

Solución.-

Para hallar la aproximación a la solución, se debe especificar el paso, h , que se va a utilizar y luego se halla la sucesión de puntos x_0, x_1, \dots, x_k , para una vez que se hayan hallado, calcular los valores de la función solución en cada uno de sus puntos de acuerdo a la fórmula (2.2).

Tal y como explicita el problema $a = 2$, y $b = 3$. Por tanto se debe elegir un número determinado de subintervalos m tal que especifique el paso a utilizar, h . Si se elige $m=5$, entonces $h=0.2$

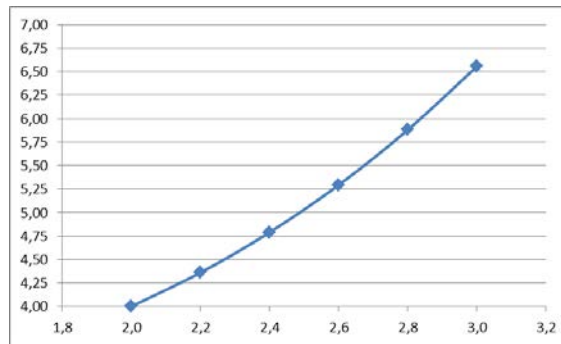
Ahora se genera la tabla de datos en Excel, tabla 1, para ello en la celda A1, se introduce “ h ” y en la celda B1, el correspondiente valor de h , en este caso 0,2. Seguidamente se le da nombre a esta celda como “ h ”, de esta forma se podrá utilizar la variable h en las fórmulas. En la segunda fila se ponen los nombres de las columnas, en nuestro caso x e y ; en la tercera fila se introducen el valor de x_0 , en la columna de las x y el valor de y_0 en la columna de las y ; en la cuarta fila se rellena la columna de las x , mediante la fórmula “ $=A3+h$ ”, de esta forma se va incrementando el valor de las x con el valor de h , de forma que se obtengan los distintos valores de x_n , y se rellena la columna de las y con la siguiente fórmula “ $=B3+h*(0.1*raiz(B3)+0.4*A3^2)$ ” que se corresponde con la fórmula dada en (2.2) y sustituyendo la f por la ecuación diferencial del PVI; para las siguientes filas nos basta con arrastrar las fórmulas y de esa forma se rellena la tabla, obteniéndose los siguientes valores:

Tabla 1. Solución del PVI con el Método de Euler

h		0,2
x	y	
2,00	4,00000000	
2,20	4,36000000	
2,40	4,78896123	
2,60	5,29352862	
2,80	5,88034396	
3,00	6,55604280	

Una vez obtenidos los datos se puede generar la curva solución, figura 1, con un gráfico XY.

Figura 1. Curva solución del PVI, Método Euler, $h=0,2$



Hallada la primera aproximación, se mejora la misma disminuyendo el valor de h .

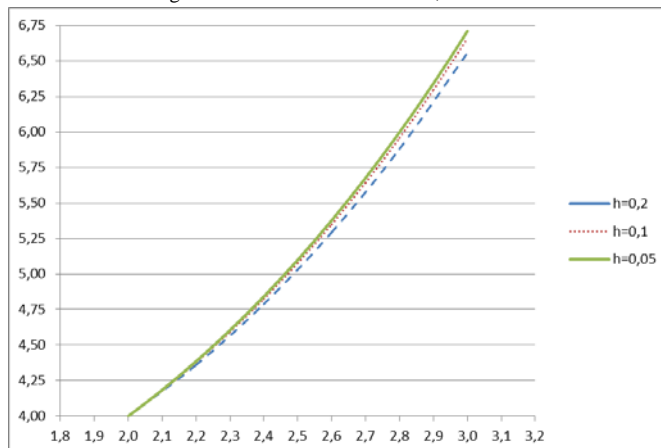
Tabla 2. Solución con el Método de Euler

h	0,20	0,1	0,05
x	y		
2,00	4,00000000	4,00000000	4,00000000
2,05			4,09000000
2,10		4,18000000	4,18416187
2,15			4,28258949
2,20	4,36000000	4,37684505	4,38538670
2,25			4,49265735
2,30		4,59136596	4,60450530
2,35			4,72103435
2,40	4,78896123	4,82439343	4,84234832
2,45			4,96855099

2,50		5,07675793	5,09974612
2,55			5,23603742
2,60	5,29352862	5,34928960	5,37752862
2,65			5,52432337
2,70		5,64281813	5,67652531
2,75			5,83423804
2,80	5,88034396	5,95817274	5,99756512
2,85			6,16661008
2,90		6,29618211	6,34147641
2,95			6,52226756
3,00	6,55604280	6,65767431	6,70908692

Para terminar el ejemplo, se muestra una gráfica dónde se recogen las tres aproximaciones de la curva solución halladas, figura 2.

Figura 2. Curva solución del PVI, Método Euler



3.2. Método de Euler mejorado.

Una vez que se ha visto en el punto anterior como hallar la solución aproximada en un punto para un PVI, en esta subsección y en las siguientes se irán mostrando otros métodos que irán refinando los resultados.

El Método de Euler mejorado consiste en elegir en cada paso en vez de la pendiente en un extremo como se hacía en el Método de Euler, el valor medio de las pendientes

asociadas a los dos extremos, el de partida y el previsto por el Método de Euler.

$$y_{k+1}^* = y_k + hf(x_k, y_k) \quad (2.3)$$

$$y_{k+1} = y_k + h \frac{f(x_k, y_k) + f(x_{k+1}, y_{k+1}^*)}{2} \quad (2.4)$$

Este método es un método de predicción y corrección, así el valor de y_{k+1}^* calculado en (2.3), predice un valor de y_k , mientras que el valor de y_{k+1} calculado en (2.4) corrige al anterior para obtener el valor buscado definitivo.

Ejemplo 2.- Dado el
$$\begin{cases} y'(x) = 0.1\sqrt{y} + 0.4x^2 \\ y(2) = 4 \end{cases} \quad \text{PVI: .}$$

Resolver y dibujar la solución para $2 \leq x \leq 3$ con el Método de Euler mejorado.

Solución.-

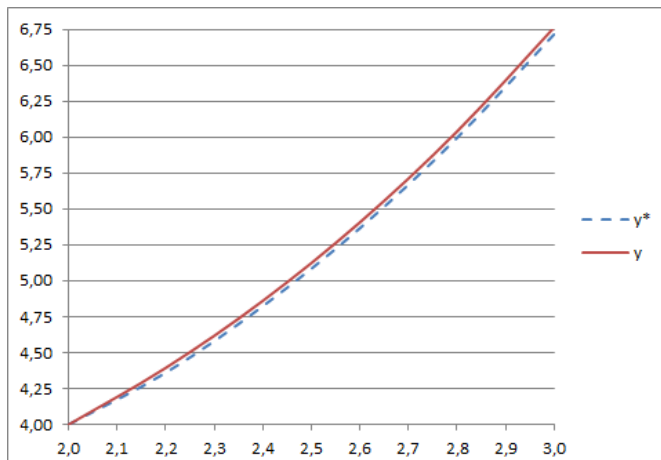
De forma análoga a como se hizo en el ejemplo 1, en una primera aproximación se toma $h=0,2$ y se genera la tabla de datos en Excel, tabla 3, para ello una vez introducidos los distintos valores de x_n , seguidamente se calcula el valor de y_1^* a partir de la fórmula (2.3), introduciendo en la hoja de cálculo la fórmula “=C3+h*(0,1*raíz(C3)+0,4*A3^2)”, de forma que se pueda corregir este valor mediante la fórmula (2.4) obteniendo el valor y_1 al introducir en la columna la fórmula “=C3+h/2*((0,1*raíz(C3)+0,4*A3^2)+(0,1*raíz(B4)+0,4*A4^2))”. Luego se dibuja la gráfica con el valor de predicción y el corregido, figura 3. Repitiendo el proceso para $h=0,1$ y $h=0,05$ se obtiene:

Tabla 3. Solución con el Método de Euler mejorado, $h=0,2$

h	0,20		0,1		0,05	
x	y^*	y	y^*	y	y^*	y
2,00		4,00000000		4,00000000	0,00000000	4,00000000
2,05					4,09000000	4,09208094
2,10			4,18000000	4,18842252	4,18624538	4,18837824
2,15					4,28681100	4,28899578
2,20	4,36000000	4,39448061	4,38528816	4,39412588	4,39180072	4,39403740
2,25					4,50131838	4,50360697
2,30			4,60868805	4,61794089	4,61546782	4,61780831
2,35					4,73435285	4,73674522
2,40	4,82360665	4,86140634	4,85103029	4,86069812	4,85807725	4,86052150
2,45					4,98674479	4,98924090
2,50			5,11314511	5,12322775	5,12045921	5,12300716

2,55					5,25932419	5,26192398
2,60	5,36630353	5,40742022	5,39586230	5,40635952	5,40344342	5,40609503
2,65					5,55292053	5,55562394
2,70			5,70001110	5,71092266	5,70785913	5,71061431
2,75					5,86836276	5,87116971
2,80	5,99472794	6,03915821	6,02642019	6,03774581	6,03453496	6,03739366
2,85					6,20647921	6,20938964
2,90			6,37591763	6,38765701	6,38429896	6,38726109
2,95					6,56809761	6,57111143
3,00	6,71550761	6,76324721	6,74933083	6,76148366	6,75797852	6,76104400

Figura 3. Curva solución del PVI, Método Euler mejorado, $h=0,2$



3.3. Método de Taylor de orden n

En este método se va a aproximar la solución $y(x)$ por el polinomio de Taylor de orden n . Como se sabe el polinomio de Taylor es:

$$y(x) = y(x_0) + y'(x)(x - x_0) + \frac{1}{2!} y''(x)(x - x_0)^2 + \dots + \frac{1}{n!} y^{(n)}(x)(x - x_0)^n + R_n(x) \tag{2.5}$$

donde se llama resto o error asociado al polinomio a

$$R_n(x) = \frac{y^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1} \tag{2.6}$$

Como los puntos x_k están equiespaciados, se tiene que $h = x_{k+1} - x_k$, y truncando el resto, se puede escribir (2.5) de la forma que sigue:

$$y_{k+1} = y(x_k) + y'(x_k)h + \frac{1}{2!}y''(x_k)h^2 + \dots + \frac{1}{n!}y^{(n)}(x_k)h^n \quad (2.7)$$

Pudiendo por tanto calcular para cada x_k el valor de la solución y . Si el orden del método es 1 se está en el Método de Euler descrito anteriormente.

3.3.1. Método de Taylor de orden 2

Se busca la solución al PVI mediante una aproximación al polinomio de Taylor de orden 2, por tanto los puntos aproximados responderán a la ecuación:

$$y_{k+1} = y(x_k) + y'(x_k)h + \frac{1}{2!}y''(x_k)h^2 \quad (2.8)$$

Ejemplo 3.- Dado el PVI: $y'(x) = 0,2xy$; $y(1) = 1$, hallar la solución para $1 \leq x \leq 2$ mediante el Método de Taylor de orden 2 con paso $h=0,1$.

Solución.-

Para resolver el PVI, en primer lugar se tiene que hallar la expresión de $y''(x)$, la cual se halla derivando a partir de la expresión de $y'(x)$ dada en el PVI, obteniéndose en este caso:

$$y''(x) = 0,2y + 0,2xy' = 0,2y + 0,04x^2y \quad (2.9)$$

Una vez hallada la expresión de $y''(x)$, se especifica el valor del paso, h , y seguidamente se procede a generar la tabla de datos en Excel, tabla 4, de forma análoga a como se ha hecho en los ejemplos anteriormente resueltos.

Se utilizan dos columnas auxiliares para calcular y' e y'' . En la columna de y' se introduce la fórmula “=0,2*A3*B3”, y en la de y'' la fórmula “=0,2*B3+0,04*A3^2*B3”. Finalmente en la siguiente fila, en la columna de las y se introduce la siguiente expresión “=B3+h*C3+h^2/2*D3”. Se rellenan el resto de filas de la tabla y se obtienen los siguientes datos:

Tabla 4. Solución con el Método de Taylor de orden 2

<i>h</i>		0,1	
<i>x</i>	<i>y</i>	<i>y'</i>	<i>y''</i>
1,00	1,00000000	0,20000000	0,24000000
1,10	1,02120000	0,22466400	0,25366608
1,20	1,04493473	0,25078434	0,26917519
1,30	1,07135904	0,27855335	0,28669568
1,40	1,10064785	0,30818140	0,30642036
1,50	1,13299809	0,33989943	0,32856945
1,60	1,16863089	0,37396188	0,35339398
1,70	1,20779404	0,41064997	0,38117980
1,80	1,25076494	0,45027538	0,41225212
1,90	1,29785374	0,49318442	0,44698083
2,00	1,34940708		

3.3.2. Método de Taylor de orden 3

Se busca la solución al PVI mediante una aproximación al polinomio de Taylor de orden 3, por tanto los puntos aproximados responderán a la ecuación:

$$y_{k+1} = y(x_k) + y'(x_k)h + \frac{1}{2!} y''(x_k)h^2 + \frac{1}{3!} y'''(x_k)h^3 \quad (2.10)$$

Ejemplo 4.- Dado el PVI: $y'(x) = 0,2xy$; $y(1) = 1$, hallar la solución para $1 \leq x \leq 2$ mediante el Método de Taylor de orden 3 con paso $h=0,05$.

Solución.-

Para resolver el PVI, en primer lugar se tiene que hallar la expresión de $y''(x)$, e $y'''(x)$, las cuales se hallan derivando a partir de la expresión de $y'(x)$ dada en el PVI, obteniéndose en este caso:

$$y''(x) = 0,2y + 0,2xy' = 0,2y + 0,04x^2y \quad (2.11)$$

$$y'''(x) = 0,2y' + 0,08xy + 0,04x^2y' = 0,12xy + 0,008x^3y \quad (2.12)$$

Una vez halladas las expresiones de la segunda y la tercera derivada, se especifica el valor del paso, h , y seguidamente se procede a generar la tabla de datos en Excel, tabla 5, de forma análoga a como se ha hecho en los ejemplos anteriormente resueltos.

Se utilizan tres columnas auxiliares para calcular y' , y'' e y''' . En la columna y' se introduce la fórmula “=0,2*A3*B3”; en la columna y'' la fórmula “=0,2*B3+0,04*A3^2*B3” y en la columna y''' se introduce la fórmula “=0,12*A3*B3+0,008*A3^3*B3”. Finalmente en la siguiente fila en la columna de las y se introduce la siguiente expresión “=B3+h*C3+h^2/2*D3+h^3/FACT(3)*E3”. Se rellenan el resto de filas de la tabla y se obtienen los siguientes datos:

Tabla 5. Solución con el Método de Taylor de orden 3

h		0,05		
x	y	y'	y''	y'''
1,00	1,00000000	0,20000000	0,24000000	0,12800000
1,05	1,01030267	0,21216356	0,24661488	0,13665455
1,10	1,02122196	0,22466883	0,25367153	0,14567527
1,15	1,03277553	0,23753837	0,26118893	0,15508880
1,20	1,04498216	0,25079572	0,26918740	0,16492326
1,25	1,05786187	0,26446547	0,27768874	0,17520837
1,30	1,07143590	0,27857333	0,28671625	0,18597556
1,35	1,08572684	0,29314625	0,29629485	0,19725811
1,40	1,10075863	0,30821242	0,30645120	0,20909130
1,45	1,11655667	0,32380143	0,31721375	0,22151256
1,50	1,13314787	0,33994436	0,32861288	0,23456161
1,55	1,15056074	0,35667383	0,34068104	0,24828065
1,60	1,16882546	0,37402415	0,35345282	0,26271456
1,65	1,18797396	0,39203141	0,36696516	0,27791106
1,70	1,20804002	0,41073361	0,38125743	0,29392097
1,75	1,22905940	0,43017079	0,39637166	0,31079840
1,80	1,25106988	0,45038516	0,41235263	0,32860101
1,85	1,27411142	0,47142123	0,42924814	0,34739030
1,90	1,29822628	0,49332599	0,44710913	0,36723186
1,95	1,32345912	0,51614906	0,46598996	0,38819570
2,00	1,34985715			

3.3.3. Error cometido en la aproximación

En general, los valores aproximados obtenidos para y_k no son iguales a la solución real de la función evaluada en el punto x_k , $y(x_k)$. El error cometido en cada paso se llama error de truncamiento local o de discretización, es decir, suponiendo que y_k es preciso, y_{k+1} contiene error de truncamiento local, y viene dado por:

$$e(h) = y(x_k) - y_k \quad (2.13)$$

Se dice que el error cometido es de orden h^n , $O(h^n)$, si existe una constante K y un valor n positivo:

$$|e(h)| \leq Kh^n \quad (2.14)$$

Se ha supuesto que y_k es exacto, pero normalmente no lo es, así el error total al calcular y_{k+1} es una acumulación de errores dada en cada paso en que se ha calculado y_1, y_2, \dots, y_k . A este error acumulado se le llama error de truncamiento global o error global. Se puede demostrar de forma general que si un método numérico de resolución de una ecuación diferencial tiene un error de truncamiento local $O(h^{n+1})$, el error de truncamiento global es $O(h^n)$. De acuerdo a (2.6), el método de Taylor de orden n tiene un error de truncamiento local $O(h^{n+1})$ y un error global $O(h^n)$. Como el método de Euler es un método de Taylor de orden 1, sus errores de truncamiento local y global son $O(h^2)$ y $O(h)$ respectivamente.

Tabla 6. Errores de la solución aproximada respecto de la solución exacta, $h=0,2$

y'	$0,2*x*y$	$y=k*e^{(0,1*x^2)}$	
h	0,2	$k=1/(e^{0,1})$	
x	Taylor 3	RK-4	Solución Exacta
1,00	1,0000000000	1,0000000000	1,0000000000
1,20	1,0449706667	1,0449823522	1,0449823549
1,40	1,1007328428	1,1007590560	1,1007590640
1,60	1,1687815443	1,1688261847	1,1688262031
1,80	1,2510026269	1,2510709818	1,2510710194
2,00	1,3497595365	1,3498587354	1,3498588076
Error	0,0000992711	0,0000000722	

Tabla 7. Errores de la solución aproximada respecto de la solución exacta, $h=0,1$

y'	$0,2*x*y$	$y=k*e^{(0,1*x^2)}$	
h	0,1	$k=1/(e^{0,1})$	
x	Taylor 3	RK-4	Solución Exacta
1,00	1,0000000000	1,0000000000	1,0000000000
1,10	1,0212213333	1,0212220516	1,0212220516
1,20	1,0449808388	1,0449823548	1,0449823549
1,30	1,0714338014	1,0714362089	1,0714362091
1,40	1,1007556545	1,1007590636	1,1007590640
1,50	1,1331439132	1,1331484524	1,1331484531
1,60	1,1688203827	1,1688262021	1,1688262031
1,70	1,2080336769	1,2080409510	1,2080409525
1,80	1,2510620857	1,2510710173	1,2510710194
1,90	1,2982168376	1,2982276624	1,2982276654
2,00	1,3498458116	1,3498588034	1,3498588076
Error	0,0000129959	0,0000000042	

Como se puede observar comparando los resultados de las tablas 6 y 7, y sabiendo que para el método de Taylor de orden 3 el error local es $O(h^4)$ y el error global es $O(h^3)$, esto implica que cuando el paso se reduce a la mitad el error se ha de reducir entre un factor aproximado de $h^3 = 2^3$, lo cual efectivamente se puede comprobar que ocurre si se calcula la relación que existe entre los errores cometidos:

$$\frac{0,000092711}{0,0000129959} \square 7,63$$

Análogamente y como se verá en la siguiente sección, en el Método de Runge-Kutta de orden 4 el error local es $O(h^5)$ y el error global es $O(h^4)$, esto implica que cuando el paso se reduce a la mitad el error se ha de reducir entre un factor aproximado de $h^4 = 2^4$, lo cual efectivamente se puede comprobar que ocurre si se calcula la relación que existe entre los errores cometidos para el citado método en las tablas 6 y 7:

$$\frac{0,0000000722}{0,0000000042} \square 17,19$$

3.4. Método de Runge-Kutta de orden n

Se ha visto que el método de Taylor de orden n , tiene errores de truncamiento y globales de orden alto, lo cual aproxima bastante bien los valores obtenidos a la solución real, sin embargo tiene la desventaja de requerir en todo momento el cálculo y evaluación de la derivadas sucesivas de $f(x,y)$, lo cual es un procedimiento lento y complicado en gran número de ocasiones.

La idea es desarrollar fórmulas equivalentes a los polinomios de Taylor de orden n en las que no aparezcan las derivadas de f .

El desarrollo de estas fórmulas dio lugar al método de Runge-Kutta, en los cuales el orden de exactitud se va a ver incrementado por el empleo de métodos de integración de mayor orden, luego el resultado va a ser más exacto y los errores se reducirán con mayor rapidez al reducir el tamaño del paso, h . Cada método de Runge-Kutta posee un determinado número de etapas, que coincide con el número de evaluaciones de la función f necesarias para avanzar un paso h .

La idea es resolver (2.1) mediante integración, así se tiene que:

$$\int_{y_0}^y dy = \int_{x_0}^x f(x, y(x)) dx \Rightarrow y_{k+1} - y_k = \int_{x_k}^{x_{k+1}} f(x, y(x)) dx \quad (2.15)$$

por tanto ahora, se necesita calcular la integral del lado derecho, lo que se hará por alguno de los métodos numéricos de integración conocidos. Así se obtiene:

$$y_{k+1} = y_k + h\phi(x_n, y_n) \quad (2.16)$$

Siendo

$$\begin{aligned}\phi(x_n, y_n) &= \sum_{j=1}^n c_j K_j \\ K_1 &= f(x_k, y_k) \\ K_j &= f(x_k + ha_j, y_k + h \sum_{p=1}^{j-1} b_{jp} K_p), \quad j = 2, 3, \dots, k\end{aligned}\tag{2.17}$$

Si $k=1$, se obtiene nuevamente el método de Euler.

3.4.1. Método de Runge-Kutta de orden 2

Por (2.16) y (2.17) se tiene que para orden $n=2$

$$y_{k+1} = y_k + h(c_1 K_1 + c_2 K_2)\tag{2.18}$$

donde

$$\begin{aligned}K_1 &= f(x_k, y_k) \\ K_2 &= f(x_k + ha_2, y_k + hb_{21} K_1)\end{aligned}\tag{2.19}$$

Por otro lado, si para aproximar la integral dada en (2.15) se utiliza el método de integración de los trapecios, es decir:

$$\int_{x_k}^{x_{k+1}} f(x, y(x)) dx \approx \frac{1}{2} h (f(x_k, y_k) + f(x_{k+1}, y_{k+1}))\tag{2.20}$$

se tiene que para que la fórmula dada en (2.18) coincida con un Polinomio de Taylor de orden 2, debe cumplirse que:

$$c_1 + c_2 = 1; \quad a_2 c_2 = 1/2; \quad b_{21} c_2 = 1/2\tag{2.21}$$

luego, se tiene un sistema de 3 ecuaciones con 4 incógnitas, lo cual da lugar a infinitas soluciones. En función de los valores que se vayan escogiendo se estará en uno u otro método.

Si $c_1 = c_2 = 1/2$ y $a_2 = b_{21} = 1$, se está en el caso del ya estudiado Método de Euler mejorado.

Si $c_1 = 1/4$, $c_2 = 3/4$, $a_2 = b_{21} = 2/3$, se está ante el Método de Heun:

$$y_{k+1} = y_k + \frac{1}{4} h \left[f(x_k, y_k) + 3f\left(x_k + \frac{2}{3}h, y_k + \frac{2}{3}hf(x_k, y_k)\right) \right] = y_k + \frac{1}{4} h (K_1 + 3K_2)\tag{2.22}$$

Ejemplo 5.- Dado el PVI: $y'(x) = 0,2xy$; $y(1) = 1$, hallar la solución para $1 \leq x \leq 2$ mediante el Método de Heun.

Solución.-

Para resolver el PVI, y una vez especificado el valor del paso, h , se procede a generar la tabla de datos en Excel, tabla 8, de forma análoga a como se ha hecho en los ejemplos anteriormente resueltos.

Se utilizan dos columnas auxiliares para calcular K_1 y K_2 que de acuerdo a (2.19) y a (2.22) valdrán:

$$K_1 = 0,2 * x_k * y_k$$

$$K_2 = 0,2 * (x_k + \frac{2}{3}h) * (y_k + \frac{2}{3}h * K_1)$$

luego, en las columnas se introducirán respectivamente las siguientes fórmulas “=0,2*A3*B3” y “=0,2*(A3+2/3*h)*(B3+2/3*h*C3). Finalmente en la siguiente fila en la columna de las y se introduce la siguiente expresión “=B3+h/4*(C3+3*D3)”. Se rellenan el resto de filas de la tabla y se obtienen los siguientes datos:

Tabla 8. Solución con el Método de Runge-Kutta de orden 2 (Método de Heun)

h		0,1	
x	y	$K1$	$K2$
1,00	1,00000000	0,20000000	0,21617778
1,10	1,02121333	0,22466693	0,24177793
1,20	1,04496335	0,25079120	0,26895963
1,30	1,07140510	0,27856533	0,29792681
1,40	1,10071375	0,30819985	0,32890305
1,50	1,13308647	0,33992594	0,36213444
1,60	1,16874470	0,37399831	0,39789264
1,70	1,20793661	0,41069845	0,43647850
1,80	1,25093996	0,45033839	0,47822601
1,90	1,29806537	0,49326484	0,52350688
2,00	1,34966001	0,53986400	0,57273572

3.4.2. Método de Runge-Kutta de orden 3

En este caso la integración se hace por el método numérico de integración de Simpson y de las ecuaciones (2.16) y (2.17) se tiene que para $n=3$ se puede llegar a las siguientes expresiones dando ciertos valores a los parámetros

$$K_1 = f(x_k, y_k)$$

$$K_2 = f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}hK_1) \quad (2.23)$$

$$K_3 = f(x_k + h, y_k - hK_1 + 2hK_2)$$

y, por tanto:

$$y_{k+1} = y_k + \frac{h}{6}(K_1 + 4K_2 + K_3) \quad (2.24)$$

Ejemplo 6.- Dado el PVI: $y'(x) = 0,2xy$; $y(1) = 1$, hallar la solución para $1 \leq x \leq 2$ mediante el Método de Runge-Kutta de orden 3.

Solución.-

Para resolver el PVI, y una vez especificado el valor del paso, h , se procede a generar la tabla de datos en Excel, tabla 9, de forma análoga a como se ha hecho en los ejemplos anteriormente resueltos.

Se utilizan tres columnas auxiliares para calcular K_1 , K_2 y K_3 que de acuerdo a (2.23) y a (2.24) valdrán:

$$K_1 = 0,2 * x_k * y_k$$

$$K_2 = 0,2 * (x_k + \frac{1}{2}h) * (y_k + \frac{1}{2}h * K_1)$$

$$K_3 = 0,2 * (x_k + h) * (y_k - h * K_1 + 2 * h * K_2)$$

luego, en las columnas se introducirán respectivamente las siguientes fórmulas “=0,2*A3*B3”, “=0,2*(A3+1/2*h)*(B3+1/2*h*C3)” y “=0,2*(A3+h)*(B3-h*C3+2*h*D3)”. Finalmente en la siguiente fila en la columna de las y se introduce la siguiente expresión “=B3+h/6*(C3+4*D3+E3)”. Se rellenan el resto de filas de la tabla y se obtienen los siguientes datos:

Tabla 9. Solución con el Método de Runge-Kutta de orden 3

h		0,1		
x	y	$K1$	$K2$	$K3$
1,00	1,00000000	0,20000000	0,21210000	0,22493240
1,10	1,02122221	0,22466889	0,23746480	0,25109959
1,20	1,04498267	0,25079584	0,26438061	0,27892259
1,30	1,07143668	0,27857354	0,29304865	0,30861294
1,40	1,10075970	0,30821272	0,32368940	0,34040289
1,50	1,13314925	0,33994478	0,35654541	0,37454843
1,60	1,16882717	0,37402469	0,39188437	0,41133253
1,70	1,20804208	0,41073431	0,43000258	0,45106890
1,80	1,25107231	0,45038603	0,47122889	0,49410620
1,90	1,29822910	0,49332706	0,51592923	0,54083290
2,00	1,34986038	0,53994415	0,56451161	0,59168268

3.4.3. Método de Runge-Kutta de orden 4 (RK4)

Este método es uno de los más usados, por su exactitud. Por (2.16) y (2.17) se tiene que para $n=4$, hay 11 ecuaciones y 13 incógnitas. Los conjuntos de valores más usados se corresponden con los métodos de integración 1/3 de Simpson, de donde resultan las siguientes expresiones,

$$\begin{aligned}
 K_1 &= f(x_k, y_k) \\
 K_2 &= f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}hK_1) \\
 K_3 &= f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}hK_2) \\
 K_4 &= f(x_k + h, y_k + hK_3)
 \end{aligned}
 \tag{2.25}$$

y, por tanto:

$$y_{k+1} = y_k + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) \tag{2.26}$$

y del método de integración 3/8 de Simpson, de donde resultan las siguientes expresiones,

$$\begin{aligned}
 K_1 &= f(x_k, y_k) \\
 K_2 &= f(x_k + \frac{1}{3}h, y_k + \frac{1}{3}hK_1) \\
 K_3 &= f(x_k + \frac{2}{3}h, y_k + \frac{1}{3}hK_1 + \frac{1}{3}hK_2) \\
 K_4 &= f(x_k + h, y_k + hK_1 - hK_2 + hK_3)
 \end{aligned}
 \tag{2.27}$$

y, por tanto:

$$y_{k+1} = y_k + \frac{h}{8}(K_1 + 3K_2 + 3K_3 + K_4) \tag{2.28}$$

Ejemplo 7.- Dado el PVI: $y'(x) = 0,2xy$; $y(1) = 1$, hallar la solución para $1 \leq x \leq 2$ con el Método RK4.

Solución.-

Para resolver el PVI, conocido h , se genera la tabla de datos en Excel, tabla 10. Se utilizan cuatro columnas auxiliares para calcular K_1 , K_2 , K_3 y K_4 que de acuerdo a (2.27) y a (2.28) valdrán:

$$\begin{aligned}
 K_1 &= 0,2 * x_k * y_k \\
 K_2 &= 0,2 * (x_k + \frac{1}{2}h) * (y_k + \frac{1}{2}h * K_1) \\
 K_3 &= 0,2 * (x_k + \frac{1}{2}h) * (y_k + \frac{1}{2}h * K_2) \\
 K_4 &= 0,2 * (x_k + h) * (y_k + h * K_3)
 \end{aligned}$$

luego, en las columnas se introducirán respectivamente las siguientes fórmulas “=0,2*A3*B3”, “=0,2*(A3+1/2*h)*(B3+1/2*h*C3)”, “=0,2*(A3+1/2*h)*(B3+1/2*h*D3)” y “=0,2*(A3+h)*(B3+h*E3)”. Finalmente en la siguiente fila en la columna de las y se introduce la siguiente expresión “=B3+h/6*(C3+2*D3+2*E3+F3)”. Se rellena el resto de la tabla y se obtienen los siguientes datos:

Tabla 10. Solución con el Método de Runge-Kutta de orden 4

<i>h</i>		0,1			
<i>x</i>	<i>y</i>	<i>K1</i>	<i>K2</i>	<i>K3</i>	<i>K4</i>
1,00	1,00000000	0,20000000	0,21210000	0,21222705	0,22466900
1,10	1,02122205	0,22466885	0,23746476	0,23761192	0,25079598
1,20	1,04498235	0,25079577	0,26438054	0,26455035	0,27857372
1,30	1,07143621	0,27857341	0,29304852	0,29324393	0,30821297
1,40	1,10075906	0,30821254	0,32368921	0,32391362	0,33994513
1,50	1,13314845	0,33994454	0,35654516	0,35680247	0,37402518
1,60	1,16882620	0,37402438	0,39188405	0,39217873	0,41073499
1,70	1,20804095	0,41073392	0,43000218	0,43033937	0,45038696
1,80	1,25107102	0,45038557	0,47122841	0,47161400	0,49332832
1,90	1,29822766	0,49332651	0,51592866	0,51636940	0,53994584
2,00	1,34985880	0,53994352	0,56451095	0,56501458	0,59067131

4. Solución de ecuaciones diferenciales mediante VBA (Visual Basic for Applications)

En esta sección se va a implementar mediante VBA algunos de los métodos vistos para solucionar ecuaciones diferenciales, en particular se programan sencillos procedimientos para los Métodos de Euler, de Taylor de orden 3 y de Runge-Kutta de orden 4.

4.1. Método de Euler

Como se ha visto en este método se van obteniendo los puntos mediante la fórmula:

$$y_{k+1} = y_k + h f(x_k, y_k) \quad (3.1)$$

El programa va a leer los valores del PVI directamente de la hoja de cálculo, en particular se ha de conocer el valor del paso, h , que se introducirá en la celda B2, el valor de x_0 que estará en la primera fila de la tabla a generar, celda "A5" y el valor de y_0 , que estará en la celda "B5". Se necesitará también el valor de x_n en el que se quiere conocer la función solución, celda B3.

El código VBA del procedimiento Euler que soluciona el PVI dado en el Ejemplo 1, es:

```
Sub Euler()
' Halla la solución de un PVI mediante el Método de Euler
' Lee h, x0, y0 y xn
' Genera una tabla con los valores aproximados de la función y
' evaluada en los puntos x0, x1, ..., xn
```



```

' Fº Javier Palencia, 20-V-2013
' Se asignan los valores conocidos a sus respectivas variables
x0 = Range("A5").Value
y0 = Range("B5").Value
h = Range("B2").Value
xn = Range("B3").Value
' Se halla el número de subintervalos
m = (xn - x0) / h
' Se asignan los valores conocidos a las variables de trabajo
xk = x0
yk = y0
For i = 1 To m
' Se halla el valor de la solución en el siguiente punto
yk = yk + h * (0.1 * Sqr(yk) + 0.4 * xk ^ 2)
xk = xk + h
' Se escriben los valores en la tabla
Cells(5 + i, 1).Value = xk
Cells(5 + i, 2).Value = yk
Next i
End Sub

```

Se asigna el código a un botón y haciendo clic en el mismo se obtiene la tabla de resultados, tal y como se muestra en la figura 4.

	A	B	C	D
1	y'	0,1*raiz(y)+0,4*x^2		
2	h	0,2		
3	xn	3	Euler	
4	x	y		
5	2,00	4,00000000		
6	2,20	4,36000000		
7	2,40	4,78896123		
8	2,60	5,29352862		
9	2,80	5,88034396		
10	3,00	6,55604280		
11				
12				

Figura 4. Tabla generada mediante la macro Euler para el Ejemplo 1

Al implementar el procedimiento, se ha escrito específicamente la función que se está tratando, así se ha puesto en el código la línea:

$$y_k = y_k + h * (0.1 * \text{Sqr}(y_k) + 0.4 * x_k^2)$$

4.2. Método de Euler independiente del PVI

Sería interesante que el código del procedimiento fuera independiente del PVI a

solucionar, es decir que el procedimiento tome los distintos valores iniciales de la hoja de datos, inclusive la función que define la ecuación diferencial y devuelva la tabla generada.

Pues bien, esto se va a lograr haciendo unas pequeñas modificaciones en el código y en la hoja de entrada de datos.

Para que la función pueda ser evaluada, tienen que estar especificadas las variables que toman parte en la ecuación a evaluar en la hoja de cálculo, en el caso que nos ocupa x e y . Luego se ha de especificar mediante el administrador de nombres de Excel en que celda está “ x ” y en que celda se encuentra “ y ”. En nuestro caso, serán las celdas “A5”, y “B5”. Seguidamente se escribe la fórmula de la ecuación y se pone en la celda “B1”, es decir la celda ahora será “ $0.1*\text{sqrt}(y)+0.4*x^2$ ”, y finalmente en el código del procedimiento se hace una llamada a la función “Evaluate”, de forma que en ese momento se evalúe la función con los valores actuales de x e y . El código sería de la siguiente forma:

```
Sub Euler_Eval()
'Halla la solución de un PVI mediante el Método de Euler
' Lee h, x0, y0 y xn
' Fº Javier Palencia, 20-V-2013
' Se asignan los valores conocidos a sus respectivas variables
x0 = Range("A5").Value
y0 = Range("B5").Value
h = Range("B2").Value
xn = Range("B3").Value
' Se asigna la función que define la ecuación
f = Range("B1").Formula
' Se halla el número de subintervalos
m = (xn - x0) / h
' Se asignan los valores conocidos a las variables de trabajo
xk = x0
yk = y0
For i = 1 To m
'Evaluamos la función, x e y están definidas en la hoja
yk = yk + h * Evaluate(f & "+x*0")
xk = xk + h
' Se escriben los valores en la tabla
Cells(5 + i, 1).Value = xk
Cells(5 + i, 2).Value = yk
' Se actualizan los valores de x e y para reevaluar la función
Cells(5, 1).Value = xk
Cells(5, 2).Value = yk
```

```

Next i
'Se restablecen los valores iniciales en la tabla
Cells(5, 1).Value = x0
Cells(5, 2).Value = y0
End Sub

```

Se asigna el código a un botón y haciendo clic en el mismo se obtiene la tabla 11.a) de resultados. Para hallar los resultados de un nuevo PVI, el que se establece en el ejemplo 4, se cambia la ecuación a tratar en la celda "B1" que toma ahora la expresión "0.2*x*y", y se establecen los valores iniciales y finales, así se establece el valor de x_0 en la celda "A5" que viene dado por el PVI y es 1, el valor de y_0 en la celda "B5" que vale también 1 y en la celda "B3" se establece el valor donde se quiere hallar la solución, es decir x_n , que viene dado por el problema y es 2. Ahora se vuelve a ejecutar el procedimiento, obteniéndose los datos de la tabla 11.b). Luego efectivamente el procedimiento programado es independiente del PVI a solucionar.

y'	0.1*sqrt(y)+0.4*x^2	
h	0,2	
xn	3	
x	y	
2,00	4,00000000	
2,20	4,36000000	
2,40	4,78896123	
2,60	5,29352862	
2,80	5,88034396	
3,00	6,55604280	

y'	0.2*x*y	
h	0,2	
xn	2	
x	y	
1,00	1,00000000	
1,20	1,04000000	
1,40	1,08992000	
1,60	1,15095552	
1,80	1,22461667	
2,00	1,31278907	

Tabla 11 a) b) . Resultados obtenidos para distintos PVI con el Método de Euler mediante la macro Euler_Eval

4.3. Método de Taylor de orden 3

Como se ha visto en este método se van obteniendo los puntos mediante la fórmula:

$$y_{k+1} = y(x_k) + y'(x_k)h + \frac{1}{2!}y''(x_k)h^2 + \frac{1}{3!}y'''(x_k)h^3 \quad (3.2)$$

El programa va a leer los valores del PVI directamente de la hoja de cálculo, en particular se ha de conocer el valor del paso, h , que se introducirá en la celda B2, el valor de x_0 que estará en la primera fila de la tabla a generar, celda "A5" y el valor de y_0 , que estará en la celda "B5". Se necesitará también el valor de x_n en el que se quiere conocer la función solución, celda B3.

El código VBA del procedimiento Taylor_3 que soluciona el PVI dado en el

Ejemplo 4, es:

```

Sub Taylor_3()
' Halla la solución de un PVI mediante el Método de Taylor, orden 3
' Lee h, x0, y0 y xn
' Fº Javier Palencia, 20-V-2013
  'Se asignan los valores conocidos a sus respectivas variables
  x0 = Range("A5").Value
  y0 = Range("B5").Value
  h = Range("B2").Value
  xn = Range("B3").Value
  'Se halla el número de subintervalos
  m = (xn - x0) / h
  'Se asignan los valores conocidos a las variables de trabajo
  xk = x0
  yk = y0
  For i = 1 To m
    'Se calcula el valor de y'
    y1 = 0.2 * xk * yk
    'Se calcula el valor de y"
    y2 = 0.2 * yk + 0.004 * xk ^ 2 * yk
    'Se calcula el valor de y'''
    y3 = 0.12 * xk * yk + 0.008 * xk ^ 3 * yk
    'Se halla el valor de la solución en el siguiente punto
    yk = yk + y1 * h + 1 / 2 * y2 * h ^ 2 + 1 / 6 * y3 * h ^ 3
    xk = xk + h
    'Se escriben los valores en la tabla
    Cells(5 + i, 1).Value = xk
    Cells(5 + i, 2).Value = yk
  Next i
End Sub

```

Se asigna el código a un botón y haciendo clic se obtiene la tabla de la figura 5.

	A	B	C
1	y'	0,2*x*y	
2	h	0,1	
3	xn	2	Taylor3
4	x	y	
5	1,00	1,00000000	
6	1,10	1,02122133	
7	1,20	1,04498084	
8	1,30	1,07143380	
9	1,40	1,10075565	
10	1,50	1,13314391	
11	1,60	1,16882038	
12	1,70	1,20803368	
13	1,80	1,25106209	
14	1,90	1,29821684	
15	2,00	1,34984581	
16			
17			

Figura 5. Tabla generada mediante la macro Taylor_3 para el Ejemplo 4

4.4. Método de Runge-Kutta de orden 4

Como se ha visto en este método se van obteniendo los puntos mediante las fórmulas:

$$\begin{aligned} K_1 &= f(x_k, y_k) \\ K_2 &= f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}hK_1) \\ K_3 &= f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}hK_2) \\ K_4 &= f(x_k + h, y_k + hK_3) \end{aligned} \quad (3.3)$$

y, finalmente:

$$y_{k+1} = y_k + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) \quad (3.4)$$

El programa funciona de forma análoga a los anteriores, si bien previamente al cálculo de y_k , va a ir obteniendo los valores de K_1, K_2, K_3 y K_4 .

El código VBA del procedimiento RK_4 que soluciona el PVI dado en el Ejemplo 7, es:

```
Sub RK_4()
'Halla la solución de un PVI mediante el Método de Runge-Kutta 4
'Lee h, x0, y0 y xn
'Fº Javier Palencia, 20-V-2013
'Se asignan los valores conocidos a sus respectivas variables
x0 = Range("A5").Value
y0 = Range("B5").Value
h = Range("B2").Value
xn = Range("B3").Value
'Se halla el número de subintervalos
m = (xn - x0) / h
'Se asignan los valores conocidos a las variables de trabajo
xk = x0
yk = y0
For i = 1 To m
'Se calculan los valores auxiliares para cada etapa K1...K4
K1 = 0.2 * xk * yk
K2 = 0.2 * (xk + h / 2) * (yk + h / 2 * K1)
K3 = 0.2 * (xk + h / 2) * (yk + h / 2 * K2)
K4 = 0.2 * (xk + h) * (yk + h * K3)
'Se halla el valor de la solución en el siguiente punto
yk = yk + h / 6 * (K1 + 2 * K2 + 2 * K3 + K4)
xk = xk + h
'Se escriben los valores en la tabla
```

```

Cells(5 + i, 1).Value = xk
Cells(5 + i, 2).Value = yk
Next i
End Sub

```

	A	B	C
1	y'	0,2*x*y	
2	h	0,1	
3	xn	2	RK-4
4	x	y	
5	1,00	1,00000000	
6	1,10	1,02122205	
7	1,20	1,04498235	
8	1,30	1,07143621	
9	1,40	1,10075906	
10	1,50	1,13314845	
11	1,60	1,16882620	
12	1,70	1,20804095	
13	1,80	1,25107102	
14	1,90	1,29822766	
15	2,00	1,34985880	
16			

Figura 6. Tabla generada mediante la macro RK-4 para el Ejemplo 7

5. Conclusiones

Se ha podido comprobar que mediante la hoja de cálculo Excel es extremadamente sencillo encontrar soluciones aproximadas tanto numéricas como gráficas para problemas de valor inicial de ecuaciones diferenciales ordinarias, a partir de las fórmulas de determinados métodos numéricos: Euler, Taylor, Runge-Kutta. El grado de exactitud de las aproximaciones obtenidas viene dado por los distintos métodos numéricos que se han utilizado.

Asimismo se ha mostrado que con una sencilla implementación en VBA, se pueden hallar también excelentes aproximaciones a las soluciones de los PVI. La programación se ha realizado de la forma más simple posible de forma que se pueda replicar fácilmente el programa para cualquier ecuación dentro de un PVI que sea planteado.

6. Referencias bibliográficas

1. M. Amelot, VBA Excel 2007, *Ediciones ENI*, (Barcelona, 2007)
2. J.J Bernal García, Aportaciones para la mejora de la presentación grafica de datos cuantitativos en Excel, *Revista Rect@*. Vol. 16: 41242-412, XVI Congreso ASEPUMA y IV Encuentro Internacional. (Cartagena, 2008).
3. J.J. Bernal García, Representación automática de funciones en Excel y su aplicación docente, *Revista Rect@*. Vol. 12, pp: 141 a 157. (2011)
4. R.L. Burden, J.D. Faires, *Análisis numerico (7ª ed.)*, Thomson Learning, (México, 2002)
5. C. H. Edwards, D.E. Penney, *Ecuaciones diferenciales (4ª ed)*, Pearson Prentice-Hall, (México, 2001)

6. C.F. Gerald, P.O. Wheatley, Análisis numérico con aplicaciones (6ª ed.), *Pearson Prentice-Hall*, (México, 2000)
7. J. H. Mathews, K.D. Fink, Métodos numéricos con Matlab (3ª ed.), *Pearson Prentice-Hall*, (Madrid, 2000).
8. Shoichiro Nakamura, Análisis numéricos y visualización gráfica con Matlab, *Pearson Prentice-Hall*, (México, 1997)
9. J. Walkenbach, Excel 2010. Los mejores trucos, Anaya multimedia, (Madrid, 2011)
10. J. Walkenbach, Excel 2010. Programación con VBA, Anaya Multimedia-Wiley (Madrid, 2011)
11. D.G, Zill, M.R. Cullen, Ecuaciones diferenciales con problemas de valor en la frontera (6ª ed), *Thomson Learning*, (México, 2006)

RANKING DE PRODUCTIVIDAD EN INVESTIGACIÓN DE LAS UNIVERSIDADES PÚBLICAS ESPAÑOLAS

MARÍA CONCEPCIÓN PÉREZ-CÁRCELES

mcperez@ucam.edu

*Universidad Católica San Antonio / Administración y Dirección de Empresas
Campus de Los Jerónimos. 30107. Guadalupe (Murcia)*

JUAN CÁNDIDO GÓMEZ-GALLEGO

jcandido@ucam.edu

*Universidad Católica San Antonio / Administración y Dirección de Empresas
Campus de Los Jerónimos. 30107. Guadalupe (Murcia)*

MARÍA GÓMEZ-GALLEGO

mggallego@ucam.edu

*Universidad Católica San Antonio / Ciencias de la Salud-Neurociencias
Campus de Los Jerónimos. 30107. Guadalupe (Murcia)*

JUAN GÓMEZ-GARCÍA

jgomezg@um.es

*Universidad de Murcia / Métodos Cuantitativos para la Economía y la Empresa
Campus de Espinardo. 30100. Espinardo (Murcia)*

RESUMEN: La creación del Espacio Europeo de Educación Superior ha incrementado la relevancia de la evaluación de la calidad científica de las universidades. Se hace necesario utilizar métodos que garanticen la fiabilidad, validez y utilidad de las evaluaciones y clasificaciones que se establezcan. El objetivo del trabajo es evaluar la producción investigadora de las universidades públicas españolas aplicando el Análisis Envoltante de Datos. Esta metodología produce rankings que no dependen de la especificación del modelo aplicado, establecen criterios objetivos de ponderación en la definición de los indicadores globales y obtienen ordenaciones según determinadas orientaciones productivas. El método se aplica a una muestra de 48 universidades públicas españolas, con datos referidos al año 2010 mediante el software Frontier Analyst. Se obtienen ordenaciones totales según la eficiencia investigadora global y la eficiencia en innovación.

Palabras claves: Universidad pública, Eficiencia en investigación global, Innovación, Análisis Envoltante de Datos, DEA Frontier

ABSTRACT: The creation of the European Space for Higher Education has increased the importance of the evaluation of the scientific quality of the universities. It is necessary to use methods that ensure reliability, validity and usefulness of evaluations and rankings. The aim of the study was to evaluate researching activity of Spanish public universities using Data Envelopment

Analysis. This methodology produces rankings that do not depend on the specification of the model applied, establish objective criteria weighting in the definition of global indicators and obtain rankings under certain productive orientations. The method is applied to a sample of 48 Spanish public universities in 2010 by DEA Frontier software. Rankings are obtained according to overall research efficiency and innovation.

Keywords: Public universities, Overall research efficiency, Innovation, Data Envelopment Analysis, DEA Frontier

1. Introducción

La Universidad es una institución que ha resultado ser esencial para el progreso y modernización de la sociedad. Tanto por su función docente, que posibilita la difusión del conocimiento más avanzado a través de la formación de los estudiantes, como por su función investigadora, que se centra en la generación del conocimiento que está en la base de la resolución de los problemas específicos de las empresas e instituciones (Buesa et al., 2009). El sector de la educación superior ha experimentado en las últimas décadas una demanda creciente, a la vez que una internacionalización progresiva. Esto significa mayor posibilidad de escoger entre una universidad u otra y mayor competitividad entre instituciones. Los rankings de universidades surgen para valorar la calidad de las mismas a la vez que posibilitar la toma de decisiones de los potenciales estudiantes.

Tres han sido básicamente las metodologías aplicadas para la evaluación de la actuación de las universidades: los sistemas basados en prestigio, el uso de indicadores y los modelos frontera (Giménez y Martínez, 2001). Los primeros consisten en rankings orientados a servir de utilidad en el proceso decisorio de un estudiante potencial acerca de dónde cursar los estudios universitarios. Si bien reflejan con facilidad las distintas necesidades que plantea la sociedad, tienen como inconvenientes la multitud y subjetividad de los rankings existentes. Bowden (2000), afirma que los rankings que se han realizado hasta la fecha no proporcionan una guía demasiado útil a los futuros estudiantes para buscar la mejor universidad, ya que no se ajustan a sus necesidades individuales. Sin embargo, Sarrico et al. (1997) afirma que los rankings realmente han servido como guía a los estudiantes dado que la mayoría presentaban un patrón de comportamiento muy similar. Desde la perspectiva del alumno potencial, lo deseable es que se proceda a realizar los rankings teniendo en cuenta determinados requisitos o necesidades de cada estudiante. En este sentido, concluye que: “la aproximación Análisis Envoltante de Datos (DEA), con su mayor flexibilidad, permite producir rankings más consistentes con los requerimientos de diferentes tipos de estudiantes potenciales”.

Los indicadores son “aquellas medidas objetivas, usualmente cuantitativas, de los logros de una institución o de todo un sistema de educación superior” (Ball y Halwachi, 1987). Así, los indicadores miden ciertos aspectos de los inputs de las universidades, de

su funcionamiento y de sus resultados (Kells, 1991). Sus aplicaciones han sido muy numerosas, por ejemplo, en el Plan Nacional para la Evaluación de la Calidad de las Universidades en España. La ventaja más clara de los indicadores, una vez alcanzado un consenso en su definición, es su objetividad y su fácil interpretación individualizada. Buela-Casal et al. (2010) mantiene que son necesarias clasificaciones unidimensionales que ofrezcan información de cada uno de los criterios. Un elemento de vital importancia a la hora de elaborar rankings es la determinación del conjunto de indicadores a utilizar, puesto que de ello dependerá el resultado final en la ordenación establecida, que condiciona en parte la elección final del estudiante. Los indicadores suelen clasificarse en tres grupos: medidas de recursos, medidas de proceso y medidas de resultados.

En tercer lugar, los modelos frontera se basan en fundamentos de importante rigor económico, estableciendo estrategias en la mejora de la gestión de las unidades ineficientes. Una ventaja especialmente importante en el estudio de instituciones que proveen servicios de educación universitaria es su capacidad de calcular una única medida de eficiencia en procesos donde intervienen una multiplicidad de inputs y de productos. Se trata de una metodología apropiada para calcular la eficiencia en sectores caracterizados por la ausencia de precios, como el sector público, ya que fija de forma objetiva los valores óptimos para cada unidad evaluada al maximizar su eficiencia productiva relativa. Entre las debilidades del DEA para la realización de rankings, se cita el alto nivel promedio de eficiencia relativa de las unidades consideradas. Esto puede deberse a que calcula medidas de eficiencia relativa y, por lo tanto, la frontera de producción puede no ser la verdadera frontera a alcanzar si las unidades fueran realmente eficientes en el uso de los recursos (Johnes, 2005). De este modo, las puntuaciones de eficiencia podrían estar sobreestimadas, aunque las comparaciones entre unidades de producción seguirían siendo válidas. Por último, la aparición de numerosas unidades eficientes puede deberse a la cantidad de inputs y outputs considerados, dado que, por las características propias del modelo, cuanto mayor sea el número de variables, mayor será la cantidad de unidades eficientes (Martín, 2006).

En el contexto internacional existen rankings, generales y específicos sobre la calidad de las universidades. Pagani et al. (2006) y Buela-Casal et al. (2007) han descrito algunos de los más importantes, que se centran de manera prioritaria en los resultados de la investigación. En el caso de las universidades españolas, existen algunos rankings que están orientados a la calidad docente y/o a la productividad investigadora (Buesa et al., 2009 y Buela-Casal et al., 2009, 2010).

El objetivo del trabajo es establecer rankings de las universidades públicas españolas sobre productividad investigadora. Para ello se propone el método basado en los modelos DEA, que evita los inconvenientes señalados para esta metodología. Los resultados no dependen de la especificación del modelo aplicado, la eficiencia se obtiene para determinadas orientaciones productivas concretas y las puntuaciones obtenidas permiten establecer un ranking para el conjunto total de unidades. El trabajo está organizado de la forma siguiente: en la sección 2 se presenta el desarrollo

metodológico del procedimiento propuesto. En la sección 3 se muestran los resultados y en la sección 4 se exponen las conclusiones.

2. Metodología

La metodología DEA surge a partir de una investigación referente a la educación en varios centros públicos estadounidenses desarrollada en 1978 por Charnes, Cooper y Rhodes. Es una técnica no paramétrica que permite la construcción de una superficie envolvente, frontera eficiente o función de producción empírica, a partir de los datos disponibles del conjunto de unidades de las mejores prácticas. El método que proponemos parte de los resultados generados al aplicar modelos DEA sobre una muestra de unidades evaluables. Por ello, se plantea el modelo básico DEA sobre el que se construye el procedimiento. Supongamos una muestra de n DMUs¹ tales que cada DMU_j ($j = 1, 2, \dots, n$) produce S *outputs* y_{rj} ($r = 1, 2, \dots, S$), utilizando m *inputs* x_{ij} ($i = 1, 2, \dots, m$). La eficiencia técnica de cada una de las unidades se define como el cociente entre la suma ponderada de los *outputs* y la suma ponderada de los *inputs*.

El modelo CCR *output* orientado en su formulación multiplicativa viene dado por las siguientes ecuaciones:

$$\text{Min } h_0(u, v) = \sum_{i=1}^m v_i x_{i0} + v$$

$$\text{s. a. } \sum_{r=1}^S \mu_r y_{r0} = 1$$

$$\sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^S \mu_r y_{rj} \geq 0$$

$$v_i, \mu_r \geq 0; \quad j = 1, \dots, n; \quad r = 1, \dots, S; \quad i = 1, \dots, m; \quad v \text{ libre}$$

El programa lineal selecciona las ponderaciones que minimizan el *input* virtual de la unidad evaluada ($v_i x_{i0}$), condicionadas a que su *output* virtual ($\mu_r y_{r0}$) sea igual a la unidad, así como a que la aplicación de dichas ponderaciones al resto de unidades evaluadas no permita que su *input* virtual exceda del *output* virtual. La unidad será eficiente si su *input* virtual es la unidad. El cálculo de los índices de eficiencia resulta más sencillo si se utiliza la forma dual del modelo anterior, presentado a continuación, a través de la cual se construye una aproximación lineal por tramos a la verdadera frontera de producción.

Max φ

$$\text{s.a. } \sum_{j=1}^n \lambda_j x_{ij} \leq x_{i0} \quad i = 1, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq \varphi \cdot y_{r0} \quad r = 1, \dots, S$$

¹Decision Making Unit

$$\lambda_j \geq 0 \quad j = 1, \dots, n$$

En este caso, si $\phi = 1$, la unidad evaluada se considera eficiente, pues no existe otra que produzca más o que consiga el mismo nivel de producción con menores recursos que ella.

El modelo anterior asigna a todas las unidades eficientes el mismo valor, por tanto no es útil a la hora de establecer un *ranking*. Además, los resultados del modelo dependen de su especificación. El procedimiento que se propone evita ambas debilidades de la técnica. El método consiste en resolver tantos modelos DEA como posibles especificaciones haya por combinación de los *inputs* y *outputs* seleccionados. La aplicación del Análisis de Componentes Principales (ACP) a la matriz de eficiencias de todos los modelos resueltos permite reducir la dimensión en un número menor de factores con suficiente capacidad explicativa de la varianza total. Tales factores tienen una interpretación que se realiza en base al tamaño y signo de las cargas factoriales. Las correlaciones entre las componentes principales y las variables originales (modelos con diferentes especificaciones) permiten definir, formular e interpretar las componentes principales extraídas. Cada unidad evaluada tiene una puntuación en cada uno de los factores. En consecuencia, es posible establecer una ordenación total en el conjunto de DMUs, tomando como criterio las valoraciones en cada una de las componentes. Lo anterior permite construir *rankings* globales unidimensionales bajo diferentes orientaciones productivas.

3. Datos y variables

La muestra de unidades evaluadas se compone de 48 universidades públicas españolas. Los datos proceden del observatorio IUNE, que utiliza como fuentes de información la plataforma Web of Science, la Red OTRI, INVENES y el Ministerio de educación/INE, entre otras. Las universidades se pueden agrupar en cuatro clusters, según su orientación productiva (Gómez y Pastor, 2011): ciencias de la salud, universidades grandes y consolidadas, universidades de reciente creación y universidades politécnicas. Según Martín Rivero (2007), la principal dificultad para la modelización de la actividad de las instituciones universitarias es la multidimensionalidad del producto educativo, así como su contextualización y posterior medición, lo que hace que sea de crucial importancia el proceso de selección de los *inputs* y *outputs* que caracterizan las actividades universitarias. La revisión de los trabajos empíricos previos sobre la evaluación de la eficiencia de las universidades puede, sin duda, ayudar a avanzar en este conocimiento.

Gómez Sancho y Mancebón (2005) realizan un exhaustivo estudio de la idoneidad de los distintos indicadores para la investigación resaltando como resultados de la

investigación universitaria: artículos, libros, tesis, ponencias y patentes. Según Buesa et al. (2009), se observa que la categoría que es tomada en consideración con mayor frecuencia en la construcción de rankings internacionales es la investigación y la producción científica. Finalmente, para la realización de este estudio se ha seleccionado como input *profesores* (número de profesores en 2010) y, como outputs, cuatro indicadores de entre los criterios establecidos en Buéla-Casal et al. (2011): *publicaciones* (número de artículos publicados en 2010), *tesis doctorales* (número de tesis doctorales en el período 2006-2010), *proyectos* (número de proyectos de investigación europeos y del Plan Nacional obtenidos en convocatorias públicas competitivas en 2010) y *patentes* (número de patentes concedidas en 2010).

4. Resultados

De acuerdo con lo expresado en la sección de metodología, se han resuelto 15 modelos DEA-CCR orientación output que corresponden a las 15 especificaciones posibles a partir de un input (*profesores*, A) y de cuatro outputs (*publicaciones*, 1; *tesis*, 2; *proyectos*, 3 y *patentes*, 4). El modelo se representa por los símbolos que identifican los inputs y outputs utilizados. Esta múltiple estimación de eficiencias bajo las distintas variantes del vector de input-output tiene dos ventajas fundamentales. En primer lugar, el hecho de calcular distintas puntuaciones de eficiencia puede resaltar cómo éstas dependen de la especificación utilizada. En segundo lugar, una ordenación global de las unidades puede derivarse de la aplicación de un análisis multivariante de las eficiencias obtenidas bajo las distintas combinaciones, ver tabla 1, tal y como se muestra en el trabajo de Gómez-Gallego et al. (2012). En un primer análisis de la tabla 1 se aprecia que para todas las combinaciones de input y output, en al menos dos ocasiones se obtiene una puntuación eficiente por parte de las universidades, destacando los modelos A3, A12, A13, A14, A23 y A123 con 5 unidades eficientes, los modelos A134, A23 y A24 con 6 unidades eficientes y los modelos A1234, A124 y A234 con 7 unidades eficientes. Si centramos la atención en las 48 observaciones, es de destacar que tan solo el 15% obtiene una puntuación eficiente en alguno de los modelos. Así, la Universidad de La Rioja aparece como 100% eficiente en todos los modelos y la Autónoma de Barcelona y la Pompeu Fabra únicamente presentan una eficiencia por debajo del 100% con la inclusión exclusiva del output *patentes*, es decir, en el modelo A4. A continuación, la Universidad de Barcelona y la Complutense de Madrid no alcanzan la eficiencia máxima en los modelos que incluyen *publicaciones*, *patentes* y ambas en el caso de la primera, y *tesis*, *patentes* y ambas, en la segunda. Las universidades Politécnica de Catalunya y Autónoma de Madrid también obtienen puntuaciones eficientes bajo 8 y 4 modelos, respectivamente. En el primer caso el factor común que presentan es la aparición del output *patentes*. En el caso de la universidad Autónoma de Madrid, sin embargo, las puntuaciones eficientes al 100% se obtienen en los modelos A1234, A124, A234 y A24. Teniendo en cuenta la interpretación en cuanto a la

orientación productiva de los outputs, los resultados son consistentes con Buela-Casal et al. (2011), ya que las universidades Complutense de Madrid y Barcelona figuran como las de mayor producción y la Pompeu Fabra y Autónoma de Barcelona destacan entre las de mayor productividad.

La matriz de resultados (ver tabla 1) puede ser tratada como un conjunto de datos multivariantes: las especificaciones de los modelos son las variables y las unidades. Se tienen así 15 variables y 48 observaciones. Existen correlaciones claras entre las distintas variables, y por tanto la reducción es posible. Escogemos el análisis de componentes principales entre varias técnicas multivariantes ($KMO=0,84$). En el análisis de componentes principales, Jolliffe (1972) indica que los autovalores son relevantes si son mayores que 0,8. La primera componente está asociada a un autovalor de 12,86 y explica el 85,73% de la variabilidad total, mientras que la segunda componente está asociada a un autovalor de 1,29 y tiene un poder explicativo del 8,60%. En suma, ambas componentes principales explican el 94,33% de la varianza total. Las componentes rotadas tienen un poder explicativo del 62,41% y 31,92% respectivamente. La matriz de componentes, tabla 2, presenta las correlaciones entre componentes principales y los modelos.

Las correlaciones entre los modelos y la primera componente son muy elevadas y con signo positivo en todos los modelos, salvo en el modelo A4. La mayor carga factorial se presenta para el modelo A1234, con 0,978 y la menor correlación (0,533) se obtiene con el modelo A4. En consecuencia, la primera componente es interpretada como una medida de eficiencia en investigación global sobre todas las posibles especificaciones, como sugiere Serrano-Cinca y Mar-Molinero (2004). La segunda componente se interpreta como una medida de eficiencia en innovación. Cuando se extraen las componentes rotadas, se mantiene la interpretación de ambos factores.

A efecto de reflejar cómo la eficiencia de las unidades, depende de las distintas combinaciones de inputs y outputs que se han obtenido en la especificación del DEA, utilizamos la técnica Property-Fitting (ver gráfico 1 en el Anexo).

En el eje horizontal se representan las cargas que ofrece la primera componente principal con los distintos modelos, y cuya interpretación se expresa como eficiencia investigación global. Las unidades que se sitúan en el primer y cuarto cuadrante ofrecen una mayor eficiencia global, destacando las universidades de Barcelona, Autónoma de Barcelona, Pompeu Fabra y Complutense de Madrid. En el extremo inferior del segundo y tercer cuadrante se posicionan universidades de menor tamaño y creación más reciente: Burgos, Almería, Huelva, Rey Juan Carlos, entre otras.

El eje de ordenadas representa la medida de eficiencia en innovación. Destacan en la parte superior del gráfico las universidades politécnicas de Cataluña, Madrid y Cartagena, bien posicionadas en esta orientación productiva pero con diferentes resultados de eficiencia en investigación global. Las últimas posiciones del ranking en innovación las ocupan la universidad de León, Extremadura y la UNED, entre otras.

Tabla 1. Eficiencias bajo las 15 especificaciones DEA

Universidad	A1	A2	A3	A4	A12	A13	A14	A23	A24	A34
EHU	42,71	49,00	64,95	44,19	49,20	64,95	65,52	64,95	59,95	74,00
UA	33,74	41,99	27,99	24,72	41,99	33,74	42,40	41,99	47,79	32,63
UAB	100	100	100	23,26	100	100	100	100	100	100
UAH	35,93	58,02	26,56	43,64	58,02	35,93	52,42	58,02	68,85	43,64
UAL	26,18	32,33	34,46	46,08	32,33	34,46	48,78	34,46	49,18	49,67
UAM	75,33	91,03	64,04	53,49	91,03	75,33	95,01	91,03	100,00	76,95
UB	100	98,53	100	20,93	100	100	100	100	98,53	100
UBU	15,29	24,66	12,49	35,70	24,66	15,29	37,69	24,66	40,38	36,39
UC3M	49,16	54,65	57,57	18,88	54,65	57,57	51,67	57,57	56,23	58,30
UCA	20,25	30,11	27,74	42,87	30,11	27,74	42,87	30,11	44,08	42,91
UCLM	39,46	29,79	35,73	16,40	39,46	39,46	43,53	35,73	33,59	38,01
UCM	78,46	100	100	37,21	100	100	92,34	100	100	100
UCO	43,90	63,54	36,70	29,31	63,54	43,90	52,67	63,54	68,69	41,22
UDC	25,55	31,61	31,67	22,04	31,61	31,67	33,31	31,67	37,01	35,11
UDG	46,31	34,39	52,29	9,60	46,31	52,29	46,31	52,29	34,39	52,29
UDL	35,74	42,62	42,74	10,32	42,62	42,74	35,74	42,74	42,62	42,74
UGR	54,08	66,43	77,79	32,56	66,43	77,79	66,73	77,79	69,56	82,08
UHU	24,24	26,07	29,35	26,63	26,07	29,35	32,38	29,35	32,54	32,94
UIB	48,82	25,99	39,38	14,57	48,82	48,82	49,41	39,38	28,34	40,03
UJAEN	31,90	26,04	13,95	27,81	31,90	31,90	40,17	26,04	33,10	27,81
UJI	36,22	37,44	31,89	15,75	37,44	36,22	38,33	37,44	39,10	32,96
ULL	34,55	28,08	32,59	13,95	34,55	34,55	37,81	32,59	30,98	34,41
ULPGC	18,03	25,16	17,80	12,62	25,16	18,03	22,33	25,16	27,80	19,95
UM	36,78	46,47	32,12	16,28	46,47	36,78	41,26	46,47	48,95	34,73
UMA	24,98	31,80	36,64	37,21	31,80	36,64	42,35	36,64	43,58	46,01
UMH	50,07	59,74	28,14	41,77	59,74	50,07	58,37	59,74	66,14	46,96
UNAVARRA	31,23	31,30	30,68	41,14	31,30	31,23	48,52	31,30	47,04	47,21
UNED	17,10	54,48	21,62	3,56	54,48	21,62	17,10	54,48	54,48	21,62
UNEX	25,46	31,42	32,00	5,93	31,42	32,00	25,66	32,00	31,42	32,00
UNICAN	48,43	38,33	52,70	55,97	48,43	52,70	66,64	52,70	58,64	61,91
UNILEON	20,51	43,12	26,75	5,53	43,12	26,75	20,71	43,12	43,12	26,75
UNIOVI	38,71	36,87	44,20	37,21	38,71	44,20	54,19	44,20	47,66	53,23
UNIRIOJA	100	100	100	100	100	100	100	100	100	100
UNIZAR	47,72	48,97	61,11	32,56	49,11	61,11	59,19	61,11	55,24	67,11
UPC	52,63	66,57	76,74	100	66,57	76,74	100	76,74	100	100
UPCT	30,83	30,14	34,16	63,75	30,83	34,16	71,27	34,16	67,36	73,29
UPF	100	100	100	41,92	100	100	100	100	100	100
UPM	31,66	31,00	62,47	90,70	31,66	62,47	90,70	62,47	90,70	91,00
UPO	31,80	50,81	28,29	48,38	50,81	31,80	58,71	50,81	67,48	56,95
UPV	39,99	54,53	61,61	44,19	54,53	61,61	59,37	61,61	64,86	71,41
URJC	26,07	26,06	20,81	19,39	26,07	26,07	32,37	26,06	30,88	24,12
URV	64,91	61,27	57,05	16,27	64,91	64,91	64,91	61,27	61,27	57,13
US	41,98	58,21	61,26	44,19	58,21	61,26	65,12	61,26	66,69	70,59
USAL	30,80	57,22	41,06	11,63	57,22	41,06	33,16	57,22	57,22	41,70
USC	49,78	48,36	66,95	37,21	49,78	66,95	63,66	66,95	56,98	74,12
UV	70,60	60,73	65,03	16,28	70,60	70,60	71,05	65,03	60,73	65,41
UVA	23,31	35,86	30,60	11,63	35,86	30,60	26,76	35,86	36,92	32,12
UVIGO	44,82	36,49	34,85	15,81	44,82	44,82	47,78	36,49	39,38	36,86

Universidad	A123	A124	A134	A234	A1234
EHU	64,95	66,10	74,00	74,00	74,00
UA	41,99	47,79	42,40	47,79	47,79
UAB	100	100	100	100	100
UAH	58,02	68,85	52,42	68,85	68,85
UAL	34,46	49,18	49,67	49,67	49,67
UAM	91,03	100	95,01	100	100
UB	100	100	100	100	100
UBU	24,66	40,38	37,69	40,38	40,38
UC3M	57,57	56,23	58,30	58,30	58,30
UCA	30,11	44,08	42,91	44,08	44,08
UCLM	39,46	43,53	43,53	38,01	43,53
UCM	100	100	100	100	100
UCO	63,54	68,69	52,67	68,69	68,69
UDC	31,67	37,01	35,11	37,01	37,01
UDG	52,29	46,31	52,29	52,29	52,29
UDL	42,74	42,62	42,74	42,74	42,74
UGR	77,79	70,88	82,08	82,08	82,08
UHU	29,35	32,54	32,94	32,94	32,94
UIB	48,82	49,41	49,41	40,03	49,41
UJAEN	31,90	40,17	40,17	33,10	40,17
UJI	37,44	39,10	38,33	39,10	39,10
ULL	34,55	37,81	37,81	34,41	37,81
ULPGC	25,16	27,80	22,33	27,80	27,80
UM	46,47	48,95	41,26	48,95	48,95
UMA	36,64	43,58	46,01	46,01	46,01
UMH	59,74	66,14	58,37	66,14	66,14
UNAVARRA	31,30	48,52	48,52	47,21	48,52
UNED	54,48	54,48	21,62	54,48	54,48
UNEX	32,00	31,42	32,00	32,00	32,00
UNICAN	52,70	66,64	66,64	61,91	66,64
UNILEON	43,12	43,12	26,75	43,12	43,12
UNIOVI	44,20	54,19	54,19	53,23	54,19
UNIRIOJA	100	100	100	100	100
UNIZAR	61,11	59,19	67,11	67,11	67,11
UPC	76,74	100	100	100	100
UPCT	34,16	71,27	73,29	73,29	73,29
UPF	100	100	100	100	100
UPM	62,47	90,70	91,00	91,00	91,00
UPO	50,81	67,48	58,71	67,48	67,48
UPV	61,61	64,86	71,41	71,41	71,41
URJC	26,07	32,37	32,37	30,88	32,37
URV	64,91	64,91	64,91	61,27	64,91
US	61,26	67,27	70,59	70,59	70,59
USAL	57,22	57,22	41,70	57,22	57,22
USC	66,95	63,66	74,12	74,12	74,12
UV	70,60	71,05	71,05	65,41	71,05
UVA	35,86	36,92	32,12	36,92	36,92
UVIGO	44,82	47,78	47,78	39,38	47,78

Tabla 2. Matriz de componentes

DEA-Profit

Modelo	Componente	
	1	2
A123	0,967	- 0,212
A12	0,910	- 0,360
A1	0,894	- 0,306
A2	0,907	- 0,315
A23	0,968	- 0,189
A3	0,931	- 0,125
A13	0,954	- 0,151
A134	0,964	0,186
A1234	0,978	0,132
A124	0,974	0,127
A14	0,950	0,215
A24	0,951	0,144
A34	0,945	0,204
A4	0,533	0,807
A234	0,971	0,144

5. Conclusiones

El procedimiento DEA-Profit posibilita obtener un ranking total para el conjunto de universidades de la muestra, superando así la ordenación incompleta generada a partir del modelo DEA convencional. Este procedimiento permite establecer ordenaciones

libres de la dependencia de la especificación particular del modelo DEA aplicado. Además obtiene rankings de universidades cuando se tiene en cuenta exclusivamente la productividad en investigación ó la productividad en innovación. El procedimiento DEA-Profit posibilita obtener ordenaciones en cualquier dirección resultado de una combinación particular de inputs y outputs.

El método DEA-Profit, además de permitir la ordenación total de la muestra de universidades, aporta información sobre las estrategias adecuadas para mejorar los resultados y sobre las universidades que pueden servir de referencia para obtener mejores posiciones en el ranking.

Referencias Bibliográficas

1. R. Ball and J. Halwachi, Performance indicators in higher education, *Higher Education*, 16 (1987) 393-405,
2. R. Bowden, Fantasy Higher Education: University and college league tables, *Quality in Higher Education*, 6-1 (2000) 41-60,
3. G. Buéla-Casal, O. Gutiérrez, M. P. Bermúdez and O. Vadillo, Comparative study of international academic ranking s of universities, *Scientometrics*, 71 (2007) 349-365,
4. G. Buéla-Casal, M. P. Bermúdez, J. C. Sierra, R. Quevedo-Blasco and A. Castro, Ranking de 2008 en productividad en investigación de las universidades públicas españolas, *Psicothema*, 21-2 (2009) 304-312,
5. G. Buéla-Casal, M. P. Bermúdez, J. C. Sierra, R. Quevedo-Blasco and A. Castro, Ranking de 2009 en investigación de las universidades públicas españolas, *Psicothema*, 22-2 (2010) 171-179,
6. G. Buéla-Casal, M. P. Bermúdez, J. C. Sierra, R. Quevedo-Blasco, A. Castro and A. Guillén-Riquelme, Ranking de 2010 en investigación de las universidades públicas españolas, *Psicothema*, 23-4 (2011) 527-536,
7. M. Buesa, J. Heijs and O. Kahwash, La calidad de las universidades en España. Elaboración de un índice multidimensional, in *Minerva*, eds, (Consejo Económico y Social, Comunidad de Madrid, 2009),
8. A. Charnes, W. W. Cooper and E. Rhodes, Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2-4 (1978) 429-444,
9. V. Giménez and J. L. Martínez, Eficiencia en costes en la universidad. Una aplicación a los departamentos de la UAB, in *X Jornadas de la Asociación de la Economía de la Educación*, (2001) 461-472,
10. J. Gómez and J. M. Pastor, La actividad productiva de las universidades españolas: influencia de la especialización y de la tipología de universidades, Lecturas sobre economías de la educación, in *Ministerio de Educación, Subdirección General de Información y Publicaciones*, eds, (2011) 127-150,
11. J. C. Gómez-Gallego, M. C. Pérez-Cárceles and J. Gómez-García, Métodos multivariantes de obtención de rankings. Aplicación al caso de las universidades públicas españolas, *Regional and Sectoral Economic Studies*, 12-3 (2012) 161-178,
12. J. M. Gómez Sancho and M. J. Mancebón, Algunas reflexiones metodológicas sobre la evaluación de la eficiencia productiva de las instituciones de educación superior, *Ekonomiaz*, 60-1 (2005) 141-166,

13. J. Johnes, Data envelopment analysis and its application to the measurement of efficiency in higher education, *Economics of Education Review*, 25 (2005),
14. I. T. Jolliffe, Discarding variables in a principal component analysis I: Artificial data, *Applied Statistics*, 21 (1972) 160–173,
15. H. R. Kells, The inadequacy of performance indicators for higher education: The need for a more comprehensive and development construct, *Higher Education Management*, 2-3 (1991) 258-270,
16. R. Martín Rivero, La eficiencia productiva en el ámbito universitario: aspectos claves para su evaluación, *Estudios de Economía Aplicada*, 25-3 (2007) 793-812.
17. E. Martín Vallespín, An application of the Data Envelopment Analysis methodology in the performance assessment of the Zaragoza University Departments, WP 2003-06 (2006),
18. R. Pagani, O. Vadillo, G. Buena-Casal, J. C. Sierra, M. P. Bermúdez, O. Gutiérrez-Martínez, D. Agudelo, J. Bretón and I. Teva, Estudio inter.acional sobre criterios e indicadores de calidad de las universidades, ACAP (Madrid, 2006),
19. C. S. Sarrico, S. M. Hogan, R. G. Dyson and A. D. Athanassopoulos, Data Envelopment Analysis and university selection, *The Journal of the Operational Research Society*, 48-12 (1997) 1163-1177,
20. C. Serrano Cinca and C. Mar Molinero, Selecting DEA specifications and ranking units via PCA, *Journal of the Operational Research Society*, in *Palgrave MacMillan*, eds, 55 (2004) 521-528,

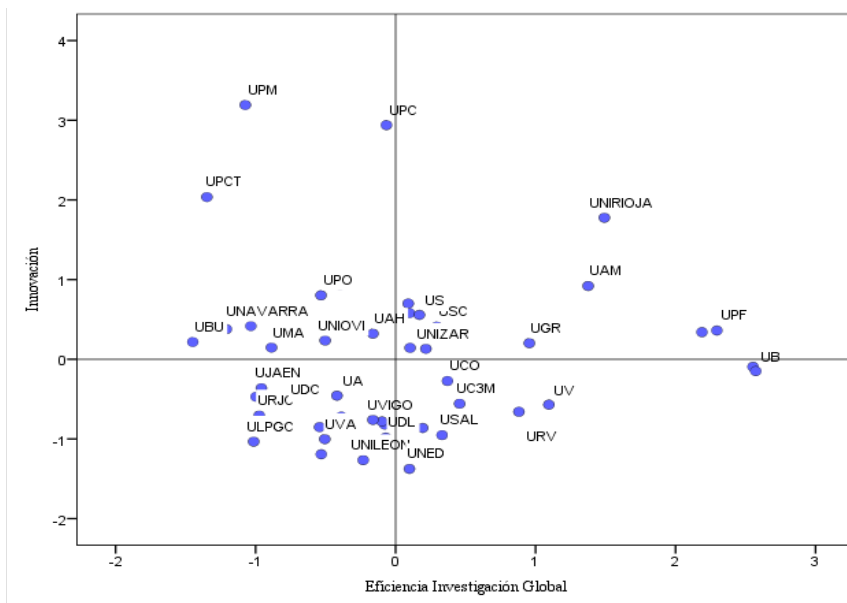
Anexo A

Tabla 3. Listado de universidades

Universidades	
EHU	Universidad del País Vasco
UA	Universidad de Alicante
UAB	Universidad Autónoma de Barcelona
UAH	Univ Alcalá de Henares
UAL	Universidad de Almería
UAM	Universidad Autónoma de Madrid
UB	Universidad de Barcelona
UBU	Universidad de Burgos
UC3M	Universidad Carlos III de Madrid
UCA	Universidad de Cádiz
UCLM	Universidad de Castilla-La Mancha
UCM	Universidad Complutense de Madrid
UCO	Universidad de Córdoba
UDC	Universidad de A Coruña
UDG	Universidad de Girona
UDL	Universidad de Lleida
UGR	Universidad de Granada
UHU	Universidad de Huelva
UIB	Universidad Internacional de Andalucía
UJAEN	Universidad de Jaén
UJI	Universidad Jaume I de Castellón
ULL	Universidad de La Laguna
ULPGC	Universidad de Las Palmas de Gran Canaria

UM	Universidad de Murcia
UMA	Universidad de Málaga
UMH	Universidad Miguel Hernández de Elche
UNAVARRA	Universidad Pública de Navarra
UNED	Universidad Nacional de Educación a Distancia
UNEX	Universidad de Extremadura
UNICAN	Universidad de Cantabria
UNILEON	Universidad de León
UNIOVI	Universidad de Oviedo
UNIRIOJA	Universidad de la Rioja
UNIZAR	Universidad de Zaragoza
UPC	Universidad Politécnica de Catalunya
UPCT	Universidad Politécnica de Cartagena
UPF	Universidad Pompeu Fabra
UPM	Universidad Politécnica de Madrid
UPO	Universidad Pablo de Olavide
UPV	Universidad Politécnica de Valencia
URJC	Universidad Rey Juan Carlos
URV	Universidad Rovira i Virgili
US	Universidad de Sevilla
USAL	Universidad de Salamanca
USC	Universidad de Santiago de Compostela
UV	Universidad de Valencia
UVA	Universidad de Valladolid
UVIGO	Universidad de Vigo

Gráfico 1. Property Fitting



DISTRIBUCIÓN ESPACIAL DE LA ACTIVIDAD ECONÓMICA EN LA COMARCA DE CARTAGENA. UN ANÁLISIS BASADO EN LA DEMOGRAFÍA EMPRESARIAL USANDO ARCVIEW Y GEODA.

FRANCISCO TOMAS CERDÁN

ftomas.cerdan@upct.es

Departamento de Métodos Cuantitativos e Informáticos
Universidad Politécnica de Cartagena
C/Real 3 30201-Cartagena

FERNANDO A LÓPEZ HERNÁNDEZ

fernando.lopez@upct.es

Departamento de Métodos Cuantitativos e Informáticos
Universidad Politécnica de Cartagena
C/Real 3 30201-Cartagena

RESUMEN: El objetivo del presente estudio es realizar un análisis exploratorio de datos espaciales para describir la estructura económica del Campo de Cartagena desde la perspectiva espacial, utilizando la información disponible en la base de datos SABI (Sistemas de Análisis de Balances Ibéricos) que permite conocer la localización exacta de las empresas ubicadas en esta región. Se han utilizado los programas ArcGIS y GeoDa para realizar este análisis. En la primera parte del trabajo se hace una breve descripción del más popular de los estadísticos de autocorrelación espacial, el índice de Moran, junto con su versión gráfica, el diagrama de dispersión de Moran. En la segunda parte, se aplican estos indicadores a la información sobre demografía empresarial disponible en la base de datos SABI. Se concluye en este análisis que tanto para los sectores agrícola, industrial y construcción hay una fuerte estructura de dependencia espacial, mientras que para el sector servicios tanto comerciales como no comerciales la localización de empresas no presenta estructura espacial.

Palabras claves: Autocorrelación espacial, índice de Moran, demografía empresarial, aglomeración empresarial

ABSTRACT: The aim of this paper is present an exploratory analysis of spatial data to describe the economic structure of the small area in Murcia Region named Campo de Cartagena. We used the ArcGIS and GeoDa software to perform this analysis using information available in the databank SABI (Analysis Systems Iberian Balance) which give the exact location of companies located in this area. In the first part of the paper a brief description of the most popular statistical spatial autocorrelation, Moran index, is showed. Moreover, we present the graphical version: the Moran scatterplot. In the second part, we apply both indicators to Business Demography to the set of data available in SABI. We conclude from this analysis that both the agricultural, industrial and construction sectors there is a strong spatial

dependence structure, while the service sector, commercial or noncommercial, no spatial structure is founded.

Keywords: spatial autocorrelation, Moran index, firm demography, firm agglomeration.

1. Introducción

Existe un amplio consenso en el hecho de que las empresas resultan más competitivas cuando se encuentran agrupadas espacialmente, de esa forma explotan las ventajas productivas asociadas a la proximidad espacial con otras y con aquellos agentes económicos con los que interactúan en su actividad productiva (ver por ejemplo Puga 2010). Por esta razón, el análisis de la distribución espacial de la actividad económica presenta un interés creciente para los investigadores. Tradicionalmente este tópico ha sido explorado utilizando grandes unidades administrativas (p.e. países, regiones europeas, provincias) pero poco se ha explorado cuando el nivel de desagregación espacial es pequeño (p.e. distritos, secciones censales). Este enfoque a nivel micro es una nueva perspectiva del análisis econométrico de datos espaciales que Arbia (2011) ha tildado como '*spatial microeconomic*' y que está tomando fuerza debido a la facilidad para disponer de grandes bases de datos con referencia espacial.

Por otra parte, el concepto de independencia es una de las ideas centrales en el desarrollo teórico de la Estadística, pues en la mayoría de los fenómenos físicos, económicos o sociales es raro encontrar dos sucesos que no estén de alguna manera relacionados. Es por ello necesario dar mayor importancia al fenómeno de la dependencia estadística, así como al desarrollo de nuevos instrumentos teóricos que se compadezcan más con la realidad. Los trabajos en Econometría de Series Temporales (por ejemplo, Box y Jenkins 1970) han demostrado que la mayoría de fenómenos económicos no se modelizan por un esquema de ruido blanco, y que, por contra, la inmensa mayoría de estos fenómenos siguen esquemas autorregresivos y/o de dependencia temporal. Cuando se estudian datos de corte transversal, es decir, ubicados en un mismo momento, pero repartidos por el espacio, se pierde la secuencialidad propia de las series temporales (cada dato es continuador de otro y le sigue su siguiente, lo que genera una estructura ordinal clara). Esta diferente naturaleza de los datos de naturaleza espacial, obliga al desarrollo de nuevas herramientas de análisis de *autocorrelación espacial*, partiendo de un principio básico, enunciado por Tobler, W (1970), que se ha convertido en la Primera Ley de la Geografía, "*cualquier cosa está relacionada con cualquier cosa, pero las cosas cercanas están más relacionadas que las distantes*"

En concreto son fundamentalmente tres las diferencias entre los datos de naturaleza temporal y espacial:

Cuando se analiza información con referencia temporal, se plantea unidireccionalmente, en una dimensión, mientras que en el espacio la dependencia es multidireccional descompuestas en tantas dimensiones como tenga el espacio que estemos estudiando. Esta diferencia es sustancial y separa de forma importante el estudio teórico de ambos tipos de procesos.

En las series temporales existe un sólo punto frontera, mientras que en caso de los procesos espaciales el número de puntos frontera es mucho mayor. (Por ejemplo en una retícula regular $n \times m$ hay $2n+2m-4$ fronteras).

Finalmente, otra de las diferencias que debe destacarse, es que los modelos autorregresivos en series temporales pueden especificarse mediante las covarianzas, la función de distribución conjunta o las condicionales siendo estos tres tipos de especificación equivalentes. En el caso de los modelos espaciales esto no es cierto.

La autocorrelación espacial en los modelos donde subyace una estructura de dependencia espacial, implica que si se encuentra una determinada observación en un punto localizado geográficamente es más fácil (o más difícil) encontrar observaciones semejantes en puntos próximos a éste que en puntos alejados. Se entiende por tanto que hay latente en la localización de las observaciones una compleja información sobre las observaciones del entorno.

Los trabajos de Moran (1948) y Geary (1954) fueron pioneros en la detección de dependencia entre observaciones referenciadas espacialmente. En ambos trabajos se desarrollan índices que detectan la presencia de autocorrelación espacial que permiten identificar, tanto el nivel de intensidad de la relación como el sentido de la autocorrelación espacial entre las observaciones.

Pero nuestro trabajo tiene una vocación fundamentalmente económica y más en concreto empresarial. La herramienta de la que nos serviremos es el análisis exploratorio de datos como una fase previa a la econometría espacial. Este término, econometría espacial, fue acuñado por Jean Paelinck (1985) al comienzo de la década de los setenta, aunque la definición elegida en este trabajo es debida a Anselin (1992):” *es el conjunto de métodos para tratar adecuadamente las características especiales de los datos georeferenciados y de los modelos de economía espacial*”. El manual de Anselin (1988) es con diferencia la referencia en este tópico.

En los últimos veinte años han sido muchos los autores que han abordado la importancia de la localización en variables económicas clave como la productividad, la eficiencia o el crecimiento. Entre otros destacamos los trabajos de Krugman (1993, 1997) donde se resalta el papel de las externalidades espaciales en los modelos de comercio internacional y crecimiento. Se han multiplicado los modelos

que estudian la influencia del espacio sobre la localización de empresas, desarrollo de complejos industriales, difusión del conocimiento y la tecnología.

El objetivo de este trabajo es analizar la presencia o no de dependencia espacial, tanto a nivel global como sectorial en la distribución de empresas dentro del ámbito geográfico de la Comarca de Cartagena (CC) a la vez que mostrar el uso de dos de los programas más populares para el análisis de este tipo de procesos. Para alcanzar estos objetivos nos hemos servido de la Base de Datos SABI que posee información bastante exhaustiva de alrededor de 8300 empresas ubicadas en la zona. Los 22 distritos censales en los que se divide la CC son las unidades geográficas sobre las que se ha apoyado el estudio.

La herramienta estadística que se ha utilizado ha sido el Índice I de Moran, del que se hace una somera descripción en el capítulo siguiente. Para el análisis de este estadístico, así como para la elaboración de los mapas incluidos en este trabajo hemos recurrido a dos software: GeoDa y ArcGIS. Ambos son dos sistemas de información geográfica. El primero de ellos, GeoDa, tiene un enfoque principalmente hacia el Análisis Exploratorio de Datos Espaciales (AEDE) mientras que el segundo tiene una orientación más general aunque las últimas versiones incorporan técnicas básicas de AEDE

2. Metodología

Con el fin de identificar la presencia o ausencia de autocorrelación espacial en un conjunto de datos se han desarrollado un buen número de estadísticos. Estos contrastes plantean como hipótesis nula la independencia de las observaciones frente a una hipótesis alternativa de correlación espacial según una determinada estructura de vecindades especificada a priori por el investigador. El contraste puede plantearse como:

H_0 : Independencia espacial.

H_1 : Hay autocorrelación espacial según la estructura de vecindades determinada.

Rechazar la hipótesis nula implica asumir la presencia de un modelo de dependencia espacial definido bajo la estructura de vecindades definida.

En lo que sigue presentaremos sólo el más popular de los estadísticos de autocorrelación espacial dejando al margen otros contrastes alternativos (p.e. Geary 1954, Getis 1992, Pinkse 2002, López et al. 2010).

2.1 El estadístico I de Moran

Como se plantea en Chasco (2003) las interacciones espaciales suelen estar detrás de la existencia de correlación espacial. El valor que una variable toma en un

determinado lugar, x_i es función de los valores que esta toma en el resto de lugares donde se ha hecho el estudio, esto es:

$$x_i = f(x_1, \dots, x_n)$$

Pero esta expresión además de no ser operativa, no contempla necesariamente la estructura interna del espacio estudiado. Dicho espacio está dotado de una estructura de vecindades mucho más compleja que la existente en las series temporales, donde cada dato es limítrofe sólo con otros dos, el anterior y el siguiente.

El concepto de vecindad no es único, pues apelando únicamente a la estructura espacial podríamos hablar de que dos regiones son vecinas si, por ejemplo: comparten frontera común, comparten un punto o se hayan separados por menos de una determinada distancia. Incluso podríamos apelar a vecindad entre dos regiones sin usar conceptos geográficos, por ejemplo, si planteamos que dos regiones son vecinas si tienen una renta similar.

El elemento que nos permite reflejar esta estructura de vecindades es la matriz de interacciones espaciales. El ejemplo más simple de matriz es aquella que coloca el valor 1 en la posición i, j si las regiones i y j son vecinas (habiendo previamente definido este concepto) o 0 en caso contrario o cuando $i=j$, esto es, en la diagonal de la matriz. Este concepto tan simple de contigüidad fue extendido por Cliff y Ord (1973, 1981). Se pretendía incluir una medida general de interacción entre dos unidades espaciales. Esta extensión vendrá expresada mediante la matriz de contactos W ,

$$W = \begin{bmatrix} 0 & w_{12} & w_{13} & \cdot & w_{1n} \\ w_{21} & 0 & w_{23} & \cdot & w_{2n} \\ w_{31} & w_{32} & 0 & \cdot & w_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ w_{n1} & w_{n2} & w_{n3} & \cdot & 0 \end{bmatrix}$$

$W = (w_{ij})$ es una matriz cuadrada no estocástica que cuantifica el efecto de la zona j sobre la i . Los valores de la matriz difieren en función del criterio que se fije para definir la proximidad entre las áreas. En el caso de zonas irregulares pueden ponderarse las fronteras para diferenciar entre distintos grados de proximidad, distinta longitud de la frontera común, etc. Es también una práctica común en aquellas situaciones para las que sea posible, dividir cada elemento de la matriz W por la suma de la fila correspondiente. De esta forma todas las ponderaciones que

afectan a la zona i suman la unidad. Este procedimiento tiene como ventaja el que se pondera a todas las regiones por igual, aunque la matriz resultante no será simétrica, generando ciertos problemas de cálculo.

La forma en que se definan los w_{ij} es un elemento exógeno al problema, estando en manos del investigador dicha definición. Por ejemplo, Cliff y Ord (1.981) proponen una combinación entre la distancia euclídea que une dos zonas (o los centroides de estas zonas) y la longitud relativa de la frontera común entre las zonas. Esta idea implica generalmente una matriz W no simétrica. Otro ejemplo de gran importancia, y que será utilizado en la siguiente sección del trabajo, es el de la inversa de la distancia entre regiones: $w_{ij}=d_{ij}^{-1}$. Como en el caso de la matriz binaria, la diagonal estaría formada por ceros. En Chasco (2003, página 59) se cita una relación bastante exhaustiva de definiciones de pesos, así como el autor que los desarrolló.

Pero si el primer paso era conocer cual es la estructura preexistente (las vecindades) en el espacio estudiado, el objetivo principal de un estudio de dependencia espacial es ver si la distribución de una variable sobre el territorio presenta algún grado de dependencia con la estructura de vecindades, es decir, si regiones o puntos vecinos presentan valores similares (disimilares).

En los últimos años se han desarrollado diferentes herramientas estadísticas que pretenden analizar el grado de dependencia espacial, tanto de forma global, que se manifiestan en la presencia de tendencias en la distribución de una variable sobre el espacio; como de forma local, entendida como el estudio de la presencia de patrones en determinadas áreas geográficas. En Chasco (2003, página 63) encontramos una enumeración de algunos de estos estadísticos.

El primero de los índices globales que se desarrolló, y quizás el más intuitivo de todos, es el I de Moran (1950). Su fundamento yace en el estudio de la correlación entre el valor de la variable en una región, y_i , y el promedio del valor de ésta en las regiones vecinas, $W y_j$.

$$I = \frac{n}{S_0} \frac{\sum_{r=1}^n \sum_{s=1}^n w_{sr} (y_r - \bar{y})(y_s - \bar{y})}{\sum_{r=1}^n (y_r - \bar{y})^2} \quad (1)$$

$$S_0 = \sum_{r=1}^n \sum_{s=1}^n w_{rs} \quad (2)$$

Donde n es el número de elementos de la muestra.

Observemos que cuando la matriz W está estandarizada, el cociente $\frac{n}{S_0} = 1$

Debemos destacar que este coeficiente tiene la forma clásica de una medida de autocorrelación. El numerador es una medida de la covarianza y el denominador la varianza. Bajo la hipótesis nula, el momento de primer orden del estadístico de Moran, es:

$$E(I) = -\frac{1}{n-1} \quad (3)$$

Su distribución, bajo condiciones relativamente débiles (Sen, 1976), es asintóticamente normal. Esta distribución asintótica del estadístico puede ser utilizada para contrastar la hipótesis nula utilizando el clásico procedimiento inferencial. Por otra parte, son muchas las situaciones en las que la normalidad de la variable es de difícil justificación. En estos casos, ha ganado popularidad debido a su elevada potencia en un amplio rango de situaciones el método de inferencia basado en bootstrapping permutacional. Este método, básicamente consisten en describir la distribución del estadístico I mediante la obtención de diferentes valores del índice permutando las realizaciones de la variable por las diferentes regiones. La implementación de este método es posible gracias a software como GeoDa y ArcGIS, que además permiten elegir el número de permutaciones (boots en terminología inglesa) deseadas. El valor de 999 boots es el más común.

La hipótesis alternativa, a la que antes hacíamos alusión, se refiere a la existencia de autocorrelación o dependencia espacial en la variable analizada bajo la estructura de vecindades definida por W . Hablaremos de autocorrelación espacial positiva cuando se constituyen clusters de valores similares en nuestra variable. Por el contrario, se habla de autocorrelación espacial negativa cuando se detectan clusters de valores disímiles.

Este estadístico alcanza valores positivos y negativos dependiendo del signo de la correlación espacial, acercándose a 1 cuando esta es positiva y a -1 cuando la autocorrelación es negativa.

En el caso de la distribución de empresas, Guillain y Le Gallo (2010) y Arbia (2001) sugieren utilizar tres esquemas diferentes para tratar de identificar la distribución espacial de las actividades económicas de referencia: (i) alta concentración y baja dependencia espacial; (ii) alta concentración y alta dependencia espacial; y (iii) baja concentración y alta dependencia espacial.

2.1.1 Diagrama de dispersión de Moran.

El diagrama de dispersión de Moran es un gráfico mediante el cual se representa en el eje de abscisas las observaciones de la variable X normalizada y el eje de

ordenadas el retardo espacial de esta variable también normalizado, suponiendo que la matriz W está estandarizada. De esta forma aparecerá una nube de puntos (x_i, wx_i) de tal forma que si se encuentra dispersa sobre el plano será un indicador de ausencia de autocorrelación espacial, mientras que si la nube de puntos es próxima a la recta $y=x$ (o a la recta $y=-x$) indicará la presencia de una fuerte dependencia espacial positiva (resp. negativa), esto se corresponde con la regresión lineal entre las variable original X y su retardo espacial, WX . En este gráfico los cuatro cuadrantes se corresponden con un tipo distinto de asociación espacial. Así aquellos valores situados en el primer (o tercer) cuadrante se corresponden con observaciones que están por encima (respectivamente por debajo) de la media para las que en su entorno se encuentran valores por encima de la media (respectivamente por debajo de la media), y esto será más cierto en tanto en cuanto esté más próximo de la diagonal ($y=x$). Los valores situados en el segundo y cuarto cuadrante se corresponderán con aquellos en los que el valor observado en un área sea diferente del valor medio del entorno. Los valores situados en el segundo cuadrante se corresponde con observaciones de la variable que siendo inferior a la media en su entorno se encuentran valores superiores a ésta, por el contrario los puntos situados en el cuarto cuadrante indican valores superiores a la media en cuya vecindad se da un valor medio inferior a la media.

Con este tipo de gráficos es posible obtener una información detallada de las observaciones que contribuyen de forma positiva a la autocorrelación espacial y cuales contribuyen negativamente.

Otra utilidad de este tipo de gráficos es la posibilidad que ofrecen para detectar observaciones espacialmente atípicas (conocidas como *outliers* espaciales) que pueden desfigurar el índice de Moran, bien con contribuciones excesivamente elevadas a este índice, que indiquen autocorrelación espacial cuando realmente no la hay, o bien con contribuciones elevadas y de signo negativo que induzcan a pensar que no existe asociación espacial cuando realmente si la hay.

2.2. Software para el análisis de datos espaciales

Existe una gran oferta de programas que gestionan información georreferenciada. En Chasco (2003, páginas 32 y 33) podemos ver alguno de ellos. En este trabajo se ha utilizado principalmente ArcGIS, y de manera complementaria GeoDa.

ArcGIS, no es un software libre, pero quizás es el software de tratamiento de datos georeferenciados más potente del mercado. Es un programa producido y distribuido por ESRI (Environmental Systems Research Institute) y que agrupa varias aplicaciones para la gestión de la información geográfica: ArcGIS Desktop, Arcglobe, Arcmap, Arccatalog, Arctoolbox, Arcscene.

Para nuestros propósitos, ArcGIS ha sido utilizado en una doble dirección:

Creación y tratamiento de mapas: Este programa trabaja superponiendo capas (raster o de píxeles y vectoriales). ArcGIS nos permite trabajar de forma independiente sobre cada capa además de resolver los posibles problemas de georreferenciación ligados a la disparidad de sistemas de referencia que pueden tener las distintas capas.

Cálculo: Nos ha permitido crear Diagramas de dispersión de Moran, Mapas Lisa (en la sección siguiente se hablará de ellos) y la distribución del Índice I de Moran junto con los p-valores y el valor de dicho índice.

Por otro lado GeoDa es un software con menos potencialidad como Sistema de Información Geográfica y carece de muchas de las opciones que permite manejar este tipo de información, pero por el contrario está específicamente desarrollado para realizar un AEDE, permitiendo la obtención, tanto de matrices de conectividades (W) como de los mas populares indicadores globales y locales de autocorrelación espacial. Dispone también de un output gráfico ideal para explorar la vertiente espacial de cualquier conjunto de datos, tanto de observaciones puntuales como reticulares. En el caso de la aplicación que se mostrará en la siguiente sección, utilizaremos el diagrama de dispersión de Moran.

3. Estudio de la distribución espacial de empresas en el campo de Cartagena

El objetivo de esta sección es explorar la distribución espacial de la actividad económica de la Comarca de Cartagena. Para alcanzar este objetivo utilizaremos como variable proxy o indicador de la actividad económica el número total de empresas. Sobre esta variable utilizaremos las herramientas del AEDE descritas en la sección anterior. En concreto, mediante el uso de la información disponible en SABI, que aporta la ubicación geográfica de cada empresa y los programas ArcGIS y GeoDa, se estudiará el grado de dependencia espacial de esta variable, tanto a nivel global como para los distintos sectores de actividad económica. Y lo haremos desde una doble vertiente:

- Cuantitativa, mediante el calculo del Índice de Moran, su distribución y el Diagrama de Dispersión de Moran.
- Gráfica, con análisis de los diferentes mapas presentados.

3.1. Descripción del ámbito geográfico: La Geografía de la Comarca de Cartagena

La Comarca de Cartagena (Figura 1) está situada en el Sureste de La Región de Murcia. Es una zona fundamentalmente plana con la excepción de la zona litoral donde se encuentran las mayores elevaciones de la Comarca. Lo anterior, junto con una ligera inclinación NO-SE hace que la mayoría de los cauces (con algunas excepciones) desemboquen en la zona Este, donde se encuentra el Mar Menor, una laguna salada que es sin duda el elemento geográfico más sobresaliente de la zona. Otro elemento a destacar, sobre todo por su importancia económica, es la riqueza minera de esta área centrada sobre todo en municipios como La Unión o Mazarrón.



Figura 1. Mapa de la Comarca de Cartagena

La Comarca de Cartagena está dividida en 8 municipios, cuya extensión supone algo más del 13% de la superficie regional. En cuanto a su población, como vemos en la Tabla 1, su peso es significativamente mayor en el contexto regional, con una población de 390.983 que alcanza una densidad de 263,83 habitantes por Km² que más que duplica el dato regional.

Tabla 1. Datos demográficos de la Comarca

Municipios	Población	Superficie	Densidad	% Población Comarca	% Población Región
Los Alcázares	16.217	19,82	818,2	4,15%	1,10%
Cartagena	214.918	558,08	385,1	54,97%	14,62%
Fuente Álamo	15.873	273,52	58,0	4,06%	1,08%
Mazarrón	35.473	318,87	111,2	9,07%	2,41%
San Javier	32.366	75,10	430,9	8,28%	2,20%
San Pedro del Pinatar	24.093	22,37	1.077,0	6,16%	1,64%
Torre-Pacheco	33.218	189,40	175,3	8,50%	2,26%
La Unión	18.825	24,79	759,3	4,81%	1,28%
Comarca de Cartagena	390.983	1.482	263,8		
% Sobre total regional	26,60%	13,10%			
Región de Murcia	1.470.069	11.313,91	129,93		

3.2. La Base de Datos SABI, Sistema de análisis de balances ibéricos

La base de datos SABI suministrada por Informa S.A recopila información económica, fundamentalmente sobre datos contables, de más de 1.230.000 empresas en España. Se dispone de información referente a los principales registros contables de la empresa, así

como otras magnitudes relevantes tales como su tamaño empresarial, sector de actividad, localidad y lo que es mas importante para nuestro estudio: La exactas coordenadas de latitud y longitud de cada empresa que permite incorporar esta información en un Sistema de Información Geográfica para su análisis espacial. Esta base de datos no es exhaustiva. Aproximadamente una tercera parte de las empresas españolas están aquí registradas. La cobertura de la base de datos es desigual para los distintos tamaños, comunidades o sectores de actividad si se considera como referencia los datos suministrados del DIRCE (Directorio Central de Empresas). Por tamaño empresarial se contienen valores de mas de un tercio de las microempresas, y para más de las tres cuartas partes de las pequeñas empresas (10 a 49 trabajadores), porcentaje que se va incrementado cuanto mayor es el tamaño empresarial. Por comunidades autónomas, SABI presenta una cobertura para Madrid y Cataluña superior al 40%, mientras que para el resto de comunidades el grado de cobertura es sensiblemente inferior. Sectorialmente, el mayor grado de cobertura se obtiene en las actividades extractivas, energía y agua alcanzando casi el 80% y el menor en el sector servicios en el que escasamente se supera el 30% de las empresas.

En el año 2009, se disponía de información de 35.892 empresas ubicadas en la Región de Murcia. Con la información bruta de SABI, para nuestros objetivos de análisis las empresas se han agrupado por municipios y distritos censales¹ de ubicación, atendiendo también al sector económico de actividad. Así, se han agrupado las empresas dependiendo del tipo de actividad económica que realizan en: Agricultura, Pesca, Industria, Construcción, Servicios Comerciales (SC) y No Comerciales (SNC), en estos últimos se incluye la Hostelería. Las variables en las que hemos concentrado nuestra atención son el número de empresas y el empleo. Respecto al número de empresas se han considerado sólo aquellas que se encuentran activas. A continuación, se ha agrupado los datos de dichas variables por municipios de instalación y distrito censal.

Como se apuntó anteriormente, nuestro análisis se centrará sólo en los municipios que componen la Comarca de Cartagena como unidad económica. En total en la base de datos SABI se dispone de información para 8.314 empresas, donde más de la mitad se encuentra clasificada dentro del sector servicios, tanto comerciales como no comerciales. Sólo 40 empresas están clasificadas en el sector de la Pesca que hemos mantenido en este estudio por la importancia histórica de la actividad en la Comarca, aunque no formará parte del análisis espacial de la sección siguiente debido a que su escaso número restaría significatividad a los datos obtenidos.

3.3 La demografía empresarial de la Comarca de Cartagena

En la Tabla 2 se presenta la distribución porcentual de empresas agrupadas por sector de

¹ En España, el distrito censal es la unidad administrativa inmediatamente inferior al municipio. Por debajo de este se encuentra la sección censal. En todo el país había en 2.005 algo más de 8.100 municipios que se dividían en 10.521 distritos (en la inmensa mayoría de los pueblos más pequeños el municipio coincide con el distrito) y 34.800 secciones.

actividad de la información recogida en SABI. Destacamos en este caso el municipio de Los Alcázares por su actividad en servicios no comerciales. Fuente Álamo por su actividad Agrícola e Industrial. Mazarrón por la actividad pesquera y La Unión por su posición destacada en actividades asociadas a la Construcción.

Tabla 2: Distribución número empresas por municipio y sector de actividad

Municipio	Agrícola a	Pesc a	Industria a	Construcción c	SC	SNC	Total *
Cartagena	3,10%	0,50 %	9,50%	17,40%	22,40 %	47,00 %	4.123
Fuente Álamo	14,70%	0,00 %	16,80%	16,80%	20,70 %	31,10 %	341
La Unión	2,40%	0,50 %	15,50%	27,20%	14,60 %	39,80 %	257
Los Alcázares	2,70%	0,00 %	4,50%	17,70%	22,70 %	52,30 %	266
Mazarrón	8,00%	1,90 %	6,40%	21,40%	19,30 %	43,00 %	681
San Javier	7,80%	0,00 %	7,70%	15,70%	22,30 %	46,50 %	818
San Pedro del Pinatar	4,20%	0,60 %	8,30%	15,80%	18,80 %	52,30 %	596
Torre Pacheco	12,00%	0,10 %	9,00%	20,80%	21,10 %	37,00 %	1.232
Total Comarca	482	40	768	1.516	1.780	3.728	8.314
Total Región de Murcia	1.553	74	5.271	6.206	8.626	14.161	35.891

* Totales en valores absolutos

3.4 Dependencia espacial en el número total de empresas

Antes de comenzar este estudio, debemos precisar que el número de distritos censales que componen la comarca es reducido, 22, y que este pequeño tamaño muestral puede condicionar alguno de los resultados que se van a presentar. En todo caso debemos advertir igualmente que este número es superior a 16 y que hacer un estudio basado en unidades espaciales más pequeñas (secciones censales) podría sesgar más aún los resultados debido a que un gran número de estas unidades espaciales no contienen ninguna empresa.

Los mapas que se van a presentar reflejan por una parte, la ubicación de las empresas y por otra, el número de éstas que están instaladas en cada uno de los distritos censales, marcado esto último con el color de cada distrito: cuanto más oscuro, mayor número de empresas. La Figura 2 muestra la distribución del número total de empresas en la Comarca de Cartagena utilizando ArcGIS. Cada uno de los puntos del mapa se corresponde con la exacta localización de las empresas atendiendo a la información

obtenida en SABI. La *unión espacial* de los mapas de los distritos censales por un lado y de los puntos con las empresas por otro, permite contabilizar cuántas empresas y de que sector económico hay en cada unidad espacial. Esta información no está disponible si no se utilizan este tipo de sistemas informáticos.

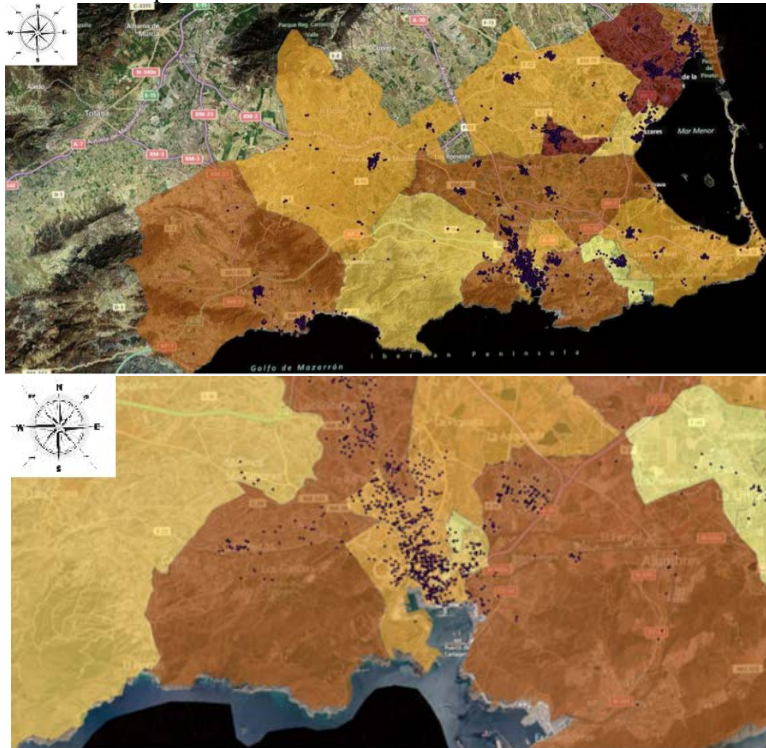


Figura 2: Distribución de empresas del Campo de Cartagena. Obtenido con ArcGIS.

Para complementar la información de los mapas procedemos al cálculo del estadístico I de Morán. Tal y como ya se apuntó en la sección anterior, utilizando tanto el programa ArcGIS como GeoDa. Con el primero de ellos obtendremos la significatividad del estadístico utilizando la versión asintótica, mientras que con el segundo la versión permutacional. En los gráficos que se muestran en la Figuras 3 y 4 aparecen estos gráficos para la totalidad de las empresas de la Comarca. Los resultados aquí obtenidos son en principio contradictorios cuando utilizamos ArcGIS puesto que si se usa el principio de contigüidad (la matriz cuyo valor $w_{ij}=1$ si las regiones i y j tienen frontera común y 0 en caso contrario) el z-valor del índice de Moran es 1,55 con un p-valor de 0,121, presumiéndose una distribución aleatoria de las empresas. Por el contrario, si el principio utilizado es el de la inversa de las distancias entre regiones, el z-valor es en este caso 2,197 con un p-valor asociado de 0,028, concluyéndose ahora sí, la existencia de correlación espacial en la distribución estudiada. Similares resultados obtenemos al utilizar GeoDa ahora evaluando la significatividad mediante permutaciones. Las Figuras 5 y 6 vienen a confirmar esta tesis.

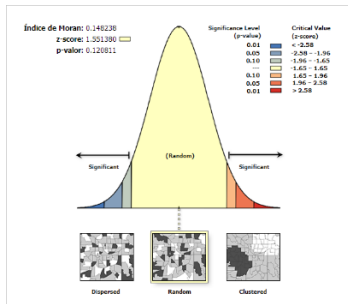


Figura 3: I de Moran (contigüidad)

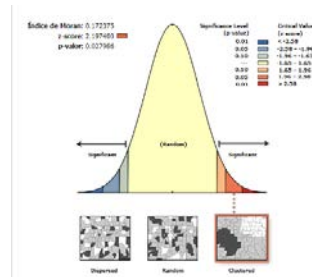


Figura 4: I de Moran, (inversa de la distancia)

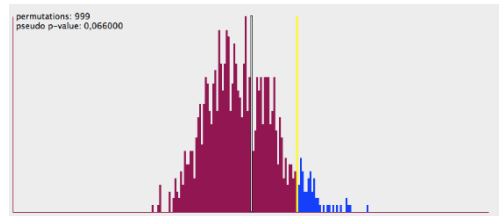
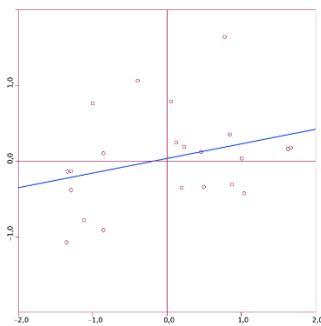


Figura 5. Diagrama de dispersión de Moran con principio de contigüidad

En la Figura 5 se muestra el diagrama de dispersión de Moran usando GeoDa. En este gráfico se representa, en eje horizontal, la variable número de empresas en los distritos censales normalizada. En el eje vertical aparece la variable, también normalizada, promedio de empresas en la vecindad de cada distrito. En este diagrama se ha aplicado la matriz de pesos basada en el principio de contigüidad de la torre: la posición i, j de la matriz es uno si dos regiones comparten frontera y 0 en caso contrario. Los valores de la diagonal son 0. A la derecha de este gráfico se muestra el pseudovalor del estadístico I de Moran obtenido esta vez por bootstrap junto con la distribución de los 999 replicas del estadístico I obtenidos para cada una de las permutaciones. Los resultados en significatividad por ambos métodos (asintótico y permutacional) aunque no iguales si que suministran la misma información. No hay claros síntomas de dependencia espacial.

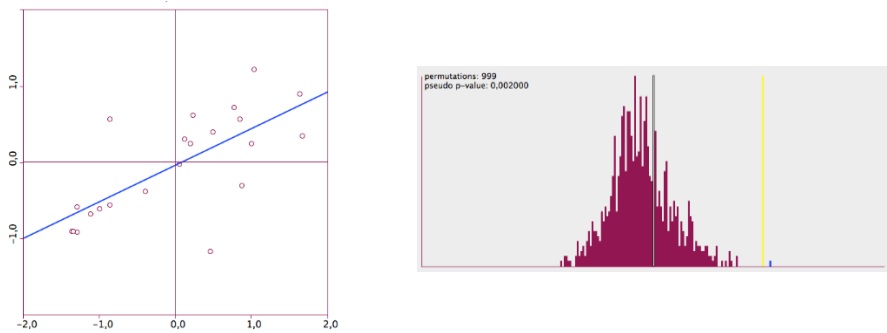


Figura 6. Diagrama de dispersión de Moran con principio del inverso de la distancia

En la Figura 6 los ejes representan las mismas variables que en la figura anterior, aunque en este caso se ha aplicado la matriz de pesos basada en el principio del inverso de la distancia: la posición i, j de la matriz es el inverso de la distancia entre las regiones i y j y los valores de la diagonal son 0. Como ya se podía ver en la Figura 4, el grado de correlación es mayor también cuando se utiliza la versión boots del estadístico I.

La aparente contradicción anterior se explica si tenemos en cuenta que: (a) Considerar los pesos de los inversos de las distancias permite contemplar la influencia de cualquier región en todas las demás, a diferencia del principio de contigüidad que solo permite analizar la influencia entre regiones vecinas. (b) Estamos estudiando una comarca no muy extensa, además de bien comunicada, especialmente en su parte oriental. Esto viene a justificar que la influencia de “*todos sobre todos*” sea un hecho. (c) Se ha de tener en cuenta el efecto que tienen los distritos censales de la ciudad de Cartagena, mucho más pequeños que el resto y con mayor número de vecindades que los demás, además de concentrar una gran cantidad de empresas en su seno. (d) Como advertíamos al principio, puede que el número de distritos en los que hemos basado el estudio explique parte de esta discrepancia.

Lo anterior nos lleva a concluir la conveniencia de usar la matriz de pesos formada por los inversos de las distancias, y como derivada de lo anterior, deducir la existencia de correlación espacial en la distribución de las empresas dentro de los distritos censales de la Comarca.

Desde un punto de vista meramente descriptivo, podemos concluir, de observar la Figura 2, varios elementos interesantes sobre la taxonomía empresarial de la zona:

- (a) Los cascos urbanos de las ocho ciudades son los grandes cluster de esta área. (b) Zonas como Escombreras o el Polígono de Los Camachos, no logran desbancar ni de lejos a la ciudad de Cartagena como base de la mayoría de empresas de la ciudad. (c) Caso aparte merece el Polígono Industrial Cabezo Beaza, donde vemos que sí se han instalado un número importante de empresas. (d) Por último

reseñar el papel de las vías de comunicación como elementos sobre los que se instalan empresas, sobre todo la N-301a, con un trazado Norte-Sur, sobre la que han surgido gran número de empresas desde su salida de Cartagena hasta Santa Ana.

Pasamos ahora a analizar la distribución del número de empresas en los cinco sectores en los que las hemos clasificado. En el primero de ellos, el agrario, mostrado en la Figura 7, debemos apuntar que:

- El Norte de la Comarca (Torre Pacheco, Fuente Álamo, San Javier y el Norte de Cartagena) son los principales focos de atracción de empresas.
- La ubicación de la sede de estas empresas está en su inmensa mayoría en los núcleos poblacionales. Esta aparente paradoja, recordemos que hablamos de un sector que desarrolla su actividad en el campo, se explica porque las sedes suelen instalarse donde se desarrollan las tareas administrativas.
- A diferencia de lo que ocurre para el conjunto de las empresas, las del sector agrario no demuestran estar condicionadas por las grandes vías de comunicación.
- El municipio de Los Alcázares y el distrito censal donde se ubica el núcleo poblacional de Torre Pacheco forman un par de *cluster*. Para concluir debemos apuntar que este sector sí presenta un alto grado de correlación espacial pues su p-valor es prácticamente nulo

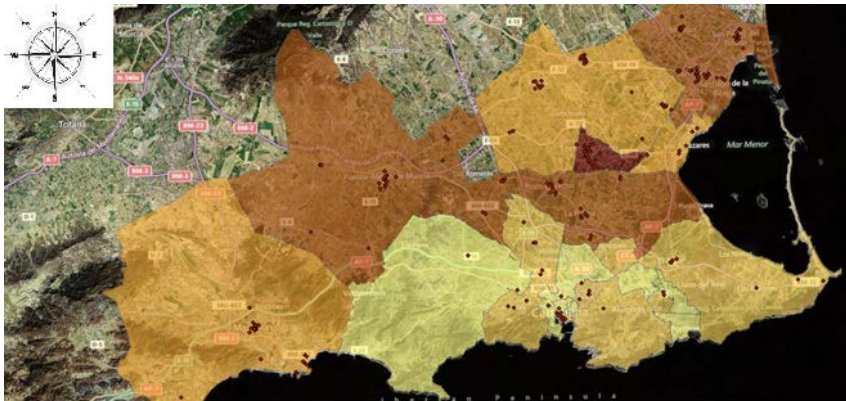


Figura 7. Mapa de distribución de empresas agrarias

Si el Norte de la Comarca es el gran foco de atracción de la actividad agraria, tanto a nivel comarcal como regional, el municipio de Cartagena concentra la mayor cantidad de empresas industriales de la Comarca. En 2010, concentraba más de la mitad de las empresas activas de la zona. Casualmente, idéntica proporción se daba en el número de trabajadores dedicados a esta actividad.

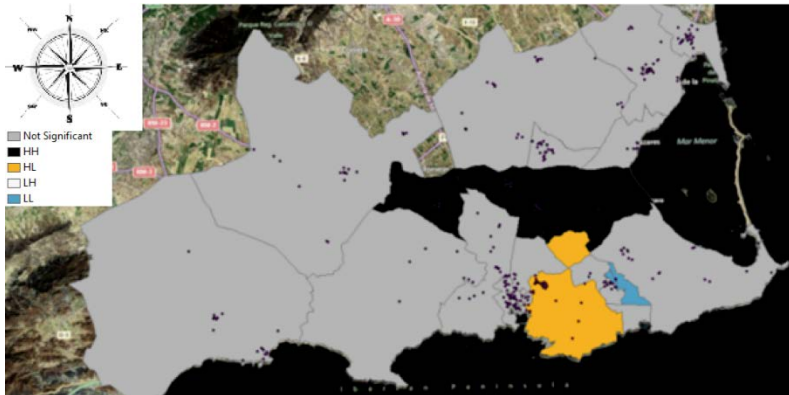


Figura 8. Mapa LISA del número de empresas industriales

Por razones didácticas hemos decidido usar para este sector un mapa LISA expuesto en la Figura 8 (ver Chasco 2003 para información detallada de estos mapas). En este tipo de mapas se representan las relaciones entre el valor de una variable en una región y el de las regiones vecinas. Más en concreto, el color amarillo nos informa de que en esa región se da un valor alto de la variable pero bajo en las regiones vecinas. El negro viene a significar un alto valor de la variable en la región y en sus vecinas.

La impresión que se desprende de esta figura nos permite ubicar como protagonistas a dos distritos del Cartagena: La Palma y Escombreras. El primero (en negro en el mapa) podría ser considerado como un *cluster* pues posee una alta concentración de empresas industriales y los distritos colindantes también. Por el contrario el Distrito de Escombreras, pese a tener también un alto número de este tipo de empresas, está rodeado de distritos poco desarrollados. Es este un hecho interesante puesto que son los distritos situados al Este del citado valle, los del municipio de La Unión, quienes sufren esta carencia (uno de ellos en azul claro en el mapa). Podríamos estar ante un caso en el que la existencia de dos grandes focos de atracción de empresas industriales, el Valle de Escombreras y el Polígono Industrial Cabezo Beaza (situado también en este distrito), impiden el desarrollo de La Unión, explicando en gran medida que sea éste el municipio de La Región con mayor tasa de desempleo (cerca de 3.000 de sus 18.00 habitantes estaban en paro en septiembre de 2012).

También en este sector, como en el agrario, obtenemos un p-valor casi nulo (recordemos que se está aplicando la matriz de los inversos de la distancia) lo que nos lleva a concluir que se da una fuerte dependencia en la distribución espacial de las empresas de este sector.

En el tercero de los sectores, el de la construcción, la Base de Datos SABI tenía registrada a comienzos de 2010, unas 768 empresas. La grave crisis económica que sufrimos se está cebando principalmente con este sector, lo que hace que cualquier dato deba ser puesto en cuarentena. Para visualizar este hecho valgan los siguientes datos: en la Región de Murcia hay, en 2012, un 37% menos de empresas de construcción (que son

las catalogadas en los grupos 41,42 y 43 en la clasificación CENAE-2009) que en 2008.

Añádase a lo anterior el enorme peso que las microempresas tienen en este sector, ya que casi 5 de cada 8 son autónomos sin asalariados (casi el doble del porcentaje de este tipo de empresas que en el conjunto del país). Esto añade más incertidumbre a los datos pues en general las empresas pequeñas tienen una vida media inferior a las grandes. Además, la pequeña dimensión de la mayoría de ellas hace que en muchos casos no aparezcan en la Base de Datos SABI

Entrando en el análisis del sector, mostrado en la Figura 9, debemos reseñar que:

- Los municipios costeros, sobre todo Mazarrón, con 145 empresas, San Javier, con 129, y Torre Pacheco con algo más de 250 (en esta localidad más del 20% del tejido empresarial se dedica a la construcción y emplea a cerca de 2.700 personas), poseen la mayor concentración de empresas constructoras.
- Centrándonos en el municipio de Cartagena, debemos indicar que el polígono industrial Cabezo Beaza es otro foco de concentración de empresas de construcción pues da cobijo a 66 (sobre un total de 394 empresas) en las que trabajan cerca de 1.000 personas.
- Finalizamos apuntando que, como en el sector agrario y en el industrial, este sector también presenta un alto grado de dependencia espacial a juzgar por el p-valor obtenido con la matriz de pesos de los inversos de la distancia.

Todo lo anterior viene a justificar un hecho por todos conocidos, y es que la actividad constructora en La Comarca se ha centrado en la costa y en la construcción de *Resorts* en el interior. Aquí ha jugado un papel fundamental *Polaris World*, paradigma de empresa dedicada a la construcción de urbanizaciones entorno a campos de golf, alejadas de la costa, ubicadas en antiguos terrenos de uso agrícola recalificados. La mayor parte de sus desarrollos se ha centrado en el municipio de Torre Pacheco, donde además tiene su sede social.

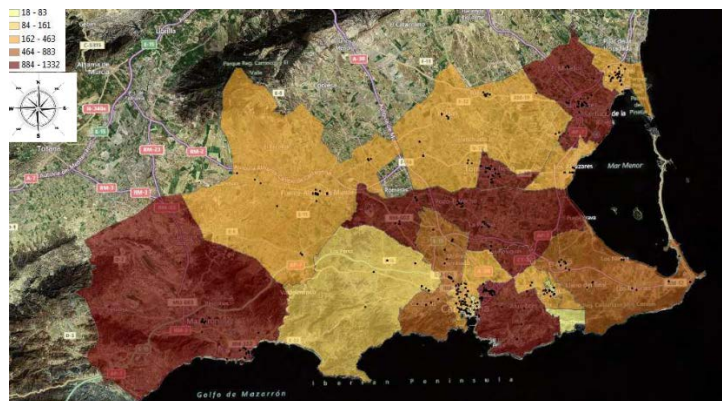


Figura 9. Mapa de distribución de empresas del sector de la construcción

En cuanto a la distribución de empresas de servicios comerciales indicaremos que algo más de la mitad de las empresas del sector, 924, se asientan en el municipio de

Cartagena. Además es interesante indicar el papel que desempeña La Carretera Nacional *N-301a* así como el del polígono industrial Cabezo Beaza, donde se encuentran 127 de estas empresas (de un total de 1.449 en el sector) y trabajan 1.430 personas, justamente el 10% del total de trabajadores de este sector en toda La Comarca.

En la categoría de empresas de servicios no comerciales se incluyen todas aquellas dedicadas a sectores como la educación, la sanidad, las actividades asociativas, recreativas, culturales y deportivas. En este sector se percibe una menor influencia de los polígonos industriales como atractores de estas empresas. Por poner un ejemplo, el polígono más importante de la comarca, el Cabezo Beaza, incluye al 10% del tejido de empresas de servicios comerciales. En las no comerciales, apenas alcanza el 5%. La naturaleza de este tipo de negocio obliga a ubicarlo en el centro de las ciudades, un ejemplo de esto es que en 2010, sólo en la ciudad de Cartagena había 1.923 empresas que daban trabajo a cerca de 10.000 personas.

Un hecho que hace singular a este sector respecto al resto es el peso de “Lo Público”, sobre todo en sanidad y educación. Como este trabajo se centra únicamente en la actividad empresarial, deja fuera el papel que las diferentes administraciones tienen en la oferta de servicios a los ciudadanos.

Para finalizar debemos indicar que en estos dos últimos sectores no se aprecia dependencia espacial en la distribución de las empresas, a diferencia de los tres sectores anteriores.

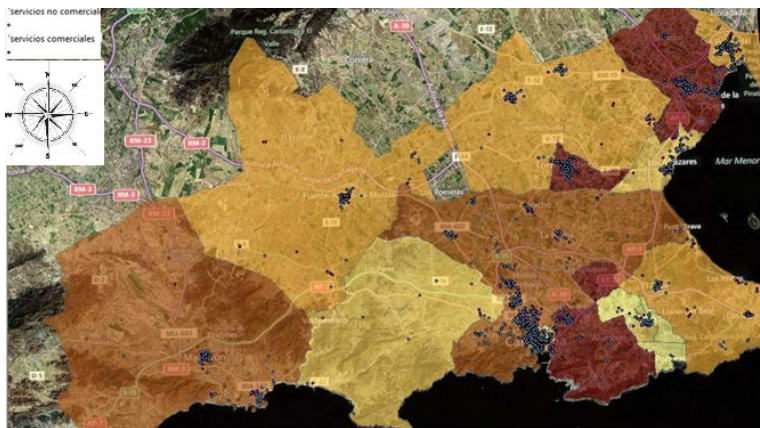


Figura 10. Mapa de distribución conjunta de empresas de servicios comerciales y no comerciales

4. Conclusiones y futuras líneas de investigación

El objetivo de este trabajo ha sido doble. Por un lado se ha presentado uno de los indicadores de dependencia espacial más utilizados en el ámbito de la econometría

espacial: El índice I de Moran. Por otro, se ha explorado la distribución espacial de la economía de la Comarca de Cartagena utilizando la demografía empresarial como herramienta de análisis. Ambos análisis se han realizado utilizando dos software específicos para el análisis de la información con referencia espacial ArcGIS y GeoDa.

Aunque la Comarca de Cartagena no es una unidad geográfica oficialmente establecida, sí que hay razones históricas y económicas que invitan a considerarla como una región homogénea que engloba a 8 municipios de la Región de Murcia y cuya población supone más del 25% del total regional.

En lo que se refiere a la demografía empresarial, sólo hay cifras oficiales a nivel de Comunidad Autónoma ofrecidas por el INE a través de la base de datos DIRCE. No hay por tanto información demográfica de las empresas a un nivel de desagregación espacial inferior. Esto nos ha obligado a descartar esta base y acudir a la información disponible en la base de datos SABI, menos exhaustiva que la anterior (recordemos que DIRCE tiene naturaleza censal) pero que posee datos sobre la ubicación espacial de cada empresa, elemento éste básico en nuestra investigación.

La información disponible en SABI nos permite elaborar el primer mapa de la distribución geográfica de las empresas en La Comarca, y por extensión el primer análisis del mismo. En este mapa se observa con detalle la localización exacta de todas y cada una de las más de 8.000 empresas, permitiendo identificar patrones espaciales de localización. El uso de Sistemas de Información Geográfica ha sido una herramienta imprescindible para este estudio.

A partir de este mapa y utilizando la potencia del SIG, hemos podido asignar cada empresa al Distrito Censal donde se encuentra ubicada. Esta asignación ha permitido explorar la distribución espacial de las empresas utilizando como unidad espacial de análisis el Distrito Censal. De los resultados de este análisis podemos extraer las siguientes conclusiones:

- Si consideramos todas las empresas, hay estructura de dependencia espacial confirmada por el índice de Moran con un valor estandarizado de 2,19 (p -valor=0,027). Esta estructura no es más que el fruto de la fuerte agrupación de empresas en Cartagena, dividida en 5 Distritos Censales de pequeño tamaño. También están condicionadas por las grandes vías de comunicación.
- Si consideramos las empresas por sectores de actividad, tanto en Agricultura como en Industria se identifican patrones de dependencia espacial, frente a los sectores de servicios donde no se observa esta estructura. El patrón espacial en Agricultura se debe a la fuerte presencia de empresas de este sector en el Norte de la Comarca (Torre Pacheco, Fuente Álamo, San Javier y el Norte de Cartagena). Las empresas agrícolas no están condicionadas por las vías de comunicación más importantes.
- La aglomeración de empresas en el Sector Industrial es fundamentalmente debida a la presencia de dos distritos en Cartagena con elevados porcentajes de actividad en este sector, el de la Palma y el de Escombreras.

- El sector de la Construcción también presenta estructura de dependencia espacial. Los municipios costeros, sobre todo Mazarrón y San Javier, con casi 300 empresas entre ambos, y Torre Pacheco con más de 250 (en Torre Pacheco más del 20% del tejido empresarial se dedica a la construcción y empleaba a más de 2700 empresas) poseen una mayor concentración de empresas constructoras.
- Los sectores de servicios no presentan estructura de dependencia espacial. La principal razón es que estas empresas se distribuyen fundamentalmente en los cascos urbanos y en ninguno de ellos se observa un nivel de actividad superior al que cabe esperar en caso de equidad en la distribución.

Notemos finalmente que son varias las limitaciones de este estudio. En primer lugar la falta de representatividad muestral de la información que ha servido de base a este trabajo basada en SABI. Esta limitación es insalvable ya que no se dispone de otra base de datos que suministre información a un nivel de desagregación inferior al provincial. Esperemos que en un futuro se disponga de información de mayor calidad que permita contrastar los resultados de este análisis. En segundo lugar el número de unidades espaciales con las que se ha trabajado es muy pequeño, lo que podría inducir a una falta de consistencia de los resultados. Pese a lo anterior, el método de inferencia utilizado, basado en bootstrapping permutacional en vez de en la distribución asintótica del estadístico, confiere robustez a los resultados obtenidos.

Referencias Bibliográficas

1. D. Puga, The magnitude and causes of agglomeration economies. *Journal of Regional Science* 50 (2010) (1):203–219.
2. G. Arbia, G. A lustrum of SEA: recent research trends following the creation of the Spatial Econometrics Association (2007–2011). (2011). *Spatial Economic Analysis*, 6(4), 377–395.
3. G.E.P. Box y G.M. Jenkins Time Series Analysis, Forecasting and Control (Holden-Day, San Francisco 1970).
4. W. Tobler, A computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46 (1970) 234–240.
5. P. Krugman *Geografía y comercio*. Editorial Antoni Boschv (1993)
6. P. Krugman *Desarrollo, geografía y teoría económica*. Editorial Antoni Boschv (1997)
7. P. Moran, The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society B*, 10, (1948) 243–251.
8. R. Geary, The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5, (1954) 115–145.
9. J.H.P. Paelinck, *Elements d'Analyse Economique Spatiale*. (1985) París: Economica

10. L. Anselin, Space and applied econometrics, Anselin (ed.), Special Issue. *Regional Science and Urban Economics*, 22. (1992)
11. L. Anselin. *Spatial econometrics: Methods and models*, (1988) Kluwer Academic Press.
12. A. Getis, A. y J.K. Ord. The analysis of spatial association by use of distance statistics, *Geographical Analysis* (1992) 24, 189-206.
13. J. Pinkse, y Brett, C., Spatial price competition a semiparametric approach. *Econometrica* (2002) 70, 1111–1153.
14. FA López, M Matilla, J Mur y M Ruiz A Non-Parametric Spatial Independence Test Using Symbolic Entropy, *Regional Science and Urban Economic*. (2010) 40, 106-115
15. Chasco, C. (2003) Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales, Consejería de Economía e Innovación Tecnológica de la Comunidad de Madrid. Descargable desde: http://www.madrid.org/cs/Satellite?idConsejeria=1109266187242&idListConsj=1109265444710&c=CM_Publicaciones_FA&pagename=ComunidadMadrid%2FEstructura&sm=1109266101003&idOrganismo=1109266227074&pid=1109265444699&language=es&cid=1114175317832
16. A.D. Cliffy J.K. *Spatial Autocorrelation* (Pion, London, 1973)
17. A.D. Cliff y J.K *Spatial Processes*. Models & Applications. (Edit. Pion Limited, 1981).
18. A.K. Sen, Large sample-size distribution of statistics used in testing for spatial correlation. *Geographical Analysis* 8, (1976) 175-184
19. R. Guillain y J. Le Gallo, Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis, *Environment and Planning B: Planning and Design*, 37, (2010) pp. 961-981
20. G. Arbia, Modelling the geography of economic activities on a continuous space, *Papers in Regional Sciences*, (2001)80(4), 411-424

Software

1. Microsoft Excel 2000, Microsoft Corporation <Http://microsoft.com/spain/Office>
2. ArcGIS 10.0, Environmental Systems Research Institute, www.esri.com/software/arcgis
3. Geoda, Anselin, Luc, Ibnu Syabri and Youngihn Kho (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38 (1), 5-22. <https://geodacenter.asu.edu>

GEODASPACE: A RESOURCE FOR TEACHING SPATIAL REGRESSION MODELS

CORO CHASCO

coro.chasco@uam.es

*Departamento de Economía Aplicada, Universidad Autónoma de Madrid
Avenida de Francisco Tomás y Valiente 5, 28049 Madrid*

ABSTRACT: Spatial econometrics has become a prominent topic in the recent scientific literature. For this reason, it is being used in research as well as teaching both undergraduate and graduate econometrics courses. *GeoDaSpace* is a software package for the estimation and testing of spatial econometric models in an intuitive and easy-to-use point and click environment. It is still an alpha release freely downloadable from the GeoDa Center (Arizona State University), which incorporates a wide range of estimation methods (OLS, 2SLS, ML, GM/GMM) and models (spatial lag, spatial error, spatial lag and error, spatial regimes), with options for spatial and non-spatial diagnostics, non-spatial endogenous variables and heteroskedasticity/HAC covariance estimators. *GeoDaSpace* is a very useful teaching resource that can be used by both teachers and students.

Keywords: spatial autocorrelation models, endogenous variables, heteroskedasticity, spatial regimes, GeoDaSpace.

RESUMEN: La econometría espacial se ha convertido en un tema de gran relevancia en la literatura científica reciente. Por este motivo, se está empleando no sólo en la investigación sino también en la docencia, tanto de cursos de pregrado como de posgrado. *GeoDaSpace* es un paquete informático especializado en la estimación y contraste de los modelos econométricos espaciales dentro del entorno intuitivo y amigable del “apuntar-y-cliquear”. Se trata aún de una versión alfa, que está disponible gratuitamente desde el GeoDa Center (Universidad del Estado de Arizona), que incorpora una amplia gama de métodos de estimación (MCO, MC2E, MV, GM/GMM) y modelos (retardo espacial, error espacial, retardo y error espacial, regímenes espaciales), con opciones para la obtención de contrastes espaciales y no espaciales, tratamiento de regresores estocásticos no espaciales y de estimaciones robustas a la heteroscedasticidad en los errores (HAC). *GeoDaSpace* es un recurso docente útil que puede ser utilizado tanto por estudiantes como profesores.

Palabras claves: modelos de autocorrelación espacial, variables endógenas, heteroscedasticidad, regímenes espaciales, GeoDaSpace.

1. Introduction

The explosive diffusion of geographic information systems (GIS) technology and the associated availability of geo-coded socioeconomic data sets have created a need for specialized methods to deal with the distinguishing characteristics of such geographic data. For this reason, spatial econometrics has become a prominent topic specialized in the treatment of spatial data in all branches of social sciences and economics, particularly in regional science, urban and real estate economics, economic geography and environmental economics (see for example Anselin et al., 2004 for a review). Accordingly spatial regression techniques are now becoming an established component in the applied econometrics toolbox, as witnessed by the increasing attention given to this topic in standard econometrics textbooks (Maddala, 2001; Wooldridge, 2002; Gujarati, 2003; Kennedy, 2003; Baltagi, 2008). As a matter of fact these methods are also being implemented in the formal curricula of undergraduate and postgraduate university programs as part of the econometric courses in economics, business administration, marketing, geography, environmental studies or epidemiology, between others.

GeoDaSpace is a very useful teaching resource that can be used by both teachers and students in the classroom. This software package, which is still an alpha release freely downloadable from the GeoDa Center at Arizona State University (<https://geodacenter.asu.edu>), is a stand-alone program that has been designed for the estimation and testing of spatial econometric models in an intuitive and easy-to-use point and click environment. This package is part of the PySAL project developed by the professors Luc Anselin and Serge Rey (Rey and Anselin, 2010), from the GeoDa Center, as an open source cross platform modular library of spatial analytical functions (running in Windows, Macintosh and Linux), which are written in the Python scripting language (<http://pysal.org>). It is conceived as the foundational framework to deliver spatial analytical functionality in many different forms. Particularly, *GeoDaSpace* is the front-end GUI (Graphical User Interface) for the spatial econometric routines contained in the PySAL library (Anselin, 2012). These routines include the functionalities for input/output data formats (csv, dbf, shp), the creation and transformation of spatial weight matrices and the regression modules, which includes the state-of-the-art spatial econometrics estimation methods (OLS, 2SLS, ML¹, GM/GMM) and models (spatial lag, spatial error, spatial lag and error, spatial regimes), with options for spatial and non-spatial diagnostics, non-spatial endogenous variables and heteroskedasticity/HAC covariance estimators.

Since *GeoDaSpace* is intended for the user who is somewhat familiar with the methods, but prefers a point and click environment to the command line, the number

¹ Maximum Likelihood estimation is not present in the last version (version 0.8.1, May 2013), though it will come soon.

of available options are those that are most common but less than in other spatial econometrics routines present in Stata (Drukker et al. 2013), Matlab (Le Sage and Pace 2009) or R (Bivand 2002, Piras 2010).

The rest of the paper is organized as follows. Section 2 provides some background information on basic characteristics of the design and functionality of *GeoDaSpace*. Sections 3 and 4 present the spatial weight matrix and spatial regression menus, which are the bulk of this program. Section 5 illustrates the previous items with an empirical application. Section 6 provides a brief summary and some concluding remarks.

2. Design and functionality

GeoDaSpace is geared to the econometric analysis of discrete geospatial data, that is, objects characterized by their location in space either as points (point coordinates) or polygon (polygon boundary coordinates). Differently from its predecessor –*GeoDa*– the design of *GeoDaSpace* does not consist of an interactive environment combining maps with statistical graphs, using a technology of dynamically linked windows. The full set of functions of this software is listed in Table 1 and can be classified into three categories:

- spatial data utilities: data input and output,
- spatial weight matrix: creation and manipulation,
- spatial regression: estimation and diagnostics of linear spatial regression models.

Table 1. *GeoDaSpace* functionality overview

Category	Functions
Spatial data	. Data input from shape file (point, polygon) . Data input from dBase (dbf), Comma Separated Values (csv), model specification files (mdl) or results text files (txt). . Data output to model specification files (mdl) or results text files (txt).
Spatial weight matrices	. Spatial weights creation (contiguity rook and queen, binary distance bands, inverse distance, k-nearest neighbors and kernel weights), higher order spatial weights. . Spatial weights properties (editor, point-map viewer, connectedness histogram, transformations). . Spatial weights reading/writing from/to other extensions: ArcGIS (dbf, swm, txt), dat, gal, geoBUGS, gwt, kwt, MatLab, MatrixMarket, STATA text files.
Spatial regression	. OLS with diagnostics (e.g., Jarque-Bera, Breusch-Pagan, LM tests, Moran's I, etc.) . 2SLS for non-spatial endogeneity with diagnostics (Anselin-Kelejian test) . Spatial 2SLS+GMM spatial lag, spatial error and spatial lag+error models with

HAC and spatial HAC (Kelejian-Prucha) estimation

. Spatial regimes for basic and spatial models

. Output text files for predicted values and residuals, variance-covariance matrix and complete model specification

The functionality of *GeoDaSpace* is invoked by clicking toolbar buttons located in different parts of the main screen, as illustrated in Figure 1. Once a data file is invoked, the program opens a new menu with the complete set of variables, which must be dragged to the corresponding boxes in the specification menu.

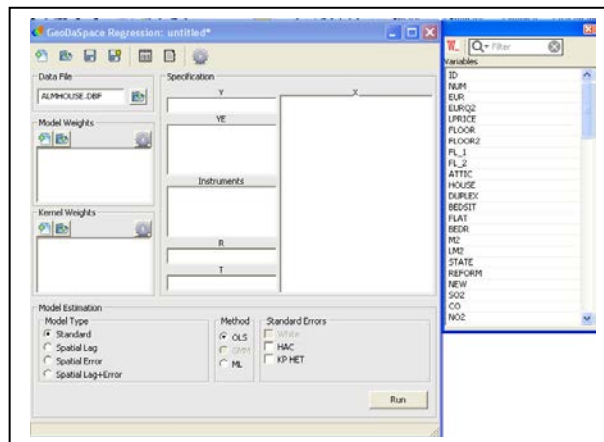


Figure 1. The opening screen with menu items and toolbar buttons.

There are six toolbars with different subdivisions:

- (1) Main menú, in the upper part of the main screen:
 - (a) Open (new or existing) model files.
 - (b) Save model files.
 - (c) Open de variable list.
 - (d) Show the results window.
 - (e) Show advanced settings for the model estimation:
 - (i) Computation of the standard deviation of the coefficients.
 - (ii) GMM settings.
 - (iii) Instruments settings.
 - (iv) Output files options.
 - (v) Spatial regime models.
 - (vi) Other estimation options.
- (2) Data file menu, which allows opening dBase (dbs) and Comma Separated Values (csv) input databases.
- (3) Model spatial weights matrix menu:
 - (a) Creation of a spatial weights matrix:
 - (i) Input map file (shp).
 - (ii) Select/add an ID variable for the weights file.

- (iii) Contiguity matrix: queen, rook, higher orders of contiguity, inclusion of lower orders.
- (iv) Distance-based matrix: distance metrics (Euclidean, arc distance in miles and kilometres), k-nearest neighbors (number of neighbors), binary distance band (cut-off point) and inverse distance (power and cut-off point).
- (b) Open an existing spatial weights matrix: ArcGIS (dbf, swm, txt), dat, gal, geoBUGS, gwt, kwt, MatLab, MatrixMarket, STATA text files.
- (c) Properties for selected weights: name, transform (B: binary, R: row-standardization, which is the default, D: double standardizations, V: variance stabilizing, O: restore original transformation), islands, neighbors of each observation, cardinalities, ids, histogram and neighborhood map viewer.
- (4) Kernel spatial weights matrix menu:
 - (a) Creation of a kernel weights matrix:
 - (i) Input map file (shp).
 - (ii) Select/add an ID variable for the weights file.
 - (iii) Adaptive kernels: distance metrics (Euclidean, arc distance in miles and kilometres), kernel function (uniform, triangular, Epanechnikov or quadratic, quartic or bisquare, Gaussian), number of neighbors.
 - (b) Open an existing spatial weights matrix: ArcGIS (dbf, swm, txt), dat, gal, geoBUGS, gwt, kwt, MatLab, MatrixMarket, STATA text files.
 - (c) Properties for selected weights: name, transform (B: binary, R: row-standardization, which is the default, D: double standardizations, V: variance stabilizing, O: restore original transformation), islands, neighbors of each observation, cardinalities, ids, histogram and neighborhood map viewer.
- (5) Model specification menú:
 - (a) Dependent variable box.
 - (b) Independent variables box.
 - (c) Endogenous variables (regressors) box.
 - (d) Instrumental variables box.
 - (e) Regime indicators (dummy) variable box.
- (6) Model estimation:
 - (a) Model type: standard, spatial lag, spatial error, spatial lag+error.
 - (b) Method: OLS (2SLS), spatial GMM (and ML in a near future).
 - (c) Standard errors: White, HAC, Kelejian-Prucha HET.

In the following sections, some applications are highlighted, focusing on some distinctive features of this software: spatial weight matrices, specification of the model and estimation methods.

3. Spatial weight matrix

One of the major distinguishing characteristics of spatial data analysis (as opposed to a mere non-spatial analysis) is that the spatial arrangement of the observations is taken into account. This is formally expressed in a spatial weights matrix, \mathbf{W} , with elements w_{ij} , where the ij index corresponds to each observation pair. The conceptual idea of spatial weights is that in this $n \times n$ matrix the diagonal elements (w_{ii}) are set to zero by definition and the rest of the non-zero cells (w_{ij}) capture the potential of spatial interaction. That is to say, for a spatial data set composed of n

locations (points, areal units, network edges, etc.), the spatial weights matrix expresses the potential for interaction between observations at each pair ij of locations. There is a rich variety of ways to specify the structure of these weights, and *GeoDaSpace* supports the creation, manipulation and analysis of spatial weights matrices across three different general types (see <http://pythonhosted.org/PySAL/users/tutorials/weights.html>):

- Contiguity based weights: ij locations interact when sharing a common border.
- Distance based weights: ij locations interact when being within a critical distance band.
- Kernel weights: only ij locations within a critical distance band interact, though following a distance-decay function (triangular, uniform, quadratic, etc.)

For the moment, the first two types (contiguity and distance based) are the only ones that can be used –in *GeoDaSpace*– for the specification of spatial models (spatial lag, spatial error, spatial lag+error) and the LM tests. The kernel spatial weights matrices can be used in the specification of the spatial HAC covariance matrix.

Spatial weight matrices tend to be fairly sparse (i.e. many cells contain zeros) and hence a full $n \times n$ array would not be an efficient representation, especially when dealing with large datasets because of computer memory constraints. For this reason, *GeoDaSpace* increases the computation speed by storing the spatial weights matrix in a compact format such that the spatial neighbors of each observation (all them identified by its ID) are only listed with their corresponding weight values. For example, in the case of a simple contiguity weight matrix if observation '03' has three neighbors sharing a common border (observations '01', '02' and '04') their corresponding weight value is 0.33. Instead of including the complete set of zero-one values for all the observations present in the system, the compact spatial weight matrix only lists the following sentence for observation '03': {'01': 0.33, '02': 0.33, '04': 0.33}

3.1. Contiguity based spatial weights matrix

When working with polygonal spatial data, contiguity is defined as having a common border. In the specific case of data that are arranged on a regular square or rectangular lattice (or grid), the contiguity structure can be defined in three ways: having a common border (rook criterion), having a common corner (bishop criterion), and having either a border or corner in common (queen criterion). Typically, this results in four contiguous grids according to the rook or bishop criteria, and eight contiguous grids following the queen criterion. *GeoDaSpace* only includes the rook and queen options, since the bishop matrix can be computed as the difference between the queen and the rook cases.

To illustrate this issue, we construct a contiguity matrix for the set of seven districts located in downtown Madrid (Spain). In the model weights menu of *GeoDaSpace*,

we must select the name of the map file where the districts polygons are stored, as well as the option for the weights matrix: contiguity – queen. Finally, the familiar Windows dialog requests the file name of the GAL (Geographic Algorithm Library) weights matrix: ALMONDIS.gal. And the name of this file will appear in the model weights menu box. In the weights properties editor menu, it is possible to obtain (and edit) some useful information about this selected spatial weights matrix, as shown in Figure 2.

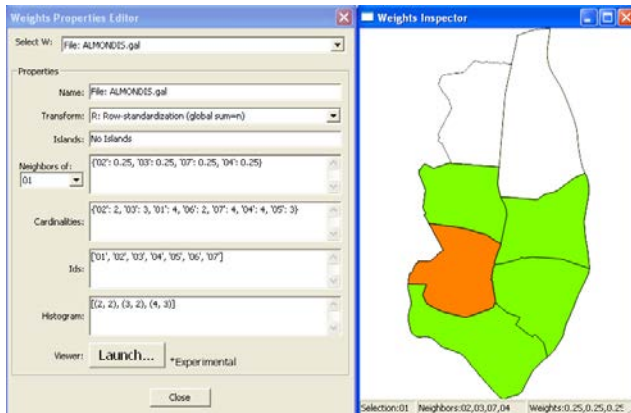


Figure 2. The spatial weights properties menu with the neighborhood map viewer.

Besides the name of the file and ID variable, the menu shows the type of transformation of the weight matrix, which is by default R: row–standardization (global sum=n). The row–standardized of the weight matrix is necessary to yield a meaningful interpretation of the results of many spatial statistics, and particularly for the estimation of spatial regression models, since they are all built with spatial weight matrices. The row standardization consists of dividing each element in a row by the corresponding row sum. Each element in the new matrix thus becomes:

$$w_{ij}^* = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} \quad . \quad (1)$$

Next for each observation, the menu indicates its corresponding neighbors. Since there are 7 districts in downtown Madrid, we could be thinking that the weights are stored in a 7×7 matrix. Nevertheless due to the sparseness of these matrices, the contiguity relations are stored in a compact form and only the set of non-zero elements (and weights values) of each row are shown. In our example, district ‘01’ has 4 neighboring districts with either a border or corner in common (queen criterion): ‘02’, ‘03’, ‘04’ and ‘07’, so as $w_{ij}=0.25$ (one fourth each). This interaction between district ‘01’ and its 4 neighbors can be visualized by the neighbourhood map viewer

(Figure 1).

It is also possible to have information about the existence of islands (unconnected observations) and the cardinality of the neighbour relations; i.e. the number of neighbors of each district: {'02': 2, '03': 3, '01': 4, '06': 2, '07': 4, '04': 4, '05': 3}. In this case, there are no islands since all the districts have at least one neighbour (in fact, the less connected districts are '02' and '06' with two neighbors). Finally, there is also a connectivity histogram attribute, which is a set of tuples indicating the cardinality of the neighbor relations. In this case: [(2, 2), (3, 2), (4, 3)], which means that there are 2 districts with 2 neighbors ('02' and '06'), another 2 districts with 3 neighbors ('03' and '05'), and 3 districts with 4 neighbors ('01', '04' and '07'). This histogram is very important to detect strange features of this distribution, which may affect spatial autocorrelation statistics and spatial regression specifications. Two features in particular warrant some attention. One is the occurrence of islands, the other a bimodal distribution, with some locations having very few (such as one) and others very many neighbors (Anselin, 2005).

3.2. Distance based spatial weights matrix

Interactions or neighborhood between spatial units can also be defined in function of the distance that separates them using information on latitude (X coordinate) and longitude (Y coordinate)². Two units are then considered to be neighbors if these points are less than a specified critical distance apart. In practice, when working with polygonal spatial units, this distance is computed between their centroids or other meaningful points (e.g. mean centers, capital cities, etc.). *GeoDaSpace* provides two distance metrics: Euclidean distance (based on the straight line) and Arc Distance on a sphere or great circle distance (in miles or kilometres), which is the appropriate measure when the spatial scale of the data analysis is global. The arc distance d_{ij} between two locations i and j is calculated as follows:

$$d_{ij} = 3959.0 \cdot \arccos \left[\cos |Y_i - Y_j| \cdot \sin X_i \cdot \sin X_j + \cos X_i \cdot \cos X_j \right], \quad (2)$$

where the X and Y are first transformed to radians, as $X = (90 \cdot \text{LAT} + \pi) / 180$ and $Y = \text{LON} + \pi / 180$.

In *GeoDaSpace*, three types of distance based spatial weights matrices can be constructed:

- (1) Distance band weights: it yields a simple contiguity matrix by using a critical distance cut-off point; i.e. the neighbour set for each spatial unit is defined as those units falling within a threshold distance (or distance band) of the focal

² The coordinates must be stored as a decimal number. Therefore, if you start from information on the coordinates in degrees you must first convert it to decimal format with the following linear combination: decimal = degrees × 1 + minutes × 0.01666667 + seconds × 0.00027778.

unit. The software proposes the minimum nearest neighbour distance, which is given by the pair of units that are the furthest apart, to assure of at least one neighbour per unit. In this case, the number of neighbors is likely to vary across observations.

- (2) K-nearest neighbors (knn) weights: it also yields a simple contiguity matrix by considering as neighbors to a given unit a set of k locations (1, 2, 3, etc.) previously defined. It is more appropriate when the minimum nearest neighbour distance is driven by two pairs of points which are significantly distant each other. In these cases, the minimum distance will not be representative for the rest of the distribution, but the knn weight matrix will assure the same number of neighbors for all the observations.
- (3) Inverse distance weights: it uses any integer power of the inverse distance between two observations as the weights. This matrix can be full (i.e. computed for the whole set of units) or it can also be specified as a distance band weights matrix taking on continuous values (rather than binary, as in the previous case), with the values set to the inverse distance separating each pair within a given threshold distance. This can be used to construct measures of potential interaction between two observations as squared inverse distance weights, in accordance with the gravity model of spatial interaction: $w_{ij} = 1/d_{ij}^2$

3.3. Kernel spatial weights matrix

Distance band weighting schemes suffer from the problem of discontinuity over the study area: it seems unnatural that the spatial association between the units ends so abruptly farther apart from a specific cut-off point. In order to solve this problem, it is possible to specify a w_{ij} as a continuous and monotone decreasing function of d_{ij} . The kernel functions or kernels has been suggested for constructing this kind of weights, in which a constant value (h) provides some control of the range of the circle of influence of each observation i (Chasco et al., 2008). They could be considered as a combination of distance based thresholds together with continuously valued weights. The distance threshold is here equivalent to the maximum distance of a k -nearest neighbour distance for each observation, which constitutes what is called the ‘bandwidth’ attribute: $h = \max(d_{knn}), \forall i$ where d_{knn} is a vector of the k -nearest neighbor distances (the distance to the k^{th} nearest neighbor for each observation). The weights of the kernel functions are, then, defined as $w_{ij} = -d_{ij}/h$. As it can be seen, h produces a decay of influence with distance.

In this case, the bandwidth (or threshold distance) h is fixed across observations and one optimum spatial kernel is determined and applied uniformly across the study area. Such approach, however, suffer from the potential problem that in some parts

of the region, where data are sparse, the local regressions might be based on relatively few data points. To offset this problem, spatially adaptive weighting functions can be incorporated in order to define different bandwidths (distances), expressing the number or proportion of observations to retain within the weighting kernel “window”, irrespective of distance: on the one hand, relatively small bandwidths in areas where the data points are densely distributed and on the other hand, relatively large bandwidths where the data points are sparsely distributed. In other words, they are able to adapt themselves in size to variations in the density of the data so that the kernels have larger bandwidths where the data are sparse and have smaller ones where the data are plentiful. In these cases, the bandwidth is adaptive in size.

GeoDaSpace specifies adaptive bandwidths from a predefined number of k -nearest neighbour observations (this number of neighbors is possible to be endogenously changed by the user). In this case, the weights of the kernel functions are defined as $w_{ij} = d_{ij}/h_i$, being h_i the adaptive bandwidth for observation i , which produces a decay of influence with distance. The form of the kernel function determines the distance decay in the derived continuous weights. The program includes the following kernel functions:

- (1) Uniform (the default in *GeoDaSpace*): $K(w) = |w|$ if $|w| \leq 1$, and 0 otherwise
- (2) Triangular: $K(w) = 1 - |w|$ if $|w| \leq 1$, and 0 otherwise
- (3) Quadratic or Epanechnikov: $K(w) = 3/4 \cdot (1 - w^2)$ if $|w| \leq 1$, and 0 otherwise
- (4) Quartic or bisquare: $K(w) = 15/16 \cdot (1 - w^2)^2$ if $|w| \leq 1$, and 0 otherwise
- (5) Gaussian: $K(w) = 2\pi^{-1/2} \cdot \exp(-w^2/2)$ if $|w| \leq 1$, and 0 otherwise

If i and j coincide, the weighting of data at that point will be the unity: $w_{ii} = 1$. The weighting of other data will decrease according to the kernel curve as the distance between i, j increases. For data a long way from i , the weighting will fall to virtually zero. Changing the bandwidth (e.g. by a change in the number of nearest neighbors) results in a different exponential decay profile, which in turn produces weights that vary more or less rapidly over space.

4. Spatial regression

GeoDaSpace includes two menus for spatial regression, “specification” and “model estimation”, as well as the submenu “show advanced settings” in the main menu. The specification menu presents a set of boxes or windows in which insert different kind of variables: dependent, independent, endogenous, instrumental and spatial regimes (see Figure 1). The model estimation menu includes three boxes: one for the model type (spatial lag, spatial error and spatial lag+error), other for the estimation method (OLS/2SLS, spatial GMM and ML in a near future), and the last

one for different methods to deal with both heteroskedasticity and spatial autocorrelation in the residuals (White, HAC and Kelejian-Prucha HET).

These menus focus on the implementation of tests for spatial autocorrelation in models that may include endogenous variables and on general method of moments (GMM) estimation methods. The suite of routines includes implementation of Kelejian and Prucha's recent techniques to deal with both spatial autocorrelation and heteroskedasticity, including a heteroskedastic autocorrelation consistent (HAC) estimator.

4.1. *Non-spatial models: estimation methods and diagnosis*

The traditional basic econometric model is specified as a linear relationship between a dependent variable (y) and a set of explanatory variables (\mathbf{X}) as follows:

$$y = \mathbf{X}\beta + \varepsilon \text{ for } \varepsilon \sim \mathbf{N}(0, \sigma^2 \mathbf{I}) , \quad (3)$$

where y is the dependent variable (in vector form, with N rows), \mathbf{X} is a matrix with observations on K explanatory variables (with N rows and K columns), β is a vector with K regression coefficients (i.e., of dimension K by 1), ε is a random error term (in vector form, with N rows), σ^2 is the population error variance, and \mathbf{I} is an identity matrix of dimension N by N .

4.1.1. Ordinary least squares (OLS) estimation of the basic model

According to a number of criteria, the method of ordinary least squares (OLS) estimation accomplishes the main objectives of finding a good match or fit between predicted values $\mathbf{X}\hat{\beta}$ and observed values of the dependent variable and determining the variables that explains it significantly in the linear relationship. The OLS estimators ($\hat{\beta}$), which are found by minimizing the sum of the squared prediction errors, are BLUE (Best Linear Unbiased Estimators). In order to achieve these good properties in the estimators, we must make certain assumptions about the random error:

- The random error has mean zero (i.e., there is no systematic misspecification or bias in the population regression equation: $E(\varepsilon_i) = 0, \forall i$.
- The random error terms are uncorrelated and have a constant variance (homoskedastic): $E(\varepsilon_i \cdot \varepsilon_j) = 0, \forall i \neq j$ and $E(\varepsilon_i \cdot \varepsilon_i) = \sigma^2, \forall i = j$
- The random error term follows a normal distribution: $\mathbf{N}(0, \sigma^2)$

The first two assumptions are crucial to obtain the unbiasedness and efficiency of the OLS estimates while the third one is needed in order to carry out hypothesis tests and to assess significance of the regression coefficients. These assumptions introduce an additional parameter to be estimated (in addition to the regression

coefficients β , i.e., the error variance σ^2 . *GeoDaSpace* reports both an unbiased and a maximum likelihood estimate for σ^2 , as well as their square root (the standard deviation for the error term). It also reports some measures of fit like the R^2 , adjusted R^2 and some measures based on the maximum likelihood (ML) estimation method (log-likelihood, Akaike and Schwartz information criterion).

We can practice with *GeoDaSpace* estimating by OLS a linear model of a production function for a sample of 1,171 European NUT3 regions, based on the application presented in Chasco et al. (2012), in which the ratio of the Gross Domestic Product or GDP per area for 2006, in logarithms (LG06) is explained by a set of “first nature” geographical variables, such as Southerly latitude (FSLAT), Westerly latitude (FSWEST), elevation over the sea level (FSEA) and a dummy variable for the presence (or not) of mineral extraction sites (miner), as well as two second nature (man-made agglomeration) variables (also in logarithms): population (PI06) and productivity or GDP per employee (DEL06). As a benchmark, we begin with the OLS estimation of a base model which must be performed as shown in Figure 3: after opening the dBase file of the map (NUTS306.DBF), we must drag the variables into their corresponding box in the specification menu of *GeoDaSpace*.

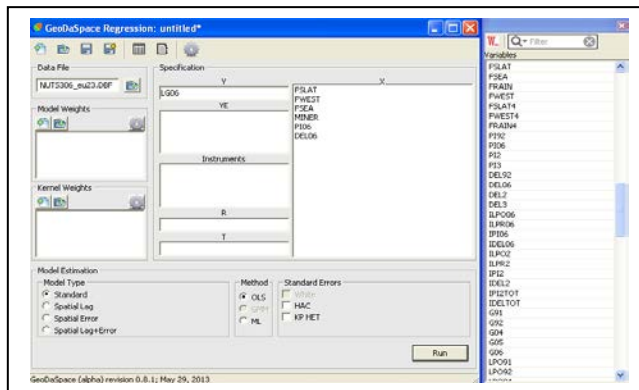
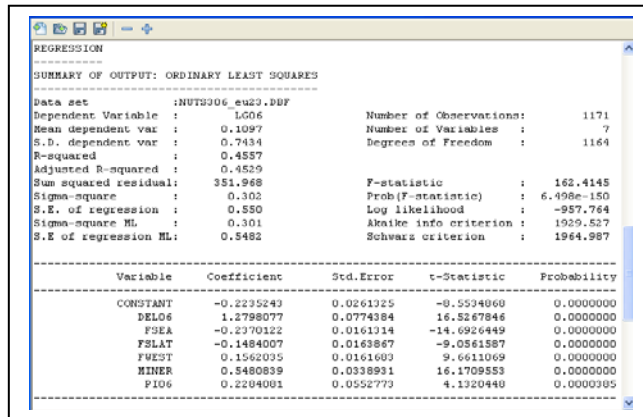


Figure 3. Specification of a basic linear model in order to be estimated by OLS.

Pressing the button “Run”, we can easily obtain a full summary of the results of an OLS regression in an output file nicely formatted and ready to be printed (Figure 4). We obtain a very significant relationship between each explanatory variable and the log of GDP per area as pointed out by the high values (in absolute terms) of their corresponding t-statistic results (and low p-values). Regarding the measures of fit, the R^2 indicates that only a 45.57% of the dependent variable variance is captured by these regressors.



REGRESSION

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES

Data set : NUTS2006_eu23.DBF

Dependent Variable : LG06 Number of Observations: 1171

Mean dependent var : 0.1097 Number of Variables : 7

S.D. dependent var : 0.7434 Degrees of Freedom : 1164

R-squared : 0.4537

Adjusted R-squared : 0.4529

Sum squared residual: 351.968 F-statistic : 162.4145

Sigma-squared : 0.302 Prob(F-statistic) : 6.498e-150

S.E. of regression : 0.550 Log likelihood : -957.764

Sigma-squared ML : 0.301 Akaike info criterion : 1929.527

S.E. of regression ML: 0.5482 Schwarz criterion : 1964.987

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	-0.2235243	0.0261325	-8.5594068	0.0000000
REL06	1.2798077	0.0774384	16.5267846	0.0000000
F2EA	-0.2370122	0.0161114	-14.6926449	0.0000000
FSLAT	-0.1484007	0.0163867	-9.0561587	0.0000000
FWEST	0.1562035	0.0161603	9.6611069	0.0000000
MINER	0.5480839	0.0338931	16.1709553	0.0000000
PI06	0.2204001	0.0552773	4.1320448	0.0000385

Figure 4. First part of the OLS estimation results of a basic linear model

4.1.2. Specification diagnostics of the OLS estimation of the basic model

GeoDaSpace contains some statistics to test for the existence of potential misspecification problems in the OLS estimation of spatial linear regression models, such as multicollinearity, non-normality, heteroskedasticity and spatial autocorrelation.

- (1) Multicollinearity takes place with the existence of a strong linear relation (i.e. correlation) between the explanatory variables included in the regression specification, which in principle, should be totally uncorrelated. As a consequence, the OLS estimates will have very large estimated variances and, therefore, very few coefficients will be found to be significant (low t statistics), even though the regression as a whole may seem to achieve a reasonable fit (high R^2). *GeoDaSpace* includes the condition number that, as a rule of thumb, should take values lower than 20 or 30 (a total lack of multicollinearity yields a condition number of 1).
- (2) Non-normality in the error terms is the basis for most hypothesis tests and a large number of regression diagnostics. The *GeoDaSpace* regression output includes the results of the Jarque and Bera test (the statistic and its associated probability): a low probability indicates a rejection of the null hypothesis of a normal error.
- (3) Heteroskedasticity is a common situation in cross-section regression models where the random regression error does not have a constant variance over all observations. As a consequence, while the OLS estimates are still unbiased, they will no longer be most efficient and, more importantly, inference based on

the usual t and F statistics will be misleading, and the R^2 measure of goodness-of-fit will be wrong. In spatial data analysis, this problem is frequently encountered when using data for irregular spatial units (with different area) or when there are systematic regional differences in the relationships you model (i.e., spatial regimes). Hence, an indication of heteroskedasticity may point to the need for a more explicit incorporation of spatial effects, in the form, for instance, of spatial regimes. *GeoDaSpace* includes some tests against heteroskedasticity, in which the null hypothesis is always homoskedasticity: the Breusch-Pagan Lagrange Multiplier (LM) test, which is not powerful for non-normal errors in small samples), its studentized version, the Koenker-Basset test (the best option when dealing with non-normal errors), and the White test, which is robust to any unspecified form of heteroskedasticity.

- (4) Spatial autocorrelation, or more generally, spatial dependence, is the situation where the dependent variable or error term at each location is correlated with observations on the dependent variable or values for the error term at other locations. The consequences of ignoring spatial autocorrelation in a regression model, when it is in fact present, depend on the form for the alternative hypothesis: either a spatial lag model (OLS estimates are biased and inference will be incorrect, as for the omitted variables problem) or a spatial error model (OLS estimates are unbiased but no longer efficient, as for heteroskedasticity). *GeoDaSpace* contains six tests for spatial dependence, three of which pertain to the spatial error case: an extension of the Moran's I for the residuals, a simple LM test against the presence of a spatial error model and its robust version. There are also two LM tests against the presence of a spatial lag model (simple and robust version) and a last LM test against the joint presence of both a spatial lag and a spatial error model (SARMA test).

```

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          2.950472

TEST ON NORMALITY OF ERRORS
TEST
Jarque-Bera                                DF      VALUE      PROB
                                           2      29.017344  0.000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST
Breusch-Pagan test                          DF      VALUE      PROB
                                           6      155.879796 0.000000
Koenker-Bassett test                        6      142.450274  0.000000

SPECIFICATION ROBUST TEST
TEST
White                                       DF      VALUE      PROB
                                           26     240.774017 0.000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST
Moran's I (error)                          MI/DF   VALUE      PROB
                                           0.3688 17.827113  0.000000
Lagrange Multiplier (lag)                  1      305.662691 0.000000
Robust LM (lag)                             1      17.498271  0.0000288
Lagrange Multiplier (error)                1      306.499516 0.000000
Robust LM (error)                          1      18.335097  0.0000185
Lagrange Multiplier (SARMA)                2      323.997787 0.000000

===== END OF REPORT =====

```

Figure 5. Diagnosis of the OLS estimation results of a basic linear model

In Figure 5, we present the specification diagnosis for the OLS estimation of the GDP per area model, which has no multicollinearity problems. However the residuals are clearly non-normal, heteroskedastic and exhibit high degree of spatial autocorrelation in both spatial lag and spatial error forms (though the slightly higher value of the robust LM test-error seems to point out to a spatial error model). Anyway, the high non-normality of the error terms could put the LM tests under suspect, though the sample size is not very low in this model. In order to compute the LM tests on spatial autocorrelation we have first created and specified –in the “model weights” menu of *GeoDaSpace*– a contiguity (queen) spatial weights matrix.

4.1.3. Two-stage least squares estimation (2SLS) of the basic model with endogenous regressors.

Other misspecification problem for the OLS estimation in linear regression models is the non-deterministic nature of one or more explanatory variables, which will be endogenous (or stochastic regressors), as follows:

$$y = X\beta + \gamma Y + \varepsilon, \quad (4)$$

where Y is a matrix with observations on G endogenous explanatory variables (with N rows and G columns) and γ is a vector with G regression coefficients (of dimension G by 1).

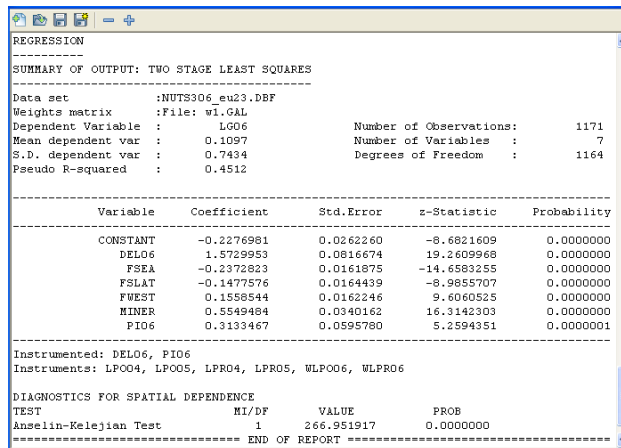
The inclusion of endogenous variables on the regressors in systems of simultaneous equations invalidates the OLS estimations, which will be biased. The instrumental variables estimator (IV) or two-stage least squares estimator (2SLS) is consistent,

but not necessarily very efficient. The principle of instrumental variables estimation is based on the existence of a set of instruments (Q) that are strongly correlated with the original endogenous explanatory variables (Y), but asymptotically uncorrelated with the error term. Once these instruments are identified, they are used to construct a proxy for the endogenous variables (Y), which consists of their predicted values (\hat{Y}) in a regression on the instruments (Q) and the exogenous variables (X). This proxy (\hat{Y}) is then used in a standard least squares regression. The instruments (Q) are the other "excluded" exogenous variables or the time-lag of the exogenous variables, since they are contemporaneously uncorrelated with the error terms.

Besides the 2SLS estimation results and measures of fit, GeoDaSpace also reports the Anselin-Kelejian test (Anselin and Kelejian 1997), which allows us to test for the presence of remaining spatial autocorrelation in the residuals of the 2SLS estimation. This is a version of the classical Lagrange Multiplier error test adapted for the case of residuals from a 2SLS regression. It is computed as the Moran's I statistic for the residuals from a 2SLS estimation, which is distributed as a χ_1^2 .

In the previously shown example for the EU regions, one problem is that the second-nature variables, population (PI06) and productivity (DEL06) are endogenous and determined simultaneously with GDP leading to a simultaneity bias in the OLS estimators. For this reason, we propose using as instruments both time-lagged variables (LPO04, LPO05, LPR04, LPR05) and spatial-lagged variables (WLPO06, WLPR06), since they are highly correlated with the stochastic regressors but also asymptotically uncorrelated with the error terms.

The 2SLS estimators are quite similar to the OLS ones except for the stochastic regressors that are higher in value. The analysis of the residuals with the Anselin-Kelejian test does not allow us to accept the null of spatial autocorrelation, as in the OLS case.



```

REGRESSION
-----
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES
-----
Data set          :NUTS306_eu23.DBF
Weights matrix   :File: w1.GAL
Dependent Variable :      LG06          Number of Observations:      1171
Mean dependent var :      0.1097       Number of Variables      :      7
S.D. dependent var :      0.7494       Degrees of Freedom      :     1164
Pseudo R-squared  :      0.4512
-----
Variable      Coefficient      Std.Error      z-Statistic      Probability
-----
CONSTANT      -0.2276981      0.0262260      -8.6821609      0.0000000
DELO6         1.5729953      0.0816674      19.2609968      0.0000000
FSEA         -0.2372823      0.0161875      -14.6583255      0.0000000
FSLAT        -0.1477576      0.0164439      -8.9855707      0.0000000
FWEST         0.1558544      0.0162246      9.6060525      0.0000000
MINER         0.5549484      0.0340162      16.3142303      0.0000000
PIO6          0.3133467      0.0595780      5.2594351      0.0000001
-----
Instrumented: DELO6, PIO6
Instruments: LPOO4, LPOO5, LPRO4, LPRO5, WLPOO6, WLPRO6

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST      HI/DF      VALUE      PROB
Anselin-Kelejian Test      1      266.951917      0.0000000
=====
END OF REPORT

```

Figure 6. 2SLS estimation and diagnosis of a basic linear model with endogenous regressors

4.1.4. OLS/2SLS plus spatial HAC of the basic model without/with endogenous regressors.

When spatial heteroskedasticity and autocorrelation in the error terms are very significant in a model, it is possible to implement a robust inference on the estimated OLS or 2SLS residuals. It is similar to the well-known adjusted White variance, which consists of estimating the covariance matrix of OLS parameter estimators, from the covariance matrix of the heteroskedastic error terms, in order to perform an inference which is robust to this problem. Kelejian and Prucha (2007) develop a spatial HAC estimator for a situation in which the error terms are not only heteroskedastic but spatially autocorrelated. They model spatial dependence in terms of a spatial weighting matrix. In spatial HAC estimation literature, an economic distance is commonly employed to characterize the decaying pattern of the spatial dependence.

The covariance of random variables at locations i and j is a function of d_{ij} , the economic distance between them; as the economic distance increases, the covariance decreases in absolute value and vice versa. The existence of such an economic distance enables *GeoDaSpace* to use the (adaptive) kernel method for the standard error estimation, which is a weighted sum of sample covariances with weights depending on the relative distances, that is: $w_{ij} = -d_{ij}/h_i$ for some bandwidth parameter h_i .

This way the robust inference based on the 2SLS estimates presented in Figure 6 presented by *GeoDaSpace* leads to the results in Figure 7. Though the 2SLS coefficients are the same, their estimated standard errors are different and robust to the already demonstrated existence of both heteroskedasticity and spatial autocorrelation. The HAC standard errors have been computed with the help of an

REGRESSION
SUMMARY OF OUTPUT: TWO STAGE LEAST SQUARES

Data set : NUTS306_eu23.DBF
Dependent Variable : LG06
Mean dependent var : 0.1097
S.D. dependent var : 0.7434
Pseudo R-squared : 0.4512
Number of Observations: 1171
Number of Variables : 7
Degrees of Freedom : 1164

HAC Standard Errors: Kernel Weights: File: kernelgauss.kwt

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	-0.2276981	0.0267493	-8.5123044	0.0000000
DELO6	1.5729953	0.1095743	14.3555072	0.0000000
FSEA	-0.2378283	0.0291398	-8.1429045	0.0000000
FSLAT	-0.1977576	0.0240972	-6.1317197	0.0000000
FWEST	0.1558544	0.0279095	5.5842728	0.0000000
MINER	0.5549484	0.0365592	15.1794332	0.0000000
PIO6	0.3133467	0.0803179	3.9013321	0.0000957

Instrumented: DELO6, PIO6
Instruments: LPO04, LPO05, LPR04, LPR05, WLPO06, WLPR06
===== END OF REPORT =====

adaptive Gaussian kernel function of the distances between the EU NUTS3 region centroids, for 11 observations (kernelgauss.kwt file).

Figure 7. 2SLS plus spatial HAC estimation of a basic linear model with endogenous regressors

4.2. Spatial models

4.2.1. Spatial two-stage least squares (S2SLS) estimation of the spatial lag model without/with endogenous regressors

The spatial lag model or mixed regressive spatial autoregressive model includes a spatial lagged dependent variable, Wy , as one of the explanatory variables:

$$y = \rho Wy + X\beta + \varepsilon, \quad (5)$$

where Wy is a N by 1 vector of spatial lags for the dependent variable, ρ is the spatial autoregressive coefficient, and ε is a N by 1 vector of normally distributed random error terms, with means 0 and constant (homoskedastic) variances σ^2 .

The presence of the spatial lag is similar to the inclusion of endogenous variables on the RHS in systems of simultaneous equations. For this reason, this model is therefore often referred to as the simultaneous spatial autoregressive model. Typically the ρ coefficient is unknown and must be estimated jointly with the regression coefficients (β). The main consequence of the inclusion of Wy on the RHS of the specification is that OLS no longer achieves consistency. This is similar

to what happens in systems of simultaneous equations. Instead of OLS, estimation must be based on an instrumental variables approach, proposed by Anselin (1988), leading to what is called spatial two-stage least squares (S2SLS), or an explicit maximization of the likelihood function when the error terms are normally distributed (this method will be implemented in *GeoDaSpace* soon).

The S2SLS estimation method consists of the construction of a proper instrument for the spatial lag variable. As in 2SLS, the resulting estimate ($\hat{\rho}$) is consistent, but not necessarily very efficient. In the spatial lag model a number of suggestions have been formulated for the choice of the best instruments for the ρ parameter. Kelejian and Robinson (1993) showed that the proper set could be a series of spatially lagged exogenous variables, for first order and higher order contiguity matrices (though in practice, this series may be truncated and only the first order spatially lagged explanatory variables may be included).

By the inclusion of W_y in addition to other explanatory variables, the spatial lag specification can be interpreted in two different ways. First it can be a way to assess the degree of spatial dependence, while controlling for the effect of these other variables; hence, the main interest is in the spatial effect. Alternatively, it allows assessing the significance of the other (non-spatial) variables, after the spatial dependence is controlled for. In this second case, it is important to note that the interpretation –and inference– of the coefficients of the exogenous variables in a spatial lag model is not straightforward, as they do not correspond to marginal effects (LeSage and Pace 2009 for a further review).

In the application for the EU regions, we introduce the spatial lag of the endogenous variable (GDP per area) with the aim to assess the significance of the first and second nature variables controlling by this variable (Figure 8).

```

REGRESSION
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
-----
Data set      : NUTS306_eu23.DBF
Weights matrix :File: w1.GAL
Dependent Variable :   LO06      Number of Observations:   1171
Mean dependent var :   0.1097      Number of Variables :       8
S.D. dependent var :   0.7434      Degree of Freedom :   1163
Pseudo R-squared :   0.5363
Spatial Pseudo R-squared : 0.4562
-----
Variable      Coefficient      Std. Error      t-Statistic      Probability
-----
CONSTANT      -0.2027911      0.0255074      -7.9502923      0.0000000
DEL06         1.2085524      0.1178137      10.3272084      0.0000000
FSEA         -0.1050440      0.0233413      -4.5286195      0.0000000
FSLAT        -0.1176348      0.0183246      -6.4189399      0.0000000
FWEST         0.1159315      0.0202632      5.7156519      0.0000000
MINER         0.5124206      0.0344036      14.8946316      0.0000000
FID6          0.2713452      0.0559021      4.7690036      0.0000019
W_LO06        0.2213418      0.0763753      2.8980824      0.0037545
-----
Instrumented: DEL06, FID6, W_LO06
Instruments: LPO06, LPO05, LPRO4, LPRO5, WLPO06, WLPRO6, W_FSEA, W_FSLAT,
             W_FWEST, W_MINER
-----
DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEXT      RI/RF      VALUE      PROB
-----
Anselin-Kelejian Test      1      12.691554      0.0003673
-----
***** END OF REPORT *****

```

Figure 8. S2SLS estimation of a spatial lag model with endogenous regressors

Besides the previously defined instrumental variables for the second nature variables (population and productivity), the S2SLS method needs to include the spatially lagged exogenous variables as new instruments for the spatial autoregressive parameter (ρ): W_FSEA, W_FSLAT, W_FWEST and W_MINER). As we can see in the output window of GeoDaSpace, though the autoregressive coefficient seems to be relevant to GDP per area, it cannot completely capture the spatial autocorrelation effect in the residuals, since the Anselin-Kelejian test is still very significant. This result confirms the outcome of the LM spatial autocorrelation tests in the OLS estimation of the basic model, which pointed out to a joint spatial lag and spatial error model (Figure 5).

4.2.2. S2SLS plus GM/GMM/spatial HAC estimation of the spatial error model without/with endogenous regressor

The spatial error model is a special case of a so-called non-spherical error model, i.e., a regression specification for which the assumptions of homoskedastic (constant variance) and/or uncorrelated errors are not satisfied. The spatial dependence in the error term can take on a number of different forms. In the current version of *GeoDaSpace*, only a spatial autoregressive process for the error term can be estimated. This model is the standard regression specification with a spatial autoregressive error term:

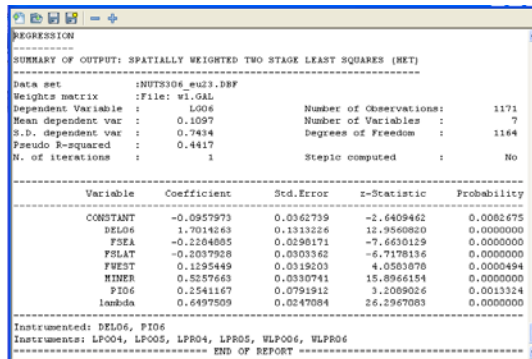
$$y = X\beta + u \text{ for } u = \lambda Wu + \varepsilon, \quad (6)$$

where $W\varepsilon$ is the spatial lag of the errors, λ is the autoregressive parameter and ε is a "well-behaved" error, with mean 0 and variance matrix σ^2 .

The consideration of non-spatial endogenous explanatory variables (Y) is model implies the estimation of a new group of γ coefficients, leading to the following expression:

$$y = X\beta + \gamma Y + u \text{ for } u = \lambda Wu + \varepsilon, \quad (7)$$

The principles of the implementation of general methods of moments (GMM) estimation of the spatial error model were originally presented in Kelejian and Prucha (1998, 1999), and more recently generalized in a series of papers by Kelejian and Prucha (2010), Arraiz et al. (2010) and Drukker et al. (2011), jointly referred to in what follows as K-P-D. The estimation strategy outlined by K-P-D consists of two major steps. The first one has to do with the estimation of the model coefficients using a feasible generalized least squares approach (Spatially Weighted Least Squares), which is a kind of spatial Cochrane-Orcutt transformation, in order to obtain an consistent (though not still efficient) spatial autoregressive parameter λ . The second step deals with obtaining an efficient estimate for λ (see Anselin et al., 2012 for technical details). This is the method employed by GeoDaSpace (Figure 9) in order to estimate a spatial error model with endogenous explanatory variables by S2SLS and GMM (GS2SLS). Additionally, we have estimated a robust inference of the estimator covariance matrix in presence of both spatial heteroskedasticity and autocorrelation (KP-HET), proposed by Kelejian and Prucha (2010).



```

REGRESSION
-----
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-----
Data set      : NUTS306_eu23.DBF
Weights matrix : File: WL06L
Dependent Variable : LOG6           Number of Observations: 1171
Mean dependent var : 0.1097         Number of Variables : 7
S.D. dependent var : 0.7434         Degree of Freedom : 1164
Pseudo R-squared : 0.4417
N. of iterations : 1                Stepic computed : No
-----
Variable      Coefficient      Std. Error      z-Statistic      Probability
-----
CONSTANT      -0.0957973       0.0362739      -2.6409462       0.0082675
DELO6         1.7014263       0.1313226     12.9560820       0.0000000
FSEA          -0.2204685       0.0298171      -7.6630129       0.0000000
FSLAT         -0.2037928       0.0303362     -6.7178136       0.0000000
FRET          0.1235449       0.0319203       4.0503978       0.0000494
MINER         0.5257663       0.0330741     15.8966154       0.0000000
PID6          0.2541167       0.0791912       3.2089026       0.0013324
lambda        0.6497509       0.0247084     26.2967083       0.0000000
-----
Instrumented: DELO6, PID6
Instruments: LPO04, LPO05, LPRO5, WLPRO6, WLPRO6
-----
END OF REPORT

```

Figure 9. S2SLS estimation of a spatial error model with endogenous regressors

4.2.3. S2SLS plus GM/GMM/spatial HAC estimation of the joint spatial lag and spatial error model without/with endogenous regressors

The joint spatial lag and spatial error model has the following specification:

$$y = \rho W y + X \beta + u \quad \text{for } u = \lambda W u + \varepsilon, \quad (8)$$

and the inclusion of additional endogenous explanatory variables leads to the following model:

$$y = \rho W y + X \beta + \gamma Y + u \quad \text{for } u = \lambda W u + \varepsilon, \quad (9)$$

The estimation strategy outlined by K-P-D consists of three major steps (see Anselin, 2011 for technical details). The first one has to do with the estimation of the model coefficients using a feasible generalized least squares approach (Spatially Weighted Least Squares), which is a kind of spatial Cochrane-Orcutt transformation, in order to obtain an consistent (though not still efficient) spatial autoregressive parameter λ . The second step deals with obtaining an efficient estimate for λ by means of a weighting matrix that is necessary to obtain the optimal (consistent and efficient) GMM estimate of λ in the second iteration. A third step consists of estimating the regression coefficients (β and ρ) in a spatially weighted regression, using filtered variables that incorporate the optimal GMM estimate of λ .

Again, it is also possible to perform a robust inference of the coefficient covariance matrix in presence of both spatial heteroskedasticity and autocorrelation (KP-HET), proposed by Kelejian and Prucha (2010).

As shown in Figure 10, both ρ and λ autoregressive parameters are very significant (especially λ), as well as the rest of coefficients, which do not experience great changes with respect to previous estimations.


```

REGRESSION
-----
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (NET)
-----
Data set      :NUTS306_eu23.DBF
Weights matrix :File: W1.GAL
Dependent Variable :   LQ06           Number of Observations:   1171
Mean dependent var :   0.1097           Number of Variables :     8
S.D. dependent var :   0.7454           Degrees of Freedom :   1163
Pseudo R-squared :   0.5320
Spatial Pseudo R-squared: 0.4371
N. of iterations :   1                   Step(s) computed :     No
-----
Variable      Coefficient      Std.Error      z-Statistic      Probability
-----
CONSTANT      -0.1354982       0.0316159      -4.2857565       0.0000182
DELO6         1.4521218       0.1418977      10.2335851       0.0000000
FSEA          -0.1691917       0.0280889      -6.0234381       0.0000000
FSLAT         -0.1582139       0.0206079      -5.5150053       0.0000000
FWEST         0.0788477       0.0303645       2.5967109       0.0094121
MINER         0.5145669       0.0295140      17.4346961       0.0000000
PIO6          0.2817859       0.0789549       3.5889470       0.0003584
W_LQ06        0.2182126       0.1020184       2.1389523       0.0324295
lambda        0.6014930       0.0466184      12.9024844       0.0000000
-----
Instrumented: DELO6, PIO6, W_LQ06
Instruments: LPO04, LPO05, LPRO4, LPRO5, WLPO06, WLPRO6, W_FSEA, W_FSLAT,
              W_FWEST, W_MINER
-----
END OF REPORT
  
```

Figure 10. S2SLS estimation of a spatial lag+spatial error model with endogenous regressors

4.3. Spatial regimes model

This model is suitable in certain instances in which the assumption of a fixed relation between the explanatory variables and the dependent variable across the complete system is not tenable. Instead, heterogeneity may be present, in the form of different intercepts and/or slopes in the regression equation for subsets of the data. This is often referred to as structural instability or structural change in the econometric literature and may be expressed in the form of switching regression models. When the different subsets in the data correspond to regions or spatial clusters, it is called spatial regimes model (Anselin, 1988).

In *GeoDaSpace*, models of spatial regimes are implemented by jointly estimating the coefficients for all the predefined spatial regimes. An augmented matrix of observations on the explanatory variables is constructed, of dimension N by MK (with M as the number of regimes), by transforming each explanatory variable into as many new variables as there are regimes. The new variables are zero for all observations that do not fall in the regime to which they correspond.

In all other respects, a model with spatial regimes is treated as a regression model, allowing the full range of estimation methods (OLS, ML, 2SLS, S2SLS, GS2SLS, spatial HAC, KP-HET, etc.). The only condition is the introduction of the variable name for a categorical indicator variable, with the spatial regimes information, in the corresponding box of the *GeoDaSpace* “specification” menu. This indicator variable may only take on consecutive integer values, with each value corresponding to a regime. The software internally allocates the observations to their proper regimes.

SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET) - REGIMES

 Data set :NUTS306_eu23.DBF
 Weights matrix :File: wl.GAL
 Dependent Variable : LG06
 Number of Observations: 1171

Mean dependent var :	0.1097	Number of Variables :	
29		Degrees of Freedom :	
S.D. dependent var :	0.7434		
1142			
Pseudo R-squared :	0.6870		
Spatial Pseudo R-squared:	0.6366		
N. of iterations :	1	Stepic computed :	
No			

Probability	Variable	Coefficient	Std.Error	z-Statistic

	1_CONSTANT	-0.0954984	0.0364761	-2.6181079
0.0088419				
	1_DELO6	1.3371885	0.2381016	5.6160407
0.0000000				
	1_FSEA	-0.0394074	0.0332461	-1.1853216
0.2358903				
	1_FSLAT	-0.2013935	0.0357169	-5.6386084
0.0000000				
	1_FWEST	-0.0697590	0.0360135	-1.9370222
0.0527426				
	1_MINER	0.5596100	0.0402790	13.8933440
0.0000000				
	1_PI06	0.0606050	0.0925870	0.6545737
0.5127423				
	2_CONSTANT	-0.3040614	0.2073275	-1.4665754
0.1424916				
	2_DELO6	0.5396302	0.6255001	0.8627179
0.3882926				
	2_FSEA	0.2171961	0.1323955	1.6405097
0.1008992				
	2_FSLAT	0.6562792	0.2620386	2.5045139
0.0122620				
	2_FWEST	0.1224703	0.0720441	1.6999357
0.0891430				
	2_MINER	0.4242440	0.0855304	4.9601547
0.0000007				
	2_PI06	0.9594303	0.2892873	3.3165312
0.0009114				
	3_CONSTANT	-0.2510336	0.0675991	-3.7135645
0.0002044				
	3_DELO6	1.9819971	0.2293848	8.6404889
0.0000000				
	3_FSEA	0.0087520	0.0701995	0.1246733
0.9007822				
	3_FSLAT	-0.0284556	0.0833487	-0.3414043
0.7327992				
	3_FWEST	-0.0973907	0.0488206	-1.9948706
0.0460570				
	3_MINER	0.5264621	0.0703571	7.4827095
0.0000000				
	3_PI06	0.5961976	0.1806479	3.3003303
0.0009657				
	4_CONSTANT	-0.3607690	0.0542661	-6.6481476
0.0000000				
Rect@	4_DELO6	0.1109832	0.1593945	0.6962800
0.4862535				
		-----	-----	-----

Figure 10. S2SLS estimation of a spatial lag+spatial error model with endogenous regressor and spatial regimes

GeoDaSpace also implement a test on the stability of the regression coefficients over the regimes. This is a spatial Chow test on the null hypothesis that the coefficients are the same in all regimes: $\beta^1 = \beta^2 = \dots = \beta^M$. In the case of the spatial models, only one autoregressive parameter (ρ and/or λ) is estimated for the whole model. This test is implemented for all coefficients jointly, as well as for each coefficient separately.

Figure 10 shows the complete set of estimates of a joint spatial lag and spatial error spatial regimes model of the log of GDP per area, with two endogenous non-spatial explanatory variables (population and productivity) and KP-HET robust inference with heteroskedasticity and spatial autocorrelation. The spatial regimes are 4 groups of European regions located in the center-core of the EU (regime #1), Northern periphery (regime # 2), Eastern periphery (regime # 3) and Southern periphery (regime # 4). The spatial Chow test is very significant both in its global version and for each of the coefficients separately, which justifies the consideration of the regimes in order to control for spatial heterogeneity in this spatial system.

The spatial autoregressive coefficients are also very significant. Regarding the coefficients for the second nature variables, population (PI06) and productivity (DEL06), they are also very significant in the four regimes, unless they suffer important changes in size. Whereas the population coefficients are specially high in the Northern and Southern peripheries, the productivity ones are higher in the center-core and Eastern periphery, revealing different patterns of man-made agglomeration effects in the location of production in the EU.

5. Conclusions

The expansion of spatial analysis and particularly, spatial econometrics has been outstanding in all branches of social sciences and economics. Accordingly, spatial regression techniques are now becoming an established component in the applied econometrics toolbox and even, in standard econometrics textbooks. For this reason, this methods are also being part of the econometric and statistics courses in economics, business administration, marketing, environmental studies, epidemiology and many other sciences.

GeoDaSpace is a recent software packages that has revealed as very useful for both teachers and students in the classroom, since it is an easy-to-use point a click program, which is freely downloadable from the GeoDa Center. It is still an alpha release which is constantly incorporating new items. The new announced methods for the near future are the implementation of Maximum Likelihood for the

estimation of the spatial models and the estimation and diagnosis of spatial panel data models.

Acknowledgements

The author acknowledges financial support from the Spanish Ministry of Economics and Competitiveness (Grant No. ECO2012-36032-C03-01), Project UAM-Santander and Xunta de Galicia (Grant No. 10SEC201032PR).

References

1. L. Anselin, R.J.G.M. Florax and S. Rey, Econometrics for spatial models: Recent advances, in *Advances in spatial econometrics methodology tools and applications*, eds L. Anselin, R.J.G.M. Florax and S. Rey (Springer Verlag, New York, 2004), pp 1-25.
2. G. Maddala, "Econometrics" (McGraw-Hill, New York, 2001).
3. G. Woolridge, "Econometric analysis of cross section and panel data" (MIT Press, Cambridge, MA, 2002).
4. D. Gujarati, "Basic Econometrics", fourth edition (McGraw-Hill, New York, 2003).
5. P. Kennedy P, "A guide to econometrics", fifth edition (Blackwell Publishers, Oxford, 2003).
6. B. Baltagi, "Econometric analysis of panel data" (John Wiley and Sons, Chichester, 2008).
7. S. Rey and L. Anselin, PySAL: A Python library of spatial analytical methods, in *Handbook of applied spatial analysis: Software tools, methods and applications*, eds. M. Fischer and A. Getis (Springer, Berlin, 2010).
8. L. Anselin, From SpaceStat to CyberGIS. Twenty Years of Spatial Data Analysis Software, *International Regional Science Review* **35** (2012) 131-157.
9. D. M. Drukker, P. Egger and I. Prucha, On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressor, *Econometric Reviews* **32** (2013) 686-733.
10. J. LeSage and R.K. Pace, "Introduction to spatial econometrics" (Chapman and Hall/CRC, Boca Raton, FL., 2009).
11. R. Bivand, Spatial econometrics functions in R: Classes and methods, *Journal of Geographical Systems* **4** (2002) 405-421.
12. G. Piras, Spatial models with heteroskedastic innovations in R, *Journal of Statistical Software* **35** (2010) 1-21.
13. L. Anselin, "Exploring Spatial Data with GeoDaTM: A Workbook" (University of Illinois at Urbana-Champaign, 2005).
14. C. Chasco, J. Vicéns and I. García, Modeling spatial variations in household disposable income with geographically weighted regression, *Estadística Española* **50** (2008) 321-360.
15. C. Chasco, A. López and R. Guillain, The influence of geography on the spatial agglomeration of production in the European Union, *Spatial Economic Analysis* **7** (2012) 247-263.
16. L. Anselin and H. H. Kelejian, Testing for spatial error autocorrelation in the presence of endogenous regressor, *Int. Regional Sci. Rev.* **20** (1997) 153-182.
17. H. H. Kelejian and I. R. Prucha, HAC estimation in a spatial framework, *J. Econometrics* **140** (2007) 131-154.
18. L. Anselin, "Spatial econometrics: Methods and models" (Kluwer Academic Press, Dordrecht, 1988).

19. H. H. Kelejian and D. Robinson, A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model, *Pap. Reg. Sci.* **72** (1993) 297-312.
20. H. H. Kelejian and I. R. Prucha, A generalized spatial two-stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances, *J. Real Estate Financ.* **17** (1998) 99-121.
21. H. H. Kelejian and I. R. Prucha, A generalized moments estimator for the autoregressive parameter in a spatial model, *Int. Econ. Rev.* **40** (1999) 509-533.
22. H. H. Kelejian and I. R. Prucha, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *J. Econometrics* **157** (2010), 53-67.
23. I. Arraiz, D. M. Drukker, H. H. Kelejian and I. R. Prucha, A spatial Cliff-Ord type model with heteroskedastic innovations: Small and large sample results, *J. Regional Sci.* **50** (2010) 592-614.
24. D. M. Drukker, I. R. Prucha and R. Raciborski, "A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables", Technical Report (Stata Corp, College Station, TX, 2011).
25. L. Anselin, "GMM estimation of spatial error autocorrelation with and without heteroskedasticity", Note (GeoDa Center, Arizona State University, 2011).
26. L. Anselin, P. V. Amaral and D. Arribas-Bel, "Technical aspects of implementing GMM estimation of the spatial error model in PySAL and GeoDaSpace", Working Paper 2/12 (GeoDa Center, Arizona State University, 2012).

APLICACIÓN DE LA HERRAMIENTA FUZZY LOGIC AL PROCESO DE SELECCIÓN DE UN CANDIDATO POLÍTICO

M^a CARMEN LOZANO GUTIÉRREZ

carmen.lozano@upct.es

*Universidad Politécnica de Cartagena/ Departamento de Economía Financiera y Contabilidad
C/ Real nº 3. C.P. 30201 (Cartagena-Murcia)*

FEDERICO FUENTES MARTÍN

federico.fuentes@upct.es

*Universidad Politécnica de Cartagena/ Departamento de Economía
C/ Real nº 3. C.P. 30201 (Cartagena-Murcia)*

RESUMEN: En el presente artículo diseñamos un modelo difuso de clasificación de candidatos a un puesto de responsabilidad política, a partir de la valoración cualitativa de las habilidades sociales y cualidades o atributos psicológicos propios de la llamada “*Inteligencia Emocional*”, presentes en cada candidato, y su comparativa respecto a las que son consideradas como deseables. Para la resolución del modelo hemos utilizado la herramienta Fuzzy Logic Toolbox de Matlab.

Palabras claves: política; sistemas difusos; inteligencia emocional; liderazgo.

ABSTRACT: The present article tries to design a fuzzy model in order of choosing the best candidates to a position of political responsibility. The starting point is the qualitative valuation of the social abilities and qualities or psychological attributes typical of the “*Emotional Intelligence*” and proper to each candidate. Right after we proceed to compare all data selecting those candidates that are considered as desirables. For the development of the model we have used the Fuzzy Logic Toolbox of Matlab.

Keywords: politics; fuzzy systems; emotional intelligence; leadership

CLASIFICACIÓN JEL: C63; Z00

1. La valoración de algunos aspectos de inteligencia emocional en la elección de un candidato político

Siempre ha sido motivo de debate y reflexión la elección de las virtudes de los buenos gobernantes y políticos en general que garanticen el mejor desempeño posible de las tareas del gobierno aunque en las democracias modernas se está constatando que las virtudes de un buen candidato no son necesariamente las mismas que las de un buen gobernante pues la capacidad de comunicación persuasiva, determinante para el éxito de una campaña, no es suficiente para el manejo competente del gobierno y de la administración de un país.

En la sociedad moderna, deseosa de alguien que la conduzca, el líder está obligado a ser un constante agente de cambio por lo que se requiere de una nueva generación de políticos, capaces de afrontar los problemas que se presentan en el país, inspirar una visión compartida y canalizar positivamente la inconformidad, así como alentar a una sociedad para que se implique en el proceso de desarrollo.

Cuando una persona tiene la capacidad de movilizar o inspirar a la gente para que alcance ciertos objetivos de una manera satisfactoria para el grupo de personas al que representa, se puede decir que tiene liderazgo político. Este tipo de liderazgo, para ser efectivo, debe cumplir con varias características, las cuales son obligatorias, es decir que, si no se cumplen, nadie puede ser considerado un buen líder político. La primera de ellas es sin duda la honestidad.

En el presente artículo nos centraremos en las habilidades sociales y cualidades o atributos psicológicos deseables (aunque no de obligatoria presencia) en el candidato político encuadradas en lo que podríamos denominar "*Inteligencia Emocional*". El concepto de Inteligencia Emocional, aunque esté de actualidad, tiene a nuestro parecer un claro precursor en el concepto de Inteligencia Social del psicólogo Edward Thorndike (1920) quien la definió como "*la habilidad para comprender y dirigir a los hombres y mujeres, muchachos y muchachas, y actuar sabiamente en las relaciones humanas*". Hoy, pocas personas de los ambientes culturales, académicos o empresariales ignoran el término o su significado. Esto se debe, fundamentalmente, al trabajo de Daniel Goleman, investigador y periodista del *New York Times*, quien llevó el tema al centro de la atención en todo el mundo, a través de su obra '*La Inteligencia Emocional*' (1995).

El concepto de la inteligencia emocional postula ampliar la noción de inteligencia (tradicionalmente ceñida a una serie de habilidades racionales y lógicas) incorporando una serie de habilidades emocionales (Mestre, J.M; Fernández Berrocal, P., 2007). Para Goleman la Inteligencia Emocional es una forma de interactuar con el mundo que tiene

muy en cuenta los sentimientos, y engloba habilidades tales como el control de los impulsos, la autoconciencia, la motivación, el entusiasmo, la perseverancia, la empatía, la agilidad mental, etcétera. De todos esos factores elegiremos para nuestro estudio aquellos que con mayor importancia (Maxwell, John, 2000) deben presentarse en un candidato político:

1. Entusiasmo. El entusiasmo con el que el político se debe enfrentar a su labor implica una primaria inteligencia emocional. Se considera que es una cualidad deseable en un candidato que debe transmitir a su equipo (partido) y a su electorado (Linz, J. J, 1997). Habilidad que se puede "entrenar" como, de hecho, hacen los actores que son capaces de autogenerarse el estado emocional más apropiado para representar un papel.
2. Empatía. Es el nombre que recibe la aptitud para reconocer las emociones en los demás. Proviene del griego *empathia* "sentir dentro", es decir, percibir lo que el otro siente en su interior. Los sentimientos no suelen expresarse verbalmente sino a través del tono de voz, los gestos, las miradas, etc. Un político que se precie debe saber "*ponerse en la piel del otro*" y que su electorado así lo aprecie.
3. Liderazgo. Se puede definir el liderazgo como el conjunto de actividades y de relaciones y comunicaciones interpersonales, que permiten a una persona ejercer diversos niveles de influencia sobre el comportamiento de los miembros de un grupo determinado, consiguiendo que este grupo defina y alcance de manera voluntaria y eficaz sus objetivos (Delgado Fernández, S. 2004).
4. Resistencia al estrés. Una de las cualidades deseables en un candidato político es la de ofrecer una alta resistencia al estrés, si se tiene en cuenta que deberá someterse a debates, mociones, críticas...y ruedas de prensa que deberá afrontar con serenidad y mente fría.
5. Dependencia emocional-afectiva. Reconocer nuestras emociones consiste en aprender a controlarlas; de modo análogo, una instancia ulterior a la empatía estriba en manipular las emociones de los demás. El candidato político debe "*atraer*" a su electorado creando un clima afectivo adecuado pero a su vez debe aprender a controlar su dependencia emocional cuando de tomar decisiones se trata (Sterrett, E. 2002).

Estas cinco variables han sido incorporadas como variables de entrada en el sistema difuso¹ del módulo Fuzzy Logic Toolbox del programa Matlab. Esta herramienta permite elegir el método de inferencia, el método de defusificación, definir los

¹ Las cinco variables de entrada se presentan en los conjuntos de clasificación difusos y sus respectivas funciones de pertenencia (membership function).

conjuntos de entrada y salida e incorporar reglas que definen el comportamiento del control. Con el fin de mostrar al lector el funcionamiento del algoritmo y las salidas del mismo que servirían de conclusión a su aplicación, hemos hecho una simulación basándonos en las valoraciones obtenidas de cuatro hipotéticos candidatos de entre los cuáles debe salir elegido el candidato ideal. A continuación describimos la metodología empleada.

2. Metodología

En la elección de un candidato político para ocupar la presidencia de un partido político, resulta especialmente significativa la complejidad en la que se han de tomar las decisiones relativas a la puesta en práctica de determinadas líneas de actuación, basadas fundamentalmente en previsiones de comportamientos de voto que a veces resultan anti-intuitivos y carentes en muchas ocasiones de pautas repetitivas, lo que impide al decisor el poder hacer uso de la experiencia e invalida la efectividad de la intuición (Little, J. 2009), razón por la cual el empleo de la lógica difusa resulta un candidato natural para su representación en ordenadores (Quintana, M.A. 2003), al permitirnos captar la realidad con toda su imprecisión. La existencia de no linealidad en las relaciones entre variables es otro de los problemas que nos podemos encontrar al diseñar un modelo de decisión, ya que nos impide analizar dichas variables de un modo aislado sino como un todo, de forma que la alteración de una de ellas puede provocar la variación del comportamiento conjunto, es por ello que un enfoque sistémico del sistema no sólo puede ser interesante sino casi obligado ya que es el único modo capaz de reflejar las interdependencias que existen en un sistema, revelando la existencia de una gran variedad de decisiones alternativas y susceptibles de ser aplicadas ante un determinado suceso. El gran potencial de la dinámica de sistemas se encuentra en que se trata de un modelamiento dinámico (Sugianto, L. 2001), es decir, que se centra en el tiempo, permitiendo observar las consecuencias que pueden producirse a corto, medio y largo plazo de las decisiones adoptadas (Forrester, J.W. 1961).

El método empieza con la identificación de la situación problemática y el propósito del modelo, y termina con el diseño de una solución o la implantación de cambios, es decir, con la decisión. Los pasos a seguir para el establecimiento del modelo, son los siguientes (Vennix, 1996):

1.- Identificación del problema y propósito del modelo, lo que supone la identificación de variables de entrada y salida, y determinación de conjuntos difusos. Esta primera fase es fundamental y determinante, ya que si identificamos correctamente todos los aspectos implicados lograremos elaborar un modelo dinámico que sea representativo del comportamiento real del sistema, permitiendo de ese modo, la simulación de distintas políticas alternativas y en consecuencia, conocer cuál de ellas resulta más idónea para su

posterior aplicación.

2.- Conceptualización del sistema, en la que se seleccionará un método para fusificación y defusificación. Deberán especificarse cuáles son los límites del sistema, es decir, discernir qué factores se van a incluir en el modelo y cuáles se excluirán, seleccionando aquellos que, intuitivamente parecen ser los más significativos (Senge, P.1995).

3.- Formulación del modelo y estimación de parámetros, creando una base de conocimiento mediante la utilización de reglas del tipo “*Si_entonces_*”

4.- Análisis del comportamiento del modelo mediante el diseño de un mecanismo de inferencia

5.- Evaluación del modelo

6.- Uso del modelo o implantación y realimentación de la información. Hay que señalar que esta definición no hace referencia exclusivamente a decisiones conscientes o explícitas sino también a decisiones inconscientes o implícitas. La utilidad que se desprende del estudio de la realimentación informativa es la posibilidad de comprender de qué modo el volumen de las acciones correctivas y de las demoras en el tiempo entre acción y efecto en los componentes interconectados de un sistema pueden originar comportamientos fluctuantes o inestables. El modelo se acaba perfeccionando mediante un proceso de aprendizaje que tiene lugar tanto durante la realización del modelo como en posteriores simulaciones. Mediante la simulación ponemos al modelo “*en movimiento*”, es decir que en un contexto virtual reproducimos el comportamiento esperado de la organización o problema a fin de observar el impacto en los indicadores relevantes de decisiones, políticas, cambios de escenarios etcétera, y permite ver en un breve lapso los efectos en el tiempo y espacio de decisiones. Hoy en día disponemos de herramientas informáticas que han facilitado la posibilidad de desarrollar modelos y simularlos (Shachmurove, Y. 2002).

7.- El último paso, consiste en aplicar las políticas consideradas óptimas en la fase anterior.

Hemos seleccionado cinco variables representativas de aspectos que deben ser tenidos en cuenta a la hora de identificar y clasificar el perfil de los candidatos políticos y tomar una decisión de elección de entre ellos. Dichas variables consideradas son: empatía, entusiasmo, liderazgo, resistencia al estrés y dependencia emocional-afectiva. En fase experimental del modelo se han utilizado cuatro candidatos con puntuaciones de 0 a 1 en cuanto a la presencia de las variables consideradas en su carácter o personalidad. El sistema difuso quedaría pues formado por 5 entradas (variables) y 1 salida (la elección del candidato ideal). Para definir los conjuntos de entrada con el control se utiliza la ventana de edición Membership Function Editor (Gulley, N. Jang, J. 2000). En esta ventana se elige el número de conjuntos y el tipo de distribución (triangular, trapezoidal), en el renglón de parámetros se definen los puntos que forman el conjunto.

En las figuras 1,2,3,4,5 se muestra el editor de funciones de pertenencia². A la hora de determinar una función de pertenencia, podíamos optar por funciones lineales (triangular y trapezoidal) o no lineales (funciones gamma, gaussiana, sigmoial, pseudo exponencial); hemos optado por elegir una función lineal sencilla triangular que nos ha parecido la adecuada por dos motivos principalmente: para que los cálculos no sean complicados, así como por su estructura lógica a la hora de definir un valor lingüístico asociado.

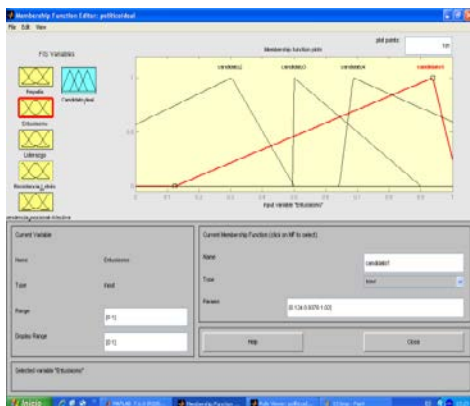


Figura 1: valoración entusiasmo en los candidatos

Figura 2: valoración empatía en los candidatos

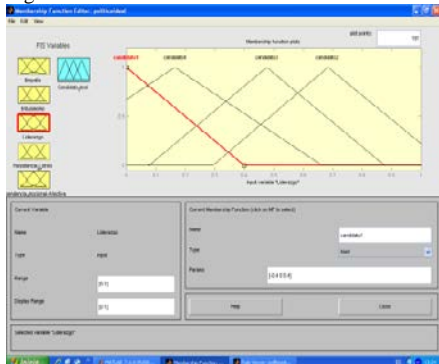


Figura 3: valoración liderazgo en los candidatos

Figura 4: valoración resistencia al estrés en los candidatos

² Los datos utilizados no se han obtenido de una aplicación real del modelo sino que responden a una simulación ofrecida al lector con fines ilustrativos para facilitar su comprensión.

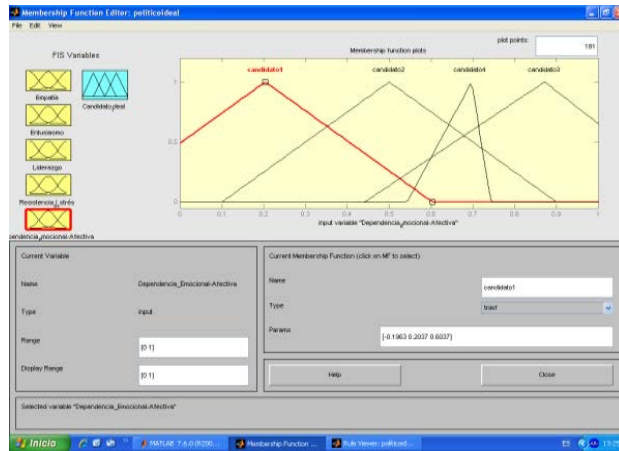


Figura 5: valoración de la dependencia emocional-afectiva de los candidatos

Para programar el controlador hay que definir las reglas que regirán su comportamiento y para ello se emplea el “Rule Editor”. Las reglas establecidas para la clasificación de candidatos de acuerdo con el grado de cumplimiento de los parámetros actitudinales elegidos por un grupo de expertos se muestran en la figura 6. Cada regla se forma combinando los conjuntos de las variables con el conector “and” u “or” relacionándolos con los conjuntos de salida.

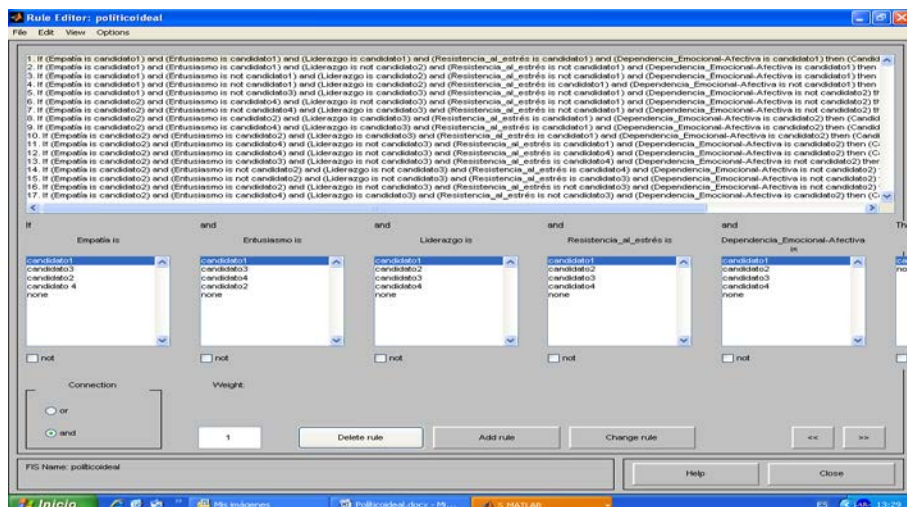


Figura 6: bases de la relación entre las variables de entrada y salida (estructura de reglas)

A continuación (figura 7) mostramos una ventana de salida del módulo Fuzzy de Matlab en la que se presenta la situación de cada uno de los candidatos respecto a la media y al perfil marcado como ideal. Se observan las variables de entrada así como los conjuntos que las conforman en color amarillo. Las líneas rojas representan los valores de entrada que se ingresan al controlador y se pueden observar los grados de pertenencia a los que corresponden dentro del conjunto difuso. Las variables de salida con sus conjuntos difusos en color azul muestran las reglas que se cumplieron y en la parte inferior de la columna se puede apreciar el polígono generado en la inferencia. El valor de salida que produce la defusificación es un valor “crisp” que se puede apreciar en la parte superior de la columna.

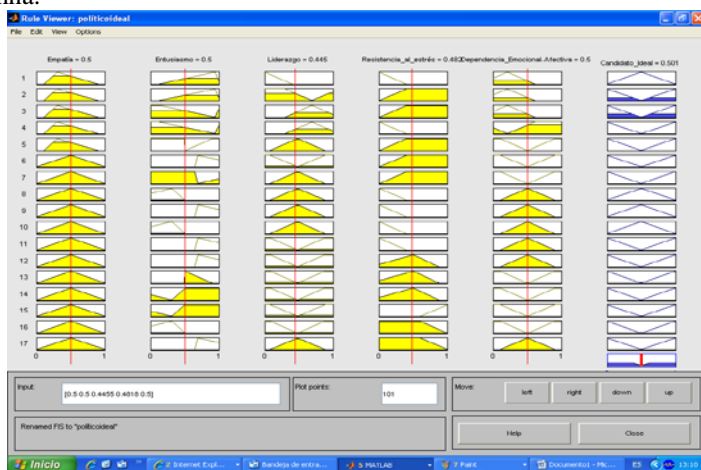


Figura 7: situación de cada candidato respecto a las valoraciones señaladas como perfil ideal

El realizar la determinación de los parámetros de entrada en el modelo dinámico con la ayuda de algoritmos de clasificación difusa facilita y enriquece el análisis en la medida que entrega resultados continuos para la categorización de los candidatos permitiendo establecer los grados de mayor o menor acercamiento de una categoría a otra y al perfil de candidato político ideal.

La relación entre las variables dos a dos es representada gráficamente a continuación a través de la superficie difusa que surge de los niveles de correlación existentes entre las mismas, la cual nos será de utilidad para analizar el grupo de candidatos en el que se concentra la mayor presencia de alguna/s de las variables. (Figura 8, 9, 10, 11, 12)

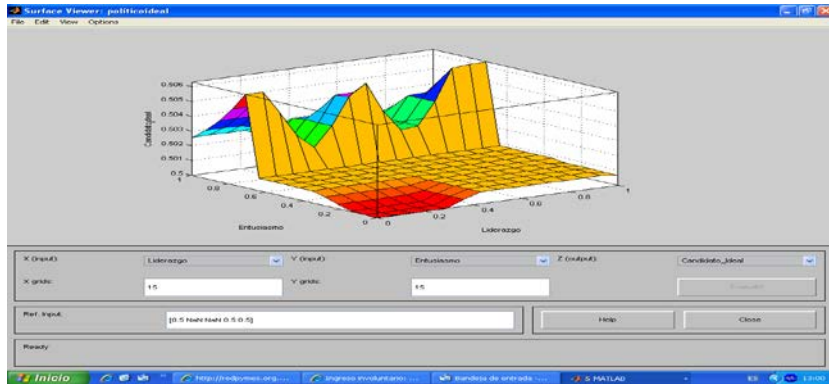


Figura 8: liderazgo-Entusiasmo

Según el [Diccionario](#) de la [Lengua](#) Española (1986), [liderazgo](#) se define como la [dirección](#), jefatura o conducción de un partido político, de un [grupo](#) social o de otra colectividad. El Diccionario de [Ciencias](#) de la [Conducta](#) (1956), lo define como las "cualidades de [personalidad](#) y capacidad que favorecen la guía y el [control](#) de otros individuos". Para algunos especialistas en Inteligencia Emocional, el liderazgo se define como sinónimo de administración y lo relacionan estrictamente con la motivación, con el esfuerzo voluntario y entusiasta en el cumplimiento de objetivos en equipos, con el ahínco, la pasión, formalidad e intensidad en la ejecución del trabajo. Observamos en la Figura 8 que la correlación Liderazgo-Entusiasmo resulta baja para la mayor parte de los candidatos y solo hay uno de ellos que presentan altos niveles de liderazgo y entusiasmo.

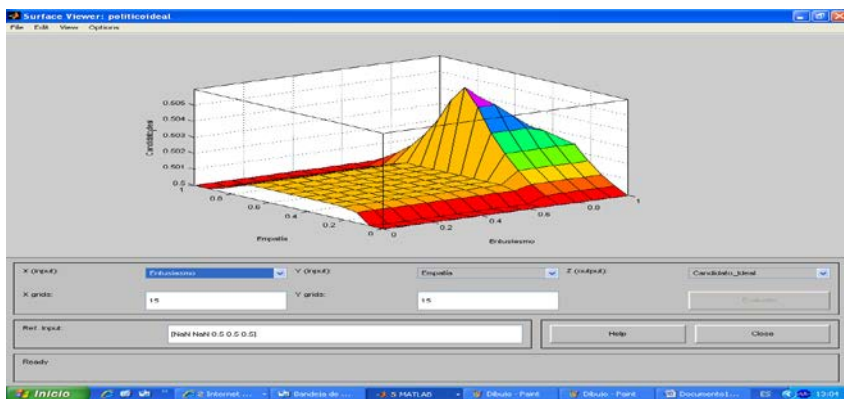


Figura 9: entusiasmo-Empatía

La empatía es un sentimiento de comprensión hacia un problema ajeno. Cada uno de nosotros influye en el estado de ánimo de los demás. Es perfectamente natural influir en el estado emocional de otra persona, para bien o para mal, lo hacemos constantemente,

“contagiándonos” las emociones. En los políticos se valora de forma muy positiva la capacidad de interrelación emocional sobre la capacitación técnica siempre que ésta vaya acompañada en una relación de equilibrio de un sentimiento de empatía. Los partidos políticos son “redes de participación” y un líder político debería inyectar a su partido y transmitir a su electorado entusiasmo y compromiso, dos cualidades que los candidatos deberían tener, pero que no se pueden imponer. Observamos que el grupo de candidatos seleccionados (figura 9) no manifiesta la presencia de correlaciones entre ambas. Sólo dos de ellos muestran cualidades de empatía cercanas a 1 aunque sus niveles de entusiasmo resultan claramente más bajos.

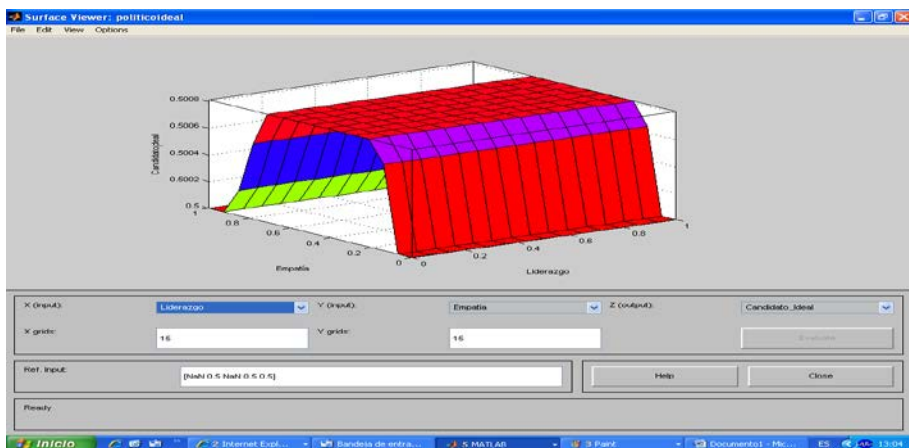


Figura 10: liderazgo-Empatía

Observamos en la figura 10 que para la totalidad de los candidatos el liderazgo guarda una clara correlación con la empatía.

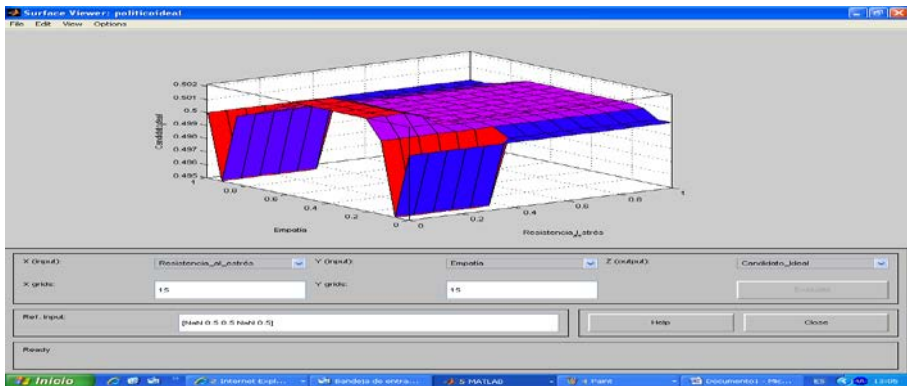


Figura 11: resistencia al estrés y Empatía

En la figura 11 se aprecia una correlación alta entre estas dos variables, y por tanto, la comprensión del problema ajeno permite aumentar los niveles de resistencia al estrés.

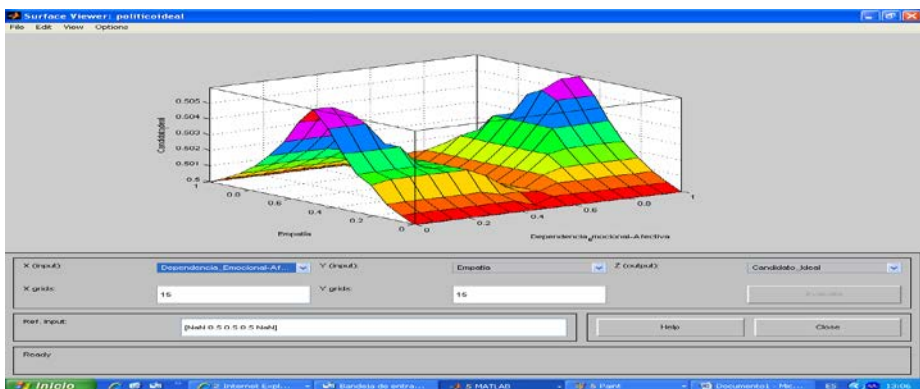


Figura 12: dependencia Emocional-Afectiva y Empatía

En esta figura se aprecia una división clara del grupo de candidatos, los más emotivos que presentan unos niveles más altos de empatía y por otro lado aquellos que procuran aislarse de los problemas ajenos (baja empatía) ya que consideran que la dependencia emocional-afectiva no es una cualidad para un político sino incluso un defecto.

Conclusiones

Con la ayuda de los algoritmos de clasificación difusa hemos podido enriquecer el análisis dado que nos han facilitado la categorización de los candidatos en función de la

distribución de grados de mayor o menor presencia o acercamiento al cumplimiento de los atributos psicológicos-personalidad que debería poseer (en opinión del decisor o los expertos) el candidato ideal llamado a liderar un partido político.

La elección de un candidato llamado a liderar un partido político o a ocupar un cargo representativo del mismo en el contexto actual resulta muy complejo. Adoptar criterios como la antigüedad en el partido, número de apoyos, cargos anteriores...etcétera, está demostrado que no garantiza el éxito en la elección. En el presente artículo se han barajado variables pertenecientes a la llamada "Inteligencia Emocional" que conforman las virtudes o cualidades deseables para un candidato líder, como la empatía y el entusiasmo en su labor y su carisma. Virtudes todas ellas que se echan de menos en el panorama político actual y que, de poseerlas un candidato, indudablemente despertaría en los ciudadanos una renovada ilusión, sin olvidar que éstas virtudes han de ir respaldadas de un proyecto político con calidad técnica y profesional y de un equipo de colaboradores de calidad que acompañen al candidato.

Referencias Bibliográficas

1. Barnabas, B., *Mathematics of Fuzzy Sets and Fuzzy Logic*. Ed. Springer (2013)
2. Blocher, E., *The role of emotional intelligence in political candidate image*. Ed. University of Kansas. Disponible online en: <http://search.proquest.com/docview/304859276>
3. Delgado Fernández, S., Sobre el concepto y el estudio del liderazgo político. *Psicología Política*, nº 29, pp.7-29 (2004)
4. Domínguez Machuca, J.A., Los modelos y juegos de Dinámica de Sistemas como instrumentos para la Dirección de Empresas. *Alta Dirección*, nº 147 y 150 (1990)
5. Forrester, J.W., *Industrial Dynamics*. The MIT Press Cambridge, Massachusetts. Traducción (1972): *Dinámica Industrial*. Ed. Ateneo. Buenos Aires. (1961)
6. Goleman, D., *Inteligencia Emocional*. Ed. Kairos.(1996)
7. Gulley, N. Jang, J., *Fuzzy Logic Toolbox User's Guide*, The Math Works Inc. Massachusetts. (2000)
8. Heifetz, R.A. y Linsky, M., *Pensar en clave política*, en Heifetz, R.A. y Linsky, M.(2003): *Liderazgo sin límites*. Ed. Paidós, p.p. 85.
9. Linz, J. J., *El liderazgo innovador*. Ed. Alcántara, M. y Martínez, A.(1997).
10. Little, J., Decision Support Systems for Marketing Managers. *Journal of Marketing*. Vol.43(1979)
11. Maxwell, J., *Las 21 cualidades indispensables de un líder*. Ed. Caribe-Betania (Thomson) (2000)
12. Mestre Navas, J.M., Fernández Berrocal, P. *Manual de Inteligencia Emocional*. Ed. Pirámide (2007)
13. Quintana, M.A., *Modelo híbrido para los procesos de Data Mining en el apoyo a la toma de decisiones basados en tecnologías inteligentes conexionistas y difusas*. Publicado en las actas del XVII Simposio Brasileiro de Redes Neurais. (2003)
14. Senge, P.: *La quinta disciplina*. Ed. Granica. (1995)

15. Shachmurove, Y., Applying Artificial Neural Networks to Business, *Economics and Finance*. University of Pennsylvania, CARESS Working Paper: 02/08(2002)
16. Sterrett, E., *Guía del directivo para la inteligencia emocional: de la gestión al liderazgo*. Ed. Centro de Estudios Ramón Areces. (2002)
17. Sugianto, L., Management of Data Uncertainty in Dynamic Programming. *Studies in Fuzziness and Soft Computing*, vol. 73. (2001)
18. Vennix, J., *Group Model Building: facilitating team learning using system dynamics*. Ed. John Wiley (1996)

COMPARACIÓN DE DIFERENTES DISTRIBUCIONES TRAPEZOIDALES, MEDIANTE SUS MOMENTOS Y DIAGRAMAS DE COCIENTE DE MOMENTOS

JOSÉ MANUEL HERRERIAS VELASCO

jmherrer@ugr.es

*Universidad de Granada / Departamento de Métodos Cuantitativos para la Economía y la Empresa
Facultad de Ciencias Económicas y Empresariales / Campus de Cartuja s/n / 18.071 - Granada*

RAFAEL HERRERIAS PLEGUEZUELO

rherreri@ugr.es

*Universidad de Granada / Departamento de Métodos Cuantitativos para la Economía y la Empresa
Facultad de Ciencias Económicas y Empresariales / Campus de Cartuja s/n / 18.071 - Granada*

RESUMEN: A partir de la distribución trapezoidal, usada como modelo probabilístico en muchos y variados problemas reales, que presentan una primera fase de crecimiento, seguida de una fase de estabilidad, para terminar con una tercera fase de amortiguamiento y utilizando las tres típicas estimaciones subjetivas del método PERT, sobre los valores mínimo, máximo y más probable del campo de variación de la variable objeto de estudio, se han determinado, de forma alternativa, tres distribuciones trapezoidales diferentes.

El objetivo principal de este trabajo, es realizar un análisis comparativo entre estas tres distribuciones, utilizando sus características estocásticas, para seleccionar la que mejor se ajuste al problema real planteado: predicción de la duración de un trabajo, selección de la inversión más rentable, valoración de un activo, etc.

Como herramienta complementaria para este análisis comparativo, se estudian los diagramas de cocientes de momentos de tales distribuciones, para determinar cuál de ellas es más flexible.

Debido a la complejidad en el cálculo de los coeficientes de variación de Pearson, de asimetría y curtosis, por una parte, y a las propias representaciones gráficas de los diagramas de los cocientes de momentos, por otra, ha sido imprescindible la implementación informática en la realización del análisis comparativo mencionado anteriormente.

Palabras claves: modelos probabilísticos trapezoidales, incertidumbre, método PERT, diagramas de cociente de momentos.

ABSTRACT: The trapezoidal distribution is used as probabilistic model in many and varied problems that present a first phase of growth, followed by a phase of stability and finished with a third phase of buffer. This paper introduces three different trapezoidal distributions that use the three typical subjective estimates of the PERT method, about the minimum, maximum, and most likely values of a variable.

The main objective of this paper is to make a comparative analysis between these three distributions, using its stochastic characteristics, to select the one that best fits to a particular

problem: prediction of the duration of work, selection of the most profitable investment, valuation of assets, etc.

Moment ratio diagrams are studied, as a complementary tool, to determine which of them enjoys greater flexibility.

The computer implementation has been essential in the comparative analysis, due to the complexity of the calculation of the coefficients of variation of Pearson, skewness and kurtosis and the own graphic representations of the moment ratio diagrams.

Keywords: trapezoidal probabilistic models, uncertainty, PERT method, moment ratio diagram.

1. Introducción

La distribución trapezoidal pertenece a un amplio grupo de distribuciones de probabilidad continuas univariantes, tales que la representación gráfica de su función de densidad es una figura geométrica determinada: rectángulo, triángulo, trapecio, parábola, etc., de la que se deriva la denominación de la propia distribución: rectangular, triangular, trapezoidal, parabólica, etc.

Esta distribución ha sido utilizada, como modelo probabilístico, en muy diferentes campos científicos: análisis de inversiones, Herrerías y Calvete (1987), valoración de activos, Herrerías et al. (2001), conjuntos borrosos, Kaufman y Gil Aluja (1986), control educativo, Akcayol et al. (2004), etc., y generalmente está relacionada con la incertidumbre de los datos disponibles en tales campos. Por ello, en los últimos años, se han publicado varios trabajos con el objetivo principal de especificar una distribución trapezoidal tetraparamétrica, $T_p(A, M1, M2, B)$, a partir de las tres típicas estimaciones subjetivas del método PERT, sobre los valores pesimista, optimista y más probable. En esta línea de trabajo se enmarcan los artículos de Callejón et al. (1996), Vivo y Franco (2006), y Herrerías y Herrerías (2009).

Según Kotz y van Dorp (2004), el primer antecedente en el uso de la distribución trapezoidal es el de Pouliquen (1970), relativo a un análisis del riesgo en un proyecto de valoración. El interés, en esta última década, por el estudio de esta distribución es creciente, véase van Dorp et al. (2007) y la generalización de la distribución trapezoidal introducida por van Dorp y Kotz (2003).

En España, Herrerías y Calvete (1987) la utilizan como modelo probabilístico para los flujos de caja de una inversión, dos años más tarde, Herrerías y Miguel (1989), determinan una expresión más sencilla para su varianza, que es la comúnmente usada en los trabajos de Callejón, Pérez y Ramos (1996), Vivo y Franco (2006), y Herrerías y Herrerías (2009).

Centrándonos en el presente trabajo, debe decirse que su objetivo principal es la comparación de las tres diferentes distribuciones trapezoidales presentadas por Callejón et al. (1996), por Vivo y Franco (2006), y por Herrerías y Herrerías (2009). Éstas han sido especificadas a partir de los tres valores de la metodología PERT: menor (A), modal (M) y mayor (B). Este estudio comparativo permitirá averiguar cuál de los tres

modelos probabilísticos trapezoidales es mejor o más adecuado en la aplicación a un problema real.

Como algunas de las aplicaciones de las distribuciones trapezoidales están relacionadas con el Análisis de Inversiones, con la estimación del tiempo a emplear en un trabajo complejo, con los métodos de valoración, etc., se debe atender al análisis comparativo de sus medias y varianzas para determinar en qué tramos un modelo es más moderado que los otros, ó menos incierto en la especificación del intervalo modal, ó más ó menos conservador en sus conclusiones debido a su mayor o menor varianza. Estas cualidades de moderación, incertidumbre del intervalo modal y conservadurismo ó poca aversión al riesgo, se cuantifican con la media, con la longitud del intervalo modal y con la varianza respectivamente.

Para complementar el análisis comparativo de las tres distribuciones trapezoidales, se han obtenido dos diagramas de cociente de momentos, utilizando los coeficientes de variación, de asimetría y de curtosis, de las mismas.

El desarrollo del trabajo es el siguiente: en el punto 2 se resume la clásica distribución trapezoidal estandarizada, esto es, el recorrido de la variable (A, B) se reduce al intervalo (0,1), junto con sus principales características estocásticas, en el punto 3 se presentan las tres distribuciones trapezoidales triparamétricas, citadas anteriormente. El punto 4, se dedica a realizar un primer análisis comparativo de estas tres distribuciones trapezoidales, utilizando las longitudes de sus intervalos modales y sus características estocásticas: medias, varianzas y coeficientes de variación. En el punto 5 se profundiza el análisis comparativo, atendiendo a los diagramas de cociente de momentos. Finalmente, en el punto 6, se recopilan las principales conclusiones del trabajo.

2. La distribución trapezoidal

La distribución trapezoidal $Tp(A, M_1, M_2, B)$ responde a la siguiente función de densidad:

$$f(x) = \begin{cases} K \frac{x-A}{M_1-A} & \text{si } A < x \leq M_1 \\ K & \text{si } M_1 \leq x \leq M_2 \\ K \frac{B-x}{B-M_2} & \text{si } M_2 \leq x < B \end{cases} \quad (1)$$

donde $K = 2(B - A + M_2 - M_1)^{-1}$ juega el papel de constante normalizadora.

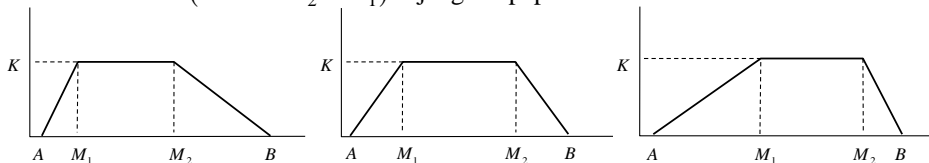


Figura 1. Distribución trapezoidal simétrica y con asimetría a la derecha e izquierda

Por comodidad de cálculo y para facilitar la comparación de las distribuciones, es conveniente estandarizar el recorrido (A, B) de la variable X al intervalo $(0, 1)$, mediante el cambio de variable lineal:

$$Z = \frac{X - A}{B - A} \quad (2)$$

resultando que esta nueva variable estandarizada tiene una función de densidad:

$$f(z) = \begin{cases} k \frac{z}{m_1} & \text{si } 0 < z \leq m_1 \\ k & \text{si } m_1 \leq z \leq m_2 \\ k \frac{1-z}{1-m_2} & \text{si } m_2 \leq z < 1 \end{cases} \quad (3)$$

donde se ha notado por m_i ($i=1, 2$), los valores estandarizados de M_i ($i=1, 2$), y por $k = 2(1 + m_2 - m_1)^{-1}$ a la constante normalizadora resultantes del cambio de variable propuesto en (2).

De (1) y (3) se deducen las respectivas funciones de distribución de las variables aleatorias ordinaria (X) y estandarizada (Z), que pueden consultarse en Herrerías et al. (2012).

Mediante (4) se obtienen los momentos centrados respecto al origen para la distribución trapezoidal estandarizada:

$$\alpha_n = E[Z^n] = \frac{k}{(n+1)(n+2)} \left[\frac{1-m_2^{n+2}}{1-m_2} - m_1^{n+1} \right] \quad (4)$$

$$\text{Para } n=1: \quad \alpha_1 = E[Z] = \frac{1}{3} \left(\frac{m_2^2 + m_2 + 1 - m_1^2}{1 + m_2 - m_1} \right) \quad (5)$$

$$\text{Para } n=2: \quad \alpha_2 = E[Z^2] = \frac{m_2^3 + m_2^2 + m_2 + 1 - m_1^3}{6(1 + m_2 - m_1)} \quad (6)$$

Por lo que la varianza de la variable estandarizada es:

$$\sigma^2 = \alpha_2 - \alpha_1^2 = \frac{1}{18} \left[1 + m_1^2 + m_2^2 - (1 + m_2)m_1 \right] - \frac{1}{9} \frac{(m_2 - m_1)m_2(1 - m_1)}{(1 + m_2 - m_1)^2} \quad (7)$$

expresión que coincide con la obtenida por Herrerías y Miguel (1989).

3. Distribuciones trapezoidales reparametrizadas

1.1. 3.1. La distribución trapezoidal CPR

La distribución trapezoidal CPR fue introducida por Callejón, Pérez y Ramos (1996). Estos autores, a partir de los tres típicos valores "A", "M" y "B" del método PERT, considerando la asimetría de la distribución, de acuerdo a que "M" sea menor o

mayor que el punto central del intervalo (A, B) , $C = (A+B)/2$, determinan los extremos inferior y superior del intervalo modal de la distribución trapezoidal mediante:

$$\begin{aligned}
 & - M_1 = C \text{ y } M_2 = M, \text{ si } C < M, \text{ o bien} \\
 & - M_1 = M \text{ y } M_2 = C, \text{ si } C > M
 \end{aligned} \tag{8}$$

Nótese que el punto medio del intervalo modal en la distribución trapezoidal CPR es $(M + C)/2$, luego no coincide con M , salvo que $M = C$.

Aplicando sobre (8) el cambio de variable propuesto en (2) se tiene:

$$\begin{aligned}
 & - m_1 = c \text{ y } m_2 = m, \text{ si } c = 0,5 < m, \text{ o bien} \\
 & - m_1 = m \text{ y } m_2 = c, \text{ si } c = 0,5 > m
 \end{aligned} \tag{9}$$

En tal caso, el punto medio del intervalo modal en la distribución trapezoidal CPR estandarizada es $(2m + 1)/4$, que tampoco coincide con m , salvo que m sea 0,5.

1.2. 3.2. La distribución trapezoidal VF

Vivo y Franco (2006) establecen que una variable aleatoria $Tp(A, M_1, M_2, B)$ puede representarse de forma equivalente mediante $Tp(A, M, W, B)$, donde M representa el punto medio del intervalo modal, $M = (M_1 + M_2)/2$, y W la mitad de su amplitud, $W = (M_2 - M_1)/2$. Finalmente suponen que $M_1 = M - W$ y $M_2 = M + W$.

A partir de los tres típicos valores “A”, “M” y “B” del método PERT, denotando por $C = (A+B)/2$ al punto central del intervalo (A, B) , determinan los extremos inferior y superior del intervalo modal de la distribución trapezoidal, de la siguiente forma:

$$\begin{aligned}
 & - M_1 = A \quad \text{si } A < M \leq (4A + B)/5 \\
 & \quad M_2 = 2M - A \\
 & - M_1 = (5M - A - B)/3 \quad \text{si } (4A + B)/5 < M < C \\
 & \quad M_2 = (A + B + M)/3 \\
 & - M_1 = (A + B + M)/3 \quad \text{si } C < M < (A + 4B)/5 \\
 & \quad M_2 = (5M - A - B)/3 \\
 & - M_1 = 2M - B \quad \text{si } (A + 4B)/5 \leq M < B \\
 & \quad M_2 = B
 \end{aligned} \tag{10}$$

Nótese que la distribución trapezoidal propuesta en (10) tiene como punto medio del intervalo modal (M_1, M_2) , al propio valor de M , es decir, la distribución trapezoidal obtenida por (10) centra el valor modal suministrado por el experto.

Aplicando sobre (10) el cambio de variable propuesto en (2) se tiene:

$$\begin{aligned}
 & - m_1 = 0 \quad \text{si } m \leq 1/5 \\
 & \quad m_2 = 2m \\
 & - m_1 = (5m - 1)/3 \quad \text{si } m \in]1/5, 1/2[\\
 & \quad m_2 = (1 + m)/3
 \end{aligned}$$

$$\begin{aligned}
 & - \begin{matrix} m_1 = (1+m)/3 \\ m_2 = (5m-1)/3 \end{matrix} \quad \text{si } m \in]1/2, 4/5[\\
 & - \begin{matrix} m_1 = 2m-1 \\ m_2 = 1 \end{matrix} \quad \text{si } m \geq 4/5
 \end{aligned} \tag{11}$$

Al igual que la distribución trapezoidal propuesta en (10), la distribución trapezoidal estandarizada propuesta en (11) tiene como punto medio del intervalo modal, (m_1, m_2) , el valor de m .

1.3. 3.3. La distribución trapezoidal HH

Con la misma información de partida: “A”, “M” y “B”, Herrerías y Herrerías (2009) determinan, en principio, los extremos superior e inferior del intervalo modal de la distribución trapezoidal de la siguiente forma:

$$\begin{aligned}
 & - M_1 = M - (B - M)/2 \text{ y } M_2 = M + (B - M)/2 \text{ si } C < M \\
 & - M_1 = M - (M - A)/2 \text{ y } M_2 = M + (M - A)/2 \text{ si } C > M
 \end{aligned} \tag{12}$$

Siendo $C = (A + B)/2$ el punto central del intervalo (A, B)

Herrerías y Herrerías (2009) proponen determinar el intervalo (M_1, M_2) de la distribución trapezoidal, $\text{Tp}(A, M_1, M_2, B)$ de la siguiente manera:

$$\begin{aligned}
 & - \begin{matrix} M_1 = M - (M - A)/2 \\ M_2 = M + (M - A)/2 \end{matrix} \quad \text{si } M < (3A + B)/4 \\
 & - \begin{matrix} M_1 = (3M - C)/2 \\ M_2 = (M + C)/2 \end{matrix} \quad \text{si } M \in [(3A + B)/4, C] \\
 & - \begin{matrix} M_1 = (M + C)/2 \\ M_2 = (3M - C)/2 \end{matrix} \quad \text{si } M \in [C, (3B + A)/4] \\
 & - \begin{matrix} M_1 = M - (B - M)/2 \\ M_2 = M + (B - M)/2 \end{matrix} \quad \text{si } M > (3B + A)/4
 \end{aligned} \tag{13}$$

Nótese que la distribución trapezoidal propuesta en (13) tiene como punto medio del intervalo modal (M_1, M_2) , al propio valor de M , es decir, la distribución trapezoidal obtenida por (13) centra el valor modal suministrado por el experto.

A partir de (2), se tiene que (12) puede expresarse de la siguiente forma:

$$\begin{aligned}
 & - m_1 = (3m-1)/2 \text{ y } m_2 = (m+1)/2 \text{ si } c = 0,5 < m \\
 & - m_1 = m/2 \text{ y } m_2 = 3m/2 \text{ si } c = 0,5 > m
 \end{aligned} \tag{14}$$

Aplicando la estandarización considerada en (2) sobre el resultado (13) se obtiene la propuesta para determinar una distribución trapezoidal estandarizada, $\text{Tp}(0, m_1, m_2, 1)$:

$$\begin{aligned}
 & - \begin{matrix} m_1 = m/2 \\ m_2 = 3m/2 \end{matrix} \quad \text{si } m < 1/4
 \end{aligned}$$

$$\begin{aligned}
 & - \begin{cases} m_1 = (6m-1)/4 \\ m_2 = (2m+1)/4 \end{cases} \quad \text{si } m \in [1/4, 1/2] \\
 & - \begin{cases} m_1 = (2m+1)/4 \\ m_2 = (6m-1)/4 \end{cases} \quad \text{si } m \in [1/2, 3/4] \\
 & - \begin{cases} m_1 = (3m-1)/2 \\ m_2 = (m+1)/2 \end{cases} \quad \text{si } m > 3/4
 \end{aligned} \tag{15}$$

Al igual que la distribución trapezoidal propuesta en (13), la distribución trapezoidal estandarizada propuesta en (15) tiene como punto medio del intervalo modal, (m_1, m_2) , el valor de m .

4. Análisis comparativos de las tres distribuciones

1.4. 4.1. Análisis de las longitudes del intervalo modal

En este apartado se comparan las longitudes de los intervalos modales estandarizados de las tres distribuciones trapezoidales. Sus longitudes respectivas, en función de los valores de m , son:

- Distribución trapezoidal CPR:

$$\begin{cases} 1/2 - m & \text{si } 0 < m \leq 1/2 \\ m - 1/2 & \text{si } 1/2 < m < 1 \end{cases}$$

- Distribución trapezoidal VF:

$$\begin{cases} 2m & \text{si } 0 < m \leq 1/5 \\ 2/3(1-2m) & \text{si } 1/5 < m \leq 1/2 \\ 2/3(2m-1) & \text{si } 1/2 < m < 4/5 \\ 2(1-m) & \text{si } 4/5 \leq m < 1 \end{cases}$$

Distribución trapezoidal HH:

$$\begin{cases} m & \text{si } 0 < m \leq 1/4 \\ 1/2 - m & \text{si } 1/4 < m \leq 1/2 \\ m - 1/2 & \text{si } 1/2 < m < 3/4 \\ 1 - m & \text{si } 3/4 \leq m < 1 \end{cases}$$

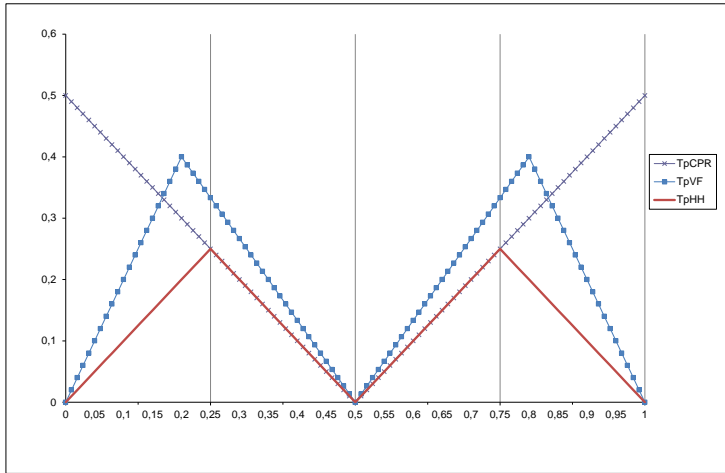


Gráfico 1: Representación de las longitudes de los intervalos modales, en función de m .

Comparando las longitudes de los intervalos modales de las distribuciones trapezoidales estandarizadas, véase gráfico 1, se tiene que:

- Cuando $0 < m < 1/6$ ó $5/6 < m < 1$, la distribución que presenta un intervalo modal con menor longitud es la distribución TpHH, seguida de la TpVF y por último la TpCPR.
- Cuando $m = 1/6$ ó $m = 5/6$, coinciden las longitudes de los intervalos modales de las distribuciones TpCPR y TpVF. Aunque la distribución TpHH sigue presentando el intervalo modal con menor longitud.
- Cuando $1/6 < m < 1/4$ ó $3/4 < m < 5/6$, la distribución que presenta un intervalo modal con menor longitud es la distribución TpHH, seguida ahora de la TpCPR y por último la TpVF.
- Cuando $1/4 \leq m < 1/2$ ó $1/2 < m \leq 3/4$, coinciden las longitudes de los intervalos modales de las distribuciones TpCPR y TpHH, siendo, esta longitud común, menor que la que presenta la distribución TpVF.
- Cuando $m = 1/2$, la longitud de los intervalos modales de las tres distribuciones: TpCPR, TpVF y TpHH es nula, ya que dichas distribuciones trapezoidales degeneran o se transforman en distribuciones triangulares.
- Cuando $m = 0$ ó $m = 1$ la longitud de los intervalos modales de las distribuciones TpVF y TpHH es nula, mientras que la de la distribución TpCPR es igual a 0,5.

1.5. 4.2. Análisis de las medias

A partir de (5) se obtienen las expresiones de las medias estandarizadas para cada una de las tres distribuciones trapezoidales, en función de los valores de m , siendo:

- Distribución trapezoidal CPR:

$$\mu_{TpCPR} = \begin{cases} \frac{7-4m^2}{6(3-2m)} & \text{si } 0 < m \leq 1/2 \\ \frac{4m^2+4m+3}{6(1+2m)} & \text{si } 1/2 < m < 1 \end{cases}$$

- Distribución trapezoidal VF:

$$\mu_{TpVF} = \begin{cases} \frac{4m^2+2m+1}{3(2m+1)} & \text{si } 0 < m \leq 1/5 \\ \frac{5m-8m^2+4}{3(5-4m)} & \text{si } 1/5 < m \leq 1/2 \\ \frac{8m^2+m+2}{3(1+4m)} & \text{si } 1/2 < m < 4/5 \\ \frac{4m-4m^2+2}{3(3-2m)} & \text{si } 4/5 \leq m < 1 \end{cases}$$

- Distribución trapezoidal HH:

$$\mu_{TpHH} = \begin{cases} \frac{4m^2+3m+2}{6(m+1)} & \text{si } 0 < m \leq 1/4 \\ \frac{6m-8m^2+5}{6(3-2m)} & \text{si } 1/4 < m \leq 1/2 \\ \frac{8m^2+2m+3}{6(1+2m)} & \text{si } 1/2 < m < 3/4 \\ \frac{5m-4m^2+3}{6(2-m)} & \text{si } 3/4 \leq m < 1 \end{cases}$$

Comparando las medias estandarizadas de las tres distribuciones trapezoidales, véase gráfico 2, se tiene que:

- Cuando $m = 0$, se comprueba que $\mu_{TpCPR} > \mu_{TpHH} = \mu_{TpVF}$
- Cuando $0 < m < 1/2$, es fácil comprobar que $\mu_{TpCPR} > \mu_{TpHH} > \mu_{TpVF}$
- Cuando $m = 1/2$, se comprueba que $\mu_{TpHH} = \mu_{TpVF} = \mu_{TpCPR}$, como era de esperar.
- Cuando $1/2 < m < 1$, es fácil ver que $\mu_{TpCPR} < \mu_{TpHH} < \mu_{TpVF}$
- Cuando $m = 1$, se comprueba que $\mu_{TpCPR} < \mu_{TpHH} = \mu_{TpVF}$

Como puede verse en el gráfico 2, las gráficas de las medias estandarizadas de las tres distribuciones trapezoidales son crecientes. Esto mismo ocurre con las medias estandarizadas de otras distribuciones usadas en el método PERT como son:

- la distribución triangular, $\mu_T = (1+m)/3$
- la distribución beta PERT, $\mu_{BP} = (1+4m)/6$

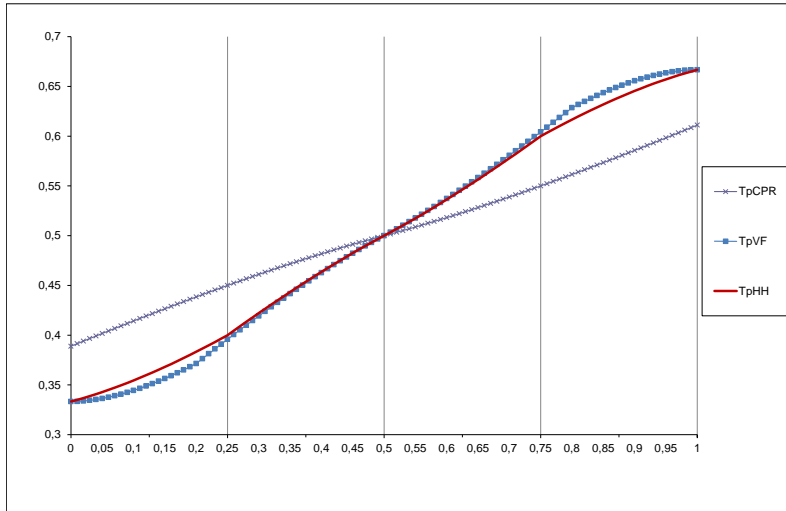


Gráfico 2: Medias de las distribuciones trapezoidales reparametrizadas, en función de m .

Las representaciones gráficas de las medias de estas distribuciones pueden verse en Herrerías y Herrerías (2009) y se puede observar que la distribución menos moderada es la distribución beta PERT ya que es la que presenta mayores oscilaciones, que van desde $1/6$ a $5/6$. Por el contrario, la distribución más moderada es la distribución uniforme, que presenta una media constante e igual a $1/2$.

Centrándonos en las tres distribuciones trapezoidales a las que se hace referencia en este trabajo, concluimos que la más moderada es la TpCPR, seguida de la TpHH y la menos moderada es la TpVF.

1.6. 4.3. Análisis de las varianzas

En este apartado, a partir de (7) se obtienen las expresiones de las varianzas estandarizadas para cada una de las tres distribuciones trapezoidales, en función de los valores de m

- Distribución trapezoidal CPR:

$$\sigma_{\text{TpCPR}}^2 = \begin{cases} \frac{37 - 90m + 112m^2 - 72m^3 + 16m^4}{72(3 - 2m)^2} & \text{si } 0 < m \leq 1/2 \\ \frac{3 + 18m - 8m^2 + 8m^3 + 16m^4}{72(1 + 2m)^2} & \text{si } 1/2 < m < 1 \end{cases}$$

- Distribución trapezoidal VF:

$$\sigma_{TpVF}^2 = \begin{cases} \frac{1+4m+16m^3+16m^4}{18(1+2m)^2} & \text{si } 0 < m \leq 1/5 \\ \frac{109-389m+627m^2-464m^3+112m^4}{54(5-4m)^2} & \text{si } 1/5 < m \leq 1/2 \\ \frac{-5+79m-93m^2+16m^3+112m^4}{54(1+4m)^2} & \text{si } 1/2 < m < 4/5 \\ \frac{37-116m+114m^2-80m^3+16m^4}{18(3-2m)^2} & \text{si } 4/5 \leq m < 1 \end{cases}$$

- Distribución trapezoidal HH:

$$\sigma_{TpHH}^2 = \begin{cases} \frac{4+6m-5m^2+18m^3+7m^4}{72(1+m)^2} & \text{si } 0 < m \leq 1/4 \\ \frac{187-576m+856m^2-576m^3+112m^4}{288(3-2m)^2} & \text{si } 1/4 < m \leq 1/2 \\ \frac{3+144m-200m^2+128m^3+112m^4}{288(1+2m)^2} & \text{si } 1/2 < m < 3/4 \\ \frac{30-78m+91m^2-46m^3+7m^4}{72(2-m)^2} & \text{si } 3/4 \leq m < 1 \end{cases}$$

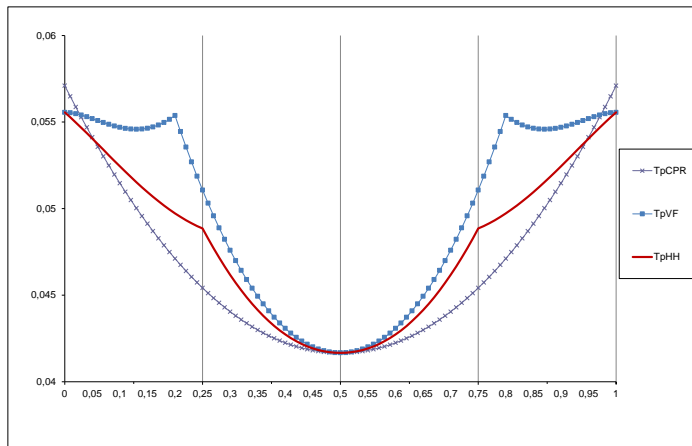


Gráfico 3: Varianzas de las distribuciones trapezoidales reparametrizadas, en función de m .

Comparando las varianzas estandarizadas de las tres distribuciones trapezoidales, véase gráfico 3, se tiene que:

- Cuando $m = 0$ ó $m = 1$, se comprueba que $\sigma_{TpCPR}^2 > \sigma_{TpVF}^2 = \sigma_{TpHH}^2$
- Cuando $0 < m < 0,02749184$ ó $0,9725082 < m < 1$, es fácil comprobar que $\sigma_{TpCPR}^2 > \sigma_{TpVF}^2 > \sigma_{TpHH}^2$
- Cuando $m = 0,02749184$ ó $m = 0,9725082$, se comprueba que $\sigma_{TpCPR}^2 = \sigma_{TpVF}^2 > \sigma_{TpHH}^2$

- Cuando $0,02749184 < m < 0,0532785$ ó $0,9467215 < m < 0,9725082$, es fácil comprobar que $\sigma_{TpVF}^2 > \sigma_{TpCPR}^2 > \sigma_{TpHH}^2$
- Cuando $m = 0,0532785$ ó $m = 0,9467215$, se comprueba que $\sigma_{TpVF}^2 > \sigma_{TpCPR}^2 = \sigma_{TpHH}^2$
- Cuando $0,0532785 < m < 1/2$ ó $1/2 < m < 0,9467215$, se comprueba que $\sigma_{TpVF}^2 > \sigma_{TpHH}^2 > \sigma_{TpCPR}^2$
- Cuando $m = 1/2$, se comprueba que $\sigma_{TpCPR}^2 = \sigma_{TpVF}^2 = \sigma_{TpHH}^2$, como era de esperar.

Atendiendo al gráfico 3, puede notarse que las gráficas de las varianzas de las tres distribuciones trapezoidales son de tipo parabólico. Al igual que ocurre con la varianza estandarizada de la distribución triangular, $\sigma_T^2 = (m^2 - m + 1)/18$, y al contrario de lo que sucede con las distribuciones beta PERT y uniforme, que tienen varianzas constantes. Sus representaciones gráficas pueden verse en Herrerías y Herrerías (2009).

De la misma forma puede comprobarse fácilmente que, en cualquier caso, la varianza de la distribución beta PERT, $\sigma_{BP}^2 = 1/36$, es menor que cualquiera de las varianzas de las distribuciones utilizadas en este trabajo, y la varianza de la distribución Uniforme, $\sigma_U^2 = 1/12$, es mayor que cualquiera de ellas.

1.7. 4.4. Análisis de los coeficientes de variación

En este apartado, a partir de los valores de las desviaciones típicas y las medias de las tres distribuciones trapezoidales estudiadas, se obtienen los correspondientes valores de los coeficientes de variación para los diferentes valores de m . Entendiendo que la distribución con menor coeficiente de variación es la distribución más adecuada, ya que es la que se comporta más homogéneamente, o mejor, bajo el doble criterio de media y varianza.

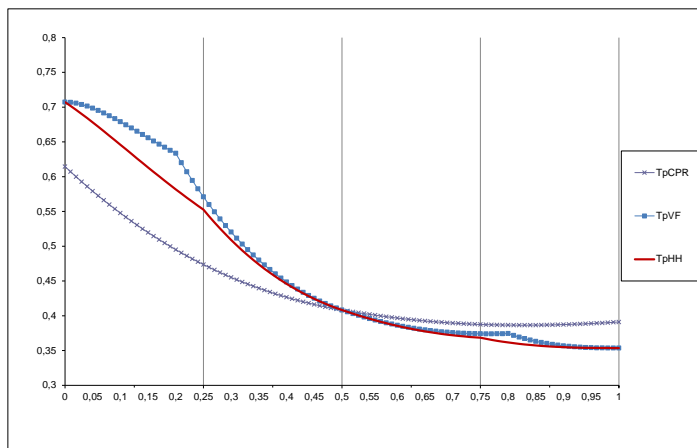


Gráfico 4: Coeficiente de variación de las distribuciones trapezoidales reparametrizadas

Comparando los valores de los coeficientes de variación de las tres distribuciones trapezoidales, véase gráfico 4, se tiene que:

- Cuando $m = 0$, se comprueba que $CV_{TpCPR} < CV_{TpVF} = CV_{TpHH}$
- Cuando $0 < m < 1/2$, es fácil comprobar que $CV_{TpCPR} < CV_{TpHH} < CV_{TpVF}$
- Cuando $m = 1/2$, se comprueba que $CV_{TpCPR} = CV_{TpVF} = CV_{TpHH}$, como era de esperar.
- Cuando $1/2 < m < 1$, es fácil comprobar que $CV_{TpCPR} > CV_{TpVF} > CV_{TpHH}$
- Cuando $m = 1$, se comprueba que $CV_{TpCPR} > CV_{TpVF} = CV_{TpHH}$

5. Análisis de los diagramas de cociente de momentos

El primer diagrama de cociente de momentos fue introducido por Craig (1936), aunque fue popularizado por Elderton y Johnson (1969), para poder estudiar mejor las distribuciones de probabilidad univariantes continuas pertenecientes al sistema de Pearson. Es decir, el constituido por todas las distribuciones de probabilidad, cuya función de densidad, $f(x)$, verifica la ecuación diferencial:

$$\frac{f'(x)}{f(x)} = \frac{x - a}{b_0 + b_1x + b_2x^2} \tag{16}$$

El diagrama suministra información, mediante una simple mirada, sobre el reparto de contribuciones de la asimetría y la curtosis de una familia de distribuciones. Su forma clásica se obtiene representando en el eje de abscisas los valores del coeficiente de asimetría, β_1 , y en el eje de ordenadas los valores del coeficiente de curtosis, β_2 , definidos mediante los cocientes de momentos centrales siguientes:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ y } \beta_2 = \frac{\mu_4}{\mu_2^2}, \text{ donde } \mu_k = \{E[X - E(X)]\}^k \text{ para } k=2, 3, 4 \quad (17)$$

Este diagrama tiene claramente un inconveniente y es que el signo de μ_3 , indicador de la asimetría a la derecha o izquierda que presenta la distribución de probabilidad, desaparece. Por ello, Kotz y Johnson (1985) proponen un ligero cambio en el diagrama de cociente de los momentos para retener esta información, consistente en representar la $\sqrt{\beta_1}$ en el eje de las abscisas, con el acuerdo de que dicha $\sqrt{\beta_1}$ retiene el signo del momento μ_3 , manteniendo los valores de β_2 en ordenadas.

Los valores de $\sqrt{\beta_1}$ y β_2 pueden calcularse utilizando la expresión general de los momentos respecto al origen:

$$\alpha_k = E(X^k) \text{ para } K = 1, 2, 3 \text{ y } 4 \quad (18)$$

siendo sus relaciones con los momentos centrales o respecto a la media:

$$\mu_2 = \alpha_2 - \alpha_1^2$$

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3 \quad (19)$$

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4$$

Sobre los ejes cartesianos $(\sqrt{\beta_1}, \beta_2)$, la región del diagrama situada por debajo de la parábola:

$$\beta_2 = (\sqrt{\beta_1})^2 + 1 \quad (20)$$

se llama región no factible, ya que en ella no pueden existir puntos del diagrama debido a que todas las distribuciones de probabilidad verifican la relación:

$$\beta_2 \geq (\sqrt{\beta_1})^2 + 1 \quad (21)$$

véase Kotz y Johnson (1985).

Dados unos datos empíricos, la forma de actuar con el diagrama de cociente de momentos es la siguiente:

1°. se obtiene el par de estimaciones muestrales de la simetría, $\sqrt{\beta_1}$, y la curtosis, β_2 .

2°. se sitúa dicho par en el diagrama de cociente de momentos.

3°. se considera como familia de distribuciones de probabilidad candidata a modelizar los datos empíricos, aquella que tenga su gráfica más próxima al par de estimaciones.

Para dar una base de comparación consistente, cada familia de distribuciones de probabilidad debe expresarse en términos de los mismos o parecidos parámetros.

Diagramas de cociente de momentos de este tipo, para familias de distribuciones beta en forma de J, campanoides, en forma de U, triangulares, uniformes, Two sided power (TSP), etc. pueden verse en Kotz y van Dorp (2004).

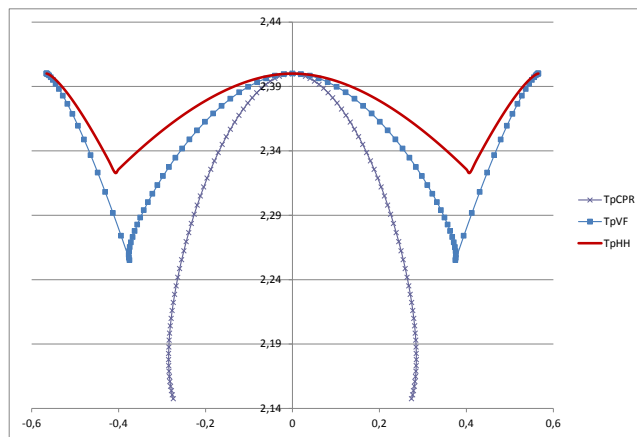


Gráfico 5: Diagrama primero de cociente de momentos (asimetría y curtosis)

Nótese que las tres distribuciones trapezoidales son platicúrticas, ya que su curtosis es inferior a la de la distribución normal. Además su variabilidad es pequeña puesto que la de mayor rango oscila en el intervalo (2.14755296, 2.4).

Otro diagrama de cociente de momentos similar al de la caracterización de las curvas de Pearson, pero que utiliza los coeficientes de variación y de asimetría, por lo que afecta a los momentos centrales 2º y 3º, fue propuesto por Müller y Vahl (1976). Este diagrama de cociente de momentos describe consistentemente la región factible para una familia de distribuciones de probabilidad, en función de los dos coeficientes seleccionados.

Sin pérdida de generalidad puede considerarse cada distribución en su forma estándar, definida por uno o dos parámetros de forma. Expresándose estos parámetros en términos de los coeficientes seleccionados se obtiene la relación o relaciones entre ellos, véase algunos ejemplos en Johnson (2002). Estas relaciones definen una región factible para cada distribución, en forma de una curva, para las distribuciones de un solo parámetro (gamma, logarítmico normal, etc.) o un área para las distribuciones de dos parámetros (beta1 en sus diferentes formas de campana, jota o U, beta2, etc.). Un clarificador y reciente trabajo sobre los dos diagramas de cociente de momentos referenciados, es el de Vargo et al. (2010).

La representación gráfica, para este segundo diagrama, se obtiene situando los valores del coeficiente de asimetría, $\sqrt{\beta_1}$, en ordenadas y los valores del coeficiente de variación, $CV = \sigma / \mu$, en abscisas.

Los resultados y los gráficos de este trabajo no se hubiesen podido obtener sin el uso de la Informática, en particular en este trabajo se han utilizado el programa Mathematica 9 y la hoja de Cálculo Microsoft Excel 2010.

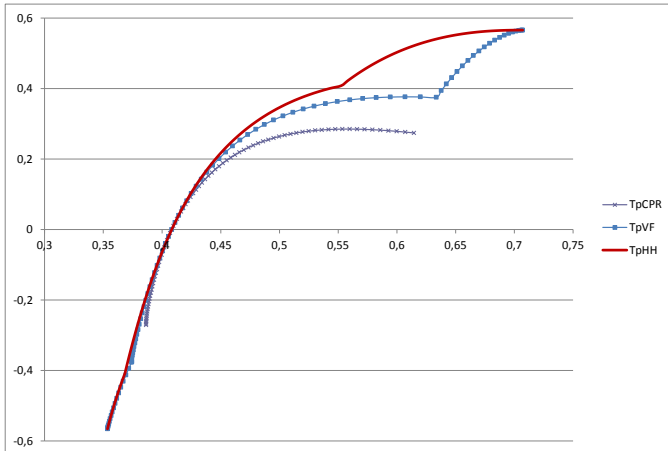


Gráfico 6: Diagrama segundo de cociente de momentos (coeficiente de variación y asimetría)

Comparando los valores de los dos diagramas de cociente de momentos, de las tres distribuciones trapezoidales, véase gráficos 5 y 6, se tiene que:

- La flexibilidad, respecto a la asimetría, de las distribuciones TpHH y TpVF es muy parecida y mayor que la de la distribución TpCPR.
- La flexibilidad, respecto a la curtosis, es superior en la distribución TpCPR, ya que tiene un mayor recorrido que las otras dos distribuciones.
- Para un valor fijo de la asimetría el orden de las distribuciones, de menor a mayor curtosis, es el siguiente: $TpCPR < TpVF < TpHH$.
- Para una asimetría determinada, la distribución trapezoidal de mayor curtosis es la TpHH, seguida de la TpVF, por el contrario, curtosis inferiores a 2.25 sólo son posibles con la distribución TpCPR, aunque deben ir apareadas con asimetrías moderadas que oscilen en el intervalo $(-0.2848, 0.2848)$.
- De forma casi general, se cumple que a medida que el coeficiente de variación aumenta también lo hace el coeficiente de asimetría. Pero a partir del gráfico 6, se comprueba que la distribución TpHH es la única que presenta una dependencia positiva entre el coeficiente de variación y el coeficiente de asimetría.
- El recorrido de los coeficientes de variación y de asimetría de las distribuciones TpVF y TpHH son parecidos y superiores al de la distribución TpCPR.
- Para coeficientes de variación superiores a 0.45, el orden de las distribuciones, de menor a mayor asimetría, es el siguiente: $TpCPR < TpVF < TpHH$.

6. Conclusiones

De acuerdo con los resultados obtenidos en los puntos anteriores de este trabajo, pueden deducirse las siguientes conclusiones:

- (i) Por centrar el valor modal del experto, las distribuciones TpVF y TpHH son preferidas a la distribución TpCPR que no centra dicho valor.

- (ii) Es claro que, para minimizar la incertidumbre del valor modal m serán preferibles los intervalos modales de menor longitud. Del gráfico 1 se concluye claramente, que la distribución TpHH es la que verifica dicha propiedad. Salvo para los valores degenerativos de m : 0, $1/2$ y 1.
- (iii) Desde un punto de vista conservador o de aversión al riesgo, el hecho de tener mayor varianza hace que una distribución sea preferida a otra, ya que se evitan los resultados excesivamente optimistas en las decisiones empresariales, véase Suárez (2007). Por ello, la distribución TpVF es la preferida en el amplio intervalo para $m \in (0,02749184; 0,9725082)$ y sólo fuera de él, la distribución TpCPR tiene una varianza mayor, por lo que será favorita en las aplicaciones.
- (iv) Debido a la falta de acuerdo unánime sobre la bondad de una distribución, usando los criterios de media y varianza, se hace imprescindible el estudio de los coeficientes de variación de estas tres distribuciones. Del gráfico 4 y de los resultados señalados en el epígrafe 4.4, se concluye que en el intervalo de $m \in (0, 1/2)$ la distribución TpCPR es la preferida, mientras que para el intervalo de $m \in (1/2, 1)$, la distribución TpHH es la favorita.
- (v) Respecto a la flexibilidad para la asimetría, las distribuciones TpVF y TpHH superan a la TpCPR, pero respecto a la flexibilidad para la curtosis, es la distribución TpCPR la más flexible.
- (vi) Debe señalarse, como conclusión final, que no se puede indicar una sola distribución óptima para todos los problemas, debido a los comportamientos diferentes señalados, para los distintos valores de $m \in (0,1)$, pero por la supremacía de la distribución TpHH, respecto al criterio de minimizar la incertidumbre del valor modal m y centrar dicho valor, puesto de relieve en las conclusiones (i) y (ii), y por lo resaltado en la conclusión (iv), es ésta la distribución a emplear cuando $m \in (1/2, 1)$. Además es más flexible, en lo que respecta a la variación y la asimetría, que la distribución TpCPR, aunque ésta sea más flexible respecto a la curtosis y tenga un mejor coeficiente de variación para $m \in (0, 1/2)$.

Agradecimientos

Este trabajo ha sido desarrollado con la financiación y colaboración del Ministerio de Ciencia e Innovación y la Unión Europea: “Proyecto I+D ECO2010-15885” y por la Junta de Andalucía a través del Grupo de Investigación FQM-150: “Modelos Probabilísticos Aplicados a las Ciencias Sociales”.

Referencias bibliográficas

1. R. Herrerías y H. Calvete, *Una ley de probabilidad para el estudio de los flujos de caja de una inversión* (Libro Homenaje al profesor Gonzalo Arnáiz Vellando. INE, Madrid 1987), pp. 279-296.

2. R. Herrerías, J. García, S. Cruz y J.M. Herrerías, Il modello probabilistico trapezoidale nel metodo delle due distribuzione della teoria generale de valutazioni. *Genio Rurale. Rivista di Scicienze Ambientali*, **LXIV** (2001) 3-9.
3. A. Kaufmann y J. Gil Aluja, *Introducción de la teoría de conjuntos borrosos a la gestión de empresas*. (Milladoiro, Santiago de Compostela, 1986).
4. M.A. Akcayol, Ç. Elmas, O.A. Erdem y M. Kurt, An educational tool for fuzzy logic controller and classical controllers. *Computer Applications in Engineering Education*, **12**, Issue 2 (2004) 126-135.
5. J. Callejón, E. Pérez y A. Ramos, La distribución trapezoidal como modelo probabilístico para la metodología PERT. *Actas en CD-Rom de la X Reunión de ASEPELT-ESPAÑA* (1996) celebrada en Albacete por la Universidad de Castilla la Mancha.
6. J.M. Vivo y M. Franco, La distribución trapezoidal en la metodología PERT. *Anales de Economía Aplicada* **XX** (2006) 100-112.
7. R. Herrerías y J.M. Herrerías, Especificación de una distribución trapezoidal, a partir de las estimaciones subjetivas del método PERT. *VIII Seminario ASEPELT sobre Modelos de Generación de Distribuciones* (2009) celebrado en Almería
8. S. Kotz y J.R. van Dorp, *Beyond Beta. Other Continuous Families of Distributions with Bounded Support and Applications* (World Scientific, 2004).
9. L.Y. Pouliquen, Risk Analysis in Project Appraisal, *World Bank Staff Occasional Papers* **11**, (1970) distributed by John Hopkins University Press, Baltimore.
10. J.R. van Dorp, S. Cruz, J. García y R. Herrerías, An Elicitation Procedure for the Generalized Trapezoidal Distribution with a Uniform Central Stage, *Decision Analysis Journal* **4** (2007) 156-166.
11. J.R. van Dorp y S. Kotz, Generalized trapezoidal distribution, *Metrika* **58** (1) (2003) 85-97.
12. R. Herrerías y S. Miguel, Expresiones alternativas para la varianza de la distribución trapezoidal. *Actas de la II Reunión de ASEPELT-ESPAÑA* (1989) celebrada en Valladolid, pp. 55-59.
13. C.G. Craig, A New Exposition and Chart for the Pearson System of Frequency Curves. *Annals of Mathematical Statistics* **7** (1936) 16-28.
14. R. Herrerías, F. Palacios y J. Callejón, *Técnicas Cuantitativas para la Inferencia*. (Delta Publicaciones 2012).
15. W.P. Elderton y N.L. Johnson, *Systems of frequency Curves*. (Cambridge University Press, 1969).
16. S. Kotz y N.L. Johnson, Moment ratio diagrams. *Encyclopedia of Statistical Sciences*. Wiley **5** (1985) 602-604.
17. P.H. Müller y H. Vahl, Pearson's system of frequency curves whose left boundary and first three moments are known, *Biometrika*, **63** (1976) 191-194
18. D. Johnson, Triangular approximations for continuous random variables in risk analysis, *Journal of the Operational Research Society*, **53** (2002) 457-467.
19. E. Vargo, R. Pasupathy y L. Leemis, Moment-Ratio Diagrams for Univariate Distributions. *Journal of Quality Technology*, **42**, n° 3 (2010) 1-11.
20. A. Suárez, *Decisiones óptimas de inversión y financiación en la empresa*. (Pirámide, 2007)

WIENER PROCESS IN MATHEMATICA

JOSE ANTONIO SEIJAS MACIAS

Antonio.smacias@udc.es

*Universidade da Coruña / Departamento de Economía Aplicada II
F. de Economía e Empresa – Campus de Elviña (1571 – A Coruña)*

RESUMEN: El proceso de Wiener (movimiento Browniano) se aplica, en la actualidad, en varios campos de la ciencia, especialmente en finanzas y matemáticas financieras. Utilizamos una construcción primaria del proceso de Wiener, basado en una secuencia adecuada de simples caminos aleatorios simétricos. Este método es una simplificación de los de F.B. Knight and P. Révész, citada en [1]. Hemos utilizado el programa de cálculo simbólico Mathematica para proporcionar una implementación de dicho proceso como recursos didáctico que ayuda a comprender dicho fenómeno. Se presentan dos implementaciones de los métodos aplicados para obtener un proceso de Wiener, uno de ellos un modelo de tiempo discreto, y el otro método, el de Lévy.

Palabras claves: Proceso de Wiener, Camino Aleatorio, Método de Lévy, Mathematica, Movimiento Browniano.

ABSTRACT: The Wiener process (Brownian motion) is commonly applied, nowadays, in several fields of science, especially in finance and financial mathematics. We use an elementary construction of the Wiener process, based on a proper sequence of simple symmetric random walks. This method is a simplification of those of F.B. Knight and P. Révész, cited in [1]. Mathematica software has been useful to provide an implementation of the method proposed. We present two implementations of methods applied to obtain a Wiener process, one being a discrete-time model, and the other, the Lévy's method.

Keywords: Wiener Process, Random Walk, Levy's Method, Mathematica, Brownian Motion.

1. Introduction.

In 1828, R. Brown published a paper, in which he summarized his observations on the motion of tiny organic and inorganic particles when immersed in fluids move randomly and, tried to find its physical explanation. However, Brown could only speculate on the causes of this phenomenon. His studies were used to, among other things; result in the so-called Brownian motion, which is reproduced as a seemingly chaotic zigzag movement of microscopically small particles in fluids. The physics of this phenomenon was explained satisfactorily only 80 years later, independently by Einstein and Smoluchowski: in suspensions a small solid particle is hit randomly by molecules of liquid form different directions, and unless the crashes are perfectly balanced, the will

cause chaotic migration of the particle through the suspension.

N. Wiener (1918) was the first to present a general mathematical treatment of the Brownian motion and presented a fully rigorous mathematical model describing the physical phenomenon. He defined and analysed a stochastic process, which has served up till now as a stochastic model of Brownian motion. This process is also referred to as the Wiener process. Nowadays the enormous importance of the Brownian motion process is above all due to the fact that it is the basis of the modern approach to financial mathematics. Wiener process plays a crucial role in mathematics of finance, and has fruitful applications in disciplines as time series analysis, operations research, communication theory and reliability theory ([2], pp. 351)

Discrete models are only a crude approximation to the way in which stock markets actually move. A better model would be one in which stock prices can change at any instant ([3], pp. 51). First attempt to evolve these models was in 1900 L. Bachelier. He proposed Brownian motion as a model of the fluctuations of stock prices. Although his proposal was neglected, several authors later resumed their principles and focused his research on the modelling process of Brownian motion or Wiener processes.

Computer algebra systems (like “Mathematica” or “Matlab”) are part of the tools that are available for teaching mathematics. These tools are introduced into courses for different studies (mathematics, engineering, etc.), basically for advanced courses. First year students are more confident into traditional tools like graphical calculators; but for more years courses the usefulness of computer algebra systems are more evident.

In this paper we discuss the basis of construction of a Wiener process. The computer algebra system “Mathematica” is used to demonstrate the basic methods for the construction of Wiener process. Our goal in this paper is to provide a representation of the construction of a Wiener process by using the computer algebra system Mathematica. We use several functions and procedures implemented in Mathematica (version 9) and, we show the techniques for constructing a Wiener process (or Brownian motion).

Our first approach will be considered Wiener process as an “infinitesimal” random walk in which smaller and smaller steps are taken at ever more frequent time intervals. Then, we will study two approaches: as a discrete-time model ([4], pp. 26) and a formal construction, due to Lévy ([3], pp. 51).

Mathematica software contains a function to simulate a Wiener process, it is WienerProcess function, and it may be used with RandomFunction function to construct a Wiener process. If we compare with the results of our implementation has better development. The symbolic computation software Mathematica has been widely used in many fields of engineering, mathematics and finance. There are functions that allow calculations of most of the financial functions (deterministic or stochastic). Our aim is to show how you can make the implementation of these functions using the basic functions of calculus and linear algebra.

In the Section 2 of this paper, we present the construction of random walks as a sequence of values of a random variable, then section three presents an approach to the distribution of random walks through the normal distribution and uniform distribution, this approach will define Wiener process as random walks. Section 4 presents the construction of the Wiener process by discrete-time model. Section 5 uses a method of Wiener process construction because Lévy. Finally, in Section 6 collect the conclusions

2. Random Walk.

The simplest model of a Wiener process is a symmetric random walk in one dimension, that is, one that represents, for example, the moving of a particle at a moment of $t, \{t=1, 2, \dots, n\}$. We have a sequence (a walk) (X_1, X_2, \dots, X_n) of independent and identically distributed random variables with

$$P\{X_n = 1\} = P\{X_n = -1\} = \frac{1}{2}, \quad (n = 1, 2, \dots) \quad (1)$$

To simulate a random walk we could use the command `Random[_]`, which is part of Mathematica, to get an integer on $[0,1]$ and construct the command `RandomWalk[k_integer]`, for a sequence of $\{1,-1\}$ with k elements.

First, we construct a random sequence of integer elements between -1 and 1 with the procedure `X[k_integer]`:

```
X[k_Integer]:=Table[2*Random[Integer,{0,1}]-1,{i,1,k}].
```

```
Example1=X[20]
```

```
{1,1,-1,-1,-1,-1,1,-1,-1,1,1,1,-1,1,-1,1,-1,-1}
```

Now, we check that the mean is zero and the variance is unitary, using a sequence with 10000 elements.

```
Example2=X[10000];
```

```
N[Mean[Example2]]
```

```
0.076
```

```
N[Variance[Example2]]
```

```
1.00004
```

Now, we consider the command `RandomWalk[k_integer]`. A random walk marks the position of the particle at time n is given by the partial sums

$$S_0 = 0, S_n = X_1 + X_2 + \dots + X_n. \quad (2)$$

The stochastic process $\{S_n\}_{n \geq 0}$ is a simple random walk under the probability measure P if $S_n = \sum_{i=1}^n X_i$, where X_i takes the values $\{-1,+1\}$ and these are independent and identically distributed under P .

Then, we define the process for simulation of a random walk.

```
RandomWalk[n_]:=NestList[#+(-1)^Random[Integer]&,0,n]
```

We graph two examples: one of them with 50 points, where the movements of the particle are very defined and the other with 2000 points, where the movements are less defined:

```
RandomWalk[50]
```

```
{0, 1, 2, 1, 0, 1, 2, 1, 2, 3, 2, 3, 2, 3, 4, 3, 2, 1, 2, 3, 2, 3, 4, 3, 2, 1, 2, 1, 1, 2, 1, 2, 3, 4, 5, 4, 5, 4, 5, 4, 5, 6, 5, 4, 3, 4, 3, 2, 3, 2, 3, 4}
```

We provide a graphical representation of a random walk.

```
ListPlot[RandomWalk[50],Joined ->True];
```

(See Figure 1)

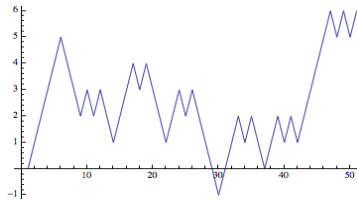


Figure 1: Random Walk 50 points

```
ListPlot[RandomWalk[2000], Joined ->True];
```

(See Figure 2)

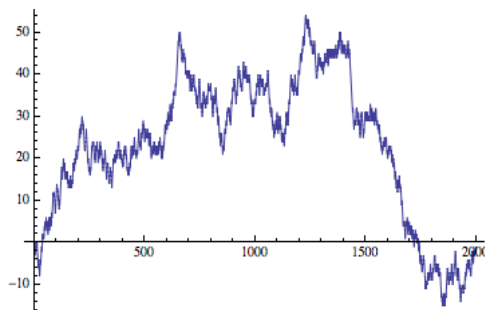


Figure 2: Random Walk 2000 points.

2.1 Construction of Random Walks.

To construct a symmetric random walk, we repeatedly toss a fair coin ([5], p. 119) where variable X_i represents the different results for every toss. The process S_n is a symmetric random walk for different values of length $n=1, 2, \dots$.

$$S_n = \sum_{i=1}^n X_i \quad (3)$$

Considering a group of random walks with n points, we choose a random number of walks from the first k points in the walk. In [3], p. 52, this process is compared to a model of the gains from repeated plays of a fair game. For each n , S_n models the net gain after n plays.

We define a procedure for calculating m random walks with minus than k points RandomWalks[m_Integer, k_Integer].

```
RandomWalks[m_Integer, k_Integer]:=
Module[{x, s=1, LWalks={}}, While[s<=m, A=Walk[k]; x=Random[Integer, {1, k}];
AppendTo[LWalks, Table[{i, S[i]}, {i, 0, x}]]; s++; LWalks]
```

```
Example1 = RandomWalks[3, 15]
```

```
{{{0, 0}, {1, -1}, {2, -2}, {3, -3}, {4, -2}, {5, -1}, {6, -2}, {7, -3}, {8, -4}, {9, -3}, {10, -2}},
{{0, 0}, {1, -1}, {2, -2}, {3, -1}, {4, 0}, {5, 1}, {6, 2}, {7, 3}, {8, 2}, {9, 1}, {10, 0}, {11, -1}},
{{0, 0}, {1, -1}, {2, -2}, {3, -3}, {4, -2}, {5, -3}, {6, -4}, {7, -5}, {8, -4}, {9, -3}, {10, -4}, {11, -3},
{12, -2}}}
```

The example above (Example1) is a group of 3 random walks, with until 15 points. The different random walks are represented graphically using function ListPlot.

```
FramesWalk1 = ListPlot[Example1, Joined -> True]
```

(See Figure 3)

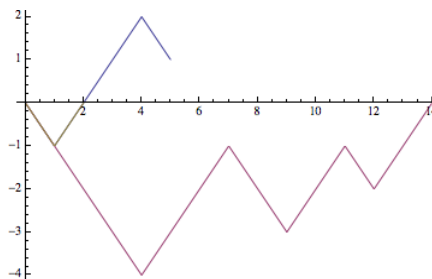


Figure 3: FramesWalk 1 (3 walks, until 15 points)

Another example (Example2), now, it is a group of 5 random walks, with until 90 points. We have 5 random walks with a common origin, the zero point, and different overall lengths.

```
Example2 = RandomWalks[5, 90];
```

```
FramesWalk2 = ListPlot[Example2, Joined -> True]
```

(See Figure 4)

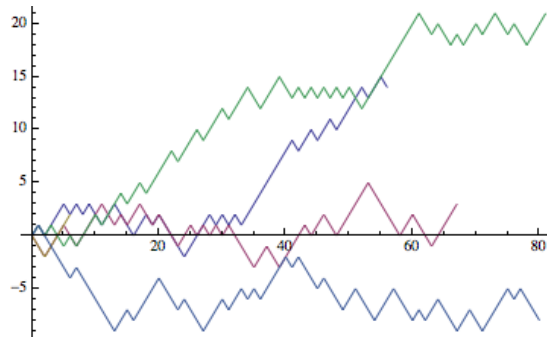


Figure 4: FramesWalk2, (5 walks, until 90 points)

The distribution of S_n is a linearly transformed symmetric binomial distribution ([1] pp. 251). Consequently, $E[S_n] = 0$ and $\text{var}[S_n] = n$.

First, we use `VectorWalk[n_, m_]` for constructing m random walks with n elements, and then, we construct the distribution of S_n with the `VectorSn[V_, n_]` function.

```
VectorWalk[n_, m_] := Table[X[n], {i, 1, m}]
```

```
V = VectorWalk[800, 10000];
```

```
VectorSn[V_, n_] := Total[Take[#, n]] & /@ V
```

The parameter n chosen for the function `VectorSn[V, n]` must be less than or equal to the first argument of `V[n, m]`. This function is very fast and we estimate 250 random walks in only 0.10 seconds

```
Timing[vA = VectorSn[V, 250];]
{0.100307, Null}
```

Now, we calculate the mean and the variance for the random walk estimated.

```
N[Mean[vA]]
```

```
-0.0808
N[Variance[vA]]
239.586
```

Another example for 1000 random walks (in 0.23 seconds).

```
Timing [vB = VectorSn2[V, 1000];]
{0.233903, Null}
N[Mean[vB]]
-0.0176
N[Variance[vB]]
1007.28
```

These examples show the speed of the program calculation process of a large number of random walks and confirm the theoretical mean and variance parameters of random walks.

3. From random walks to Wiener process.

We could think of a Wiener process as an infinitesimal random walk. In terms of a gambling game, the interval between plays is δt and the stake is δx , and we are thinking of both of these as tending to zero. We use the Central Limit Theorem with

$$\mu = 0, \sigma^2 = 1 \text{ and } \delta t = \frac{1}{n}, \delta x = \frac{1}{\sqrt{n}}.$$

Suppose that $\{S_n\}_{n \geq 0}$ is a symmetric simple random walk, then $P\left\{\frac{S_n}{\sqrt{n}} \leq x\right\}$ converges

to $\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$ as $n \rightarrow \infty$, that is, the CDF of the Normal distribution with mean zero and variance 1. Using Mathematica we have expressions for PDF and CDF of normal distribution with mean μ and variance σ .

$$\left\{ \begin{array}{l} \text{PDF[NormalDistribution}[\mu, \sigma], x], \text{CDF[NormalDistribution}[\mu, \sigma], x] \\ \left\{ \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}, \frac{1}{2} \operatorname{erfc}\left(\frac{\mu-x}{\sqrt{2}\sigma}\right) \right\} \end{array} \right\}$$

Proba[x] is a procedure for calculating the value of the CDF of a normal distribution with zero mean and variance 1. Then we have represented such distribution on a graph.

```
Proba[x_] := (1/2) * (1 + Erf[x/Sqrt[2]])
Proba[0]
1/2
ListPlot[Table[Proba[x], {x, -1.5, 1.5, 0.1}], Joined -> True]
```

(See Figure 5)

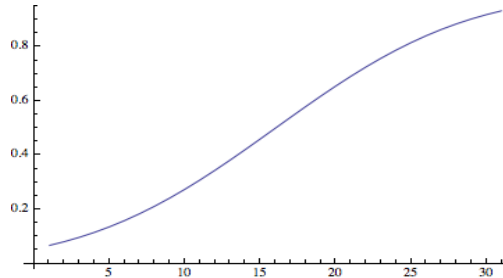


Figure 5: CDF Normal Distribution [0,1]

Now, we considered that the each instance of the random walk is divided in t periods. Then, suppose that $\{S_n\}_{n \geq 0}$ is a symmetric simple random walk, then $P\left\{\frac{S_{[nt]}}{\sqrt{n}} \leq x\right\}$ converges to $\int_{-\infty}^x \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} dy$ as $n \rightarrow \infty$, where $[nt]$ is the integer part of nt . In order to prove this statement we need to point out that $\frac{S_n}{\sqrt{n}}$ is

$$\frac{S_n - E[S_n]}{\sigma[S_n]} = Z, \text{ the standard normal distribution, which has PDF: } \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}}.$$

Now, the evolution of the random walk coincides with the CDF of a normal distribution with zero mean and variance t . Procedure `TempProba[x_, t_]` can get the values of the random walk for a given value of x and a value of t . The graph shows these random walks for t values between 0 and 10.

```
TempProba[x_,t_]:=Erf[-∞, x/ Sqrt[2t]]/2
TempProba[0.99,.5]
0.919254
ListPlot[Flatten[Table[TempProba[x,t],{x,-1.5,1.5,0.1},{t,.1,10}],Joined->True];
```

(See Figure 6)

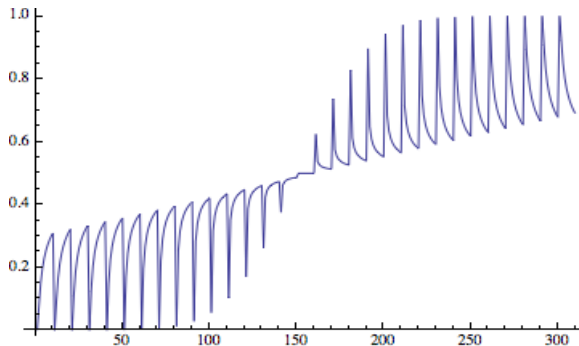


Figure 6: CDF Normal Distributions [0,t], t=0.1,...,10

From the passage to limit in the random walk we could introduce the Wiener process.

A symmetric random walk $\{W_t\}_{t \geq 0}$ is a Wiener process, under a probability P , if for some real constant σ verifies:

1. for each $s \geq 0$ and $t > 0$ the random variable $W_{t+s} - W_s$ has the normal distribution with mean zero and variance $\sigma^2 t$,
2. for each $n \geq 1$ and any time $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$, the random variables $\{W_{t_i} - W_{t_{i-1}}\}$ are independent,
3. $W_0 = 0$,
4. W_t is continuous on $t \geq 0$.

The parameter σ^2 is known as the variance parameter. By scaling of the normal distribution, it is immediate that $\{W_{t/\sigma}\}_{t \geq 0}$ is a Wiener process with variance parameter one.

The transition probabilities of a Wiener process with variance parameter one are:

$$P[W_{t_n} \leq x_n | W_{t_i} = x_i, 0 \leq i \leq n-1] =$$

$$P[W_{t_n} - W_{t_{n-1}} \leq x_n - x_{n-1}] = \int_{-\infty}^{x_n - x_{n-1}} \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} \exp\left(-\frac{u^2}{(t_n - t_{n-1})}\right) du \tag{4}$$

This is the probability density function of the random variable W_{t+s} conditional on $W_s = x$. Then, we define procedure `TransitionProb[x_,y_,t_]` that evaluates the PDF of the random variable W_{t+s} conditional on $W_s = x$, where t is the value of the increment δt and δx is $x-y$. In these situation the transition probability follows a normal distribution with mean y and variance t . In the graph, an example for several values of y and t is representing.

```
TransitionProb[x_,y_,t_]:=1/Sqrt[2π t] Exp[-((x-y)^2)/(2 t)]
TransitionProb[0.1,0.2,.5]
0.558576
```

```
ListPlot[Flatten[Table[TransitionProb[x,y,t],{x,0.01,0.99,0.1},
{y,0.01,0.99,0.1},{t,.1,.5,0.1}],Joined->True]
```

(See Figure 7)

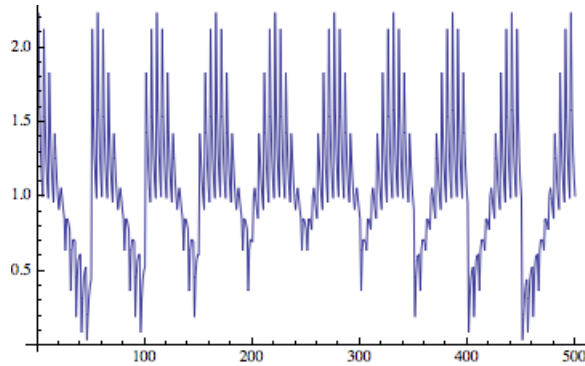


Figure 7: Probability density function random variable conditioned

The joint distributions of W_{t_1}, \dots, W_{t_n} for each $n \geq 1$ and all t_1, \dots, t_n are called the finite dimensional distributions of the process.

Behavioural characteristics of Wiener process are:

1. Although $\{W_t\}_{t \geq 0}$ is continuous everywhere, it is differential nowhere.
2. Wiener process will eventually hit any and every real value no matter how large or how negative. No matter how far above the axis, it will be back down to zero at some later time.
3. Once Wiener process hits a value, it immediately hits again infinitely often.
4. It doesn't matter, what scale you examine Wiener process on, it looks just the same.

Two different approaches were taken to construct a Wiener process using the random walk process we have defined before.

4. 4. Construction of a Wiener process as a discrete-time model

For the construction of a Wiener process using a discrete-time model, let $\Delta t > 0$ be a constant time increment. For the discrete instances $t_j = j\Delta t$, the value W_{t_j} can be written as a sum of increments ΔW_k ,

$$W_{j\Delta t} = \sum_{k=1}^j (W_{k\Delta t} - W_{(k-1)\Delta t}) \quad (5)$$

The elements ΔW_k are independent and normally distributed with $\text{var}(\Delta W_k) = \Delta t$. Increments ΔW with such a distribution can be calculated from standard normally distributed random numbers Z . The implication

$$Z \sim N(0,1) \Rightarrow Z\sqrt{\Delta t} \sim N(0,\Delta t), \quad (6)$$

leads to the discrete-time model of a Wiener process:

$$\Delta W_k = Z\sqrt{\Delta t} \text{ for } Z \sim N(0,1) \text{ for each } k. \quad (7)$$

The function `Wiener [n]`, builds a Wiener process with $n+2$ points in the interval $(0,1)$ following a discrete-time model. The value of Δt will be the value of $1/n$. The intermediate values for time between 0 and 1 will be calculated as the sum of a random value of a normally distributed variable with mean 0 and variance 1 times the square root of the increment of time.

```
Wiener[n_] :=
  Flatten[ {FoldList[Plus, 0, Array[1/n &, n] ],
    FoldList[Plus, 0,
      RandomVariate[NormalDistribution[0, 1], n] Sqrt[1/n]]}, {{2}, {1}}]
```

```
Wiener[3]
{{0, 0}, {1/3, 1.27124}, {2/3, 0.159689}, {1, 0.244396}}
```

```
ListPlot[Wiener[3],Joined->True];
```

(See Figure 8)

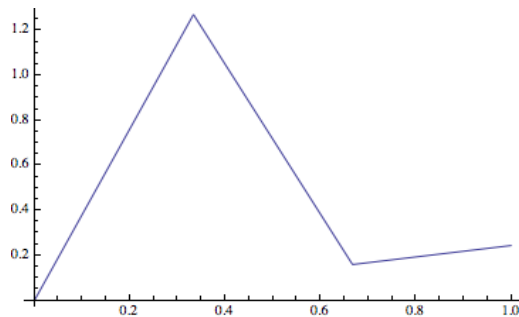


Figure 8: Wiener proces with 3 points

The next figure shows an example of a Wiener process, where the 5000 calculated points (t_s, W_s) are joined by linear interpolation. This process was executed in 0.016 seconds.

```
ListPlot[Wiener[5000],Joined->True];
```


(See Figure 9)

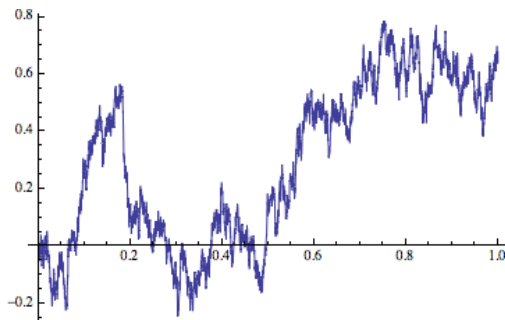


Figure 9: Wiener Process with 5000 points

When we construct a Wiener process following `Wiener[x]`, it is not a mean zero and variance one sequence, but it verifies that,

$$\sum_{i=2}^n (W_i - W_{i-1})^2 \rightarrow 1. \quad (7)$$

```
Timing[A1=Wiener[5000];]
{0.016569, Null}
```

```
A2=Transpose[A1][[2]];
μ=Mean[A2]
-0.12084
σ2=Variance[A2]
0.0517237
```

```
Total[(Rest[A2]-Most[A2])^2]
1.01733
```

5. Lévy construction of the Wiener process

The idea here is that we can produce a path for a Wiener process by direct polygonal interpolation.

Without loss of generality we take the range of t to be $[0,1]$. Lévy's construction builds a polygonal approximation to the Wiener process from a countable collection of independent, normally distributed random variables with mean zero and variance one.

We define a random variable for the interval $[-1,1]$. To get a mean zero and variance one random variable, we normalize with $\sqrt{1/3}$. When we use the Mathematica function `Random`, for a real number on $[-1,1]$, we obtain a random variable with mean zero and variance $1/3$, as follows from the properties of the Uniform Distribution.

Mean[UniformDistribution[{-1, 1}]], Variance[UniformDistribution[{-1, 1}]]
 {0, 1/3}

We use an induction process. The induction begins with

$$X_1(t) = t\xi(1), \tag{8}$$

where we index them by the dyadic points of [0,1], a generic variable being denoted by $\xi(k * 2^{-n})$ where $n \in N$ and $k \in \{0,1,\dots,2^n\}$.

Thus, X_1 is a linear function on [0,1]. The n-th process, X_n , is linear in each interval $[(k-1) * 2^{-n}, k * 2^{-n}]$, is continuous in t and satisfies $X_n(0) = 0$. It is determined by the values $\{X_n(k * 2^{-n}), k = 1,\dots,2^n\}$

We take

$$X_{n+1}(2k * 2^{-(n+1)}) = X_n(2k * 2^{-(n+1)}) = X_n(2k * 2^{-n}). \tag{9}$$

We, now, determine the appropriate value for $X_{n+1}(2(k-1) * 2^{-(n+1)})$. Conditional on $X_{n+1}(2k * 2^{-(n+1)}) - X_{n+1}(2(k-1) * 2^{-(n+1)})$ we get it should be normally distributed with mean

$$\frac{1}{2} (X_{n+1}(2k * 2^{-(n+1)}) - X_{n+1}(2(k-1) * 2^{-(n+1)})). \tag{10}$$

and variance $2^{-(n+2)}$.

Then make this inductive process in Mathematica to check its operation. Procedure X[1] represents the final value of Wiener process, is a uniform distributed variable with mean 0 and variance 1.

```
X[1] = Random[Real, {-1.0, 1.0}]/Sqrt[1/3]
1.29842
```

The following graph represents a Wiener process with five points: first point is always 0, and the final of the process is indicating for value of X[1].

```
ListPlot[{0, X[1] + 2^(-3/2 + 1)*2^(-2/2 + 1), X[1] + 2^(-2/2 + 1),
X[1] - 2^(-2/2 + 1)*2^(-3/2 + 1), X[1]}, Joined -> True]
```

(See Figure 10)

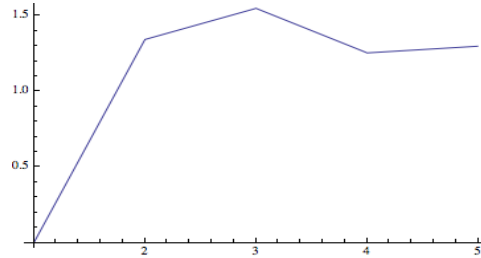


Figure 10: Lévy Process with 5 points

Now, representing a graph for a Wiener process with 7 points.

```
ListPlot[{0, X[1] + 2^-(4/2 + 1)*2^-(3/2 + 1)*2^-(2/2 + 1),
X[1] + 2^-(3/2 + 1)*2^-(2/2 + 1),
X[1] - 2^-(4/2 + 1)*2^-(3/2 + 1)*2^-(2/2 + 1), X[1] + 2^-(2/2 + 1),
X[1] - 2^-(2/2 + 1)*2^-(3/2 + 1), X[1]}, Joined -> True, PlotRange -> All]
```

(See Figure 11)

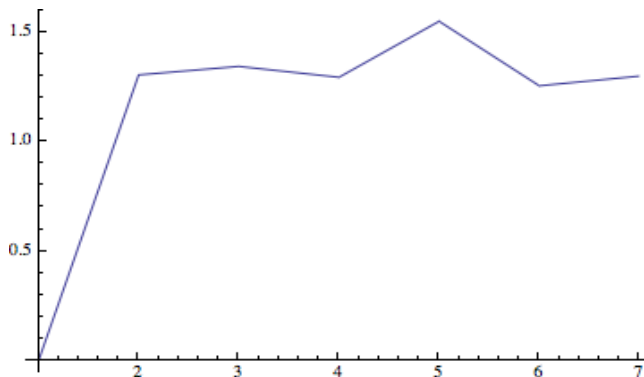


Figure 11: Lévy Process with 8 points

Both processes have the same origin and the same destination point. At each stage we introduce new intermediate points that reproduce new values in the way: a greater number of points; the greater number of turns in the walk alternative.

We produce several random walks from a Lévy's process into a $[0,1]$ interval with $\text{Interv}[n]$, where the number of points is $2^n + 1$, from 0 to 1.

```
Interv[n_]:=Table[k*2^-n,{k,0,2n}]
```

```
Interv[3]
{0,1/8,1/4,3/8,1/2,5/8,3/4,7/8,1}
```

We consider three decomposition of a random walk: first, we consider the direct path without intermediate points in the interval [0,1], then we introduce an intermediate point (1/2), and finally three intermediate points (1/3, 1/2, 2/3).

```
walks :=
{ListPlot[{{0, 0}, {1, X[1]}}, Joined -> True, PlotStyle -> RGBColor[1, 0, 0], PlotRange -> All,
DisplayFunction -> Identity],
ListPlot[{{0, 0}, {1/2, X[1] + 2^-(2/2 + 1)}, {1, X[1]}}, Joined -> True, PlotStyle ->
RGBColor[0, 1, 0], PlotRange -> All, DisplayFunction -> Identity],
ListPlot[{{0, 0}, {1/3, X[1] + 2^-(3/2 + 1)*2^-(2/2 + 1)}, {1/2, X[1] + 2^-(2/2 + 1)}, {3/4, X[1] -
2^-(2/2 + 1)*2^-(3/2 + 1)}, {1, X[1]}},
Joined -> True, PlotStyle -> RGBColor[0, 0, 1], PlotRange -> All, DisplayFunction -> Identity]]

Show[walks, DisplayFunction -> $DisplayFunction]
```

(See Figure 12)

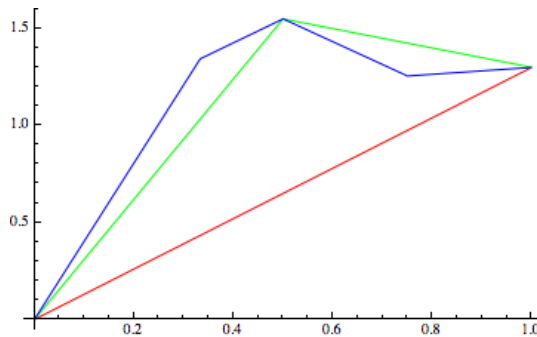


Figure 12: Decomposition for a walk

Now, we present the procedure $PWL[x]$, which builds a random walk with $2^{x-1} + 1$ points in the range [0,1], following Lévy's method. We introduce randomness by defining X_1 and $X[n]$ for values bigger than 1, as follows:

```
X[1] := (Random[Real, {-1.0, 1.0}])/Sqrt[1/3]

X[n_] :=
If[n <= 3,
Flatten[Table[(X[n-1]*(-1)^n*2^-((1/2)+1))+(-1)^(i-1)*Sum[2^-((j/2)-1)], {j,2,i}], {i,2,n}],
Flatten[Table[(X[n-1]*(-1)^n*Sum[2^-((j/2)+1)], {j,2,i}]+((-1)^(i-1)*Sum[2^-((j/2)+1)],
{j,2,i}])/X[n-1]), {i,n,n+1}]]]
```

Procedure PWL[x]

```
PWL[n_]:=Module[{A={{0,0},{1,X[1]}},s,S,B},Do[S=X[s];
B=Table[{Interv[s-1][[i*2]],S[[i]]},{i,1,Length[X[s]]}];
Do[AppendTo[A,B[[m]]],{m,1,Length[B]}],{s,2,n}];Sort[A]
```

```
A=PWL[5]
{{0, 0}, {1/16, 1.2022}, {1/8, 0.337388}, {3/16, 0.3856}, {1/4, -0.701512}, {5/16, 0.398028},
{3/8, 0.781528}, {7/16, -0.563776}, {1/2, -0.998607}, {9/16, -2.80575}, {5/8, -1.26311}, {11/16,
-1.44902}, {3/4, 1.87926}, {13/16, 1.38383}, {7/8, 1.54441}, {15/16, -1.53238}, {1, 0.263929}}
```

Representing a random walk whit a Lévy's process with $2^{5-1} + 1 = 17$ intermediate points into range $[0,1]$.

```
ListPlot[A,Joined->True];
```

(See Figure 13)

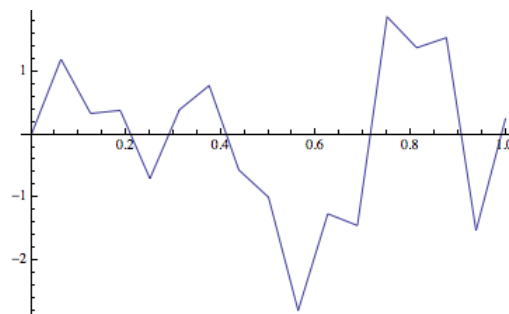


Figure 13: Wiener Process with PWL

The PWL process is very slow and produces unsatisfactory results, hence we test to obtain another function, a little more efficient, allowing us to get a better response. Time to get more than 4000 points over 2 minutes.

To carry out this program, we begin with a countable collection $\{\xi_k^{(n)}, k \in I(n), n = 0, 1, \dots\}$ of independent, standard (zero mean and unit variance) normal random variables. Here $I(n)$ is the set of odd integers between 0 and 2^n ; i.e. $I(0) = \{1\}$, $I(1) = \{1, 3\}$, $I(2) = \{1, 3\}$.

We introduce the procedure RandomEpsilon:

RandomEpsilon[x], builds a vector of standard normal random variables with 2^x elements.

```
RandomEpsilon[n_]:=Module[{A={Random[Real,{-1.0,1.0}]/Sqrt[1/3]}},
Do[A=Append[A,Table[Random[Real,{-1.0,1.0}]/Sqrt[1/3],{i,1,2j,2}],{j,1,n}];A]
```

```
RandomEpsilon[3]
{-1.42724, {-1.7191}, {-1.44253, 0.34502}, {-0.438418, -0.778514, -1.3552, 0.629436}}
```

The Wiener process associated to the random walk is WienerLevy[x], builds a Wiener process with $2^x + 1$ points, for $x > 0$, in the interval [0,1], following Lévy's method.

```
WienerLevy[n_]:=Module[{B={0,0},A1=RandomEpsilon[n]},
B=Append[B,{1,A1[[1]]}];Do[k=Table[h,{h,0,2i-1}];Ine=Table[h,{h,1,2i,2}];f=Length[Ine];
Do[t=(Ine[[j]]+1)/2i;s=(Ine[[j]]-1)/2i;μ=1/2(B[[j,2]]+B[[j+1,2]]);
σ=;B=Append[B,{Ine[[j]]/2i,μ+σ*A1[[i+1]][[j]]}, {j,1,f}];B=Sort[B,{i,1,n}];B]
```

Two examples of the procedure WienerLevy: first of then, we consider WienerLevy[3] that makes a $2^3 + 1 = 9$ points into the range [0,1], and second one, we consider a procedure for Wiener process (WienerLevy[7]) with $2^7 + 1 = 129$ points.

```
WienerLevy [3]
{{0, 0}, {1/8, -0.424782}, {1/4, -0.455172}, {3/8, -0.207726},
{1/2,-0.395019}, {5/8, -0.505499}, {3/4, 0.156079}, {7/8, 0.0333154},
{1, 0.502679}}
```

```
ListPlot[WienerLevy[7], Joined->True]
```

(See Figure 14)

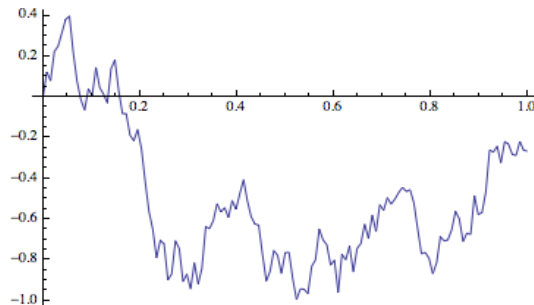


Figure 14: Wiener Process using WienerLevy

In this case, the Wiener process, producing with this method, is not a collection with mean zero and variance one, as in the previous case where we use the discrete time model.

```
Timing[A=WienerLevy[12];]
{0.295630 Second , Null}
```

```
B=Table[A[[i,2]],{i,1,Length[A]}];
```

```
Mean[B]
0.0589245
```

```
Variance[B]
0.231372
```

but rather it verifies that

$$\sum_{i=2}^n (B_i - B_{i-1})^2 \rightarrow 1. \quad (11)$$

We found that process is slower than the one presented in the previous section, now for more than 4000 points exceeds time 0.29 seconds, compared to the previous case where we got 5000 points in less than 0.016 seconds.

```
A1=WienerLevy[14];
```

```
A2=Transpose[A1][[2]];
```

```
Total[(Rest[A2]-Most[A2])^2]
0.989613
```

The sequence of WienerLevy[] is a Wiener Process. It suffices to prove that the increments $(B_{t_j} - B_{t_{j-1}})_{j=1}^n$ are independent, normally distributed, with mean zero and

$$\text{variance } t_j - t_{j-1} = \frac{1}{2^n}.$$

```
WL=WienerLevy[10];
```

```
WL1=Table[WL[[i,2]],{i,1,Length[WL]}];
```

```
WL2=Table[WL1[[i]]-WL1[[i-1]],{i,2,Length[WL1]}];
```

```
Mean[WL2]
0.00041186
```

```
Variance[WL2]
0.000967595
```

```
N[1/(2^10)]
0.000976563
```

Finally, comparing the 3 procedures implemented clear efficiency observed in discrete time model implemented in section 4. Three procedures represent three types of programming: functional programming, procedural programming and recursive programming.

Number of points	Functional Programming Wiener	Procedural Programming WienerLevy	Recursive Programming PWL
4097	0.018 seconds	2.34 seconds	> 1 minute
127	0.000631 seconds	0.013 seconds	0.1762 seconds

6. Conclusions.

The study by Robert Brown of the motion of a particle in a fluid medium in the early nineteenth century had significant unintended consequences on numerous fields. The creation of a mathematical model that would allow modelling this movement by Norbert Wiener had an impact in fields as diverse as particle physics or financial mathematics.

In this work we have made an approach to the Wiener process by various techniques from various authors. To do this, we used the new tools at our disposal. In particular, computer algebra systems (Mathematica) have significant potential when using them to show the formation and creation of a Wiener process by a random walk.

The program Mathematica (version 9) has a function to simulate a Wiener process, but in this work we have tried that students not only use the program-implemented method but different methods of approximation of the process.

We have used different programming styles, at the same time; analyze the impact on the performance in computing time each.

The results of all three methods for constructing a Wiener process are similar and our goal was to demonstrate a very basic approach to Wiener process using three programming methods: functional, procedural and recursive with Mathematica.

7. References

1. T Szabados, An Elementary Introduction to Wiener Process and Stochastic Integrals, *Studia Scientiarum Mathematicarum Hungarica*. 31 (1996) 249-297.
2. F. Beichelt, *Stochastic Process in Science Engineering and Finance* (Chapman & Hall, London, 2006)

3. A. Etheridge, *A Course in Financial Calculus*, (Cambridge University Press, Cambridge, 2002).
4. R. Seydel, *Tools for Computational Finance* (Springer, Berlin, 2004).
5. S. E. Shreve, *Stochastic Calculus for Finance I*, (Springer-Verlag, NY, 2003).

SOBRE LAS PRÁCTICAS DE LABORATORIO EN LA ESTADÍSTICA DE UN GRADO EN TURISMO

J. CASTELLÓ

castelló@uji.es

Universitat Jaume -I / Departamento de Matemáticas-Castellón

C. GALINDO

galindo@uji.es

Universitat Jaume -I / Departamento de Matemáticas-Castellón

V. MARTÍNEZ

martinez@uji.es

Universitat Jaume -I / Departamento de Matemáticas-Castellón

J. MOMPALER

momparle@uji.es

Universitat Jaume -I / Departamento de Matemáticas-Castellón

RESUMEN: El turismo es una actividad económica que representa aproximadamente el 10% del PIB en España. El Grado de Turismo contiene en su primer curso una asignatura de Estadística básica que en la Universitat Jaume I realiza prácticas de laboratorio utilizando el paquete de software estadístico R. En este trabajo, mostramos como pueden realizarse prácticas de laboratorio utilizando, de modo sencillo, un software aparentemente complicado. Para ello realizamos una actividad relacionada con la titulación y de actualidad. Además, mostramos un caso práctico relacionado con el futuro desarrollo profesional del estudiante.

Palabras claves: Estadística, Grado en Turismo, software R.

ABSTRACT: Tourism is an economic activity that represents approximately 10% of GDP in Spain. The Degree of Tourism in its first course contains a course in basic statistics at the Universitat Jaume I laboratory practice using the statistical software package R. In this paper, we show how labs can be performed using, so simple, seemingly complicated software. We carry out an activity related to the degree and topical. Moreover, we show a case study related to the student's future professional development.

Keywords: Statistical. Degree of Tourism. R-Software.

1. Introducción

El turismo es una actividad económica de carácter social que busca el desarrollo de actividades que puedan generar momentos especiales para los seres humanos. En

general, los turistas pretenden visitar un lugar atractivo con el objeto de divertirse, negociar o llevar a cabo cualquier otra actividad relacionada con el ocio o la cultura.

En las últimas décadas, el turismo se ha convertido en una actividad habitual de una buena parte de los habitantes de los países industrializados y como consecuencia es una gran fuente de recursos para los países receptores de turistas. En particular en países como España, esta actividad representa una buena parte del producto interior bruto (PIB) del país. Así en el año 2010 el turismo en España alcanzó el 10,23% de su PIB. Cifras análogas pueden darse dentro de la Comunidad Valenciana (lugar dónde ejercen la docencia, los firmantes de este artículo). La actividad turística se concentra especialmente en las provincias de Alicante y Castellón, donde además del tradicional turismo de sol y playa están cobrando fuerza otras modalidades turísticas que contribuyen a generar empleo estable y a reducir la estacionalidad.

No es pues de extrañar la existencia de títulos universitarios cuyo objetivo sea formar profesionales que puedan incrementar la calidad del servicio. Por lo que respecta a España, según datos de la Estadística Universitaria en España (<https://www.ine.es>) los estudiantes matriculados en la Diplomatura de Turismo (16.364), y los matriculados en el Grado de Turismo según la administración (29.532), permiten estimar que en el curso 2010/11 había 45.896 alumnos inscritos en las titulaciones de Turismo. Además, dentro del ámbito de las Ciencias Sociales y Jurídicas los estudios de Turismo ocupaban el tercer lugar, tras Administración y Dirección de Empresas y Derecho.

Es obvio que un Graduado en Turismo debe tener la capacidad de reunir e interpretar datos relevantes para poder emitir juicios que incluyan una reflexión sobre un tema relevante como éste. A modo de ejemplo, los países receptores de turismo tienen un gran interés en tener información veraz de datos económicos relacionados con el turismo, como pueden ser:

- Los movimientos de los flujos turísticos.
- El gasto promedio que produce.
- Su influencia en la balanza de pagos.
- La inversión que necesita.
- El efecto cambiario de la actividad turística de la que venimos hablando.

A nivel más reducido, cualquier operador turístico necesita también conocer datos similares a los anteriores (con la excepción del último) para que su actividad sea lucrativa y pueda obtener los beneficios esperados mediante la inversión en los lugares óptimos y la oferta de productos que se adecúen a las expectativas de los turistas. Parece pues evidente que éste es un objetivo básico para que el sector crezca y se desarrolle correctamente.

Por consiguiente será obligación de los gobiernos de los estados poner a disposición de los operadores turísticos los datos que les permitan realizar su tarea del mejor modo posible. En este sentido, existen organizaciones nacionales y supranacionales dedicadas a esta tarea. Así a nivel global, la Organización Mundial del Turismo (OMT) es reconocida por las Naciones Unidas como la organización competente para recoger, analizar, publicar, uniformar y mejorar las estadísticas de turismo. La OMT se ha ocupado de liderar iniciativas encaminadas a medir la contribución económica del turismo a las economías nacionales. Destaca entre todas ellas la Cuenta Satélite de Turismo (CST), un marco para reunir estadísticas de turismo y organizar los datos de forma coherente con las cuentas nacionales y, por lo tanto, de manera fiable y comparable en el plano internacional. En particular en 2010 se celebró el Día Mundial de la Estadística, que fue propuesto por la División de Estadística de las Naciones Unidas y proclamado por la Asamblea General de las Naciones Unidas, y tuvo por objeto sensibilizar de los numerosos logros de las estadísticas oficiales. También a nivel español está el Instituto Nacional de Estadística que ofrece, entre otra información, datos referentes a la actividad turística del Reino de España.

Por último, cabe resaltar que durante este año 2013 se está celebrando el Año Internacional de la Estadística, que tiene por objeto el reconocimiento a nivel mundial de la contribución que la Estadística ha realizado a lo largo de los años al progreso de nuestra sociedad. Los principales objetivos de Statistics2013 son:

- Incrementar el conocimiento sobre el impacto que la Estadística tiene en todos los aspectos de la sociedad.
- Promocionar la estadística como profesión, especialmente entre los jóvenes.
- Promover el desarrollo de las Ciencias Estadísticas.

Resulta ya claro a estas alturas que un estudiante de un Grado de Turismo debería poder sacar conclusiones de los datos que se le ofrecen tanto a nivel global como nacional. Esto podrá alcanzarse a través de un conocimiento de los procedimientos y métodos de la Estadística y de la capacidad de entender los grandes números estadísticos que se le pone a su alcance. Así, de poco le servirá entender con razonable profundidad el significado de los diferentes valores y coeficientes numéricos que ofrece la Estadística si no es capaz de tener en sus manos datos fiables para interpretarlos y, recíprocamente, si es capaz de recolectar esos datos e incluso de obtener los valores estadísticos más interesantes pero no es capaz de interpretarlos, tampoco se habrá logrado el objetivo.

En este artículo se pretende explicar cómo se enseña el doble proceso, antes mencionado, a los alumnos de la asignatura *Introducción a la Estadística* (TU0907) del Grado en Turismo de la Universitat Jaume I de Castellón.

2. La asignatura Introducción a la Estadística

Esta es una asignatura de 6 créditos que se imparte en el segundo semestre del primer curso de los futuros Graduados en Turismo. En ella, y según su guía docente, se pretende que el estudiante sepa identificar la necesidad de información estadística que requiere el sector turístico, así como manejar las fuentes estadísticas existentes. Además el alumno también debería ser capaz de analizar, sintetizar y resumir la información estadística mencionada. Aún más, se le pide que sobre ello sea capaz de elaborar y presentar un informe de los resultados obtenidos.

Una lectura detallada de lo anterior, deja bien claro que existe un fuerte desajuste entre los objetivos de la asignatura y el muy escaso tiempo para su desarrollo en el aula. Si acudimos de nuevo a la guía docente, nos encontramos con los siguientes contenidos:

- Descripción de muestras de datos univariantes.
- Descripción de muestras de datos bivariantes.
- Descripción de series de datos temporales.
- Números índices.
- Probabilidad.
- Variable aleatoria.
- Modelos probabilistas de variable aleatoria.
- Inferencia Estadística.

Su desarrollo en cierta profundidad claramente requiere mucho más de 6 créditos. Al final de este artículo aparecen como referencias aquellos textos que se sugiere a los alumnos para seguir la asignatura [1,2,3,4]. No obstante la referencia esencial es [1]. En todos se refleja el objetivo primordial que pretendemos mostrar en este artículo. Se trata de enseñar a los estudiantes a extraer de las fuentes públicas o privadas de que dispongan los datos que necesiten y luego hacer un análisis de ellos. En nuestra opinión es mejor que dominen los cálculos estadísticos sencillos y sepan interpretarlos correctamente. Sabiendo que cuando la magnitud del problema sea muy alta o la interpretación menos clara convendrá acudir a expertos en Estadística que puedan aconsejarle con las muchas y sofisticadas herramientas que se conocen en la actualidad. Con respecto a éstas, al menos el futuro graduado debería saber que existen.

3. La herramienta informática: el programa R

R es un lenguaje y un entorno de programación para análisis de datos y gráficos estadísticos. Es un proyecto GNU y se ha desarrollado en los Laboratorios Bell (antes AT&T, ahora Lucent Technologies) por John Chambers principalmente. R ofrece una gran variedad de datos relacionados con la Estadística. Trabaja con modelos lineales y no lineales, realiza tests estadísticos clásicos, análisis de series temporales, clasificación, clustering y un largo etcétera. También trata con técnicas gráficas. Este lenguaje es muy utilizado en el campo estadístico en general y tiene la ventaja de usar código abierto. Puesto que es de uso común por parte de la comunidad universitaria internacional, R recibe mejoras, e implementaciones de las últimas novedades, de modo constante.

Uno de los puntos fuertes de R es que permite producir informes de calidad en los que se puede incluir símbolos y fórmulas matemáticas. Así, si queremos elaborar un documento dinámico o programable que combine texto y resultados estadísticos, tanto numéricos como gráficos, la mejor opción es utilizar LaTeX y R. Con ello, se obtiene un documento con extensión .tex, que contiene el texto original y los resultados, cálculos y gráficos demandados por el código R insertado. Posteriormente, el archivo .tex se procesará como es habitual en LaTeX para obtener el documento definitivo: .ps, .pdf, etc.

Algo importante para nosotros, por la facilidad que da al alumno en su acceso y uso, es que R está disponible como software libre. Se compila y ejecuta en una amplia variedad de plataformas UNIX y sistemas similares (incluyendo FreeBSD y Linux), Windows y MacOS.

Siendo un poco más explícitos, R es un conjunto integrado de servicios de software para la manipulación de datos, cálculo y representación gráfica. En particular, R incluye:

- Datos: un manejo eficaz y cómodo de datos y almacenamiento.
- Análisis de datos: una amplia colección de herramientas para el análisis de datos.
- Matrices: un conjunto de operadores para los cálculos de estructuras ordenadas, en particular de matrices.
- Gráficos: la posibilidad de obtener gráficos tanto en pantalla como en papel para la interpretación de los datos recibidos.
- Programación: un lenguaje de programación simple y eficaz para desarrollar los cálculos necesarios.

Finalmente, R permite a los usuarios añadir nuevos recursos adicionales mediante la adición de nuevas funciones. Para las tareas computacionalmente intensivas, puede vincularse a C, C++ y Fortran acortando el tiempo de ejecución. Los usuarios

avanzados pueden escribir código C para manipular objetos de investigación directamente. Además R se puede extender a través de paquetes. Hay alrededor de ocho paquetes suministrados con la distribución de R y muchos más están disponibles a través de la familia CRAN de sitios de Internet que cubren una gama muy amplia de la Estadística moderna.

4. La práctica concreta propuesta: Introducción a la Estadística Descriptiva y al R

4.1. La web del INE

Para empezar la práctica, lo primero que se pide al estudiante es que ingrese en Internet y luego a la página web del INE: <http://www.ine.es/>. En ella se accede a algunos contenidos: primero *INEbase*, luego *Servicios* y, finalmente, *Hostelería y Turismo*. En la figura 1 se observa el aspecto de la página.

A continuación observamos que se dispone de los *datos de las Consejerías de Turismo* de cada Comunidad Autónoma. Pasamos pues, por ejemplo, a seleccionar el *año 2008*. Y luego *las plazas estimadas por categoría y meses*. En la figura 2 se pueden observar estos datos.

Operaciones estadísticas relacionadas	Elaboradas por el INE	Tablas INEbase	Contenido	Elaboradas por otros organismos	Tablas INEbase	Contenido
Encuesta de ocupación hotelera	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Turismo receptor	<input type="checkbox"/>	<input type="checkbox"/>
Encuesta de ocupación en campings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Turismo nacional	<input type="checkbox"/>	<input type="checkbox"/>
Encuesta de ocupación en apartamentos turísticos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ingresos y pagos por turismo	<input type="checkbox"/>	<input type="checkbox"/>
Encuesta de ocupación en alojamientos de turismo rural	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Índice de precios hoteleros	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Índice de precios de campings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Índice de precios de apartamentos turísticos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Índice de precios de alojamientos de turismo rural	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Indicadores de Rentabilidad del Sector Hotelero	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Encuesta anual de servicios	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Estadística de productos en el sector servicios	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Encuesta sobre la estructura de las empresas hoteleras 1996	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Encuesta sobre la estructura de empresas de agencias de viajes 1997	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			

Operaciones estadísticas elaboradas por el INE	Ambito	Fuente	Frecuencia	Cód. IOE
Encuesta de ocupación hotelera	nacional	INE	mensual	2025

Figura 1. Aspecto de la web del INE con datos de hostelería y turismo

El siguiente paso es guardar los datos en formato Excel y salir de la página. Los usuarios del sistema operativo *MacOS* deben guardar los datos en formato de texto, ya que R con este sistema no puede leer datos desde Excel.

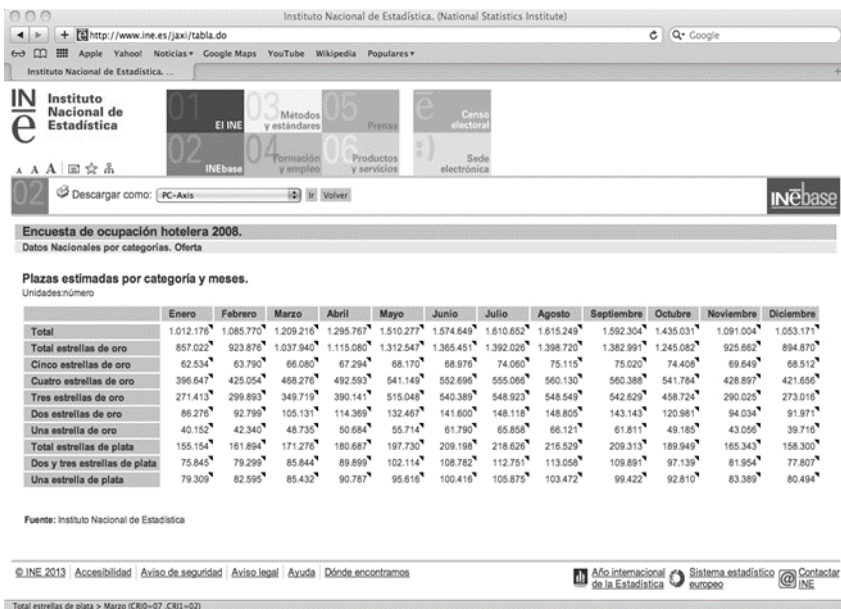


Figura 2. Datos de ocupación hotelera del año 2008 (web INE)

4.2. El paquete R Commander

Nuestra etapa siguiente consistirá en el uso de programa de software R y, en particular, trabajar con el paquete *R Commander* que utiliza ventanas. Consecuentemente cargamos este paquete *Rcmdr*. Mostramos en la figura 3 una imagen de lo que nos debe aparecer.

A partir de ahora, ya podemos trabajar con ventanas. Importamos los datos desde Excel, seleccionando primero *Datos*, para posteriormente acceder a *Importar datos desde Excel* o desde formato de texto si usamos el sistema *MacOS* (en la figura 4 se observa como deben estar cargados los datos para que R los entienda). A partir de aquí, podemos realizar cálculos estadísticos sencillos de la gran cantidad de datos reales que hemos obtenido utilizando, simplemente y de modo sucesivo, las pestañas *Estadísticos*, *Resúmenes* y *Resúmenes numéricos*. La figura 5 muestra lo que debemos ver en la pantalla de nuestro ordenador. En consecuencia, podemos calcular datos estadísticos como: medias, desviaciones, coeficientes de variación, cuartiles, etc. Y lo que es más importante, podemos dedicar nuestro tiempo a interpretar estos datos más que al puro cálculo de los mismos. Para que los estudiantes practiquen por sí mismos, mostramos un ejemplo utilizando los hoteles de 3 estrellas y luego les pedimos a los estudiantes que repitan el proceso de datos.

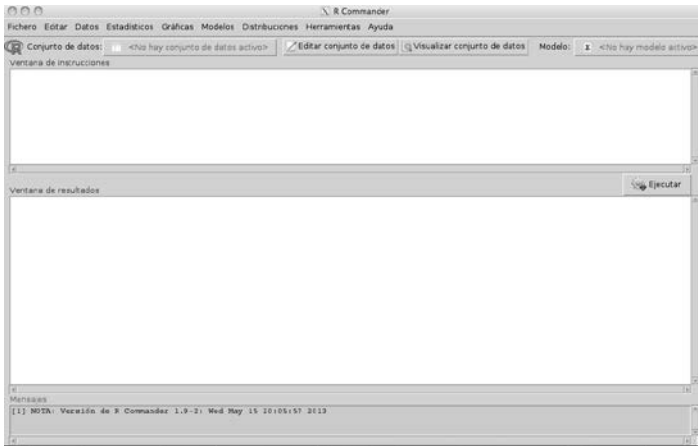


Figura 3. Aspecto del paquete R. Commander

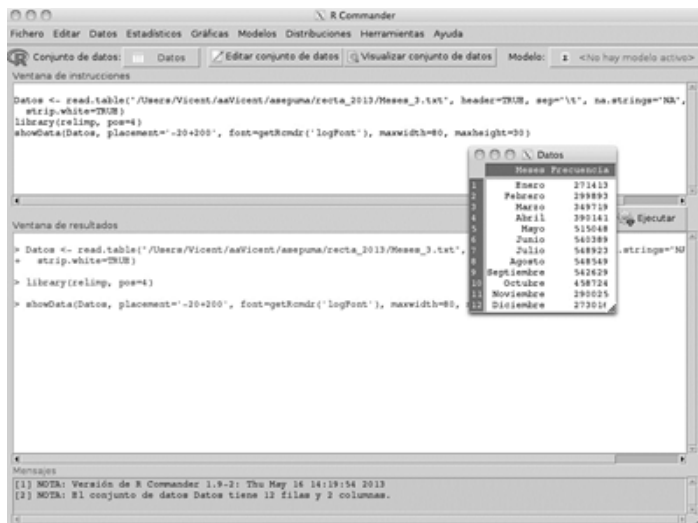
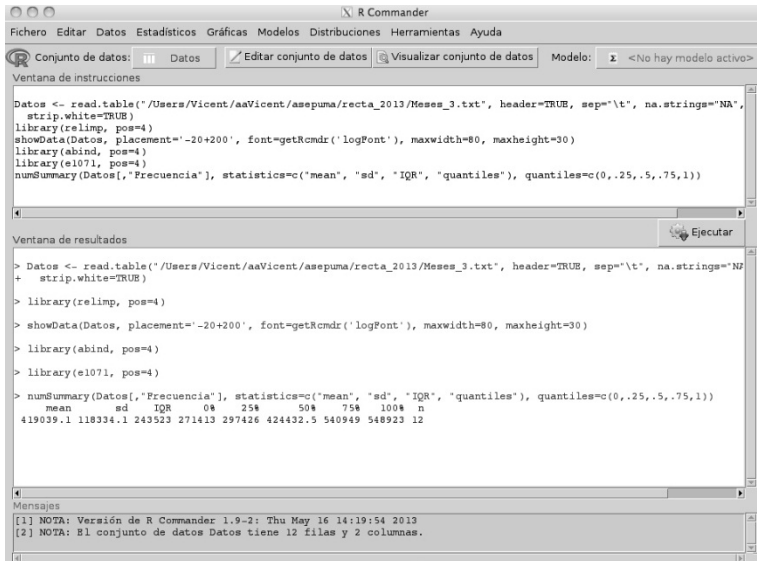


Figura 4. Aspecto de la visualización de datos en R. Commander



```

Datos <- read.table("/Users/Vicent/aaVicent/asepuma/recta_2013/Meses_3.txt", header=TRUE, sep="\t", na.strings="NA",
strip.white=TRUE)
library(relimp, pos=4)
showData(Datos, placement="-20+200", font=getRcmdr('logFont'), maxwidth=80, maxheight=30)
library(abind, pos=4)
library(e1071, pos=4)
numSummary(Datos[, "Frecuencia"], statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0.25,.5,.75,1))

```

```

> Datos <- read.table("/Users/Vicent/aaVicent/asepuma/recta_2013/Meses_3.txt", header=TRUE, sep="\t", na.strings="NA",
+ strip.white=TRUE)
> library(relimp, pos=4)
> showData(Datos, placement="-20+200", font=getRcmdr('logFont'), maxwidth=80, maxheight=30)
> library(abind, pos=4)
> library(e1071, pos=4)
> numSummary(Datos[, "Frecuencia"], statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0.25,.5,.75,1))
  mean      sd      IQR      0%      25%      50%      75%     100%  n
419639.1 118334.1 243523 271413 297426 424432.5 540949 548923 12

```

```

[1] NOTA: Versión de R Commander 1.9-2: Thu May 16 14:19:54 2013
[2] NOTA: El conjunto de datos Datos tiene 12 filas y 2 columnas.

```

Figura 5. Estadísticos que muestra R. Commander

4.3. Los gráficos

Una ventaja del programa R es la facilidad que nos ofrece a la hora de obtener gráficos para interpretar los datos de una manera más cómoda. Para ello, tenemos que darle a R los datos para que pueda trabajar con ellos. Ponemos en la ventana de instrucciones de *R Commander* las expresiones siguientes:

```
Meses<-rep(Datos$Meses,Datos$Frecuencia)
```

```
Datos2<-data.frame(Meses).
```

Con esto obtenemos un nuevo conjunto de datos: *Datos2*. Anteriormente el conjunto de datos *Datos* tenía 12 filas y 12 columnas, mientras que ahora el conjunto de datos *Datos2* tiene 5119553 filas y 1 columna. Este formato de datos es el óptimo para conseguir que R los maneje. Ahora R está preparado para obtener gráficos. Para representar un histograma, procederemos de la manera siguiente: seleccionar *Gráficas* y luego *Diagrama de barras* obteniendo así el histograma que se muestra en la figura 6. Observamos que la variable *Meses* se encuentra en orden alfabético, más tarde indicaremos como se puede obtener el orden cronológico.

Análogamente se podría obtener el diagrama de sectores que se muestra en la figura 7, actuando como sigue: seleccionar *Gráficas* y luego *Diagrama de sectores*.

Finalmente, en la figura 8 se representa la ocupación hotelera por meses ordenados cronológicamente. Para ello hemos procedido así: en primer lugar, hemos reorganizado los datos (*Modificar variables del conjunto de datos activo* y *Reordenar niveles de factor*) y definido así una nueva variable *Meses_ord*, que habrá aparecido en el desplegable de *Modificar variables del conjunto de datos activo*; y, en segundo lugar, procediendo como anteriormente (*Gráficas* y luego *Diagrama de barras*) pero con la nueva variable *Meses_ord*.

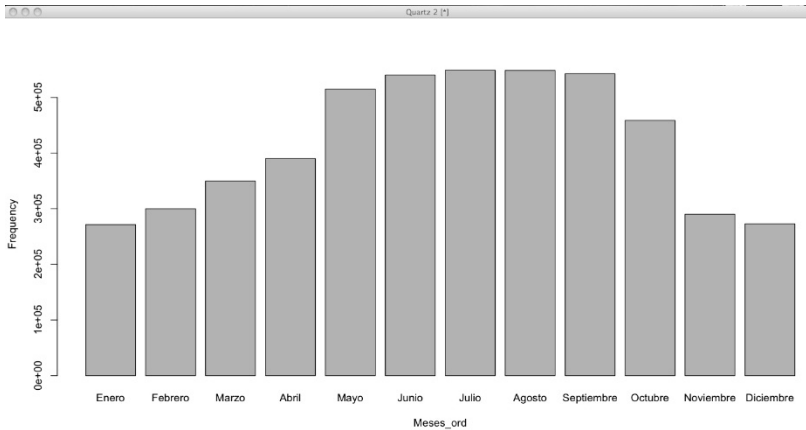


Figura 6. Histograma representando la ocupación hotelera por meses

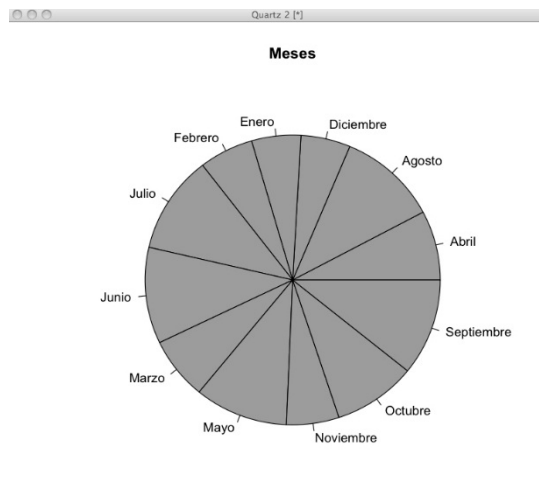


Figura 7. Diagrama de sectores mostrando la ocupación hotelera por meses

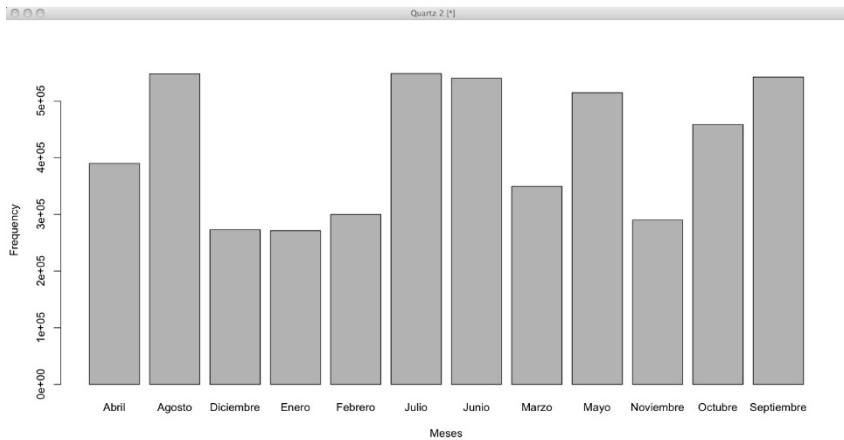


Figura 8. Histograma de ocupación hotelera por meses ordenados croológicamente

5. Conclusiones

En este trabajo se ha presentado, a modo de ejemplo, una de las prácticas que desarrollamos en la asignatura Estadística del grado de Turismo de la Universitat Jaume I de Castellón. El conjunto de estas prácticas utiliza como software el paquete R cuyas ventajas hemos comentado en secciones anteriores. Los aspectos de mejora que hemos introducido, a partir del curso 2010/11, han sido fundamentalmente los siguientes:

- Obtención y manejo de datos, disponibles en fuentes oficiales, y por tanto fiables y útiles en su ámbito profesional.
- Desarrollo de problemas fuertemente relacionados con la titulación.
- Participación en proyectos reales a los cuales tendrán que enfrentarse en su futura vida profesional.

El resultado de este cambio metodológico ha sido muy positivo. La satisfacción de los alumnos se refleja en la valoración del profesorado, la cual ha aumentado en un 18.4%, según la encuesta de evaluación del profesorado. Además, como se deduce de los datos suministrados en la Tabla 1; el número de alumnos que se presenta a las pruebas ha aumentado en 10 puntos porcentuales y lo mismo ocurre con el rendimiento académico de los mismos, cuyo porcentaje de aprobados aumenta en 11 puntos. Por tanto, el rendimiento y el interés de los alumnos en la asignatura ha aumentado considerablemente con la metodología indicada.

Tabla 1. Rendimiento de los alumnos

Resultados de evaluación en porcentajes					
		Aprobados	Suspensos	No presentados	Totales
Curso	2009/10	50.68	15.07	34.25	100
	2010/11	61.63	15.12	23.25	100
	2011/12	61.98	15.71	22.31	100

Agradecimientos

Este trabajo ha sido parcialmente financiado por Unitat Support Educatiu, Universitat Jaume I, proyecto no.10G136-334.

Referencias Bibliográficas

1. J. Alegre, M. Cladera, C. Juaneda. *Análisi quantitativa de l'activitat turística* (Piramide, 2003).
2. C. Fernández. Instruccions d'estadística descriptiva aplicada al sector turístic (Síntesi, 2004).
3. J. M. Raya. *Estadística aplicada al turismo* (Pearson 2004).
4. A. Ronquillo. *Estadística aplicada al sector turistic*. (Centro de estudios Ramón Areces, 1997).
5. <http://www.ine.es/inebmenu/mnu\%20hosteleria.htm>
6. <http://www.iet.tourspain.es/paginas/PubFrontur.aspx\%20?\option=front\&idioma=es-ES>
7. <http://unwto.org/es>
8. <http://www.comunidad-valenciana.org/opencms/opencms/turisme\es/index.jsp>
9. <http://www.educacion.gob.es/educabase>

GIS'F FRACTAL ANALYSIS WITH PIVOTING GRAPHIC

MARÍA- RAMOS ESCAMILLA

maria_ramos@my.unitec.edu.mx

*Universidad Tecnológica de México/Departamento de Posgrado
Calz Ermita Iztapalapa 557 Granjas Esmeralda, Iztapalapa, 09810 Ciudad de México,*

OSCAR -RENE VARGAS

rene_javier60@hotmail.com

*Universidad de Santiago de Compostela/Departamento de Economía del Desarrollo
Avda. do Burgo, s/n. Campus Norte, 15782 Santiago de Compostela*

RESUMEN: En este trabajo se presentan las estaciones de modelado precios geométricos fractales de Capital Markets en París, Frankfurt, Londres, Tokio, Nueva York y México, nuestro objetivo es lograr una mayor rentabilidad de las inversiones realizadas en el tiempo Ex Ante y que sirvan de referencia en tiempo iterativos Ex post, para ello utiliza la metodología de los sistemas de información geográfica y la aplicación de análisis fractal GIS'F de recurrencia vía armónica, series de Fourier, a diferencia de términos, esfera tridimensional y los campos de difracción de Fresnel de la acción del mercado para su compra y funciones de venta .

Palabras clave: Fractal, difracción, el costo, el margen; inversión de títulos.

ABSTRACT: In this paper we present geometric modeling fractal stations prices of Capital Markets in Paris, Frankfurt, London, Tokyo, New York and Mexico, our goal is to achieve a better return on investments made in time Ex Ante and to serve as references in time iterative Ex Post, for this we use the methodology of geographic information systems and implement GIS'F fractal analysis of recurrence via Harmonic, Fourier series, Unlike terms, Three-dimensional sphere and Fresnel diffraction fields of action of the market for their buying and selling functions.

Keywords: Fractal; diffraction; cost; margin; inversion de titles.

1. Introduction.

Fractal analysis is related to chaos theory because it recognizes that not all models studied are linear, as in the case of the models used to analyze financial markets. This is one of the advantages to work with Fractals applied to the financial economy; we can work with broken trends without harmony in market risks, so sticks closer to reality. The assumptions of the fractal-dimensional method, are modeled through 50 sine and cosine functions with logarithms maximum price ranges and minimum price range 100,000 times in three-dimensional scales and fourth (includes Joint leakage points), demonstrating a advantage over the use of quantitative method in which the two elements are bagged study the average Julia set's financial growth in order. Overall we can say that in times of economic boom movement intensifies, the number of participants increases, the money is readily available, the investment is made quickly, allowing growth profitability, companies are capitalized and their instruments tend to rise in price. It also suggests that when markets are efficient, adjustment to the information has to be instantaneous, hence in an efficient market can verify the relevance of information noting whether prices adjust after publication. A discrete dynamical system is a pair (x, f) where x is a field and $f : X \rightarrow X$. Given a point $x \in X$, set $\{x, f^1(x), f^2(x), f^3(x), f^4(x), \dots\}$ will be called the orbit of X , where $f^n(x) = f \circ \dots \circ f(x)$, therefore consider the classification of fixed points according to their properties in a complex dynamic system (C, f) , are as follows:

$$z_0 \in C, z_0 \text{ itself is a point attractor } |f'(z_0)| < 1 \quad (1)$$

$$z_0 \in C, z_0 \text{ itself is a repulsor point } |f'(z_0)| > 1 \quad (2)$$

$$z_0 \in C, z_0 \text{ itself is an indifferent point } |f'(z_0)| = 1 \quad (3)$$

$$z_0 \in C, z_0 \text{ is a super point attractor itself } |f'(z_0)| = 0 \quad (4)$$

The dimension¹ is then a quantitative measurement of the fractal properties of self similarity.

¹ Is sometimes used with respect to analytical processes have been divided into two parts. One dimension of cash flows could result in the separation of payment of mortgage interest and principal cash flows and direct these to different investors.

1.1. Fractal Replication of Market Prices.

The topology of the complex plane² can be designed through the equivalence set of Riemann between the sphere and the complex plane, ie the projection of the points of the radio unit sphere with center N , tangent to the complex plane on it, following a bijection. From the geometric point of view, complex numbers can be identified with the Cartesian plane points by matching the complex $Z = a + bi$ point (a, b) , as shown below:

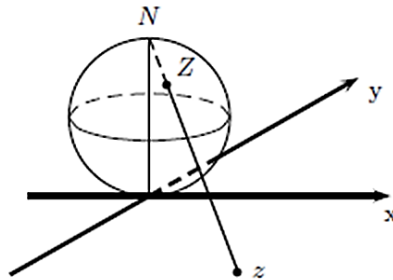


Figure 1: Riemann Sphere.

Complex numbers can be identified with the points of the Cartesian plane by matching the complex point.

There are different fractal dimensions, the simplest is the dimension of similarity Auto: $d = \frac{\log(N)}{\log(M)} \rightarrow M^d = N$; where M is the number of parts in which the object will be divided, d is the dimension of the object and N the number of resulting parts. Is used only in case the object is geometrically similar car, so that the resulting pieces are self-similar to the original object. In the case of a segment divided into three equal parts ; $d=1, M=3 \rightarrow N=3$, an area divided into three parts each side ; $d=2, M=3 \rightarrow N=9$ and a bucket ³, dividing each side into three parts; $d=3, M=3 \rightarrow N=27$.

² In a space of a single dimension (as a line), a hyperplane is one point divides a line into two lines. In a two-dimensional space (such as the xy plane), a hyperplane is a line, divides the plane into two halves. In three dimensional space, a hyperplane is a plane current, divided the space into two halves. This concept may be applied to four-dimensional space and where these objects are simply called splitters hyperplanes, since the purpose of this nomenclature is to link with the plane geometry.

³ A cube, and is a hexahedron, may also be classified as a parallelepiped rectangle straight because all sides are parallel sides and four pairs, and even as a prism with a square base and height equal to the side of the base.

The capacity dimension allows evaluation of the dimension of geometrically irregular objects. Instead of having similar auto parts resulting (N) will count the number of circles $N(r)$; where the dimension of capacity is the value of $\text{Log } N(r) / \text{Log } (1/r)$ when r tends to 0^4 . Topological dimension ⁵ describes the way in which the points of an object are connected to each other. Indicates whether the object is an edge, or a solid surface and its value is always an integer. There are several ways to determine it: i) Size of Coverage: calculate the smallest number of sets needed to cover the object, which may overlap. If each object point is covered by no more than G sets then the dimension of coverage is $d = G-1$ ii) Dimension Iterative: Based on the edges of the D -dimensional space has dimension $d-1$ as well, all three-dimensional volume can be surrounded by two-dimensional planes. Is calculated by looking for the edges of the edges up to the dimension 0 (*point*). The number of times on the operation (H) equals the dimension $d = H$.

Finally Underlying Dimension ⁶ (*embedding*): describes the space containing the fractal object. Indicates whether a line, area or volume. Its value can be an integer or a fraction and it is difficult to identify the appropriate underlying dimension. A Mandelbrot set is built according to the iterative process on a complex dynamic system with $Z_0 = 0$ and the complex constant c such that $\{f^n(z)\}_{n=1}^{\infty}$ is bounded, this follows another complex constant $|c| < 2$, otherwise $c \notin M$ and orbit $z_0 = 0$ diverges. Note that the point $Z_0 = 0$ with $c \in M$ and $f_c(z) = z^2 + c$, converted to Z_0 at one point to super attractor, as $|f'_c(0)| = |2(0)| = 0$, iterations to construct M are expressed as:

$$f^1(0) = c, f^2(0) = c^2 + c \text{ with } n = 1, 2, \dots; z_0 = 0 \ \& \ c = |c| < 2 \quad (5)$$

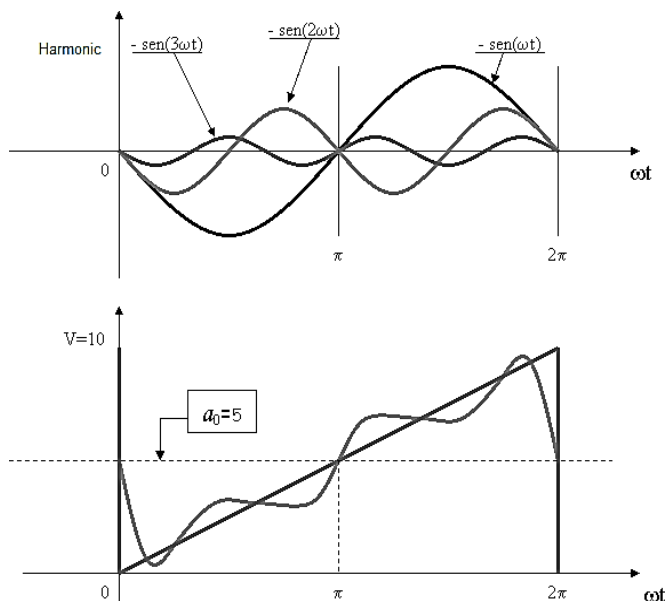
Most of the pictures of the Mandelbrot set usually appear colored depending on the speed ⁷ with each point converges to infinity, these points can be plotted according to the algorithm called "Escape time algorithm," presented below: For each point c , calculate its orbit, iterating $f_c(0)$ a number $n \propto 100$ times, if it remains bounded by the circle centered at the origin of radius 2, then we can reasonably assume that is

⁴ Marrero, J.C. Symplectic geometry and science (2002). A walk through the geometry. Department of Mathematics, University of Ohio .Pp.183.

⁵ He is interested in concepts such as proximity, number of holes, the kind of consistency (or texture) having an object, compare and classify objects and also other attributes.

⁶ It's invisible can not go in a certain factor, plus it has a monotonous process and is not completely finished it may happen that a function is not continuous throughout its domain of definition. If a function is continuous at one point, he said that the function has a discontinuity at that point and that the function is discontinuous.

⁷ Speed is a physical quantity expressing vector nature of an object displacement per unit time. She is represented by v . His unit in the International System is the m / s . By its vector character, to set the speed to be considered the direction of travel and the module, which is called speed or quickness.



within M ; if for some iteration $k < n$ "Escapes" from this circle, it is decided that does not belong to M and stops for the iteration $n + 1$, namely, $f_c^{n+1}(0)$. If each number $0 < k < n$ is assigned a color, and representing each C by Color k , for which the orbit C diverges, we get the beautiful designs that characterize the Mandelbrot sets. For there to be an aperiodic signal in prices can be represented by a harmonic series or Fourier, must respect the Dirichlet conditions:

- Having a finite number of discontinuities in the period T , in case of discontinuous.
- The average value in the period T , is finite.
- Have a finite number of positive and negative peaks.

For the case $p=0$, we have nothing but the geometric series evaluated $\frac{1}{2}$. To $p > 0$, what we have is the leading term of the derivative and of this, evaluated at that point.

Figure 2: Playing with Breasts price signals.

To find the sum, we derive p times to $1 / (1-z)$, isolate the leading term so that only depends on series of the same type with lower order powers p and then evaluate at $z =$

1/2, the analysis Fourier synthesis is called recombination of trigonometric series terms to reproduce the original signal (in our case the volatility of the price of the shares issued).

Since this signal has discontinuities, the series does not converge quickly. It is verified that with increasing the number of terms in the series, the final wave irregularities decreases and approaches the original signal.

And note that in the interval where the signal is continuous, the wave of the series converges, with some imperfections, the original signal, and places of discontinuities, the wave converges to the mean value or market stability, the series Fourier is as follows:

$$f(x) = \frac{2}{\pi} (\sin x + \frac{1}{2} \sin 2x + \frac{1}{3} \sin 3x + \frac{1}{4} \sin 4x + \frac{1}{5} \sin 5x + \dots + \frac{1}{n} \sin nx) \quad (6)$$

As in the above description, when determining the components of a vector⁸, we can determine the coefficients C_n by the inner product. Multiplying the above equation by $\phi_{m\omega(x)}$, integrating in the interval $[a, b]$ Rates of Ex Post and Ex Ante, we obtain:

$$\begin{aligned} \int_{-p}^p f(x) \phi_n(x) dx \\ = c_0 \int_{-p}^p \phi_0(x) dx + c_1 \int_{-p}^p \phi_1(x) \phi_m(x) dx + \dots \\ + c_n \int_{-p}^p \phi_n(x) \phi_m(x) dx + \dots \end{aligned} \quad (7)$$

Because of the orthogonality, each term of the right side of the last equation is zero except when $m=n$. In this case we:

$$\int_{-p}^p f(x) \phi_n(x) dx = c_n \int_{-p}^p \phi_n^2(x) dx \quad (8)$$

Then the coefficients of prices that we seek are the ranges and divide by the number of companies that have:

$$C_n = \frac{\int_{-p}^p f(x) \phi_n(x) dx}{\int_{-p}^p \phi_n^2(x) dx} \quad (9)$$

⁸ A vector field is a construction of vector calculus which associates a vector to each point in Euclidean space.

The coefficients a_0, a_1, a_2, \dots can be determined as described for the generalized Fourier series and integrating both sides, from $-p$ (cost) to p (margin), we obtain:

$$\int_{-p}^p f(x) dx = \frac{a_0}{2} \int_{-p}^p dx + \sum_{n=1}^{\infty} (a_n \int_{-p}^p \cos \frac{n\pi}{p} x dx + b_n \int_{-p}^p \sin \frac{n\pi}{p} x dx) \tag{10}$$

$n > 1$, is orthogonal to 1 in the range, the right side is reduced to a single term, and therefore:

$$\int_{-p}^p f(x) dx = \frac{a_0}{2} \int_{-p}^p dx = \frac{a_0}{2} \times \int_{-p}^p 1 dx = pa_0 \tag{11}$$

Solving a_0 is obtained:

$$a_n = \frac{1}{p} \int_{-p}^p f(x) dx \tag{12}$$

Now multiply by $\cos (m\pi x / p)$ and integrate all ranges of stock prices:

$$\int_{-p}^p f(x) \cos \frac{m\pi}{p} x dx = \frac{a_0}{2} \int_{-p}^p \cos \frac{m\pi}{p} x dx + \sum_{n=1}^{\infty} (a_n \int_{-p}^p \cos \frac{m\pi}{p} x \cos \frac{n\pi}{p} x dx + b_n \int_{-p}^p \cos \frac{m\pi}{p} x \sin \frac{n\pi}{p} x dx) \tag{13}$$

In a similar way when f is odd in the range $(-p,p)$,

$$n = 0, n=0,1,2,\dots, \quad b_n = \frac{2}{p} \int_0^p f(x) \sin \frac{n\pi}{p} x dx \tag{14}$$

Considering all the sines and cosines of the behavior of stocks we will mention, its general form for the function:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos \frac{jn\omega t}{2} + b_n \sin \frac{jn\omega t}{2}] \tag{15}$$

We are using on how to form a Fourier series expansion in very simplified:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos \frac{jn\omega t}{2} + b_n \sin \frac{jn\omega t}{2}] \tag{15}$$

$$b_n \frac{e^{j \cdot n \cdot \omega \cdot t}}{2^j} \quad (16)$$

If the Fourier series converges to: $f(x)$ for each point x where f is differentiable⁹, and b have to cost ratio (*negative*) and the function of range (*positive*):

$$f(t) = (a_0)/2 + \sum_{n=1}^{\infty} \omega_n [e^{j \cdot n \cdot \omega \cdot t} + (a_n/2 + b_n/(2^j)) + e^{-j \cdot n \cdot \omega \cdot t} + (a_n/2 - b_n/(2^j))] \quad (17)$$

We have seen therefore that the sum of multiple frequencies in the stock price are harmonically related so it gives rise to a periodic waveform having a more or less complex operation is called harmonic synthesis (Ex post or Ex ante). Conversely, a newspaper Price Range complex shape can be decomposed into several sinusoidal vibrations that are harmonically related, operation is called harmonic analysis in the Stock Market.

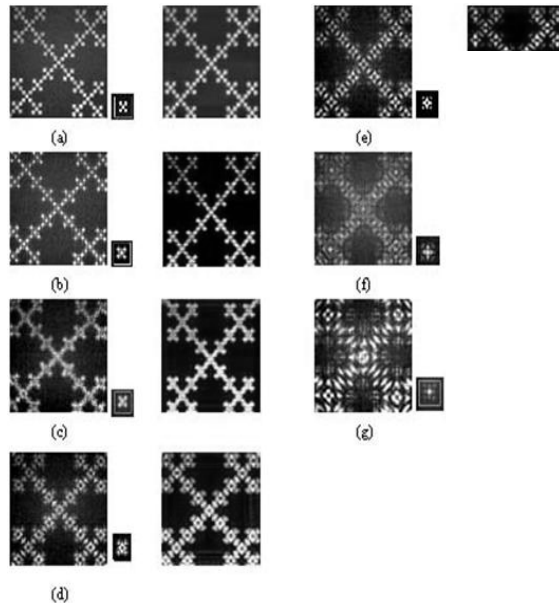


Figure 3: Sequence of the numerical propagation of a Fresnel regime.

The blade of the store and the logistic map produced some interesting patterns as the flashing if clearly by this method (as in the photo on the left), and excluded combinations can be detected, although more work than driven IFS. IFS and Kelly plots complement each other well with Fresnel regimes.

⁹ Differentiable structure is given by a maximal atlas (an atlas is a collection of cards differentiable coordinate changes). Each atlas is contained in a single maximal atlas. It is said that two maximal atlas A and A_0 on the same variety M differentiable structures defined equivalent if there exists a diffeomorphism between $(M;A)$ and $(M;A_0)$.

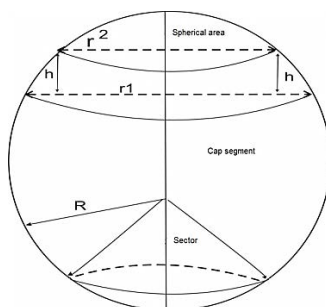
Meanwhile, some long-range correlations may be more easily seen. Furthermore, this method has more flexibility in the number of containers allocated to data. Powered

The evolution of the diffraction patterns produced by a letter Fractal ($n=4$) together with corresponding to the generator (*level fractal $n=1$*), have been obtained on the stock exchanges for the same distances z (*Ranges*), and conclude that at small distances we can reconstruct the image of fractal pivot 1-4 placing the generator providing the different positions of the initial plane.

Since what is observed is the total diffracted intensity, the phase information is not observable, and therefore can not identify the evolution of the generator with that of the entire structure, for these particular distances between the price and Ex Ante Ex Post, this indicates it is not always possible to reveal the fractal properties based only on measurements of the intensity distribution of Share Prices¹⁰.

2. Aperiodicity of flashing on the economy hypersphere global financial.

An ordinary sphere, or two-dimensional sphere consists of all points equidistant from a given point in ordinary three-dimensional Euclidean space, R^3 . A three-dimensional sphere consists of all points equidistant from a given point in R^4 (*chaos theory*)¹¹. While a two-dimensional sphere is a surface "soft" two dimensions, entirely analogous manner, it is possible to define areas of a higher number of dimensions, called hyperspheres or n-spheres. These objects are n-dimensional varieties and three-dimensional volumen (*or hiperárea*) of a hypersphere of radius r , while the four-dimensional hypervolume (The volume of the region 4 bounded by the hypersphere dimensions). The nontrivial homology groups of the hypersphere are: $H_0(S^3, Z)$ y $H_3(S^3, Z)$ are both infinite cyclic, while $H_i(S^3, Z) = \{0\}$ for all other index i .



¹⁰ The distance between two points in Euclidean space is equal to the length of the segment of straight line, expressed numerically. In more complex areas, as defined in non-Euclidean geometry, the "shortest path" between two points is a curve segment.

¹¹ It is the popular name for the branch of mathematics and physics to treat certain types of unpredictable behavior of dynamic systems.

Figure 4: Three-dimensional Sphere Capital Markets.

Homotopy groups larger ($k \geq 4$) are all finite (indicators investment margins), but also it does not follow any discernible pattern. Apparently nothing is as simple as the unit sphere in three dimensional space.

Where:

V= Operating volume.

S= Size of Market (Participation or Cognitive).

L= Size Range in Rates.

$\pi= 3.141592$

h= Maximum Price

r= Minimum Price

The fact that they can identify and R3n R2n spaces results in the odd-dimensional spheres have a fractal geometry that of even dimension. To clarify in what sense geometry is self-similar or related to another we must first introduce some concepts. The Kahler manifolds are the richest among them ¹².

Table 1: Volume and Radius of a Sphere with boundedness of S1 to Sn.

Sphere $S_{n-1} \subset R_n$	Volume	Radio=1	Volume Side
$S_1 \subset R_2$	$(\pi)r^2$	3,1415	$2(\pi)r$
$S_2 \subset R_3$	$4/3(\pi)r^3$	4,1887	$4(\pi)r^2$
$S_3 \subset R_4$	$1/2(\pi)r^4$	4,9348	$2(\pi)2r^3$
$S_4 \subset R_5$	$8/15(\pi)2r^5$	5,2637	$8/3(\pi)2r^4$
$S_5 \subset R_6$	$1/16(\pi)3r^6$	5,1677	$(\pi)3r^5$
$S_6 \subset R_7$	$16/105(\pi)3r^7$	4,7247	$16/15(\pi)3r^6$

For a circle (filled) of radius r see: P = Perimeter = $2 \cdot \pi \cdot r$, A = Area = $\pi \cdot r^2$, so that

¹² The definition of Kahler variety is as follows: it is an almost-complex endowed with a Riemannian metric g such that $g(J(X); J(Y)) = g(X; Y)$, for any vector fields X and Y, and so $rJ = 0$, be r Levy connection and then said g is a metric seal. The geometric meaning of Kahler manifolds is that the parallel transport associated with the connection of Levy commutes with the action of the almost-complex structure.

$P = 2 \cdot (\sqrt{\pi}) \cdot A^{1/2}$. Both the square and the circle have perimeters that are 1-dimensional, so that these relationships between area and perimeter.

When the angles of intersection of the circles are rational multiple 180 degrees (average market value), imposing certain relations between investments consequently establish balance in the Operating Volume of shares.

For example, it is difficult to demonstrate C_1 y C_2 intersect at an angle of $(F/N) 180$, $0 < m < n$, $(I_1 I_2)^n = \text{Identity of a Circle Fractal}$. To represent the three-dimensional sphere as a topological space and imagine that we can understand easily, using various procedures. Before mentioning some, remember certain elements that are defined in S^3 by analogy with the two-dimensional sphere, as we delimit:

- Fractal Iteration North $N = (0, 0, 0, 1)$
- Fractal Iteration South $S = (0, 0, 0, -1)$
- Fractal Iteration East $E = (0, -1, 0, 0)$
- Fractal Iteration West $O = (-1, 0, 0, 0)$

It is clear that each iteration fractal Charter is a three-dimensional ball, since:

$$\frac{x^2}{1} + \frac{x^2}{2} + \frac{x^2}{3} + \frac{x^2}{4} = 1 \tag{18}$$

$$\frac{x^2}{1} + \frac{x^2}{2} + \frac{x^2}{3} = 1 \tag{19}$$

$$-\frac{x^2}{4} \leq 1 \tag{20}$$

One way to avoid the problems that can occur when the points in a circle outside the circle are reversed (as can occur when investment circles overlap) is to prohibit all combinations involving investment from the inside out, and restricted to the limit is the limit of the orbit of a point, with the restriction that if some point in the orbit x_i disk is limited by C_j , then the next orbit point, x_{i+1} , may not $_j(x, i)$ and get 2 options:

- i) If the circles C_i bound disjoint disks, then this condition is just the familiar requirement that we have never invested in the same circle on.
- ii) If the circles intersect, this condition can be more interesting.

With this restriction, investments ¹³ will not expand the maps and the limit is restricted in the discs limited by C.

3. Conclusions.

Some instruments such as Fourier Series Harmonic and difference of terms in the Fresnel diffraction adjusting ourselves outside a Hyperspheres geometry in its segment bounded by the ranges of volume operation obtained the following result of market prices in Mexico with respect to the world (is the financial markets self-similar):

Method		Pari s	Fran fuk	Lond res	Toki o	New York	Mex ico
Harmonic	2008-	1.8	36.0	2.32	71.7	252.64	1.00
	2009	1%	0%	%	4%	%	%
	2010-	2.6	74.0	2.31	18.9	251.05	5.00
	2011	1%	0%	%	2%	%	%
Fourier series	2008-	0.7	28.0	4.67	1.73	332.53	0.03
	2009	3%	0%	%	%	%	%
	2010-	0.9	69.0	0.22	0.88	345.66	23.0
	2011	3%	0%	%	%	%	0%
Unlike terms	2008-	3.3	4.00	4.00	1.87	296.73	2.00
	2009	5%	%	%	%	%	%
	2010-	3.9	4.00	2.08	1.29	264.31	3.00
	2011	6%	%	%	%	%	%
Three- dimensional sphere	2008-	1.8	0.04	13.9	0.05	97.37	1.80
	2009	7%	%	3%	%	%	%
	2010-	1.8	0.04	7.73	0.04	87.41	2.03
	2011	3%	%	%	%	%	%
Fresnel diffraction	2008-	1.5	1.00	0.93	0.43	25.68	78.0
	2009	7%	%	%	%	%	0%
	2010-	1.5	1.00	1.03	0.66	23.15	78.2
	2011	3%	%	%	%	%	0%

Regarding the scalar quantization represents each value with an index to a table set consists of a subset of securities called codebook (*A-SF-DF-ET-DF*).

¹³ Represent loans of money upon which a company expects to obtain some future performance, either, for the realization of an interest, dividends or by selling to a higher value on acquisition cost. These are loans of money which a company or organization decides to keep for a period exceeding one year or operating cycle, counting from the date of the balance sheet.

With respect to the Harmonic Series Fourier difference Dimensional Sphere Terms and the highest yield is in New York with twice the investment and an acceptance rate of 52.64%-45.66%-64.31% and 7.41% respectively.

With respect to the Fresnel diffraction is generated the highest return our Mexican Stock Exchange with a margin of 78%, syntactic techniques for generating fractals discussed with these five methods are a pleasant and familiar natural almost fractal sets with low R^2 , although its usefulness to larger spaces is almost immediate.

One reason for its popularity is that objects are processed really are symbols related to geometric primitives rather than numerical developments may be less easy to understand (*more so in our field of action*). The idea is to generate certain default rules through a sequence of chains converging to a fractal (*Mandelbrot set*). The study of fractals is transferred in this way, regardless of the initial space dimension, the domain of infinite words.

In every transaction there is a chance that the price changes, and after a certain time horizon, there is a total change in the price. We got the price change (*since the cumulative distribution obeys a cubic law conversely, the probability distribution function for differentiation*) and obeys a law quartic (*fourth time*) reverse.

4. References.

1. Barnsley, M. *Fractals Everywhere*. Academic Press. San Diego. (1993).56-62.
2. Braun E. *Chaos, fractals and strange things*, Collection of Science in Mexico. No. 150 Fondo de Cultura Economica, Mexico. (1994).45-56.
3. Gleick, J. *Chaos*. Making a New Science. Freeman. San Francisco. (1983).12-43.
4. Guzman, M., Martín, M. ángel, Morán, M., Reyes, M. *Fractal structures*. Labor, Barcelona. (1993).32-46 .
5. Mandelbrot B. *Characteristics of Fractals*. Amsterdam. (1982).43-47.
6. Mandelbrot, B. *The behavior of markets, Fractal view of risk, ruin and reward (Basic Books)*. Washington. (2004).45-32.
7. Mantegna R.(1997). *An Introduction to Econophysics, Correlations and Markets*. University Press. Princenton.pp.67.71
8. Petters E. *Chaos and Order in Capital Markets: A New View of Cycles, Prices and Market Volatility*. Wiley. New York. (1996).57.

9. Ramos, M. *Present Economic Dynamics*. Edit. ECORFAN. Mexico. (2010).45.
10. Talanquer V. *Fractus, Fracta, Fractal, (Fractals of labyrinths and mirrors)*, Economic Culture Fund, Series Science from Mexico /147, Mexico. (1996).37.
11. Bouchaud, J. & Poters, M. The leverage effect in financial markets: retarded volatility and market panic, *Physica*. (2008).13.
12. Calvet, L. Fisher ,A. Mandelbrot, B. *Large deviations and the distribution of price changes. Cowles Foundation Discussion. Paper Núm.1165*.67.
13. Fama, E. *The distribution of daily differences of stockprices, a test of Mandelbrot's stable paretian hypothesis*. (1964) .34-47.
14. Mandelbrot, B. *Multifractal measures, especially for the geophysicist". Pure and Applied Geophysics*. (2004).89-92.
15. Mandelbrot, B. Negative fractal dimensions and multifractals. *Physica*. (2004).67.
16. Bakshi, G. Dilip,M. *What is the probability of a stock market crash?*. NYU Conference on Finance and Accounting. (1998).
17. <http://deutscheboerse.com>
18. <http://www.bmv.com.mx>
19. <http://www.euronext.com>
20. <http://www.londonstockexchange.com>
21. <http://www.worldfederationofexchanges.com>

ITERACIÓN FRACTAL DE COMPUTO IFS EN LOS MERCADOS FINANCIEROS

MARÍA- RAMOS ESCAMILLA

maria_ramos@my.unitec.edu.mx

*Universidad Tecnológica de México/Departamento de Posgrado
Calz Ermita Iztapalapa 557 Granjas Esmeralda, Iztapalapa, 09810 Ciudad de México, Distrito Federal*

MARIA JESUS- SEGOVIA VARGAS

mjsegovia@ccee.ucm.es

*Universidad Complutense de Madrid /Departamento de Economía Financiera y Actuarial
Pabellón de Quinto Curso, Campus de Somosaguas 28223- Madrid*

MARTA- MIRANDA GARCÍA

martamiranda@ccee.ucm.es

*Universidad Complutense de Madrid /Departamento de Economía Financiera y Actuarial
Pabellón de Quinto Curso, Campus de Somosaguas 28223- Madrid*

RESUMEN: En este artículo presentamos un análisis de precios fractal de las acciones emisoras que cotizan en el mercado de capitales, tomamos de herramienta matemática la modelación de sistemas de funciones iteradas, nuestro objetivo es la determinación de los cardiodes de Mandelbrot para la fijación de soportes y resistencias en las tendencias del rango de precios y sostener la hipótesis central que es la maximización del margen estocástico de los precios accionarios de la bolsa de valores en México con sus auto afines internacionales Frankfurt, Londres , Paris , Tokio y New York y representarlas con técnicas chartistas y de mapeo fractal en sus opciones de compra y venta.

Palabras clave: Fractal, media dorada, pivoteo, caos, mapeo de IFS.

ABSTRACT: In this paper we present a fractal analysis of prices of shares listed stations in the capital market, we take mathematical modeling tool iterated function systems, our goal is the determination of the Mandelbrot cardioid for setting support and resistance trends in the price range and support the central hypothesis that margin maximization is stochastic stock prices of the stock market in Mexico with international related self Frankfurt, London, Paris, Tokyo and New York and represent them with technical chartists and fractal mapping in their purchasing and selling.

Keywords: Fractal, golden mean, pivoting, chaos, IFS mapping.

Introducción.

Nuestro análisis de precios consistirá en describir y aplicar diversas técnicas fractales al mismo tiempo que establecemos las comparaciones pertinentes a los mercados de valores que son afines en cuanto número de emisoras de su actividad bursátil, tal es el caso de Frankfurt (*DAX30*), Londres (*TSE*), París (*CAC-40*), Tokio (*NIKKIE-225*) y New York (*Standar And Poor's*) respecto de México (*IPyC*). Para apreciar el valor y la diversidad del enfoque fractal-tridimensional, tomamos a [Mandelbrot, Benoit, 1982]: “Pienso que la difusión de la teoría del caos y la geometría fractal en las ciencias sociales es esencial para nuestra futura evolución, tal como el arte de contar fue esencial en los orígenes de nuestra cultura”¹. No existe una definición rigurosa que delimite con precisión matemática si un determinado conjunto es o no un fractal.

El desarrollo de este artículo presenta una alternativa de solución al problema de la determinación del costo y margen de las acciones emisoras respecto de su precio de mercado ahincando sobre el valor máximo del margen de operación estocástico, para eficientar la actividad bursátil de los inversionistas en la bolsa de valores por ello la importancia de la modelación fractal aplicada a los índices bursátiles internacionales respecto de México.

Por ello se acotan dos secciones en la primera se desarrollara la logística fractal de los mercado considerando la diversificación de su tipología en obtención de las memorias cortas y largas para las posibilidades de bifurcación del mercado accionario y en la segunda sección se aplicara el método fractal-tridimensional a las operaciones bursátiles de las acciones de los mercados internacionales, obtendremos los ruidos caóticos de la combinatoria de Mandelbrot y sus grados de anisotropía de iteración IFS Ex ante y Ex Post que parametrizan el cardiode principal además utilizaremos representaciones técnicas y chartistas para la modelación final vía matrices de complejidad del mapeo a los precios.

Finalmente las razones por las que utilizamos esta metodología es porque en el proceso de praxis de mercado, las tendencias no son lineales y las gráficas no representan la totalidad del movimiento de las acciones en teneduría como se ha venido estudiando tradicionalmente, por ello nos surge esta idea de análisis fractal para los índices de precios ya que esta modelación nos da otra alternativa de cambios absolutos y relativos para encapsular en tiempos recurrentes las tendencias de los soportes y resistencias lo que nos arroja una maximización del margen estocástico del precio de las acciones y nos ayuda a darle otra perspectiva al análisis cuantitativo del mercado de capitales.

¹ Mandelbrot, Benoit, mostró cómo los fractales pueden aparecer en ámbitos diferentes, tanto en matemáticas como en otros aspectos de la naturaleza y desarrolló no sólo nuevas ideas matemáticas, sino que también tuvo que desarrollar algunos de los primeros programas computacionales para imprimir gráficos.

1. Definición, tipología y logística fractal del mercado de capitales.

El término “fractal proviene del latín –fractus- que significa -fragmentado, fracturado, o simplemente roto o quebrado. Se aplica al conjunto de formas generadas normalmente por un proceso de repetición, se caracterizan por poseer similitud en toda escala, por no ser diferenciables y por exhibir dimensión fraccional de $\frac{3}{4}$.

El proceso de repetición al que se hace referencia, recibe el nombre de iteración. Para definir a los fractales, consideramos la original de Mandelbrot: Un objeto fractal tiene formas geométricas con una dimensión “fraccional” (*no entero*) con las siguientes características [Mandelbrot, B, 1982]:

- i) Es demasiado irregular para ser descrito en términos geométricos tradicionales.
- ii) Tiene autosimilaridad exacta, ya que posee detalle a cualquier escala de observación.
- iii) Su dimensión de Hausdorff-Besicovitch² no es entera ya que es estrictamente mayor que su Dimensión Topológica [Douketis, Constantine., Wang, Zhouhang., Haslett, Tom L., Moskovits, Martin: 1995].
- iv) Puede ser definido recursivamente, mediante los siguientes algoritmos:

IFS (Sistema de Funciones Iteradas): Unos conjuntos se reemplazan recursivamente por su imagen bajo un sistema de aplicaciones, ejemplos de esto son ; El conjunto de Cantor, la alfombra de Sierpinski, el triángulo de Sierpinski, la curva de Peano, la curva del Dragón, el copo de nieve de Koch o la esponja de Menger.

Fractales de tiempo de escape: Definidos por una relación de recurrencia en cada punto del espacio (*por ejemplo, el plano complejo*): El conjunto de Mandelbrot, conjunto de Julia, y el Fractal de Lyapunov.

Fractales aleatorios: Generados por procesos estocásticos, no deterministas; El movimiento Browniano, el vuelo de Lévy, los paisajes Fractales o los árboles Brownianos. Éstos últimos son producidos por procesos de agregación por difusión limitada.

En términos prácticos, un fractal es una estructura que está compuesta por

² Hausdorff-Besicovitch en 1914, usando el axioma de elección, obtuvo una descomposición "paradójica" de la 2-esfera como la unión disjunta de cuatro conjuntos A, B, C y Q, donde Q es numerable y los conjuntos A, B, C y BC son mutuamente congruentes. Esto inspiró más tarde la descomposición de la esfera en tres dimensiones de Banach-Tarski. Hausdorff introdujo asimismo los conceptos medida de Hausdorff y dimensión de Hausdorff, que son cruciales en el estudio de la teoría de fractales. En Análisis, resolvió lo que llamamos hoy problema del momento de Hausdorff.

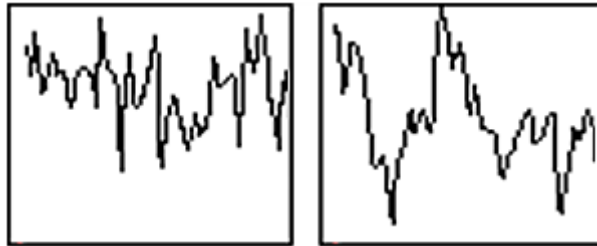
pequeñas partes, las cuales son parecidas a la Figura original, que se repiten en diferentes escalas, desde grandes (*macro*) hasta pequeñas (*micro*) [Matthieu ,W, Philippe B ,J, 2002].

El todo imita a las partes (*y viceversa*), el enfoque fractal revela que el microcosmos es similar al macrocosmos y se constituye en un nuevo campo de las matemáticas e interviene en el cambio de los paradigmas de las ciencias, en nuestro caso de la economía financiera. Las pequeñas diferencias en las condiciones iniciales pueden crecer en grandes cambios en los valores de las acciones en los mercados. Esta es la sensibilidad a las condiciones iniciales (del precio *Ex post*), es uno de los atributos del caos, su patrón es idéntico al producido por un IFS azar con ciertos pares de precios excluidos [Calvet, Laurent., Fisher, Adlai., Mandelbrot, Benoit: 1997]. Otros requieren combinaciones más excluidas (triples, cuádruples, etc.), el número de combinaciones es excluido de esta medida de la complejidad.

a) Memoria Larga

b) Memoria Corta

Figura 1: Mapa de logística fractal.



El mapa logístico está definido por una parábola, el mapa de la tienda por una línea quebrada, tanto simétrica respecto de $x = 1/2$. Para todas nuestras pruebas, utilizamos los puntos de juego del caos fijando los vértices $(0,0)$, $((\sqrt{3})/2, 1/2)$ y $(0,1)$ de un triángulo equilátero.

Las funciones de prueba para nuestra investigación de dinámica económica financiera son: el mapa de logística y el mapa de la tienda³. La altura da el parámetro del total de la actividad financiera (para nuestra investigación es el número de emisoras bursátiles que existan en el Mercado de Capitales).

³ Algoritmo de condiciones iniciales que generan un mapa de tienda en R3:

$$T^1(x,y) = (x/2, y/2)$$

$$T^2(x,y) = (x/2, y/2) + (0, 1/2)$$

$$T^3(x,y) = (x/2, y/2) + ((\sqrt{3})/4, 1/4)$$

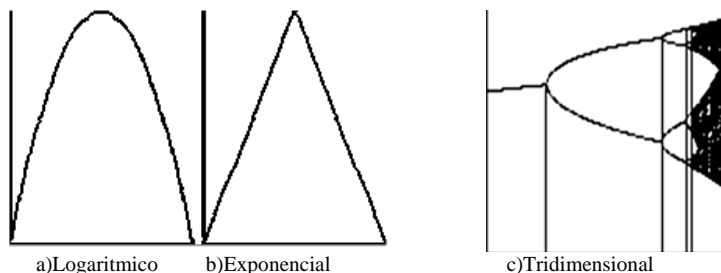


Figura 2: Parámetros de un fractal.

Para la iteración gráfica, los puntos fijos son considerados como las intersecciones de $y = f(x)$ y $y = x$.

Rompiendo el Mapa de la tienda en una función discontinua y la parte de flexión de la gráfica da un diagrama de bifurcación mucho más interesante. El lado izquierdo del diagrama se parece mucho a la de la hoja de tienda de campaña (*curva de Gauss*), pero el lado izquierdo muestra una secuencia de ventanas periódico se entretejen en el mismo orden que los discos conectado todo el cardioide del conjunto de Mandelbrot. La Iteración⁴ gráfica de un fractal se produce mediante la órbita de generación de los puntos (x_0, x) , (x_1, x_2) , (x_2, x_3) ,... en un sistema tridimensional:

$$(x, y, z) \tag{1}$$

Donde:

- x = Rango de precios máximos.
- y = Rango de precios mínimos.
- z = Logaritmo de acciones en circulación.

A partir de $x = x_0$, con una línea vertical de la gráfica $y = f(x)$ de la que se reiteramos la función, se cortan en $(x_0, f(x)) = (x_0, x_1)$, y volvemos a trazar la línea horizontal (*definiendo las fluctuaciones de mercado*) desde este punto hasta la línea diagonal (*definiendo la tendencia del mercado*) $y = x$, pues cortamos en (x_1, x_1) , para representar el precio real contra el precio logarítmico [Stewart, Ian: 2003].

El diagrama de bifurcación de mercado es un registro de los precios de las acciones de la órbita final (*trazado vertical-Prospectación del Precio*) para cada uno de una secuencia de valores de los parámetros (*trazado horizontal- Estado Ex ante del Precio*). El mapa de retorno es una gráfica de puntos (x_0, x_1) , (x_1, x_2) , (x_2, x_3) , (x_3, x_4) ,...

Para las órbitas de las funciones de $x_{i+1} = f(x_i)$, el mapa de retorno sólo produce los puntos de la gráfica $y = f(x)$. A medida que aumenta el precio mínimo de

⁴ Iteración real : $(V + i \cdot w) + (x + i \cdot y) = (v + x) + i \cdot (w + y)$
 Iteración compleja : $(v + i \cdot w) \cdot (x + i \cdot y) = (v \cdot x - w \cdot y) + i \cdot (v \cdot y + w \cdot x)$
 Iteración fractal : $(V + i \cdot w) \cdot (x + i \cdot y) = (v \cdot x - w \cdot y) + i \cdot (v \cdot y + w \cdot x)$

$f^n(x)$ disminuye hasta que toquen $y = x$, los puntos de formación de un nuevo ciclo (es decir su precio máximo). Así bifurcaciones tangente dar lugar a un par de ciclos, uno estable, la inestabilidad de otros, por esto la reivindicación de métodos Fractales.

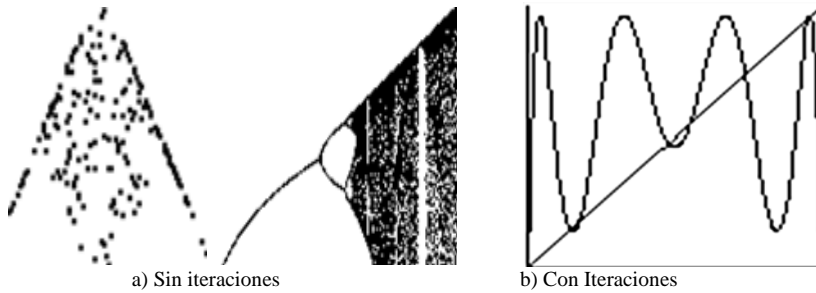


Figura 3: Diagrama de bifurcación.

El mapa de retorno es una gráfica de puntos (x_0, x_1) , (x_1, x_2) , (x_2, x_3) , (x_3, x_4) , las órbitas de las funciones de $x_{i+1} = f(x_i)$ el mapa de retorno sólo produce los puntos de la gráfica $y = f(x)$.

Las distancias entre el periodo de duplicación de bifurcaciones sucesivas llevaron al descubrimiento de un nuevo tipo de escala, y una nueva constante matemática la media dorada de 0.618 para lo que surge en una gran clase de funciones de prospectación en precios de acciones.

Como un rango de precios (máximo o mínimo) del sistema es muy variado, la dinámica puede llegar a ser más complicada por periodo de duplicación en las bifurcaciones del logaritmo de las acciones. El precio en un N -ciclo se vuelve inestable, un $2N$ estable ciclo aparece: cada punto del ciclo da lugar a dos puntos de la $2N$ -ciclo. Como el precio sigue aumentando, el $2N$ -ciclo se vuelve inestable, y cada punto da lugar a dos puntos de un estable- $4N$ ciclo, y así sucesivamente (en un plano bidimensional).

2. Geometría de la naturaleza en los precios de las acciones.

Uno de los procedimientos para caracterizar e incluso para clasificar los objetos fractales consiste en atribuir a cada uno de ellos una cantidad numérica, la dimensión fractal [Braun, E: 1996]. Por medio de este índice matemático se puede cuantificar la geometría de los objetos o de los fenómenos fractales. Así, cuantificar fractales será definir, por algún procedimiento, la proporción del espacio físico que es llenado por ellos. Encontramos una diferencia fundamental con los objetos euclidianos: si magnificamos sucesivamente un objeto euclidiano "unidimensional", observamos segmentos rectilíneos. Sin embargo, si magnificamos sucesivamente un objeto fractal, encontramos objetos con niveles de complicación comparables a los del conjunto de partida. El punto de partida es una información del mercado de capitales, se procesa y se obtiene una indexación. Éste se procesa de nuevo (se itera) y se obtiene otro rango de precios similar al anterior y se continúa haciendo lo mismo indefinidamente con cada

precio de las acciones.

La dimensión topológica, es la siguiente:

Tabla 1: Dimensión topológica.

Nivel de dimensión	Efecto dimensional
Inexistencia de dimensión	Un punto.
1° Dimensión	Una línea recta.
2° Dimensión	Un plano.
3° Dimensión	El espacio.
4° Dimensión	Un fractal.

El uso del caos para este fenómeno se busca en los bordes de los bordes hasta alcanzar la dimensión 0 (punto) hasta ir iterando y formar un fractal.

La transformación de similitud o escalamiento consiste en generar una copia similar de un precio de acción cualquiera en una escala diferente⁵. Para lograr esto, el precio *Ex ante* se debe afectar por un factor de proporcionalidad (*delta del precio = 3*), mismo que se denomina factor de escalamiento [Mantegna, R. N., Stanley, H. E: 1995]. De este modo, dos precios (*Ex ante* y *Ex post*) son similares si poseen la misma geometría, aunque tengan diferente valor nominal y número de acciones emitidas. Esto se puede expresar de una manera general, tal que si tenemos un precio y se elige modelarlo con logaritmos y se amplifica con un factor de escalamiento delta. Se observa una geometría idéntica al rango de precios (*precio máximo-precio mínimo*).

Si se toma este último rango y se amplifica nuevamente por el mismo factor de escalamiento, seguramente se obtendrá una geometría similar al precio *Ex ante*. Esta operación se puede repetir indefinidamente. La propiedad de autosimilitud en un fractal-tridimensional y se presenta en todo el intervalo de escalas. Se dice que el precio de una acción es autoafín cuando permanece invariante bajo una escala de transformación anisotrópica (*diferentes precios en todas las acciones*).

La combinatoria de Mandelbrot es el más conocido de los conjuntos fractales, y nosotros lo presentamos en \mathbb{R}^3 , definiéndolo así, en un plano complejo: Sea C un número complejo cualquiera, entonces se dice que C pertenece al conjunto de Mandelbrot, y si no, queda excluido del mismo. Por ejemplo, si $c = 1$ obtenemos la sucesión 0, 1, 2, 5, 26... que diverge. Como no está acotada, 1 no es un elemento del conjunto de Mandelbrot. En cambio, si $c = -1$ obtenemos la sucesión 0, -1, 0, -1,... que sí es acotada, y por tanto, -1 sí pertenece a la combinatoria de Mandelbrot. A menudo se

⁵ El escalamiento en \mathbb{R}^4 se obtiene con los siguientes pivoteos fractales:

$$\begin{aligned}
 1 &= 2 \cdot ((1/2)^d) + 2 \cdot ((1.4)^d) + 2 \cdot ((1/8)^d) + n \\
 &= 2 \cdot ((1/2)^d) + 2 \cdot (((1/2)^d) 2) + 2 \cdot (((1/2)^d) 3) + n \\
 &= 2 \cdot ((1/2)^d) \cdot (1 + (1/2)^d + ((1/2)^d) 2) + n
 \end{aligned}$$

representa la Combinatoria mediante el algoritmo de tiempo de escape [Ramos, María: 2010].

En ese caso, los colores de los puntos que no pertenecen al conjunto indican la velocidad con la que diverge (*tiende a 1000 iteraciones con media 0.618*) la sucesión correspondiente a tener un margen de operación en la acción que itera al precio de la acción en 50 fórmulas y genera 500,000 movimientos en el espacio al iterar 10,000 veces. A pesar de sus diferencias, en una escala de transformación, las direcciones no son completamente independientes. Si al hacer un zoom, uno de los ejes de coordenadas se transforma en un factor b , $x \rightarrow bx$, el resto de los ejes coordenados deben ser reescalados en un factor b^{α_i} , $x_i \rightarrow b^{\alpha_i}x_i$, con el objeto de preservar el conjunto invariante [Courtault, Jean Michel., Kabanov, Yuri., Bru, Bernard., Crépel, Pierre., Lebon and Arnaud Le Marchand, Isabelle :2000].

Los exponentes α_i son llamados exponentes de Hurst y nos indican cuál es el grado de anisotropía del conjunto. A este tipo de quebrados se les llama fracciones continuas.

Por lo tanto, la media dorada se obtiene también como una fracción continua [Guzmán, M., Martín, M. Ángel, Morán, M., Reyes, M: 1993], y su valor es de 0.618. Es decir, procesos matemáticos de creación de estructuras Fractales son iteraciones de reglas simples de objetos iniciales.

Pequeños cambios crean enorme variedad de patrones tanto a nivel micro como a nivel macro [Weick, K: 1979]. La creatividad de la naturaleza, según parece, viene de este procedimiento iterativo. Hay varias técnicas para determinar la dimensión Fractal-Tridimensional, las técnicas basadas en precios del mercado exhiben correlación temporal (*es decir, las acciones volátiles tienden a ocurrir con una determinada tendencia respecto de su precio*) con distribuciones de precios con cola ancha (*los rangos Ex post podrían ocurrir con mayor frecuencia que la descrita por una distribución normal*).

Por tal motivo, las técnicas tradicionales basadas en los modelos lineales no reflejan correctamente la volatilidad. A continuación detallaremos los mercados de capitales que son fuente de nuestro estudio en geometría de los precios de acciones.

Tabla 2: Perfil de los mercados de capitales.

Mercado de capitales	Índice bursátil	Número de emisoras
Bolsa de Frankfurt	DAX-30	235
Bolsa de Londres	TSE	247
Bolsa de Paris	CAC-40	138
Bolsa de Tokio	NIKKIE-225	104
Bolsa de New York	Standar And Poor's	76
Bolsa de México	IPyC	130

Estas bolsas de valores son auto afines y auto similares entre si por los valores de inscripción y mantenimiento bursátil.

Los sistemas complejos en estos mercados de capitales cumplen con la propiedad de mantener en bursatilidad procesos de precios y emisión de acciones con autocorrelación fractal y pueden ser diseñados en gran detalle, variaciones pequeñas o fluctuaciones pueden ser amplificadas mediante procesos iterativos y crean los cambios cualitativos en el nivel macro.

Una de las características más espectaculares de estos fractales, es que son no derivables en todos sus puntos. En lenguaje menos matemático: una curva cualquiera es no derivable en un punto cuando, aun existiendo ese punto, forma un pico o esquina, ya que cualquier otro punto más cercano o lejano, por la derecha o por la izquierda, sí es derivable con la Combinatoria de Mandelbrot ya que uno de nuestros objetivos va a ser modificar esas fórmulas para producir mutaciones que se ajusten a la correlación de precios.

El primer paso es considerar a $z =$ precio máximo (Ex ante y Ex post), a $c =$ precio mínimo (Ex ante y Ex post), y dar un enter a nuestro sistema para que nos arroje el resultado de las 500,000 iteraciones en el espacio.

Tabla 3: Iteraciones Ex ante de la emisora WALMEX.

Emisora	WALMEX	Z=	39.11	C=	38.56		
Z ₁ =	50.58513742	Z ₁₆ =	26.08760587	Z ₃₁ =	77.64347535	Z ₄₆ =	33.2479089
Z ₂ =	78.09506263	Z ₁₇ =	29.13950151	Z ₃₂ =	77.05041994	Z ₄₇ =	10.4805927
Z ₃ =	67.95983424	Z ₁₈ =	29.19674603	Z ₃₃ =	78.59983452	Z ₄₈ =	58.4629619
Z ₄ =	40.69433695	Z ₁₉ =	23.39856276	Z ₃₄ =	65.52795958	Z ₄₉ =	40.6411476
Z ₅ =	31.25345071	Z ₂₀ =	27.80470046	Z ₃₅ =	77.63413358	Z ₅₀ =	39.0734201
Z ₆ =	27.19941182	Z ₂₁ =	32.75610994	Z ₃₆ =	39.31394305		
Z ₇ =	45.89680474	Z ₂₂ =	45.04906488	Z ₃₇ =	40.66004884		
Z ₈ =	68.13814246	Z ₂₃ =	1.424980935	Z ₃₈ =	40.66004884		
Z ₉ =	40.69433695	Z ₂₄ =	33.13538462	Z ₃₉ =	35.69270353		
Z ₁₀ =	51.52711699	Z ₂₅ =	11.14210889	Z ₄₀ =	31.73136673		
Z ₁₁ =	79.02599595	Z ₂₆ =	30.30640645	Z ₄₁ =	13.74828976		
Z ₁₂ =	45.16280142	Z ₂₇ =	31.89019777	Z ₄₂ =	38.07663411		
Z ₁₃ =	23.40416147	Z ₂₈ =	33.01048779	Z ₄₃ =	32.90980464		
Z ₁₄ =	23.47771387	Z ₂₉ =	11.88943781	Z ₄₄ =	39.15057903		
Z ₁₅ =	38.06785611	Z ₃₀ =	40.69433695	Z ₄₅ =	12.10352767		

Posteriormente hay que determinar un rango ponderado de las iteraciones, en las 130 emisoras, para la bolsa de México:

$$P = (Z1 - Z50) / 50 \quad (2)$$

Promedio 1.08%

El resultado es 1.08%, lo que nos representa un margen positivo de 1.08-0.618 = 0.382% como utilidad del precio de la acción a nivel micro respecto de su función cognitiva, es decir solo en esta empresa, mientras que a nivel macro tenemos:

Promedio 1.66%

Tenemos 1.66%, lo que nos representa un margen positivo de 1.66-0.618 = 0.982% como utilidad del precio de la acción, pero en todo el mercado mexicano durante el periodo (2007-2010), acorde a su función de participación. Seguido de estos cálculos, vamos a representar a la combinatoria de Mandelbrot, con y sin iteraciones en el espacio, cabe destacar que la asociación con cada disco y cardioide.

Tabla 4: Cardioides y discos Ex ante de la emisora WALMEX.

1.66	Promedio
1.66^{*2}	3.32
1.66^{*3}	4.98
1.66^{*4}	6.64
1.66^{*5}	8.32
1.66^{*6}	9.96
1.66^{*7}	11.62
1.66^{*8}	13.28
1.66^{*9}	14.94

La media dorada son sus secuencias de orden jerárquico para que la relación entre el tamaño de las copias sucesivas se aproxima a un valor límite que represente los escapes en los precios de las acciones, para recorrer el proceso, escoja un precio máximo "z" complejo y un precio mínimo "c".

El límite de la combinatoria de Mandelbrot contiene infinitamente muchas copias del conjunto de Mandelbrot. De hecho, tan cerca cómo se mira a cualquier punto de la frontera, encontrará infinidad de Mandelbrots y su frontera es tan "difusa", que es 3-dimensional, supera la media dorada son sus secuencias de orden jerárquico para que la relación entre el tamaño de las copias sucesivas se aproxima a un valor límite que represente los escapes en los precios de las acciones [Kenneth ,Falconer: 1990], para recorrer el proceso, escoja un precio máximo "z" complejo y un precio mínimo "c". Luego de generar la secuencia de los números complejos z_1, z_2, z_3, \dots por:

$$\begin{aligned} z_1 &= z^{0-2} + c \\ z_2 &= z^{1-2} + c \\ z_3 &= z^{2-2} + c \end{aligned} \quad (3)$$

Y, en general:

$$z_{n+1} = z_n^2 + c \tag{4}$$

El análogo, se pueden definir para cualquier z ($n + c$), para cualquier entero $n > 0.618$. He aquí un ejemplo del efecto del precio máximo de iteraciones en la elaboración de la periferia de un halo de copias pequeñas de toda la combinatoria, cada uno de los cuales está rodeado por su propia aureola de ejemplares que se encuentran escape, y así sucesivamente, en las escalas más pequeñas, sin fin [Kenneth, Falconer: 1990]. Pese a las apariencias, estas pequeñas copias se adjuntan al cuerpo principal de la serie, a través de una secuencia de copias aún más pequeñas, ya que es conectado a los ciclos estables dentro de un comportamiento caótico y su restricción denota atención a sólo números reales.

El régimen de iteración Mandelbrot revela una interesante relación con el conocido esquema logístico mapa bifurcación.

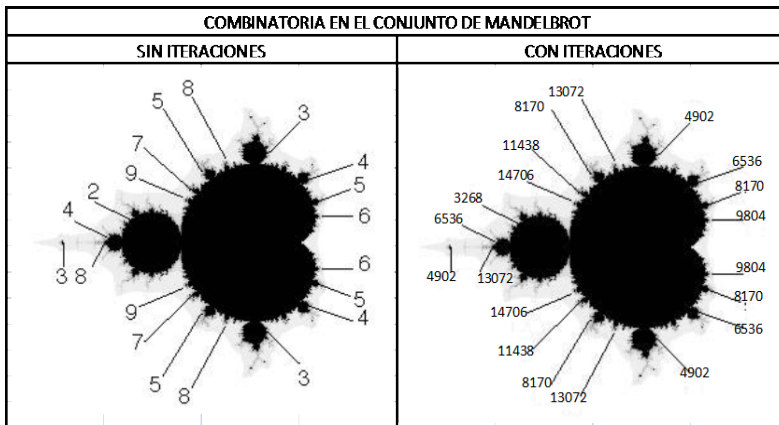


Figura 4: Escalas de la combinatoria de Mandelbrot Ex ante en bolsa mexicana.

Si $g(x)$ es una función par, $g(x)=g(-x)$, los términos b_i son nulos, si $g(x)$ es impar $g(x)=-g(-x)$, los coeficientes a_i son nulos, si $g(x)$ es alternada, $g(x+p)=-g(-x)$, la serie solamente consta de términos armónicos impares.

Las series principales del primer patrón de los ciclos de discos conectado a la cardioide principal de Mandelbrot secuencia de búsqueda de los ciclos de todos los discos de otras inherentes a la cardioide principal $\frac{1}{2}$ regla n un modelo para los tamaños de los discos adjunta a la cardioide principal, y un mapa de salida de este modelo en los ciclos de discos conectado a los discos y cardioides las normas de discos conectado a los discos y otros cardioides es lo mismo que las normas de discos conectado a la cardioide principal que corresponde a un 2 estable del ciclo, 4 ciclo, ciclo de 8, ..., y así sucesivamente. Se llama el período de duplicación en cascada para medir el diámetro de

los componentes con el fin de descubrir un patrón de cómo los componentes se encogen y sus límites de las proporciones de los diámetros de componentes sucesivas es un ejemplo de la ampliación de Feigenbaum.

El enfoque de fractal-tridimensional a esta extensión del método de Newton es difícil: la gráfica de una función compleja que es de cuatro (*real*) de objetos tridimensionales (x,y,z). Aquí la fórmula es útil, porque se puede aplicar a los números complejos, así como a los números reales.

Respecto del análisis técnico para modelación logarítmica de los precios, también llamado “chartismo”, es una técnica que trata de predecir la evolución de los precios basándose en el pasado, consideran que los cambios en la oferta y demanda pueden anticiparse observando las pautas en los movimientos de los precios de las acciones [Madan, Dilip B, Haluk Unal: 1998]. Conviene referir algunas reglas derivadas de este modelo, una básica es la regla del oscilador de la media móvil (>0.618) y la regla del filtrado. En la regla del oscilador de la media móvil, las señales de compra y venta se generan con dos medias móviles del índice, una a corto plazo (*Ex ante*) y otra a largo plazo (*Ex post*). Es decir, se recomienda comprar cuando el precio sube un tanto por ciento por encima del último pico y vender cuando baja un tanto por ciento por debajo del último mínimo. La representación de datos es la siguiente:

Tabla 5: Análisis técnico de las variaciones en el precio de CEMEX.

Variación	Máximo Ex post	Mínimo Ex post	Máximo Ex ante	Mínimo Ex ante
-0.73	14.13	13.6	33.06	5.79

Las estadísticas de operación son los rangos de los precios máximos y mínimos de teneduría e inscripción.

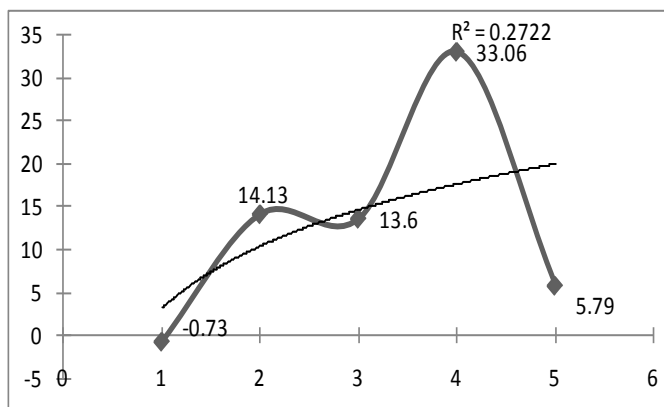
Realizamos la tendencia alcista o bajista real, es decir en términos nominales para después llevarlos al plano logarítmico, el cual nos arroja el costo o margen de aceptación, como lo mostramos en la siguiente tabla:

Tabla 6: Análisis de tendencias en el precio de CEMEX.

CEMEX	Real	Logarítmico
Variación	-0.73	-0.13667714
Máximo Ex post	14.13	1.150142162
Mínimo Ex ante	13.6	1.133538908
Máximo Ex ante	33.06	1.519302849
Mínimo Ex ante	5.79	0.762678564
Real	65.85	4.428985343
Logarítmico	Margen	8.89%

Para cada q de la ecuación $(1 p q) (r 1 \beta (q)) + \dots + (P N q) (r \beta N (q)) = 1$, si determinamos un único valor de $\beta (q)$, para cada q definir una función $g (\beta)$ por $g (\beta) = (p 1 q) (r 1 \beta) + \dots + (P N q) (r N \beta)$ ciertamente, $g (\beta)$ es una función continua.

Los valores siguen una pauta de comportamientos reconocibles, identificables y repetitivos en forma de patrones o Figuras [Rolf ,W., Banz, William., J. Breen: 1986]. Por esta razón tiene el valor de resultar predictivas y establecer la base matemática de la evolución del precio en una masa económica[Ramos, María: 2010] que pasa del optimismo al pesimismo en un determinado período de tiempo ya que el mercado sigue un determinado modelo que le permite describir pautas que son repetitivas en cuanto a su forma, pero no en cuanto a su tiempo y amplitud debido a sus movimientos en la dirección de la tendencia principal tienen una estructura de cinco ondas (*Variación, Máximo y Mínimo Ex ante al igual que Máximo y Mínimo Ex post*), como se muestra en la siguiente gráfica:



Gráfica 1: Análisis chartista con nivel de confianza al margen en el precio de CEMEX.

En la regla del oscilador de la media móvil, las señales de compra y venta se generan con dos medias móviles del índice, una a corto plazo (ex ante) y otra a largo plazo (ex post).

Los movimientos en contra de esa tendencia principal están compuestos por tres ondas, de esta forma podemos determinar cuál es la tendencia subyacente actual del precio. A su vez cada una de estas ondas se subdividen en cinco ondas de grado inferior, es decir, las ondas 1,3 y 5 están compuestas a su vez por cinco ondas 1(*variación*), 2(*máximo Ex post*), 3(*mínimo Ex post*), 4(*máximo Ex ante*) y 5(*mínimo Ex ante*). Las ondas 2 y 4 se denominan ondas correctivas, ya que son movimientos en contra de la tendencia principal. Se subdividen a su vez en tres ondas de grado inferior, denominadas x, y, z, en el método fractal-tridimensional.

4. Mapeo Fractal de las acciones en SIF'S con matices de Kelly y evidencias del MBF.⁶

Comenzamos, describiendo la relación de las variables económicas financieras que en este caso serán los precios de las acciones del mercado accionario mexicano, estos registros comprenden desde el Viernes 01 de Junio del 2007 al día Martes 01 de Junio del 2010 e incorporamos el término de perturbación de una observación no está asociado al término de perturbación asociado a otra observación:

$$E(u_i, u_j) = 0 \quad i \neq j \quad (5)$$

Mientras que la correlación fractal plantea la correlación de rezagos entre dos series diferentes (*una Ex ante y otra Ex post*):

$$(u^1, u^2, \dots, u^{10} \& v^1, v^2, \dots, v^{10}) \quad (6)$$

Cuando se traslada este concepto a precios fractales y con algún criterio geográfico, sus términos de error podrán también estar relacionados entre sí, encontrándonos entonces con el concepto de correlación espacial o autocorrelación espacial, para hablar de aquella correlación en el espacio en vez de la temporal.

Los SFI'S (*Sistemas de Información Fractal Espaciales*), identifican principalmente dos tipos de fenómenos en los precios de las acciones [Dacorogna, Michel M., Ramazan, Gencay., Müller, Ulrich., B.Olsen, Richard ., Pictet, Olivier V:2001], el primer paso es capturar los rangos de precios de las acciones (*Máximo y Mínimo*) del mercado de capitales a analizar (*considerando el mapeo Ex post o Ex ante*), dar el espacio en R^3 (*propiedad geométrica tridimensional*) y posteriormente en R^4 (*propiedad geométrica fractal*) y determinar para todos ejes geográficos la media dorada de 0.618 en x, y & z para que el programa nos arroje nuestra primer carta fractal a 10, 000 iteraciones con un mercado de 130 emisoras para el caso de la bolsa de México.

⁶ MBF es la señal o proceso con a espectro de la frecuencia tales que densidad espectral de la energía es proporcional al recíproco de la frecuencia. El ruido rosado tiene una cantidad igual de energía por octava. El nombre se presenta de ser intermedio en medio ruido blanco ($1/f_0$) y ruido rojo ($1/f_2$), conocido más comúnmente como Ruido browniano.

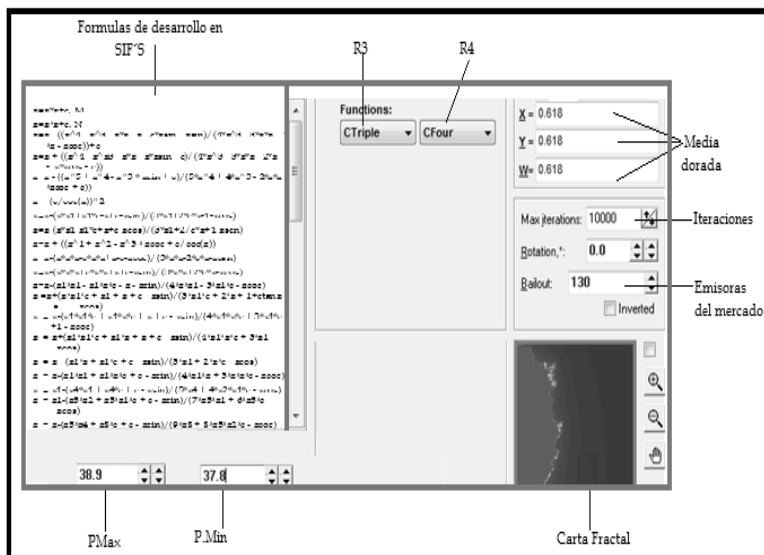


Figura 6: Construcción tecnológica de una Carta fractal de GCARSO.

Este es un prototipo de iteración de datos al software simula los precios máximos y mínimos en la función triple y cuarta en distribución de media dorada con 130 emisoras.

Guardamos la imagen para poder pintar al oleo fractal nuestro mapa, para ello tenemos que tener un esqueleto que nos sirve de guía en el vaciado de nuestras iteraciones fractales, a la que denominaremos matriz de complejidad⁷.

Tabla 7: Matriz de complejidad de 10*5 en el espacio de 50.

5-E-90'	3-E-56'	4-E-220'	6-N-145'	6-N-34'	3-N-180'	1-N-153'	9-E-45'	4-E-67'	2-E-89'
5-S-90'	3-S-56'	4-S-220'	6-S-145'	6-S-34'	3-S-180'	1-S-153'	9-S-45'	4-O-67'	2-O-89'
1-O-67'	4-O-45'	5-S-45'	7-S-24'	4-N-35'	6-N-67'	8-N-89'	9-E-41'	1-E-67'	4-N-45'
5-E-45'	7-N-90'	2-N-220'	1-E-91'	4-S-35'	6-S-67'	8-S-89'	9-O-41'	6-E-43'	8-E-32'
5-O-45'	7-O-90'	2-O-220'	1-O-91'	2-E-36'	6-E-192'	3-N-79'	9-N-169'	6-O-43'	8-O-32'

Un mapa fractal, en que cada pixel, corresponde a un valor del parámetro en precios.

Donde:

⁷ Para definir una matriz A m × n con cada entrada en la matriz A[i,j] llamada aij para todo 1 ≤ i ≤ m y 1 ≤ j ≤ n. Sin embargo, la convención del inicio de los índices i y j en 1 no es universal: algunos lenguajes de programación comienzan en cero, en cuál caso se tiene 0 ≤ i ≤ m - 1 y 0 ≤ j ≤ n - 1.

1,2,3,4,5,6,7,8,9,.....50,n	Indican el uso en espacio de la carta fractal en la Matriz.
45',25',37',280',.....360'	Indican el uso en grados de la carta fractal en la Matriz.
N	Indica desplazamiento hacia el Norte
S	Indica desplazamiento hacia el Sur
E	Indica desplazamiento hacia el Este
O	Indica desplazamiento hacia el Oeste

Se tendrá especial relevancia al quedar parametrizada en un Mapa Fractal, en que cada pixel, correspondiente a un valor del parámetro en precios, y se colorea de modo que refleje una propiedad básica del conjunto de Mandelbrot en el espacio [Barnsley, Michael: 1993].

Se ilustra el resultado de aplicar diversas transformaciones sobre la carta fractal⁸ representada en similitud con un factor de escala $r= 0.618$ y una transformación de afinidad con $r= 0.5$ y desplazamiento $x_n = x + h$, $y_n = y + k$ & $z_n = x+i$, como lo mostramos a continuación:

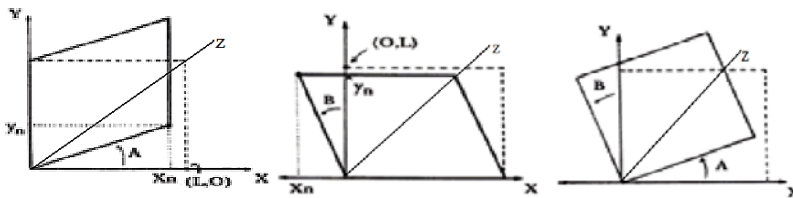


Figura 7: Movimiento estocástico de una carta fractal.

La ecuación funcional $s_1X^1+s_2X^2=sX$, junto con la relación subsidiaria. $(s1)a+(s2)a=sa$

Para ilustrarlo basta, por ejemplo, aplicar la receta anterior a las coordenadas de los cuatro vértices del cuadrado inicial:

$$\begin{aligned} (0,0) & \text{--->} (0.5*0, 0.5*0) \text{ --->} (0, 0) \\ (0,L) & \text{--->} (0.5*0, 0.5*L) \text{ --->} (0, L/2) \end{aligned}$$

⁸ El término Carta Fractal frecuentemente se considera que significa que los valores analíticos utilizados son de propiedad privada y, por lo tanto, no están a disposición del usuario o que los valores y códigos analíticos están más allá de la capacidad de comprensión del usuario y obedece a que las transformaciones o reacciones experimentadas por los Rangos de precios en su interior son desconocidos por su carácter de iterativo.

$$\begin{aligned}
 (L,0) & \rightarrow (0.5*L, 0.5*0) \rightarrow (L/2, 0) \\
 (L, L) & \rightarrow (0.5*L, 0.5*L) \rightarrow (L/2, L/2)
 \end{aligned}
 \tag{7}$$

Se dice que la transformación⁹ de todos los índices auto afines con el de México han introducido un factor de escala r=0.618, si bien la estructura de reglas geométricas como éstas se enriquece si además de reescalar la figura permite trasladarla a otro sitio o rotarla para observarla a escalas de tiempo discreto.

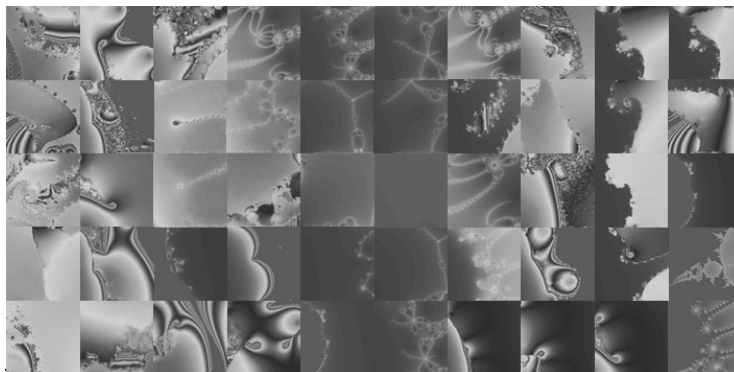


Figura 8: Mapa Fractal de GFINTER.

⁹ Realizando el mapeo de la transformación de todos las bolsas respecto al mercado de capitales mexicano, tenemos:

$$\begin{aligned}
 M_\beta &= \beta_0 + \beta_1(\text{DAX30}) + \beta_2(\text{TSE}) + \beta_3(\text{CAC40}) \\
 &\quad + \beta_4(\text{NIKKIE225}) + \beta_5(\text{S\&P}) + \xi^2 \\
 \partial M_\beta &= \beta_0 + \frac{\partial(\text{DAX30})}{\beta_1} + \frac{\partial(\text{TSE})}{\beta_2} + \frac{\partial(\text{CAC40})}{\beta_3} + \frac{\partial(\text{NIKKIE225})}{\beta_4} + \frac{\partial(\text{S\&P})}{\beta_5} \xi^2 \\
 dM &= d\beta_0 + \frac{d\beta_1(\text{DAX30})}{dx} + \frac{d\beta_2(\text{TSE})}{dx} + \frac{d\beta_3(\text{CAC40})}{dx} \\
 &\quad + \frac{d\beta_4(\text{NIKKIE225})}{dx} + \frac{d\beta_5(\text{S\&P})}{dx} + \frac{d\beta_6(\text{pc})}{dx} + \xi^2 \\
 M & \\
 M & \\
 M &= \int \begin{matrix} -1\beta_0^{360} & \left[\frac{\beta_2(\text{TSE})^{190}}{\beta_3(\text{CAC40})^{-1180}} \right] \\ -1\beta_1(\text{DAX30})^{270} & \left[\frac{\beta_6(\text{N})^{190}}{\beta_n(n)^{-1180}} \right] \\ -1\beta_4(\text{NIKKIE225})^{360} & \\ -1\beta_5(\text{S\&P})^{270} & \end{matrix}
 \end{aligned}$$

El mapa fractal es el resultado de iterar 10, 000 veces el precio de los rangos de acciones en la bolsa por emisora ejecutada en teneduría.

La trama Kelly [Coward, J: 1992] es un método de representación de datos inspirado por algunas de las pinturas de Ellsworth Kelly por el método de granulación gruesa, una secuencia de valores de datos se convierte en una secuencia de números. Y cada número se le asigna un color, que en nuestro caso el rosa-morado es un equilibrio browniano, el verde es un atractor de costos y el azul es un atractor de márgenes que nos ayudan a la construcción de una matriz cuadrada de unos cuadrados pequeños, una para cada punto de datos.

De acuerdo a nuestro mapa podemos observar que en corto plazo existen ganancias con tendencia de crecimiento y en el largo plazo existen igual ganancias pero sin crecimiento ya que pintan algunos costos en el Noroeste. Se dice en equilibrio browniano (*Ruido Kelly*) cuando los precios fijos señalan de previsibilidad a largo plazo. Si los precios del sistema hasta bastante cerca del precio Ex post, su comportamiento futuro es fácil de predecir: se acercará cada vez más cerca del precio Ex ante que en esencia de la estabilidad: las pequeñas perturbaciones desaparecen. Se dice atractor de costos (*Ruido verde*) cuando los precios fijos se comportan de la manera opuesta y situado exactamente en el rango de precios, allí se quedará. Pero pérdida en el mínimo cambio en precios (*rango*) de distancia y se sale rápidamente. Se dice atractor de márgenes (*Ruido azul*) cuando los precios fijos son los que ninguno de los casos anteriores, precios Ex post cercanos, o bien no se mueven en absoluto, o algunos se acercan, mientras que otros se mueven más lejos, asegurándonos un entorno de utilidad o ganancia absoluta. En lo que se refiere a movimientos y tendencias, el índice de una Bolsa de Valores muestra la sensibilidad, ya que refleja en gran medida lo que ocurre en el mundo económico, y es el parámetro más sensible de los hechos económicos: su sensibilidad los capta antes que sean visibles al público inversionista [Embrechts, Paul., Klüppelberg, Claudia., Mikosch, Thomas:1997].

Todos los índices son portafolios teóricos que pretenden capturar, con cierta representatividad, la evolución de un mercado específico, o lo que es lo mismo, como "benchmark" de un tipo de activo en un determinado mercado [Weick K.E: 1979]. Su fama es producto de la representatividad que tienen sobre el comportamiento de las acciones. Reflejan fielmente el comportamiento de cada acción y suelen ser los más estudiados.

3. Conclusiones.

Obtuvimos los resultados de las aplicaciones del método fractal-tridimensional y modelamos los precios con geometría fractal a escala logarítmica con \ln para las acciones que tenían peso muy grande, es decir un mayor número de emisoras en teneduría (Frankfurt y Londres) y con \log para la de peso pequeño, es decir un menor número de emisoras en teneduría (New York).

De acuerdo con la Tabla 3, el soporte quedo en 50.58% y la resistencia de los precios cerro en 39.07% por lo que la tendencia de los precios ira sobre 11.51% en nuestro método fractal-tridimensional, respecto de 10.95% del método chartista , y obtenemos un nivel de confianza aceptable de 0.56% por tanto se acepta esta tendencia para el estudio de los precios en los tiempos recurrentes de nuestro estudio pues tiene un margen de 0.06% de la media gaussiana tradicional.

Respecto de la correlación de funciones , obtuvimos que la función cognitiva (de la emisora hacia el mercado) , para todas las bolsas se mostró menor con un comportamiento del 1.08% , respecto de la función participativa (del mercado hacia la emisora) la cual obtuvo un comportamiento de 1.66% para definirse como un margen general , lo que nos demuestra que hay un sesgo de error del 0.58% y por ende que existen problemas de número de accionistas programas de mejora corporativa en su índice de actividad bursátil ó bien no tengan el porcentaje de acciones inscritas para poder participar en bolsa (resaltando , pues la amplia gama de instrumentos financieros , recordándoles que nuestro articulo solo es para mercado de capitales).

En la parte lógica definimos algunos algoritmos como los fractales de tiempo de escape, con nuestro conjunto de Mandelbrot para la fijación del soporte y la resistencia en el eje principal de la tendencia del mercado en el espacio en R^3 (*propiedad geométrica tridimensional*) y posteriormente en R^4 (*propiedad geométrica fractal*) pues los modelamos para todos ejes geográficos la media dorada de 0.618 en x, y & z que nos da el entorno perfecto de geometría no clásica en los fractales aleatorios con el movimiento browniano fraccional de ahí que no es de suma importancia el número de emisoras en el mercado , pues en la bolsa de Londres existe el mayor número de emisoras mientras que en la Bolsa de New York el mínimo de emisoras , que en representación de mercado la obtenemos en $c_n = \frac{\int_2^0 f(x)\phi_n(x)dx}{\int_2^0 \phi_n(x)dx}$, porque el Precio Ex ante está separado del Precio Ex post en el espacio libre (0.5) y nos representa otra posibilidad de maximización del margen estocástico del rango de precios según la tabla 4.

La emisora WALMEX que es la de precios más altos el promedio ponderado es de 9.13% y sigue representando otra opción más al margen sobre 10.95% de lo esperado con maximización de 1.82%.

Así mismo las Acciones en circulación de cualquier acción del mercado de capitales tiene una tendencia positiva h y con un ancho infinito se encuentra a una distancia d1 (14.3%-13.6%) Precio Ex ante y a d2 (33.06%-5.79%) Precio Ex post, otra ventaja más del método fractal-tridimensional pues para el tiempo A priori 2011-2012 tiene las diferencia al costo de 7.81% y al margen de 18.76% , lo que nos arroja como resultado la maximización del margen en 10.95%.

Ya en la tendencia la más fuerte de México la obtuvimos en la Emisora CEMEX , que se evalúa en la tabla 5 y 6 donde se arroja un margen de 8.89%. La volatilidad a la que viaja nuestro precio de la acción por arriba del promedio o rango a una distancia mayor de R^2 de 0.27% de aceptación fractal (en la gráfica 1), a 0.23% por

debajo del tradicional gaussiano de 0.5% en margen de operación.

Asumiendo que $h \ll d1$, $d2$ y $h \gg$ tendencia del Precio λ , entonces la diferencia entre la línea de tendencia y el camino difractado llamado precio prospectado (Δ), considerando en todo momento un margen de aceptación en nuestra operación bursátil, reivindicando que existe una evidencia fractal en los precios del mercado de capitales mexicano con autoafinidad y autosimilitud con el resto de las bolsas internacionales.

4. Bibliografía.

1. Alligood, Kathleen T., Sauer, Tom., Yorke, James. *Chaos, An Introduction to Dynamical Systems*. Springer, (1996).
2. Barnsley, Michael. *Fractals Everywhere*. San Diego: Academic Press. (1993)
3. Battena, Jonathan., Ellis, Craig. *Fractal structures and naive trading systems: Evidence from the spot US dollar/Japanese yen, Japan and the World Economy*, 8 (4), (1996). 411–421.
4. Braun, E. *Caos, fractales y cosas raras*. México. Colección la Ciencia en México. N° 150: Fondo de Cultura Económica. (1996).
5. Calvet, Laurent., Fisher, Adlai., Mandelbrot, Benoit. *Large deviations and the distribution of price changes*. *Cowles Foundation Discussion*. 1165.1(1), (1997). 237-253.
6. Courtault, Jean Michel., Kabanov, Yuri., Bru, Bernard., Crépel, Pierre., Lebon and Arnaud Le Marchand, Isabelle. *Louis Bachelier on the centenary of théorie de la speculation*. *Mathematical Finance*, 10(3), (2000). 341–353.
7. Cowart, J. *Method and motif: Ellsworth Kelly's Chance grids and his development of color panel painting by period: 1948-1951*, New York: Museum of Modern Art. (1992).
8. Dacorogna, Michel M., Ramazan, Gencay., Müller, Ulrich., B. Olsen, Richard., Pictet, Olivier V. *An Introduction to High-Frequency Finance*. San Diego, California: Academic Press. (2001)
9. Douketis, Constantine., Wang, Zhouhang., Haslett, Tom L., Moskovits, Martin. *Fractal character of cold-deposited silver films determined by low-temperature scanning tunneling microscopy*, *Physical Review*, 51 (16), (1995). 2-3.
10. Embrechts, Paul., Klüppelberg, Claudia., Mikosch, Thomas. *Modelling external events for insurance and finance*, New York: Springer-Verlag. (1997).

11. Falconer, Kenneth. *Fractal Geometry. Mathematical Foundations and Applications*. England: John Wiley & Sons Ltd. (1990).
12. Madan, Dilip B, Haluk Unal. "Pricing the Risks of Default", *Review of Derivatives Research*. (2), (1998).121-160.
13. Mandelbrot, Benoit. *La geometría fractal de la naturaleza*. San Francisco: WH Freeman. (1982)
14. Mantegna, R. N., Stanley, H. E.. *Scaling Behaviour in the Dynamics of an Economic Index*. *Nature* , 376, (1995). 46-49.
15. Ramos, María. *Dinámica Económica Actual*. México: ECORFAN. (2010).
16. Ramos, María. *DEFA (Dinámica Económica Financiera Actual)*, México: ECORFAN. (2010)
17. Ramos, María. *Invencción-Versión: 1.n -RNPC:03-2010-051112281100-01*, México: ECORFAN. (2010)
18. Rolf ,W., Banz, William., J. Breen. *Sample dependent results using accounting and market data some evidence*. *Journal of Finance*. 41 (4), (1986).779-793.
19. Sidney, Alexander. *Price movements in speculative markets: trends or random walks*, *Industrial Management Review*, 2 (2), (1961). 7-26.
20. Stewart, Ian. *From here to infinity: A guide to today's mathematics*. New York: Oxford University Press. (2003).
21. Weick K.E. *The social psychology of organization*. United States of America: Mc Graw-Hill, Inc. (1979).
22. Wyart , Matthieu ,Bouchaud ,Jean , Philippe. *Statistical Models for Company Growth*, *Service de Physique del Etat Condens* . [en línea] Recuperado el 16 de abril de 2012, de: <http://ssrn.com/abstract=391860> (2), (2002). 1-4.
23. <http://www.bmv.com.mx>
24. <http://www.worldfederationofexchanges.com>

ENSEÑANZA DE LA ESTADÍSTICA APOYADA EN LAS TICs. UNA VISIÓN DESDE EL OCW DE ESTADÍSTICA EN EL GRADO EN MARKETING DE LA UNIVERSIDAD DE MURCIA

ÚRSULA FAURA MARTÍNEZ

faura@um.es

*Universidad de Murcia. Departamento de Métodos Cuantitativos para la Economía y la Empresa
Campus de Espinardo, 30100 Murcia*

FUENSANTA ARNALDOS GARCÍA

arnaldos@um.es

*Universidad de Murcia. Departamento de Métodos Cuantitativos para la Economía y la Empresa
Campus de Espinardo, 30100 Murcia*

RESUMEN: Son muchos los recursos que se utilizan habitualmente para cumplir los objetivos propuestos y conseguir las competencias establecidas en el proceso de enseñanza-aprendizaje de la Estadística. En la actualidad el acceso a la red permite emplear estrategias y herramientas nuevas que mejoran o complementan las utilizadas tradicionalmente.

Entre los muchos recursos que hay en la red (libros online, calculadoras estadísticas, programas de software libre, ...) destaca la presencia de las denominadas simulaciones interactivas, que ofrecen al estudiante la oportunidad de visualizar conceptos estadísticos que les puedan resultar complejos e interactuar con ellos y aprender de la propia experiencia.

Las simulaciones interactivas pueden utilizarse en los tres grandes bloques considerados normalmente en la docencia en Estadística en las titulaciones de ciencias sociales: Descriptiva, Probabilidad e Inferencia. Dada la gran variedad y cantidad de simulaciones existentes, es aconsejable hacer una selección de las mismas y motivar su uso. Un ejemplo de integración de este tipo de recursos está disponible a través del portal OCW de la Universidad de Murcia para la asignatura Estadística del Grado en Marketing.

A lo largo del curso 2012-2013 este material ha sido utilizado por los alumnos de la asignatura, y hemos podido constatar que ha contribuido a mejorar su interés, ya que les resulta más atractiva.

Palabras claves: Estadística, OCW, TICs, Simulaciones interactivas.

ABSTRACT: Many different resources are commonly used to get the objectives and competences in the process of teaching and learning in Statistics. Nowadays the access to the web allows the use of new strategies and tools to enhance and complement the traditional ones.

Between the resources available on the web (online books, calculators statistics, free software,...) the so-called interactive simulations deserve special

recognition, for giving students the chance to visualize statistical concepts and interact with them and learn from their own experience.

Interactive simulations can be used in the three main parts normally considered when teaching Statistics in Social Sciences: Descriptive, Probability and Inference. Given the quantity and variety of existing simulations, it is advisable to make a selection of them and encourage their use. An example of the integration of these resources is available through the portal OCW of the University of Murcia for the subject "Statistics" in the degree of Marketing.

This material has been used during the academic year 2012-2013 by students of the subject, and we have found that has helped to improve their interest on it, making it more attractive for students.

Keywords: Statistics, OCW, TICs, Interactive simulations.

1. Introducción

La adaptación al nuevo Espacio Europeo de Educación Superior ha modificado el rol del profesor, de forma que debe no sólo transmitir conocimientos, sino actuar como motivador y guía del alumno en su descubrimiento de nuevas formas de aprendizaje. Esta adaptación ha supuesto también un cambio de rol para el alumno, que debe, más que nunca, gestionar su proceso de aprendizaje y adquirir no sólo conocimientos, sino competencias de diverso tipo relacionadas con la asignatura.

Las nuevas tecnologías juegan un papel importante en todo este proceso, tal y como analizan Gisbert et al. (2010). A través de la red se puede acceder a numerosos recursos, proporcionando al alumno experiencias formativas de naturaleza muy diversa que complementan a las más tradicionales. La tecnología actual permite, además, diseñar los materiales de forma interactiva, de manera que los estudiantes accedan a cada uno en el momento adecuado, y en situaciones sincrónicas o asincrónicas según se considere conveniente, haciendo las asignaturas más atractivas.

En el caso de las asignaturas de Estadística, son muchos los recursos virtuales que se pueden utilizar con esta intención. Debe destacarse, además, que el acceso libre a recursos de calidad facilita su incorporación al material de las asignaturas, acentuando el papel del profesor como discriminador de los recursos existentes y guía del alumno en su uso.

A la hora de seleccionar posibles recursos para el diseño de actividades, son muchas las opciones disponibles:

- Los libros online y las revistas docentes sobre la materia permiten solicitar a nuestros alumnos que completen cualquiera de los contenidos de las asignaturas.
- La disponibilidad de información estadística, tanto en fuentes oficiales, como en páginas donde los conjuntos de datos se agrupan por métodos de análisis, permite que se les pueda encargar su estudio.
- El acceso a software libre de análisis de datos posibilita la comparación con el software comercial de uso tradicional.
- La disponibilidad de calculadoras estadísticas ayuda a la comprobación de resultados en la resolución de ejercicios de manera tradicional.
- Los videotutoriales proporcionan acceso a las explicaciones acerca de un tema tantas veces como se considere necesario y bajo tantas versiones diferentes de explicación como autores de los mismos.
- Las simulaciones interactivas facilitan la visualización de conceptos complicados de entender, y favorecen la experimentación con el fenómeno y el aprendizaje a partir de la propia experiencia.

La lista de posibilidades es larga, teniendo en cuenta, además, que su integración en las actividades puede ser también variada, a partir de las herramientas de la Web 2.0 (Blogs, Wikies, videoconferencias, redes sociales,...), que posibilitan intercambios de naturaleza diversa con el estudiante y entre estudiantes.

En este trabajo mostramos una experiencia docente en la que hemos integrado diferentes tipos de recursos disponibles en la red de una forma interactiva. Esta experiencia se ha publicado como contenido abierto y bajo licencia Creative Commons en el portal OpenCourseWare de la Universidad de Murcia. La asignatura para la que se ha desarrollado esta experiencia docente es “Estadística”, asignatura de 6 créditos del Grado en Marketing, cuyos contenidos incluyen los bloques de Estadística Descriptiva, Probabilidad e Inferencia Estadística.

2. OCW

OpenCourseWare (OCW) es una forma de publicar contenidos docentes en abierto en la red que tienen las universidades e instituciones académicas. Aunque estos materiales están diseñados para cubrir los objetivos de alguna materia en concreto, y están pensados para los alumnos matriculados en esa materia, cualquier persona interesada puede acceder a los mismos, sin necesidad de registrarse ni matricularse.

Aunque fue una iniciativa en abril de 2001 del Instituto Tecnológico de Massachussets (MIT) ¹ se han adherido con posterioridad muchas otras

¹ <http://ocw.mit.edu/index.html>

universidades. En 2005 se creó el Consorcio OpenCourseWare², del que forman parte universidades y organizaciones de todo el mundo, y a partir de cuya creación se ha producido un crecimiento importante en el número de OCW publicados.

En España, Universia³ comenzó realizando traducciones de los OCWs del MIT, para pasar después a tener una participación más activa junto con la Universidad Politécnica de Madrid en el desarrollo de estos cursos, siendo miembros de apoyo del Consorcio OpenCourseWare. Dentro de la red de OCW-Universia⁴, hay 117 universidades, 47 españolas, siendo la Universidad de Murcia una de ellas.

La Universidad de Murcia cuenta en la actualidad en su portal OCW⁵ con 170 asignaturas distribuidas en 8 áreas temáticas (Artes, Ciencias de la Salud, Ciencias Jurídicas, Ciencias Sociales, Ciencias, Humanidades, Ingenierías y Transversales). La elaboración del material de estos cursos está adaptado al portal institucional que ofrece la Universidad de Murcia y todos ellos tienen en principio una estructura común, que consta de una presentación de la asignatura con sus objetivos, y pestañas de “Programa”, “Guía de aprendizaje”, “Material de clase”, “Bibliografía” y “Evaluación”, aunque se permite cierta flexibilidad para añadir todas aquellas opciones que se consideren necesarias para el adecuado seguimiento de la asignatura, tal y como se puede observar en la figura 1, en el que se han añadido las pestañas de “Prácticas” y “Otros recursos”.



Figura 1. Página de inicio del OCW de Estadística en el Grado en Marketing (Universidad de Murcia)

² <http://www.ocwconsortium.org/>

³ <http://www.universia.es/index.htm>

⁴ <http://ocw.universia.net/es/>

⁵ <http://ocw.um.es/>

3. OCW de Estadística en el Grado en Marketing⁶

La asignatura de Estadística en el Grado en Marketing se caracteriza porque en sólo 6 créditos realiza una revisión de los tres bloques de la Estadística (Descriptiva, Probabilidad e Inferencia). Se trata de una materia muy amplia que se imparte en poco tiempo, por lo que nos planteamos la creación del OCW como una guía de apoyo en el estudio autónomo de esta materia por parte de los alumnos. Este recurso didáctico puede, no obstante, ser utilizado por cualquier persona interesada en la materia.

Contiene material teórico con actividades variadas de aplicación de conceptos, actividades prácticas tanto de resolución de problemas como de análisis de datos con SPSS, un apartado de selección de recursos virtuales de naturaleza diversa que complementan a los materiales más tradicionales, y una sección de autoevaluación que incluye tanto cuestiones de selección múltiple como ejercicios de desarrollo. Las distintas estrategias adoptadas dependen de la ubicación de los materiales.

3.1. Material teórico

En “Material de clase” se puede acceder a un material teórico sencillo, que incluye ejemplos de aplicación y recursos de visualización y experimentación con los conceptos más complejos. Los materiales teóricos están publicados en formato HTML, lo que, además de mejorar el acceso pautado al material, posibilita la inclusión de distintas actividades interactivas:

- ejercicios resueltos que presentan la resolución a demanda del usuario, permitiendo que éste intente primero resolverlos.
- actividades de aplicación y experimentación con conceptos, en algunos casos complejos, en el momento en que estos conceptos se explican, mediante el uso de recursos interactivos procedentes de distintos proyectos de visualización disponibles en la red.

El material teórico tiene incorporados numerosos ejercicios de aplicación de conceptos. Cada uno de ellos está planteado ofreciendo al alumno la posibilidad de elegir entre trabajar con la resolución del problema desde el inicio, o resolverlo primero por su cuenta y comprobar luego dicha resolución. Por ejemplo, en el ejercicio mostrado en la figura 2, correspondiente al tema 1 (distribuciones de frecuencias unidimensionales) se estudian las distribuciones de frecuencias y, entre otros, se plantean ejercicios para practicar con los distintos tipos de frecuencia.

⁶ <http://ocw.um.es/cc.-sociales/estadistica-en-el-grado-de-marketing>

Ejercicio 1.2 A partir de los datos sobre la edad de 20 personas, complete la siguiente tabla de frecuencias:

x_i	n_i	f_i	N_i	F_i
50			4	
51		0,2		0,65
52	2			
53		0,15		0,9
54				
55				

SOLUCIÓN

Figura 2. Ejemplo de ejercicio propuesto

El ejercicio se ofrece sin que la resolución sea visible, pudiendo acceder a ella con el botón **SOLUCIÓN**, como se puede ver en la figura 3.

Ejercicio 1.2 A partir de los datos sobre la edad de 20 personas, complete la siguiente tabla de frecuencias:

x_i	n_i	f_i	N_i	F_i
50			4	
51		0,2		0,65
52	2			
53		0,15		0,9
54				
55				

SOLUCIÓN

Si $N_1 = 4$, se tiene que $n_1 = 4$, y como $N = 20$, $f_1 = \frac{n_1}{N} = \frac{4}{20} = 0,2$ y $F_1 = 0,2$.

Al ser $f_2 = 0,2$, se tiene $n_2 = 4$, $N_2 = N_1 + n_2 = 8$ y $F_2 = 0,4$.

Si $F_3 = 0,65 \Rightarrow F_3 = \frac{N_3}{N} = 0,65 \Rightarrow N_3 = 0,65N = 0,65 \cdot 20 = 13$ y $n_3 = N_3 - N_2 = 5$ con $f_3 = 0,25$.

El resto de los datos se obtienen de forma similar y en la última línea hay que tener en cuenta que $N_5 = 20$.

x_i	n_i	f_i	N_i	F_i
50	4	0,2	4	0,2
51	4	0,2	8	0,4
52	5	0,25	13	0,65
53	2	0,1	15	0,75
54	3	0,15	18	0,9
55	2	0,1	20	1
	20			

Figura 3. Ejemplo de ejercicio propuesto y resuelto

Muchos de los conceptos estudiados en la asignatura se pueden visualizar, lo que mejora su comprensión al tiempo que lo hace más atractivo. Por este motivo, hemos incorporado la visualización en el material teórico en varias ocasiones. En muchas de ellas, la herramienta empleada para visualizar el concepto es también interactiva, lo que permite experimentar con él y aprender de la propia experiencia.

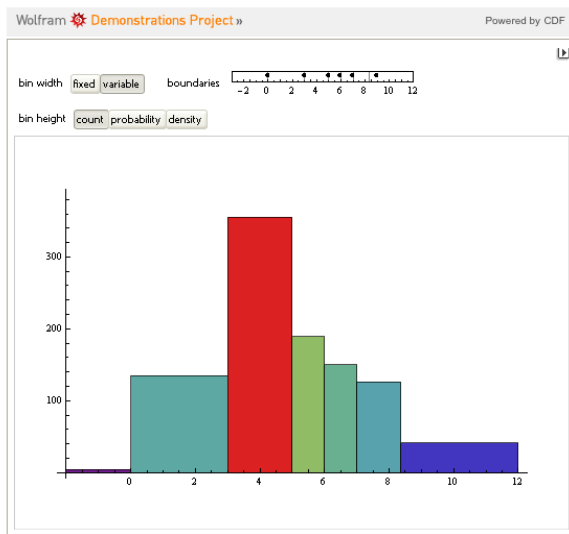
La proliferación en la red de este tipo de recursos facilita su incorporación a la docencia, aunque deben seleccionarse con precaución pensando en sus beneficios potenciales (Romero et al., 2000). Existen numerosas simulaciones disponibles para su uso, en la mayoría de los casos gratuitas y bajo licencias de uso Creative Commons. También es de destacar la existencia de proyectos de potenciación de la visualización desde grandes empresas como el caso de Wolfram Research o desde instituciones sin ánimo de lucro como las responsables del desarrollo de Geogebra. Una revisión de proyectos desarrollados que incorporan la visualización y que pueden servir de fuente de información y punto de partida para su consideración en la docencia de Estadística puede encontrarse en Arnaldos y Faura (2012).

En la selección de las simulaciones empleadas para el material teórico⁷, se ha primado, a igualdad de objetivos cubiertos, la posibilidad de insertarlas en el propio texto, frente a proporcionar un enlace externo para su uso. El Wolfram Demonstration Project posibilita insertar las simulaciones en el código HTML, lo que también ocurre con las construcciones de Geogebra, por lo que han sido ampliamente utilizadas en el curso. Las simulaciones interactivas se han empleado con diferentes intenciones. Se han usado para hacer al alumno reflexionar sobre las implicaciones de la toma de algunas decisiones, como en el ejemplo de la figura 4, relacionado con los efectos de la selección de un número concreto de intervalos en la representación gráfica de las observaciones a través de un histograma.

⁷ El uso de las simulaciones incluidas en el OCW precisa de ciertos requisitos de software. Es necesario que esté instalada la máquina virtual Java. En <http://www.java.com/es/download/index.jsp> se puede comprobar si está instalada y, en caso contrario, descargar una copia gratuita. Para trabajar con los ficheros del proyecto de demostración Wolfram es necesario tener instalado el Wolfram CDF Player. En <http://demonstrations.wolfram.com/download-cdf-player.html> se puede descargar una copia gratuita. Para utilizar una escena del Proyecto Descartes es necesario tener instalado el plug-in de Descartes Web2.0: <http://recursostic.educacion.es/descartes/web/DescartesWeb2.0/>

Actividad de apoyo...

¿Afecta la modificación del número de intervalos en la agrupación de los datos al histograma? Plantee una pregunta y utilice el siguiente recurso del Proyecto de Demostración Wolfram para comprobar si su respuesta es acertada.



Effects of Bin Width and Height in a Histogram from the Wolfram Demonstrations Project by Brett Champion

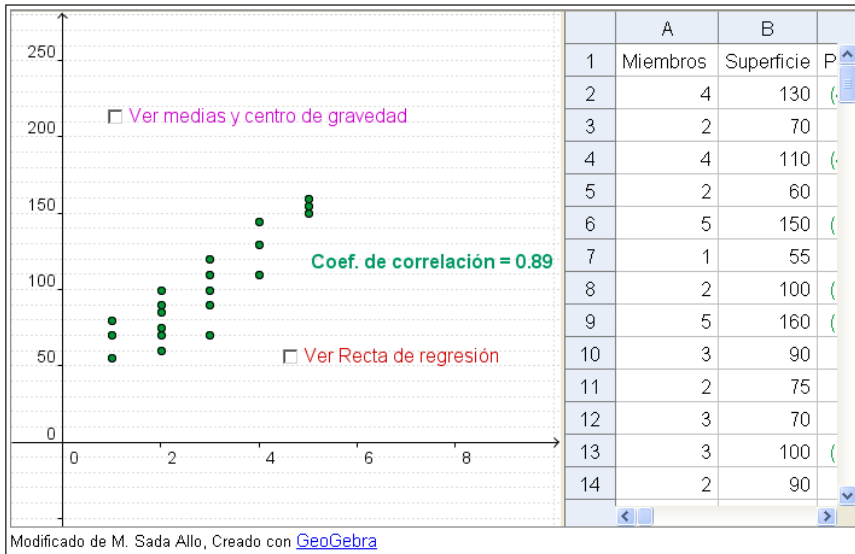
Figura 4. Ejemplo de experimentación con los efectos de la modificación del número de intervalos

También se han empleado para apoyar a ejercicios propuestos, como en la figura 5, en la que se observa una construcción de Geogebra con la que se representa la recta de regresión de mínimos cuadrados que los alumnos han calculado previamente en un ejercicio propuesto. Este applet permite, además, estudiar los efectos de la modificación de observaciones sobre la recta de regresión.

Cuando hemos considerado que otra simulación puede complementar a alguna ya incluida, hemos propuesto al alumno que la utilice también, de forma que, bajo diferentes visiones del mismo problema le permita mejorar la comprensión del problema, como se recoge también en la figura 5.

Actividad de apoyo...

En el siguiente applet de Geogebra puede visualizar la nube de puntos del ejercicio 2.7. Utilice el applet para insertar la recta de regresión, realizar y visualizar predicciones y para observar los cambios que se producen cuando modifica, añade o borra algún punto.



Puede seguir trabajando con estos conceptos en el siguiente recurso. Observe que en él puede también visualizar las sumas de los cuadrados de los residuos (diferencia entre el valor observado y el valor estimado según la recta).

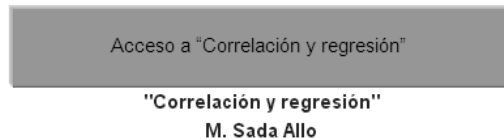


Figura 5. Ejemplo de simulaciones interactivas para apoyar ejercicios propuestos de regresión

Otro ejemplo de combinación de ejercicios tradicionales con simulaciones interactivas puede verse en la figura 6, en la que tras la resolución de ejercicios de cálculo de probabilidades con sucesos, se sugiere al alumno que compruebe los resultados acudiendo a una simulación que permite la reproducción del ejercicio, de forma que una vez el alumno conoce estos nuevos recursos, puede utilizarlos en situaciones diferentes para mejorar su comprensión de este tipo de problemas.

Ejercicio 3.3 Un alumno se examina de dos asignaturas, A y B. La probabilidad de que apruebe la asignatura A es 0,568 y la probabilidad de que apruebe la asignatura B es 0,4023. Teniendo en cuenta que la probabilidad de que apruebe alguna de las dos asignaturas es 0,7613, calcule la probabilidad de que el alumno:

- a) Apruebe las dos asignaturas.
- b) No apruebe ninguna asignatura.
- c) Sólo apruebe la asignatura A.
- d) Sólo apruebe la asignatura B.
- e) Sólo apruebe una asignatura.

SOLUCIÓN

Actividad de apoyo...

Puede utilizar el siguiente applet del Aula Virtual de Estadística de la Universidad de Córdoba para realiza este ejercicio. Elija un tamaño de los sucesos A, B y $A \cup B$ proporcional a su probabilidad y visualice los sucesos de los distintos apartados y sus probabilidades.

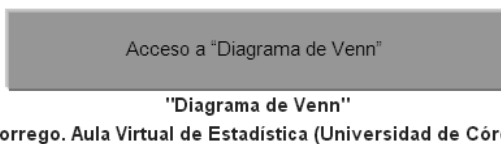
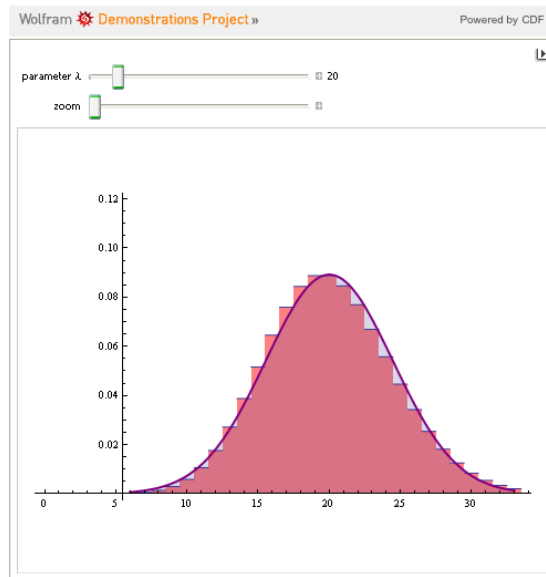


Figura 6. Ejemplo de simulación interactiva para apoyar a ejercicios propuestos de probabilidad de sucesos

Existen conceptos teóricos difíciles de transmitir y complicados de entender, en los que la experimentación puede ser de gran utilidad. Un ejemplo es la convergencia entre modelos. En la figura 7 se observa uno de los applets seleccionados en el material de clase para que los alumnos comprendan el concepto de convergencia y experimenten con las condiciones bajo las que se produce.

Actividad de apoyo...

En el siguiente recurso del Wolfram Demonstration Project puede observar la convergencia de distribución de Poisson a la Normal cuando aumenta el parámetro λ .



Normal Approximation to a Poisson Random Variable from the Wolfram Demonstrations Project by Chris Boucher

Figura 7. Ejemplo de simulación interactiva para experimentar con conceptos complejos

Estos son sólo algunas de las aplicaciones de las simulaciones interactivas consideradas en el material teórico. En la tabla 1 se muestra un listado de todas las simulaciones incluidas, con indicación del bloque de la materia para el que se han utilizado y el tema al que corresponden.

Tabla 1. Simulaciones interactivas utilizadas en el material teórico.

BLOQUE I: ESTADÍSTICA DESCRIPTIVA	
Distribuciones de frecuencias unidimensionales	“Effects of Bin Width and Height in a Histogram”. B. Champion. Wolfram Demonstrations Project. “Idea gráfica de la media”. M. Sada. Geogebra. “Medidas descriptivas”. R. Gómez. Aula Virtual de Estadística (Universidad de

	<p>Córdoba).</p> <p>“Índice de Gini” en CaEst 1.5. J.M. Lejarza</p> <p>“Income distribution by country”. VisualEconomics.</p>
Distribuciones de frecuencias bidimensionales	<p>“Understanding correlation”. R. W. West.</p> <p>“Correlations”. J. Marden. CUWU.</p> <p>“Regresion by eye”. R. W. West. StatCrunch.</p> <p>“Correlación y regresión”. M. Sada. Geogebra.</p>
BLOQUE II: PROBABILIDAD	
Fundamentos de probabilidad	<p>“Interactive Venn Diagrams”. M. Brodie. Wolfram Demonstrations Project.</p> <p>“Lanzamiento de dados”. Calvo Llorca et al. Statmedia (Universidad de Barcelona).</p> <p>“Gamblers fallacy”. D. Lane. OnlineStatbook.</p> <p>“Diagrama de Venn”. J. Borrego. Aula Virtual de Estadística (Universidad de Córdoba).</p> <p>“Conditional Probability”. C. Boucher. Wolfram Demonstrations Project.</p> <p>“Teorema de la probabilidad total”. J.A. Argote. Proyecto Descartes.</p> <p>“Teorema de Bayes”. J.A. Argote. Proyecto Descartes.</p> <p>“Combinación de dados y urnas”. Calvo Llorca et al. Statmedia (Universidad de Barcelona).</p>
Variables aleatorias unidimensionales y bidimensionales	<p>“Variable discreta y continua”. A. Bogarín. Aula Virtual de Estadística (Universidad de Córdoba).</p> <p>“Fonction de repartition (French)”. J. Zizi. Wolfram Demonstrations Project.</p> <p>“Lotería primitiva”. Calvo Llorca et al. Statmedia (Universidad de Barcelona).</p> <p>“Percentiles of Certain Probability Distributions”. C. Boucher. Wolfram Demonstrations Project.</p> <p>“Trinomial Distribution”. C. Boucher. Wolfram Demonstrations Project.</p> <p>“Bivariate Normal Distribution”. C. Boucher. Wolfram Demonstrations Project.</p>
Modelos de variables aleatorias	<p>“Binomial-Poisson”. P. Lacoste. CDNIS Math.</p> <p>“Continous Uniform Distribution”. C. Boucher. Wolfram Demonstrations Project.</p>

	<p>“Area of a Normal Distribution”. E Schulz. Wolfram Demonstrations Project.</p> <p>“Distribución Normal”. P. Bujosa. GeoGebra.</p> <p>“Funciones de distribución y cuantiles”. Aula Virtual de Bioestadística (U. Complutense de Madrid).</p> <p>“SOCR Distributions”. Statistics Online Computational Resource (UCLA).</p>
BLOQUE III: INFERENCIA ESTADÍSTICA	
Teoría de muestras	<p>“Normal Approximation to a Poisson Random Variable”. C. Boucher. Wolfram Demonstrations Project.</p> <p>“Normal Approximation to a Binomial Random Variable”. C. Boucher. Wolfram Demonstrations Project.</p> <p>“Chi-Squared Distribution and the Central Limit Theorem”. P. Falloon. Wolfram Demonstrations Project.</p> <p>“Sampling Distribution of the Mean and Standard Deviation in Various Populations”. I. McLeod. Wolfram Demonstrations Project.</p>
Métodos de estimación	<p>“Sampling Distribution of the Sample Mean”. J.R. Larkin. Wolfram Demonstrations Project.</p> <p>“Confidence Intervals for a Mean”. C. Boucher. Wolfram Demonstrations Project.</p> <p>“Estimación por intervalos”. Aula Virtual de Bioestadística (U. Complutense de Madrid).</p> <p>VESTAC. University Center for Statistics. Katholieke Universiteit Leuven.</p> <p>“Confidence Intervals: Confidence Level, Sample Size, and Margin of Error”. E. Schulz. Wolfram Demonstrations Project.</p> <p>“Confidence, Significance and Critical Values”. D. Gurney. Southeastern Louisiana University.</p> <p>“Decisions Bases on p-Values and Significance Levels”. E. Schulz. Wolfram Demonstrations Project.</p> <p>“P-Value and Critical Value Comparison”. D. Gurney. Southeastern Louisiana University.</p> <p>“Chi-squared worksheet”. PurpleMonkeyMath.com.</p>

El acceso a todas las simulaciones interactivas que aparecen en la tabla 1 está disponible en el material teórico, con distintas aplicaciones, y de forma directa a

través de la pestaña del curso denominada “Otros recursos”. En ella está disponible una página en la que, con la herramienta Pearltrees, se ha creado un árbol con recursos de naturaleza diversa. Además de contener todas las simulaciones interactivas agrupadas por temas y bloques, incluye recursos adicionales que consideramos útiles para el seguimiento de la materia, como fuentes estadísticas clasificadas por temas, vídeos explicativos de fuentes oficiales, libros online, calculadoras estadísticas, etc...

Se trata de un árbol interactivo, en el que pinchando en cada una de las “perlas” se puede visualizar la rama correspondiente. Permite mantener un sistema de archivo de páginas web, pudiendo visualizarse en su nivel más alto o ir desarrollando las ramas hasta llegar a las páginas que se encuentran enlazadas.



Figura 8. Pearltree de “Otros recursos” del OCW

3.2. Material práctico

En la pestaña “Prácticas” se han incluido distintos tipos de actividades entre las que se encuentran relaciones de problemas y prácticas de análisis de datos en el aula de informática con el software SPSS. Se proporciona también el material complementario (tablas estadísticas y de intervalos y contrastes) necesario para la realización de estas actividades.

Para cada uno de los bloques disponen de una relación de problemas en la que los enunciados se encuentran en un fichero separado de las soluciones. Para mejorar la

presentación y navegación en los ficheros pdf utilizados en esta parte, se les ha dado un efecto que permite pasar páginas del documento digital como si fuese un documento físico; es decir, simula el efecto natural de pasar página. Además, facilita la búsqueda a lo largo del documento, permite saltar a una página concreta e incluye una tabla de contenido de acceso rápido a los puntos principales. Este mismo efecto se ha empleado para el material de apoyo de las prácticas de análisis de datos con SPSS, que se han incluido, también para cada uno de los bloques, y para todo el material complementario incorporado en esta pestaña. En la figura 9 puede verse un ejemplo de este tipo de material.

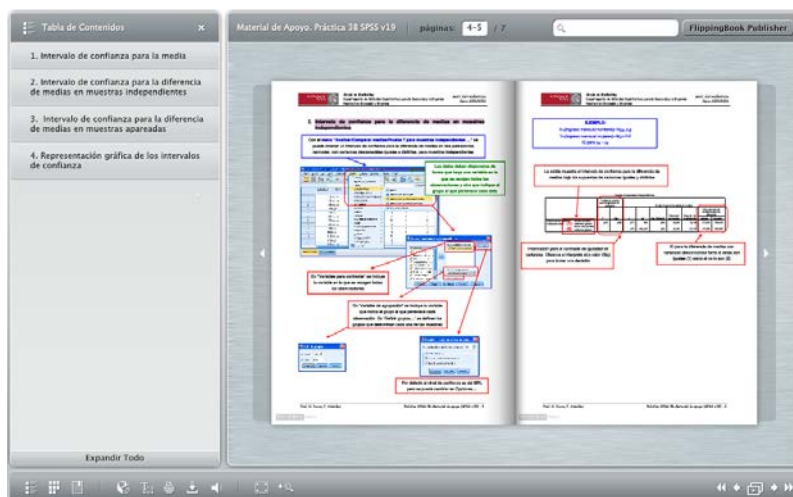


Figura 9. Ejemplo de material de apoyo para las prácticas de análisis de datos con SPSS

En el caso de las prácticas de análisis de datos, se incluyen, además, videotutoriales elaborados con Camtasia Studio (figura 10).

Material de apoyo SPSS

1. IC para la media
2. IC para la diferencia d
3. IC para la diferencia d
4. Representación gráfi
5. CH para la media
6. CH para la diferencia
7. CH para la diferencia
8. CH de bondad de aju
9. CH de bondad de aju

EJEMPLO:
 $X=(\text{Ingreso mensual})$, Contraste con $\alpha=5\%$
 $H_0: X \text{ se distribuye } N(\mu, \sigma)$ frente a $H_1: \text{no } H_0$

Figura 10. Ejemplo de videotutorial de apoyo para las prácticas de análisis de datos con SPSS.

3.3. Pruebas de evaluación

En “Evaluación” hemos incluido distintos tipos de cuestiones clasificadas según los bloques de la asignatura, que permiten al alumno comprobar su progreso en la comprensión de la materia. Separadas en cuestiones tipo test y cuestiones abiertas, se ha procurado dotarlas también de un carácter interactivo.

Como se muestra en la figura 11, en el caso de las cuestiones tipo test, al marcar unas de las opciones aparece un comentario acerca de su corrección.

Cuestiones test

Inicio
Bloque I
Bloque II
Cuestiones test
Cuestiones abiertas
Bloque III

? **Cuestión 1**

Señale la afirmación correcta:

a) $P[(A-B) \cup (B-A)] = P(A \cup B) - P(A \cap B)$

b) $A \cap \Omega = \Omega$.

c) Ninguna de las anteriores

May bien

? **Cuestión 2**

Sean los sucesos A y B con $P(A) = 0.2$, $P(B) = 0.3$ y $P(\overline{A \cap B}) = 0.56$. Señale la afirmación correcta:

a) $P(A \cap B) = 0$.

b) La probabilidad del suceso intersección entre A y B es igual a 0,44


c) Los sucesos A y B son independientes.

Figura 11. Ejemplo de autoevaluación empleando cuestiones test

Por su parte, en las cuestiones abiertas, la resolución se mantiene oculta hasta que el usuario decide comprobarla, tal y como se puede observar en la figura 12.


Cuestiones ab

Inicio
Bloque I
Bloque II
Bloque III
Cuestiones test
Cuestiones abiertas

 **Cuestión abierta 1**

El número de viajes en avión que hace una familia durante un año sigue una distribución de Poisson de media 2. Escríbanle a 500 familias determine la probabilidad de que el número total de viajes realizados en un año sea igual a 950.

[Click aquí](#)

 **Cuestión abierta 2**

El peso de un electrodoméstico se distribuye según una variable aleatoria de media 75kg y desviación típica 5kg. Calcule la probabilidad de que el peso total de 70 aparatos no supere 5350kg.

[Ocultar](#)

Sea X_i =(Peso del aparato), con $E[X_i] = 75$ y $\sigma_{X_i} = 5$ y $(X_1, X_2, \dots, X_{70})$ una m.a.s. El peso total, $\sum_{i=1}^{70} X_i$, se distribuye aplicando el TCL como $N(\mu = 70 \cdot 75, \sigma = \sqrt{70} \cdot 5) = N(\mu = 5250, \sigma = 41833)$.

$$P\left(\sum_{i=1}^{70} X_i \leq 5350\right) = P(Z \leq 2.39) = 0.99158$$

Figura 12. Ejemplo de autoevaluación empleando cuestiones abiertas

4. Conclusiones

El proyecto OCW posibilita llegar a un gran número de personas interesadas en una materia. La participación en el mismo, a través del portal institucional de una Universidad, se ciñe a una estructura predeterminada, pero dentro de ella se tiene libertad para decidir la presentación de los materiales. En el caso del OCW comentado, hemos optado por la elaboración de unos materiales interactivos, en la medida de lo posible, favorecido por el gran avance de las herramientas informáticas en la actualidad, que permite, con un esfuerzo relativo, lograr esta interactividad.

Para el desarrollo del proyecto hemos analizado numerosos recursos disponibles en la red, debiendo destacar la elevada calidad de muchos de ellos. Es labor del profesor seleccionar los recursos, planificar de forma adecuada su uso y proporcionarlos en el momento más idóneo. Esta ha sido nuestra intención en la elaboración del material de clase.

El tiempo y esfuerzo dedicado a un proyecto de este tipo se ve recompensado por la opinión de los alumnos que lo utilizan, y la actitud con la que afrontan la asignatura, en comparación a experiencias anteriores utilizando metodologías y materiales más tradicionales.

El alumno afronta la asignatura de una forma más autónoma, ya que puede decidir por su cuenta qué estudiar y cómo hacerlo. La visualización y la experimentación consiguen que los conceptos le queden más claros y tiene la posibilidad de contrastar a través de la autoevaluación el progreso en su proceso de enseñanza-aprendizaje.

Referencias Bibliográficas

1. M. Gisbert, J. Cela-Ranilla, y S. Isus, Las simulaciones en entornos TIC como herramientas para la formación en competencias transversales de los estudiantes universitarios, *TESI*, 11, 3 (2010) 352-370.
2. <http://ocw.mit.edu/index.html> (Fecha de consulta, junio 2013).
3. <http://www.ocwconsortium.org/> (Fecha de consulta, junio 2013).
4. <http://www.universia.es/index.htm> (Fecha de consulta, junio 2013).
5. <http://ocw.universia.net/es/> (Fecha de consulta, junio 2013).
6. <http://ocw.um.es/> (Fecha de consulta, junio 2013).

7. <http://ocw.um.es/cc.-sociales/estadistica-en-el-grado-de-marketing> (Fecha de consulta, junio 2013).
8. V. Romero, D. Berger, M. Healy y C. Aberson, Using cognitive learning theory to design effective on-line statistics tutorials, *Behaviour Research Methods, Instruments & Computers*, 32, 2 (2000) 246-249.
9. F. Arnaldos y Ú. Faura, Aprendizaje de los fundamentos de la probabilidad apoyado en las TICs, *@tic. revista d'innovació educativa*, 9 (2012) 131-139, <http://ojs.uv.es/index.php/attic/article/view/943/1464> (Fecha de consulta, junio 2013).
10. <http://www.java.com/es/download/index.jsp> (Fecha de consulta, junio 2013).
11. <http://demonstrations.wolfram.com/download-cdf-player.html> (Fecha de consulta, junio 2013).
12. <http://recursostic.educacion.es/descartes/web/DescartesWeb2.0/> (Fecha de consulta, junio 2013).

INFLUENCIA DE LAS TECNOLOGÍAS DE LA INFORMACIÓN SOBRE EL RENDIMIENTO EN LAS PRUEBAS DE MATEMÁTICAS PISA-2009

GONZALO SANZ-MAGALLÓN REZUSTA

sanmag@ceu.es

*Universidad San Pablo CEU / Departamento de Economía
C/Julián Romea 23*

MANUEL M. MOLINA LÓPEZ

manuelmaria.molinalopez@ceu.es

*Universidad San Pablo CEU / Instituto de Estudios de la Familia
C/Julián Romea 23*

M. CRISTINA AGUIRRE ARRABAL

aguiarr@ceu.es

*Universidad San Pablo CEU / Departamento de Matemática Aplicada y Estadística
C/Julián Romea 23*

RESUMEN: El artículo estudia la posible influencia de la disponibilidad de Tecnologías de la Información o TIC (ordenadores personales, software educativo y conexión a internet en el hogar) sobre los resultados del aprendizaje de matemáticas de los alumnos españoles. Para ello se explotan los microdatos del proyecto PISA (*Programme for International Student Assessment*) de la OCDE correspondientes a 2009, que cuantifica el nivel de conocimientos de alumnos de 15 años en matemáticas, lengua y ciencias. El análisis efectuado ha consistido en estimar una función de producción educativa en la que una de las variables explicativas es la presencia o no de TIC en el hogar. Las conclusiones del trabajo permiten afirmar que existe un considerable efecto negativo a raíz de la no disponibilidad de TIC, por lo que las políticas que faciliten estas tecnologías a las familias desfavorecidas permitirán una mejora del rendimiento educativo.

Palabras claves: Nuevas tecnologías; TIC; Eficiencia; Educación; Política educativa; Matemáticas; PISA-OCDE

ABSTRACT: The article examines the possible influence of the availability of Information Technology or ICT (personal computers, educational software and internet access at home) on mathematics learning outcomes of Spanish students. For this purpose microdata PISA (Programme for International Student Assessment) of the OECD for 2009 has been exploited. PISA quantifies the level of knowledge of pupils aged 15 in math, language and science. The analysis has estimated an educational production function in which one of the explanatory variables is the presence or absence of ICT in the

home. The conclusions of the study support the hypothesis that there is a considerable negative effect due to the unavailability of ICT at home, so policies that facilitate these technologies for disadvantaged families will improve educational performance.

Keywords: New technologies; TIC; Efficiency; Education; Educational policy; Mathematics; PISA-OECD

Introducción

A la hora de medir el rendimiento del sistema educativo se suelen tratar diversos factores como el efecto de los compañeros sobre el alumno, sus capacidades innatas, los recursos del centro y las características socioeconómicas del propio estudiante. Un recurso que no suele ser considerado en los distintos estudios que tratan de medir el output educativo es el empleo de las nuevas tecnologías (TIC) en el hogar como parte del proceso educativo.

Los trabajos realizados sobre el empleo de las TIC en el aula determinan que la contribución de este tipo de instrumentos está condicionada fundamentalmente por la actitud del profesorado hacia el uso de las mismas (Sosin et al. 2004; Webb y Cox 2004; Assar et al. 2010). Para el caso de España, destaca la ausencia de estudios que avalen de forma empírica las potencialidades o limitaciones de algunas de las nuevas tecnologías aplicadas en las aulas (Gandol et al. 2012).

En la literatura de ámbito internacional hay diversidad de opiniones en cuanto a la relación existente entre factores como el sexo y el nivel socioeconómico con el correcto aprovechamiento de las TIC. Heemskerk et al. (2009) establecen una relación positiva entre ser mujer y el mejor aprovechamiento de las TIC, así como un mayor aprovechamiento de las clases menos favorecidas de este tipo de tecnologías. Por su parte, Biagi y Loi (2012), analizando la totalidad de países participantes en el Informe PISA no encuentran una relación significativa entre el rendimiento académico y el uso de nuevas tecnologías por parte del alumno, independientemente de la clase socioeconómica a la que pertenezca.

La ausencia de este tipo de estudios aplicados al caso de España hace que sea de interés investigar la relación entre el empleo de las TIC y el rendimiento académico. Las conclusiones de esta línea de investigación podrán servir para orientar las políticas educativas emprendidas por algunas regiones españolas, y evaluar la conveniencia de subvencionar las TIC a las familias desfavorecidas como vía para la mejora del rendimiento educativo.

Tras esta introducción, el apartado segundo revisa la literatura acerca del rendimiento de

la educación y el uso de las TIC; el tercero especifica los objetivos y datos empleados; en el cuarto se muestran los resultados del análisis y el apartado 5 presenta las principales conclusiones.

1. Eficiencia en la educación y el uso de las TIC

El ámbito de la economía de la educación ha experimentado un notable avance en los últimos años, siendo las cuestiones relacionadas con los factores que miden la eficiencia los que más han sido tratados (Cordero et al. 2013). Cada vez es más habitual encontrarse con literatura acerca de la eficiencia del sistema educativo que trata de identificar las distintas variables que la explican. Una de las formas más habituales de medir la citada eficiencia sería mediante la función de producción de la educación definida por Hanushek (1979), en donde se mide el *output* del proceso educativo de un alumno en una escuela determinada (A_{ij}) en base a una serie de *inputs* escolares (S_{ij}), las características socioeconómicas del alumno (B_{ij}), la influencia de los compañeros (P_{ij}) y las capacidades innatas del alumno (I_{ij}). La función de producción es la siguiente:

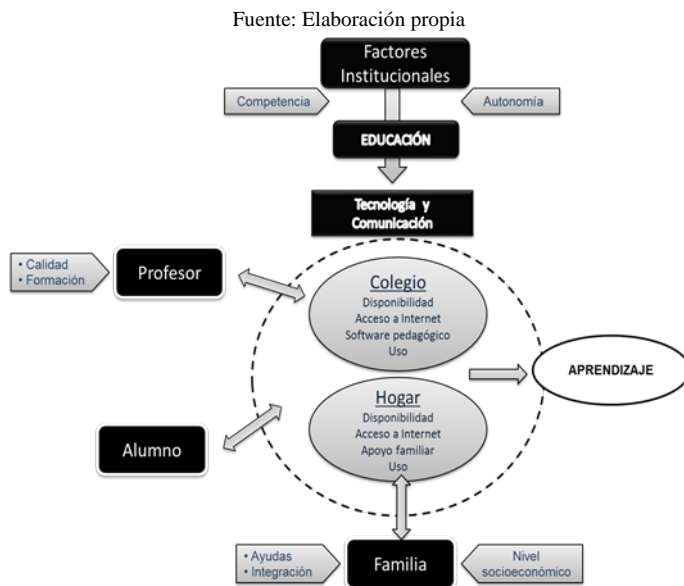
$$A_{ij}=f(S_{ij}, B_{ij}, P_{ij}, I_{ij}) \quad (1)$$

Doncel et al. (2012) analizan, para el período 2005-2009, los resultados obtenidos en las pruebas de Conocimientos y Pruebas Indispensables (CDI)¹ de sexto de primaria en 1.208 colegios de la Comunidad de Madrid, entre los que se encuentran tanto los de titularidad pública, como privada y concertada. El estudio demuestra que los centros privados y concertados obtienen un mejor resultado académico. A su vez, determina que un factor significativo de la eficiencia es el nivel socioeconómico de los padres. Este puede ser un motivo que condicione la posible tendencia de los centros privados y concertados a tener unos mejores resultados académicos (Hanushek et al. 2005). Otros estudios, controlando el posible efecto que puede tener el proceder de un entorno socioeconómico más elevado, determinan que el *output* educativo de los centros concertados es mayor que el de los públicos (Böhlmark y Lindhl, 2008).

En el rendimiento académico del alumno influyen, pues, diversos factores tanto de índoles familiar (nivel socioeconómico), institucional (existencia de competencia, política educativa de las administraciones) o personal (factores cognitivos del individuo). A continuación se muestra un gráfico que trata de relacionar los factores determinantes del rendimiento académico, incluyendo el papel que juegan las TIC en el proceso.

¹ Las pruebas CDI son realizadas a los alumnos de la Comunidad de Madrid que cursan sexto de primaria. Se llevan a cabo desde el año 2005.

Gráfico 1. Principales factores condicionantes del rendimiento académico



En cuanto al impacto de las TIC en el rendimiento académico hay divergencia de opiniones. Algunos estudios sostienen que el uso de estas nuevas tecnologías tiene un efecto positivo sobre el output educativo, estando este fundamentalmente condicionado por la actitud del profesorado hacia el uso de las mismas (Sosin et al. 2004; Webb y Cox 2004; Assar et al. 2010). Otros en cambio destacan la ausencia de tal efecto (CECE, 2011).

Gandol et al. (2012) analizan, a través de una revisión de las publicaciones más recientes, las limitaciones y potencialidades de las Pizarras Digitales Interactivas (PDI), concluyendo que el uso por sí mismo de este tipo de tecnologías no garantiza una mejora en la calidad educativa, teniendo que ir acompañado el empleo de TIC en los centros educativos de una estrategia global de todos los agentes del colegio implicados, en donde la formación del profesorado tiene un papel fundamental. Así pues, la forma mediante la cual el docente las empleó propiciará que estas puedan contribuir o no al rendimiento académico del alumno (Webb y Cox, 2004; Glover y Miller 2006).

En general, aunque el profesorado tenga medios tecnológicos a su disposición para impartir docencia, suele ser poco innovador y tiende a utilizar siempre los mismos

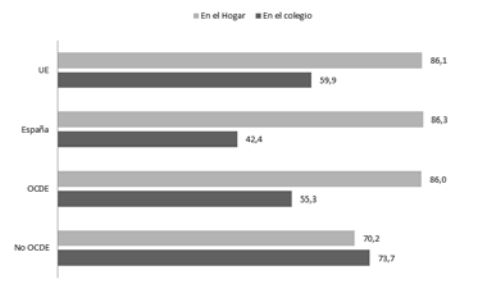
programas (Glover y Miller 2006). Esta actitud suele estar justificada en la mayoría de los casos por una posible tecnofobia del personal docente (Sáez y Jiménez, 2011). El Estudio de la Fundación Telefónica (2011) revela que gran parte del profesorado tiende a usar siempre los mismos materiales educativos, usando los de fácil ejecución frente a los que puedan ser considerados más complejos y eficaces. A su vez determina que el profesorado que emplea TIC en la docencia suele valorar de forma positiva el uso de la misma.

Por su parte, Law et al. (2008) determinan que el uso de las TIC tiene un impacto positivo en el rendimiento académico, estando sujeto a la forma de implantarse y la visión que el equipo directivo del centro tenga de las mismas. Así pues, el pleno apoyo de la dirección será necesario para que ésta tenga una repercusión positiva en el profesorado y alumnado (Eickelmann, 2011).

Otro factor condicionante del buen aprovechamiento de las nuevas tecnologías en el rendimiento académico es el sexo. Heemskerk et al. (2009) determinan que las chicas tienden a conseguir mayores beneficios que los chicos, siendo el efecto positivo para ambos sexos. Este mayor aprovechamiento por parte de las chicas, que en la mayoría de los casos parten con unos niveles de conocimiento en TIC algo inferiores, estaría motivado por una mayor concentración y proactividad en su uso.

Cabe destacar el hecho de que los estudiantes tienden a usar los ordenadores en casa con más frecuencia en comparación con los centros educativos, pese al esfuerzo económico que éstos realizan en este tipo de inversiones (Molina, 2013). Los alumnos españoles, pese a encontrarse al mismo nivel de la OCDE en cuanto a la utilización de computadoras en casa, están por debajo de muchos países en donde su uso roza la totalidad de la población. Destaca que los que más emplean el ordenador en casa tienden a hacerlo para chatear o buscar cosas en internet, siendo el uso de software una de las prácticas menos habituales (OCDE, 2010).

Gráfico 2. Alumnos de 15 años que emplean el ordenador en casa y en el hogar (En porcentaje)



En cuanto al uso de las TIC en la escuela destaca el hecho que, pese a que en un gran

número de escuelas dispone de TICs para el proceso de aprendizaje, son pocas las que realmente las utilizan eficazmente (OCDE, 2010)

Biagi y Loi (2012) estudian el impacto de las nuevas tecnologías en los países participantes en el Informe PISA 2009 de forma agregada. Para ello explotan el cuestionario realizado a las familias por PISA 2009, tratando de relacionar el uso de las TIC con el resultado en las pruebas de matemáticas, ciencias y lectura. En su trabajo se distingue la distinta aportación de las nuevas tecnologías en función del estrato socioeconómico del alumno, determinando una ausencia de relación entre uso de TIC y mejora en las pruebas realizadas, incluso después de considerar el factor socioeconómico. Cabe destacar que el anterior estudio emplea como muestra la totalidad de países participantes en el Proyecto PISA 2009, sin tener en cuenta las divergencias entre estos en lo que al uso de nuevas tecnologías se refiere.

El Gráfico 3 muestra la puntuación obtenida en las pruebas de matemáticas en los distintos países participantes en las pruebas PISA, poniéndose de manifiesto el bajo resultado obtenido por los alumnos españoles (483 puntos) en comparación con el resto. A la cabeza se encuentra Shangai con 600 puntos, seguido de otros países asiáticos como Singapur (562 puntos), China-HongKong (555 puntos) y Corea del Sur (546 puntos). El primer país europeo en la lista es Finlandia (541), seguido de Suiza (534), Países Bajos (536) y Bélgica (515).

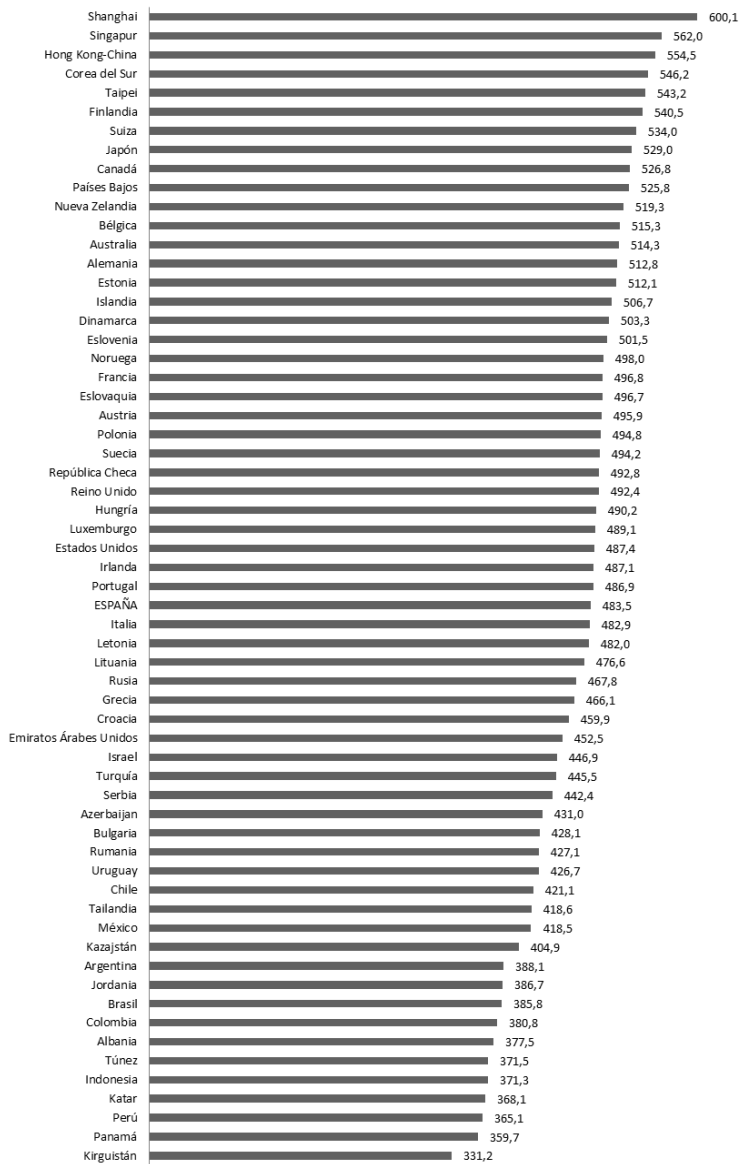


Gráfico 3. Puntuación obtenida por alumnos de 15 años en las pruebas PISA-2009 de matemáticas según país

Fuente: PISA 2009

2. Objetivos y metodología

Nuestro objetivo es estudiar la relación entre disponibilidad en el hogar de las TIC (ordenadores, software educativo y la conexión a internet) con los resultados académicos obtenidos en las pruebas de matemáticas del proyecto PISA-2009. A partir de los factores expuestos en la revisión de la literatura realizada, planteamos las siguientes hipótesis:

H₁0: Los alumnos que carezcan del acceso a las TIC en su hogar se verán penalizados en sus resultados académicos.

H₂0: Las alumnas obtienen mayores ventajas de las TIC que los alumnos.

H₃0: Para alumnos con menor nivel socioeconómicos el coste de no disponer de TIC en el hogar es mayor.

El proyecto PISA 2009, cuya muestra en España comprende 25.887 alumnos, incluye distintas variables que pueden afectar al rendimiento académico del alumno. Las competencias evaluadas son conocimientos de matemáticas, ciencias y lectura a los 15 años, caracterizándose la oleada del año 2009 por ser la de mayor participación internacional (74 países frente a los 32 del año 2000) y con mayor consolidación de los análisis regionales, permitiendo una amplia comparativa dentro del país. Así, en la oleada del año 2000 España participó como una sola muestra estatal, mientras que en el año 2009 se han analizado todas las comunidades autónomas a excepción de tres.

Nuestro análisis se realizará para el conjunto nacional, estimándose el impacto de la disponibilidad de TIC en los resultados de las pruebas de matemáticas. Para ello se incluirán como posibles variables independientes la disponibilidad de ordenador, software y conexión a internet en casa; la titularidad del centro educativo; el ratio ordenador/alumnos en el colegio; la lengua materna del alumno; y su sexo. Para estimar los resultados se han utilizado las técnicas recomendadas en el *PISA Data Analysis Manual* (2009) de la OCDE, que implica realizar las estimaciones a partir de cinco valores plausibles y ochenta réplicas de cada alumno.

3. Resultados

3.1. Descripción de la muestra

De la muestra utilizada el 60,4% de los alumnos asiste a escuelas públicas y el 39,6% a privadas. Atendiendo al sexo, la distribución es bastante similar: un 49,2% son alumnas y un 50,8% alumnos. Respecto a la lengua materna, para el 84,5% de los estudiantes coincide con la utilizada en el examen PISA, frente al 15,5% de alumnos que hablan en su casa otro idioma.

Tabla 1. Distribución de los alumnos de la muestra según sexo, lengua materna y titularidad del colegio

	Nº	%
Alumnas	12.746	49,2
Alumnos	13.141	50,8
Lengua del examen es la lengua materna	21.560	84,5
Lengua del examen distinta de la materna	3.953	15,5
En escuela pública	15.336	60,4
En escuela privada	10.034	39,6

Fuente: PISA 2009

El uso de las nuevas tecnologías en los hogares es superior al que se hace en las escuelas: un 86% frente a un 42%. En la Tabla 2 se muestra la disponibilidad de recursos en el hogar: ordenador, acceso a internet y disponibilidad de software educativo, destacando la alta disponibilidad de ordenadores en los hogares, seguida de internet y software educativo.

Tabla 2. Disponibilidad de recursos en casa del alumno

	Sí		No	
	Nº	%	Nº	%
Ordenador	24.107	94,0	1.535	6,0
Internet	22.394	87,4	3.223	12,6
Software educativo	13.149	52,0	12.114	48,0

Fuente: PISA 2009

Por su parte el nivel socioeconómico, medido en PISA a través del indicador sintético ESCS, que engloba nivel educativo, cultural y económico de la familia del alumno, toma un valor promedio en los centros públicos del -0,5, frente a 0,1 en los privados,

siendo el valor 0 la media de la OCDE. Esta distancia de 0,6 puntos entre centros privados y públicos es ligeramente inferior al promedio de países participantes en PISA. Cabe señalar, a su vez, que el estatus socioeconómico de las familias de alumnos en centros privados y concertados en España es inferior al promedio de la OCDE: 0,1 de ESCS en España, frente a 0,4 en el conjunto de países.

3.2. Resultados econométricos

A continuación se muestra el efecto de disponer en el hogar de ordenador, software educativo e internet en la puntuación obtenida en las pruebas de matemáticas de PISA. Se han realizado tres regresiones lineales para cada tipo de tecnología, tratando de contrastar el impacto de las tecnologías mencionadas en el resultado obtenido considerando el efecto de diversos factores. Las tres regresiones vienen representadas por las siguientes tres ecuaciones respectivamente,

$$A_i = C + \beta_1 \text{TIC} \quad (2)$$

$$A_i = C + \beta_1 \text{TIC} + \beta_2 \text{ESCS} \quad (3)$$

$$A_i = C + \beta_1 \text{TIC} + \beta_2 \text{ESCS} + \beta_3 \text{PRIV} + \beta_4 \text{LENGUA} \quad (4)$$

donde A_i es el resultado obtenido en matemáticas por el alumno i , β_{TIC} la no disponibilidad de una tecnología en cuestión (ordenador, internet o software educativo), β_{ESCS} el factor socioeconómico, β_{PRIV} estudiar en un centro privado y β_{LENGUA} tener una lengua materna distinta a la oficial. Las variables β_{TIC} , β_{PRIV} y β_{LENGUA} se han codificado como una dummy.

Tabla 3. Importancia de los recursos disponibles en casa en las pruebas de matemáticas

	Variable Independiente	Beta	Intervalo de confianza al 95%	
			Mínimo	Máximo
Ec.1	Constante	547,26	531,41	563,12
	No tener ordenador	-57,67	-71,63	-43,71
Ec.2	Constante	527,66	513,24	542,09
	ESCS	28,4	25,38	31,43
	No tener ordenador	-31,06	-43,88	-18,23
Ec.3	Constante	518,13	498,57	537,69
	ESCS	26,84	23,95	29,74
	Estudiar en centro privado	12,18	5,45	18,91

	Lengua materna distinta	-6,73	-12,69	-0,76
	No tener ordenador	-30,6	-43,45	-17,74
Variable Independiente		Beta	Intervalo de confianza al 95%	
			Mínimo	Máximo
Ec.1	Constante	507,38	499,70	515,06
	No disponer Software educativo	-14,56	-19,13	-9,99
Ec.2	Constante	500,32	493,51	507,13
	ESCS	29,61	26,89	32,33
	No disponer Software educativo	-3,63	-7,57	0,32
Ec.3	Constante	490,45	476,88	504,01
	ESCS	28	25,37	30,63
	Estudiar en centro privado	12,47	5,85	19,1
	Lengua materna distinta	-6,51	-12,48	-0,54
	No disponer de Software educativo	-3,51	-7,4	0,38
Variable Independiente		Beta	Intervalo de confianza al 95%	
			Mínimo	Máximo
Ec.1	Constante	536,97	527,40	546,54
	No disponer conexión a internet	-44,4	-51,3	-37,49
Ec.2	Constante	512,66	503,90	521,43
	ESCS	28,16	25,27	31,05
	No disponer conexión a internet	-15,77	-22,7	-8,84
Ec.3	Constante	504,58	489,58	519,59
	ESCS	26,8	24,08	29,51
	Estudiar en centro privado	11,07	4,11	18,04
	Lengua materna distinta	-7,22	-13,21	-1,24
	No tener conexión a internet	-15,04	-21,93	-8,15

Fuente: Elaboración propia con datos de PISA 2009

En primer lugar, sin incluir ninguna variable de control, el efecto de no tener ordenador en el hogar supone 57,7 puntos menos en las pruebas PISA (con un intervalo de confianza a nivel de significación del 95% que oscila entre 71,6 y 43,71). Controlando la disponibilidad de ordenador con el nivel socioeconómico (ESCS), la influencia sigue siendo notable, de 31,1 puntos. Similar es el efecto si añadimos en la ecuación el factor estudiar en un centro privado o el hecho de que la lengua materna sea

distinta de la oficial.

La influencia de la disponibilidad de software educativo en el hogar es bastante menor, produciendo su ausencia 14,6 puntos menos en las pruebas de matemáticas si no consideramos otros factores y reduciéndose en 3,6 al incluir el ESCS. Al igual que ocurre con el ordenador, otros factores como estudiar en un centro privado o que la lengua materna sea distinta de la oficial no modifica la contribución del software al resultado académico.

Por último, no disponer de conexión a internet en el hogar propicia obtener 44,4 puntos menos en las pruebas de matemáticas, descendiendo a 15,8 puntos al incluir el efecto del ESCS y manteniéndose similar este efecto al incluir otras variables como la titularidad del centro o la lengua materna del estudiante.

Por lo tanto, la disponibilidad de ordenador en el hogar es el elemento que potencia en mayor medida obtener mejores resultados en matemáticas, seguido de la disponibilidad de conexión a internet y del software educativo.

La Tabla 4 muestra la influencia de la disponibilidad de ordenador y del resto de variables independientes distinguiendo entre dos grupos de alumnos según el nivel socioeconómico de la familia: el más alto y el más bajo, tras dividir a los alumnos en tres niveles, atendiendo a su ESCS. Puede comprobarse que no disponer de ordenador en el hogar afecta por igual a alumnos de clase social alta y baja, mientras que el efecto positivo de estudiar en un centro de titularidad privada es mayor entre los alumnos de un nivel socioeconómico inferior.

Tabla 4. Impacto del ordenador en las pruebas de matemáticas según el estrato socioeconómico

	Nivel socioeconómico bajo			Nivel socioeconómico alto		
	Coef.	Intervalo de confianza al 95%		Coef.	Intervalo de confianza al 95%	
		Mínimo	Máximo		Mínimo	Máximo
R-Cuadrado	0,06	0,04	0,08	0,05	0,03	0,07
Constante	509,38	486,67	532,09	532,15	494,16	570,14
ESCS	25,23	17,29	33,17	28,87	21,91	35,82
Ser privado	15,53	7,30	23,76	8,16	-0,59	16,91
Lengua materna distinta a la oficial	-4,57	-12,21	3,07	-14,63	-25,00	-4,26
No tener ordenador	-30,05	-46,95	-13,16	-30,86	-58,28	-3,44

Fuente: Elaboración propia con datos de PISA 2009

Finalmente, en la Tabla 5 se muestra el impacto diferenciado sobre las pruebas de matemáticas en hombres y mujeres de las variables estudiadas: titularidad del centro, disponibilidad de TIC en casa y disponibilidad de TIC en el colegio. Puede comprobarse que la disponibilidad de TIC contribuye de forma similar a hombres y mujeres, mientras que no se puede determinar que la mayor disponibilidad de TIC en el centro educativo tenga un efecto positivo en el rendimiento académico del alumno.

Tabla 5. Impacto del uso de TIC en casa y en el hogar según el sexo

Variable Independiente	Mujeres			Hombres		
	Beta	Intervalo de confianza al 95%		Beta	Intervalo de confianza al 95%	
		Mínimo	Máximo		Mínimo	Máximo
R-Cuadrado	0,16	0,12	0,20	0,14	0,12	0,17
Constante	469,91	458,83	480,98	478,88	469,36	488,39
ESCS	27,81	23,37	32,25	26,11	23,35	28,88
TIC en casa	6,95	2,98	10,91	7,54	5,22	9,86
TIC en colegio	1,09	-2,57	4,75	1,64	-0,86	4,14
Ser privado	11,23	2,89	19,57	10,91	3,98	17,84

Fuente: Elaboración propia con datos de PISA 2009

4. Conclusiones y discusión

Según el análisis realizado, podemos afirmar que la disponibilidad de las nuevas tecnologías de la información en el hogar tiene un impacto positivo y de bastante alcance en los resultados en las pruebas de matemáticas de PISA 2009. De las distintas modalidades de TIC estudiadas, la de mayor efecto es el ordenador, situándose a continuación la conexión a internet y por último la disponibilidad de software educativo.

Las posibles razones que expliquen esta relación pueden ser de dos tipos. En primer lugar, en la actualidad es habitual que los profesores de enseñanza primaria y secundaria propongan a los alumnos la realización, a través de internet u otras plataformas, de ejercicios de matemáticas de refuerzo y complementarios a los realizados en la clase. La no disponibilidad de TIC en el hogar supone una desventaja para estos alumnos.

Una segunda posible explicación tiene que ver con las posibilidades que ofrece internet y el ordenador para aumentar la eficiencia de las tareas, deberes y trabajos a realizar para el conjunto de asignaturas. En este sentido, los alumnos que carezcan de TIC en el hogar deberán emplear más tiempo en la realización de estas tareas, quedando limitado para ellos el tiempo disponible para estudiar matemáticas y el resto de asignaturas.

En cuanto al posible impacto diferenciado de las TIC, no se aprecian influencias distintas entre las clases sociales bajas y altas, ni tampoco distinguiendo entre sexos. Por otra parte, no hay indicios de que el rendimiento de los alumnos de 15 años en las pruebas de matemáticas se vea influido por la mayor o menor disponibilidad de TIC de los centros.

Referencias bibliográficas

1. K.Sosin, B.J. Blecha, R. Argawal, R.L. Batlett and J.I. Daniel, Efficiency in the Use of Technology in Economic Education: Some Preliminary Results, *The American Economic Review*, 94(2004), 253-258
2. M.Webb and M. Cox, A Review of Pedagogy Related to Information and Communications Technology , *J. Technology, Pedagogy and Education*, 13(2004), 235-286
3. S. Assar, R. El Amrani and R. Watson, ICT and education: A critical role in human and social development, *J. Information Technology for Development*. 16(2010), 151-158

4. F. Gandol, E. Carrillo and M.A. Prats, Potencialidades y limitaciones de la pizarra digital interactiva. Una revisión crítica de la literatura, *Revista de Medios y Educación*, 40(2012), 171-183
5. I. Heemskerck, I. Volman and W. Admiraal, Gender Inclusiveness in Educational Technology and learning experience of girls and boys, *Journal of Research on Technology in Education*, 41(2009), 256-276
6. F. Biagi and M. Loi, *ICT and Learning: Results from PISA 2009* (Joint Research Centre, Institute for the Protection and Security of the Citizen, Luxembourg, 2012)
7. J.M. Cordero, E. Crespo, F. Pedraja, Educational achievement and determinants in PISA: A review of literature in Spain, *Revista de Educación*, 362 (2013)
8. E. Hanushek, Conceptual and Empirical Issues in the Estimation of Educational Production Functions, *Journal of Human Resources*, 351(1979), 351-388
9. L.M. Doncel, J.Sáinz, I. Sanz, An Estimation of the Advantage of Charter over Public Schools, *KYKLOS*, 65(2012), 442-463
10. E. Hanushek, J. Kain, S. Rivkin, G. Branch, Charter school quality and parental decision making with school choice, *National Bureau of Economic Research* , 11252(2005)
11. A. Böhlmark and M. Lindhal, Does School Privatization Improve Educational Achievement? Evidence from Sweden's Voucher Reform, *IZA*, 3691(2008)
12. CECE, *Informe de Tecnología Educativa 2011*(Instituto de Técnicas Educativas de la CECE, Madrid, 2012)
13. D. Glover and D. Miller, Running with technology: the pedagogic impact of the large-scale introduction of interactive whiteboards in one secondary school, *Journal of Information Technology for Teacher Education*, 10 (2006), 257-278
14. J.M. Sáez and P.A. Jiménez, La aplicación de la pizarra digital interactiva: un caso en la escuela rural en primaria, *ENSAYOS, Revista de la Facultad de Educación de Albacete*, 26(2011), 1-16
15. Fundación Telefónica, *Las TIC en la educación, Realidad y Expectativas* (Ariel, Madrid, 2011)
16. N. Law, M.W. Lee, A. Chan, A. Yuen, Factors Influencing the Impact of ICT-use on Students' Learning, *The Proceedings of IRC*, (2008), 1-15
17. B.Eickelman, Supportive and hindering factors to a sustainable implementation of ICT in schools, *Journal for educational research online*, 3 (2011), 75-103
18. M.M. Molina, Reformas en el sistema educativo español para potenciar la eficiencia y la libertad de elección de los padres. Resumen de las conclusiones del Workshop ELEDUCA celebrado en abril de 2012, *Documentos de Trabajo ELEDUCA. CEU Ediciones*, 2 (2013)

19. OECD, *Are the new millenium lernerers making the grade? Technology use and education performance in PISA*, (Center for Educational Research and Innovation, 2010)
20. OECD, *PISA Data Analysis Manual SPSS, SECOND EDITION*, (Programme for International Student Assessment , 2009)

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN JOB CHARACTERISTICS AND THE GENDER WAGE GAP

JAUME GARCÍA

jaume.garcia@upf.edu

Departament d'Economia i Empresa, Universitat Pompeu Fabra

PEDRO JESÚS HERNÁNDEZ

nani@um.es

Departamento de Fundamentos del Análisis Económico, Universidad de Murcia

ANGEL LÓPEZ NICOLÁS

angel.lopeznicolas@gmail.com

Departamento de Economía, Universidad Politécnica de Cartagena

ABSTRACT: This paper re-examines gender wage differences, taking into account not only worker characteristics but also job characteristics. Consideration of a wide set of “job quality” indicators can explain a fraction of the wage gap that would otherwise be attributed to pure wage discrimination. In any case, the fraction of the wage gap that remains associated to differential rewards for identical factors across sexes is still substantial. Our results suggest that in order to avoid overestimation of the fraction of the wage gap attributable to discrimination, it is necessary to control for job characteristics.

JEL classification: J7, C4

Keywords: wage differentials, gender discrimination

RESUMEN: Este artículo evalúa las diferencias salariales por género teniendo en cuenta no sólo las características de los trabajadores sino las del puesto de trabajo. La consideración de un amplio conjunto de indicadores de "calidad del trabajo" explica una fracción de la brecha salarial que de otra manera sería atribuida a la discriminación. No obstante la fracción de la brecha salarial que refleja una remuneración diferencial entre géneros para el mismo factor es sustancial. Nuestros resultados sugieren que es necesario controlar por características del puesto de trabajo en aras a evitar la sobreestimación del componente discriminatorio de la brecha salarial entre hombres y mujeres.

1. Introduction

In most gender wage gap evaluations, the analysis has focused on the worker characteristics associated with different wages for men and women, independently of the type of job done. This paper proposes a complementary approach, asking if and by how much the same job characteristic is rewarded differently for men and women. If wages are related to productivity, and productivity influences promotion up the job ladder, then some job characteristics will have explanatory power in addition to individual proxies for productivity, such as education. Indeed, many critiques in the literature on wage discrimination focus on the fact that in wage regressions, sex dummy variables (or alternatively, the differences in coefficients in separate regressions for men and women) are just picking up the effect of unobserved productivity differentials between genders. The analysis in this paper intends to disentangle these effects by means of enriching the specification of the wage equation to account for job characteristics.

Our approach is related to that adopted in studies addressing the issue of occupational segregation in that we intend to control for the fact that assignment of women to jobs is not the same as that of men. Indeed, the majority of studies on segregation find that part of the wage gap may be explained by the fact that women tend to concentrate in low wage occupations. However, even after controlling for segregation, Miller (1987b) and Hernández (1996) find that there remains a substantial proportion of unexplained wage differences. One hypothesis arising from their results is that there is further (unobserved) segregation within each of the considered occupations. The difficulty of testing this hypothesis is rooted in the fact that information on job characteristics is usually scarce in labour surveys. This study uses information from a database, the “*Encuesta sobre Estructura, Conciencia y Biografía de Clase*” (1991), comprising responses to a series of questions designed to capture precisely the job characteristics related to the commonly “unobserved” job segregation mentioned above.

The idea motivating inclusion of these variables in a wage gap decomposition exercise is clear: if, once job characteristics have been fully taken into account, there is no unexplained wage gap then the discrimination hypothesis will appear less likely. Whether access to these job characteristics is itself a result of differential treatment of men and women in promotion and job assignment is, of course, another issue. If, on the contrary, one finds differential rewards for the same job characteristic, then our inference will be quite the opposite.¹

We specify a model accounting for the fact that the wage equation can have sample selection problems due to participation and that individuals sort themselves into different occupations. Wages are determined by several job characteristics and individual variables such as age and education. From the model estimates, we shall

¹ This study is also related to studies such as Bayard *et al* (1999) and De la Rica and Felgueroso (2001) which control for the proportion of females within each occupational class, each industry and each firm. Also, work by Johnson and Solon (1986), Sorensen (1990) and Macpherson and Hirsch (1995) show that wages are negatively associated with the proportion of women within the worker’s occupational class.

implement the wage decomposition procedure proposed by Neumark (1988). Given the presence of two selection processes in our model, we shall pay special attention to how decompositions of the wage gap need to be carried out in the presence of non random assignment to different groups in the labour market. In particular we shall follow the procedures proposed by Neuman and Oaxaca (1998) in carrying out the decompositions.

Our results suggest that job characteristics are important factors in explaining wages even when controlling for individual characteristics. Moreover, when we account for job characteristics, the fraction of the gender wage gap attributable to differential rewards for men and women is reduced, reflecting the fact that men tend to be assigned to the “best” jobs. However, there remains a substantial and significant “discriminatory” component, in that the reward for job and individual characteristics is higher for men.

In section 2 we present the econometric model. Section 3 comprises discussion of the data set. The empirical results are presented in section 4, while, finally, section 5 concludes.

2. The econometric model

The objective is to estimate wage equations for both males and females in order to decompose the observed wage gap into its discrimination component and the part attributable to the different observed male and female characteristics. As usual, to obtain consistent estimates for the wage equation we have to control for the potential selectivity problem generated by the fact that wages are only observed for those who are participants.² Additionally, we also wish to control for the fact that having a particular occupation can also be endogenous, in the sense that the unobserved factors affecting the choice of occupation may be related to the unobserved productivity in the wage equation. Consequently, the model we are considering comprises three equations. The endogenous variables of these equations are w , L , and I , where w is the logarithm of the wage, L is a binary indicator for participation in the labour market and I is a categorical variable signalling occupation group within a set of $j=1, \dots, M$ possibilities. I and w are only observed when L is equal to one, i.e., when the individual is employed. The model can be written as:

² Méndez and Hernández (2001) analyse the sample selection bias that could arise in the estimation of an earnings equation in economies where unemployment is a relevant feature of the labour market, as is the case in Spain. The differences between non-participant and unemployed people point to a double sample selection mechanism.

$$w_i = X_i' \beta + u_i \quad (1)$$

$$L_i = 1(Y_i' \alpha + \eta_i > 0) \quad (2)$$

$$I_i^{*j} = Z_i' \gamma^j + v_i^j, \quad I_i = j \Leftrightarrow I_i^{*j} = \text{Max}(I_i^{*1}, \dots, I_i^{*M}), \quad j=1, \dots, M \quad (3)$$

where X , Y and Z are vectors of explanatory variables for each equation and u , η and v^j are the corresponding error terms which are assumed to have zero mean and the following covariance structure:

$$\Omega^j = \begin{pmatrix} \sigma_u^2 & \sigma_{\eta u} & \sigma_{uv^j} \\ \sigma_{\eta u} & \sigma_v^2 & 0 \\ \sigma_{uv^j} & 0 & 1 \end{pmatrix} \quad j=1, \dots, M$$

In order to estimate the model above we assume a probit structure for the participation equation (2) and a multinomial logit structure for the occupation equation (3). We estimate the wage equation (1) separately for females and males by OLS. In order to account for the selectivity and endogeneity issues mentioned above we include correction terms following the procedures proposed by Heckman (1979)³ and Lee (1983), but, except for the constant term, we assume the same model for all occupations; i.e.,

$$w_i = X_i' \beta + \sigma \lambda_i + \sum_{j=1}^M d_i^j * \varphi^j \frac{\phi[J(Z_i' \gamma^j)]}{F(Z_i' \gamma^j)} + u_i'$$

³ The majority of studies that analyse the relationship between wages and occupations do not correct for potential selectivity biases. See Trost and Lee (1984), Miller (1987a and 1987b), Reilly (1991) and Hernández (1996), among others, and Dolton *et al* (1989) as an exception.

where

$$J(Z'_i \gamma^j) = \Phi^{-1} \left(\frac{\exp(Z'_i \gamma^j)}{1 + \sum_{s=2}^M \exp(Z'_i \gamma^s)} \right)$$

$$F(Z'_i \gamma^j) = \frac{\exp(Z'_i \gamma^j)}{1 + \sum_{s=2}^M \exp(Z'_i \gamma^s)}$$

$$\lambda_i = \frac{\phi(Y'_i \alpha / \sigma_\eta)}{\Phi(Y'_i \alpha / \sigma_\eta)}$$

$$\sigma = \sigma_u \rho_{\eta u}; \quad \varphi^j = \sigma_u \rho_{\eta v^j}; \quad d^i_j = 1(i \in j)$$

where $\Phi(\cdot)$ y $\phi(\cdot)$ are the c.d.f. and p.d.f. of the standard normal distribution, respectively.

3. Data and variables

Our estimating sample is extracted from the *Encuesta sobre Estructura, Conciencia y Biografía de Clase* (1991).⁴ This survey contains abundant information about socio-economic variables and it is unique, at least as far as the Spanish labour market is concerned, in the sense that it contains a great deal of information on job characteristics.

The endogenous variable in the wage equation (1) is the (log) hourly wage and that of the participation equation (2) is a discrete variable taking the value 1 if the observation corresponds to a worker and 0 otherwise. On the other hand, when defining the occupational groups corresponding to the endogenous variable in equation (3), we have followed the official classification of occupations in Spain (*Clasificación Nacional de Ocupaciones*, 1979) with some slight changes due to small samples for some of the original cells. Namely, the classification we use is the following:

- Occupation 1:* Technicians, Directing staff in the public and private sector
- Occupation 2:* Administrative staff
- Occupation 3:* Sales representatives and retail staff
- Occupation 4:* Restaurant and hotel staff

⁴ The survey was carried out by the *Instituto Nacional de Estadística*, the *Comunidad Autónoma de Madrid* and the *Instituto de la Mujer*. See Carabaña et al. (1992) for more detailed information.

Occupation 5: Agriculture, fisheries and mining, and manual industrial workers.

The explanatory variables used in the different equations are defined as follows:

Age: Number of years

Seniority: Number of years in the present job

Indefinite contract: Dummy variable equal to one when the individual has an indefinite contract.

Education: Dummy variables for the following educational levels: illiterate, primary school, secondary school, vocational training, university degree (intermediate) and university degree (higher).

Years of education

Married: Dummy variable equal to one if the individual is married.

Head of the household: Dummy variable equal to one if the individual is the head of the household.

Number of children

Number of dependent adults

Number of income earners

Industrial sector: Dummy variables corresponding to classification in 12 sectors we have defined.

Region: Dummy variables for each of the 17 Spanish regions (*comunidades autónomas*)

Gross wage control: Dummy variable equal to one when reported wages are gross wages.

Regarding job characteristics, we have distinguished five blocks of variables: i) those related to the degree of worker autonomy in setting the working pace within the workplace, ii) those related to the degree of control by others over the worker's output, iii) those related to the degree of supervisory/directive powers over other employees, iv) those related to the power to decide on issues related to other employees and v) those related to the education mismatch. The definition of the specific variables included in each of these groups is presented in Table 1. Unless otherwise specified, these are binary indicators (Yes=1; No=0).

(TABLE 1)

Table A.1 in the Appendix contains descriptive statistics of all the variables used in the empirical analysis for wage earners split by gender. The average male hourly wage is 14.31% higher than that for women.⁵ On average, men have also accumulated more seniority (11.01 versus 7.67 years) and are more likely to have an indefinite (non fixed-term) contract than women. There are also noteworthy

⁵ When making the wage decompositions we will approximate this differential by the difference of the (log) wages, which is 13.54%.

differences between men and women in educational levels and sectoral concentration. Firstly, the level of education would appear to be higher for women than for men and, secondly, we find that women are more likely to work in the Public Administration, the so-called “Reproduction” sector (teaching, scientific research and health care) and the Social Services and Domestic Work sector. We also observe that men are more likely to be in occupation 5 whereas women mainly seem to cluster in occupations 1 and 2.⁶

As for the variables related to job characteristics, we find that men occupy the majority of directive and supervising positions. Up to 7.5% of employed men occupy these positions as opposed to only 2.2% of women. Also, while more than 15.7% of men carry out supervision tasks, only 11.5% of women do so.

The explanatory variables included in the reduced form participation equation are age and its square, marital status, worker education and education of partner, whether the individual is the head of the household, number of children, number of dependent adults, number of income earners in the household and regional dummies. The estimation results are reported in Table A.2 in the Appendix. As expected, there are substantially different patterns for the participation equations of both males and females. Education has a more important and significant effect for females. The number of children has a negative and significant effect on female participation, whereas for males the effect is positive, although not significant. This result is similar to the effect of being married, which is negative, although not significant at a 5% significance level.

The vector of explanatory variables in the occupation equation includes individual age, years of education, years of education of father and marital status. The estimation results are reported in Table A.3 of the Appendix. The number of years of education is the most relevant explanatory factor for occupational choices. There are some differences between males and females models, although the patterns are similar.

We use the estimates of the participation and occupation equations to construct the correction terms to be included in the wage equations. We consider four different specifications for the wage equation which have a common set of variables but differ on inclusion/exclusion of some individual characteristics and/or job characteristics.

The set of common variables are: having an indefinite contract, seniority, the type of occupation⁷ and gross wage control plus correction terms. The group of additional individual characteristics includes age and its square, educational level, industrial sector and region. The first two variables try to capture aspects related to productivity not considered in job evaluation, because they are individual not job characteristics, and the last two try to capture industrial and geographical differentials in wages. Finally, the set of variables related to job characteristics, which are taken into

⁶ Note that occupation 1 includes, among others, health care technicians. This is an occupation with a high percentage of women.

⁷ These dummies are included to allow for different constant terms depending on the occupation, as mentioned above.

account in a job evaluation process, are as defined in Table 1 in the previous section.

Model 1 includes the common set of variables plus those related to individual characteristics and will correspond to a standard specification of a wage equation in the analysis of gender wage discrimination. Model 2 also includes the variables related to job characteristics. By comparing models 1 and 2 we can evaluate the importance of job characteristics in the determination of wages and compare to what extent differences in the characteristics of the jobs occupied by males and females explain part of the observed wage gap. Models 3 and 4 have the same specification as models 1 and 2, respectively, but exclude the set of variables corresponding to individual characteristics. The estimation of models 3 and 4 will allow us to assess the importance of considering job characteristics when analysing gender wage discrimination in a context where the explanatory factors are basically those considered in job evaluation and those which can have an effect on wages through specific wage complements, such as seniority or having an indefinite contract. Additionally, we will be able to assess again the impact of individual characteristics on wages, but for a different reference model.

In Table 2 we present the estimates from the four specifications of the wage equation for both males and females. Having an indefinite contract brings a higher reward for females, although the difference is almost insignificant when not including the standard wage equation variables apart from the common variables. The effect of the experience in the current firm and also the overall experience proxied by age is more important for males than for females. The effect of education on wages is more significant for males than for females. By comparing models 1 and 2, we can infer a positive association between having a higher educational level and better job characteristics. On the other hand, there seems to be more variability of wages across industries for females.

Differences in wages due to type of occupation are reduced when age, education and the sectoral and regional dummies are included, in particular for males. In fact, there are no significant differences between the type of occupation for males in the model and the most complete specification (Model 2), although significant differences are found for females in all four models considered.

Including job characteristics in the wage equation increases the explanatory power of the model significantly. In the case of females, when we test for the joint significance of the job characteristic coefficients, we obtain an F statistic of 2.53 when comparing models 1 and 2, and 3.47 when comparing models 3 and 4, above the corresponding critical value (1.75) at a 5% significance level. In the case of the male equations, these values are 8.13 and 11.29, respectively, indicating that job characteristics have a more substantial effect on wages than in the case of females.

Variables related to the degree of autonomy have a higher (and significant) effect in the case of males. Also, a component of directive power over employees and power to make decisions on employee wages, is associated with higher wages for

males.⁸ Finally, being overeducated in terms of the educational requirements of the job has a positive and significant reward for females but has no effect on male wages.

Notice that when comparing models 2 and 4, i.e., when testing the significance of excluding the other individual characteristics in the most general model (Model 2), we obtain an F statistic of 4.92 for females, which is higher than the 2.53 obtained when comparing models 2 and 1, i.e., the significance of excluding the job characteristics in Model 2. We find the opposite result for males, an F statistic of 6.07 when comparing models 2 and 4 and 8.13 when comparing models 2 and 1. This means that the average increase in the residual sum of squares per each extra parameter not estimated is higher for the other individual characteristics for females, whereas in the case of males it is higher for the job characteristics.⁹

1.

(TABLE 2)

The model estimates can be used to implement the wage decomposition procedure proposed by Neumark (1988). This method is a generalisation of those proposed by Oaxaca (1973) and Blinder (1973), which does not assume either the male or the female wage structure as the non discriminatory ideal. Instead, the latter is obtained from an estimation using the joint sample.

In short, this procedure decomposes the difference between average (log)wages for men and women into differences in observable characteristics and differences in rewards for these characteristics. In our case, given the existence of the correction terms, this wage gap decomposition takes the following form:

$$\begin{aligned} \bar{w}^m - \bar{w}^f = & \left[(\bar{X}^m - \bar{X}^f)' \hat{\beta} + (\bar{\lambda}^m - \bar{\lambda}^f) \sigma^b + \sum_{j=1}^M \hat{\phi}_j^b \left\{ \overline{\left(\frac{\phi}{F} \right)_j^m} - \overline{\left(\frac{\phi}{F} \right)_j^f} \right\} \right] + \\ & \left[\bar{X}^m (\hat{\beta}^m - \hat{\beta}) + (\sigma^m - \sigma^b) \bar{\lambda}^m + \sum_{j=1}^M (\hat{\phi}_j^m - \hat{\phi}_j^b) \overline{\left(\frac{\phi}{F} \right)_j^m} - \right. \\ & \left. \bar{X}^f (\hat{\beta}^f - \hat{\beta}) - (\sigma^f - \sigma^b) \bar{\lambda}^f - \sum_{j=1}^M (\hat{\phi}_j^f - \hat{\phi}_j^b) \overline{\left(\frac{\phi}{F} \right)_j^f} \right] \end{aligned}$$

⁸ Notice that the negative and almost significant coefficient for this latter variable in the case of females is due to the very few cases in the female subsample with a job of these characteristics, as can be seen from the descriptive statistics in Table A.1.

⁹ If we look at the adjusted R² the reduction is smaller for both males and females when excluding the job characteristics. This does not contradict the F statistics analysis because the number of restrictions is different.

where \hat{b} is the Neuman estimator for the vector of coefficients associated to X_i and $\hat{\sigma}^b$ and $\hat{\varphi}^b$ are the estimates for the parameters associated to λ_i and $(\phi / F)_i$. The indices m and f refer to the subsample of men and women respectively. The first element on the right hand side of the last equation represents the part of the wage gap that can be explained by different attributes of men and women whereas the second term represents the “unexplained” part of the wage gap, that is, the part usually attributed to gender discrimination, which arises from a differential reward for the same observable characteristics.¹⁰

We follow one of the possibilities presented in Neuman and Oaxaca (1998) and consider differences in the coefficients for the selection terms (participation and occupation) as manifestations of discrimination. The differences in the average values of these terms are considered as differences in characteristics. However, as Neuman and Oaxaca indicate, a discriminatory component could be extracted from this: the difference between the correction term for females and what it would be if the parameters that govern their selection process were the same as those for males. In this case, the part attributed to discrimination would be greater than our estimation.

In Table 3 we present the decomposition of the observed wage gap between males and females (13.54%) in the two terms mentioned above: the part corresponding to differences in characteristics and the part corresponding to discrimination.

(TABLE 3)

Notice that when using the specifications without the job characteristics (models 1 and 3), the proportion of the wage gap which can be considered as discrimination exceeds 100%.¹¹ There are no substantial differences associated with inclusion or exclusion of the rest of the individual characteristics (Model 1 vs Model 3, Model 2 vs Model 4). However, when including the job characteristics as explanatory factors in the wage equation the proportion of the wage gap attributable to discrimination is reduced, although there is still a large portion of the wage gap which cannot be explained by either individual differences or differences in job characteristics (84.7% in Model 2 and 80.9% in Model 4).

5. Summary

In this paper we have attempted to examine the gender wage gap under the assumption that in addition to individual worker characteristics, job features also contribute to explaining wage differences. Our strategy was to enrich the traditional

¹⁰ Decomposition of the sample selection correction terms follows the procedures laid out in Neuman and Oaxaca (1998).

¹¹ Similar results are obtained by De la Rica and Ugidos (1995) using a data set from the same survey.

wage equation specification with an ample set of indicators accounting for important job characteristics. As expected, the results for the sequence of models that we have estimated confirm that these characteristics explain a significant proportion of wage variation. Moreover, once these characteristics are taken into account, the portion of the wage gap attributable to discrimination is reduced. This reflects the fact that a higher proportion of men attain positions of greater responsibility, autonomy, degree of control over work processes etc. than women. However, our estimates also suggest that women are rewarded differentially when they achieve these positions too.

These findings are important in that they support the view that, in the Spanish labour market, differential rewards exist for men and women even when they do the same job, i.e., there is pure wage discrimination. This complements previous Spanish evidence showing that men and women with equal characteristics (but not necessarily doing the same job) received different wages. A number of issues merit further research and they all have to do with the fact that even with this data set it is impossible to find a perfect case-control situation. The first issue is that the set of characteristics controlled for does not include all relevant work characteristics. The standard job evaluation methods suggest many variables that could pick up relevant characteristics, but unfortunately our data set does not include them. Also, the answers in our data set are provided by the worker and, as such, there may be a certain degree of subjectivity. One way to overcome these shortcomings would be construction of databases which record all relevant worker characteristics, possibly within the firm itself.

References

- Bayard, K., Hellerstein, J., Neumark, D. and Troske, K. (1999): "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data", NBER Working Paper No. W7003.
- Blinder, A.S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, 8: 436-55.
- Carabaña, J. et al. (1992): "Encuesta de Estructura, Conciencia y Biografía de Clase: Informe Técnico", Documento de Trabajo, Comunidad Autónoma de Madrid, Consejería de Economía.
- De la Rica, S. and Ugidos, A. (1995): "¿Son las diferencias en capital humano determinantes de las diferencias salariales observadas entre hombres y mujeres?", *Investigaciones Económicas*, 19: 395-414.
- De la Rica, S. and Felgueroso, F. (2001): "Intrafirm gender wage differences: A worker's matching approach". Mimeo.
- Dolton, P.J., Makepeace, G.H. and Van Der Klaauw, W. (1989): "Occupational Choice and Earnings Determination: The Role of Sample Selection and Non-Pecuniary Factors", *Oxford Economic Papers*, 41: 573-94.

- Heckman, J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47: 153-61.
- Hernández, P.J. (1996): "Segregación ocupacional de la mujer y discriminación salarial", *Revista de Economía Aplicada*, 11: 57-80.
- Johnson, G. and Solon, G. (1986): "Estimates of the Direct Effects of Comparable Worth Policy", *American Economic Review*, 76: 1117-25.
- Lee, L.F. (1983): "Generalized Econometric Models with Selectivity", *Econometrica*, 51: 507-12.
- Méndez, I. and Hernández, P.J. (2001): Labour supply, sample selection and wage discrimination, 13th EALE Annual Conference, 13-16 September 2001.
- Macpherson, D.A. and Hirsch, B.T. (1995): "Wages and Gender Composition: Why Do Women's Jobs Pay Less?", *Journal of Labor Economics*, 13(3): 426-71.
- Miller, P. (1987a): "Gender Differences in Observed and Offered Wages in Canada, 1980", *Canadian Journal of Economics*, 20: 225-44.
- Miller, P. (1987b): "The Wage Effect of the Occupational Segregation of Women in Britain", *Economic Journal*, 97: 885-96.
- Neuman, S. and Oaxaca, R.L. (1998): "Estimating Labor Market Discrimination with Selectivity Corrected Wage Equations: Methodological Considerations and An Illustration from Israel". Mimeo.
- Neumark, D. (1988): "Employers Discriminatory Behavior and the Estimation of Wage Discrimination", *Journal of Human Resources*, 23: 279-95.
- Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labour Markets", *International Economic Review*, 14: 693-709.
- Reilly, B. (1991): "Occupational Segregation and Selectivity Bias in Occupational Wage Equations: an Empirical Analysis Using Irish Data", *Applied Economics*, 23:1-7.
- Sorensen, E. (1990): "The Crowding Hypothesis and Comparable Worth", *Journal of Human Resources*, 25: 55-89.
- Trost, R.P. and Lee, L.F. (1984): "Technical Training and Earnings: a Polychotomous Choice Model with Selectivity", *Review of Economics and Statistics*, 66:151-56.

2. APPENDIX

Table 1: Definition of the job characteristic variables

<p>Variables related to the worker's autonomy</p> <p><i>Own work:</i> Can make decisions on important aspects of own work</p> <p><i>Working pace:</i> Can reduce working pace in working day</p> <p><i>Working hours:</i> Can make decisions on working hours</p>
<p>Variables related to the degree of control by others over the worker's output</p> <p><i>Quality control:</i> Degree of control over output quality (1: very easy 2: quite easy. 3: easy. 4: not easy at all)</p> <p><i>Quantity control:</i> Degree of control over output quantity (1: very easy 2: quite easy. 3: easy. 4: not easy at all)</p>
<p>Variables related to the degree of supervisory/directive power over other employees</p> <p><i>Director:</i> Director position</p> <p><i>Supervisor:</i> Supervisor position</p>
<p>Variables related to the employee's power to make decisions related to other employees</p> <p><i>Raising a worker's salary:</i> Maximum influence in the decision to raise a worker's salary.</p> <p><i>Penalising a worker:</i> Maximum influence in penalising a worker.</p> <p><i>Changing methods:</i> To make decisions regarding changes in the procedures or basic working methods</p>
<p>Variables related to the education mismatch</p> <p><i>Education more than sufficient:</i> Education more than sufficient for the present work.</p> <p><i>Education sufficient:</i> Educational level sufficient for the present work.</p>

Table 2: Estimation results for the wage equation

	Model 1		Model 2		Model 3		Model 4	
	Female	Males	Female	Males	Female	Males	Female	Males
Variables	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Common variables								
Constant	4.91 (21.1)	5.30 (9.21)	4.91 (20.0)	5.53 (9.92)	5.51 (41.8)	4.98 (9.12)	5.29 (36.0)	5.24 (10.6)
Indefinite contract	0.11 (2.86)	0.06 (1.76)	0.09 (2.39)	0.05 (1.62)	0.12 (3.13)	0.11 (0.7)	0.10 (2.61)	0.08 (2.54)
Seniority	0.00 (1.5)	0.00 (4.8)	0.00 (1.5)	0.00 (5.8)	0.00 (1.5)	0.00 (4.6)	0.00 (2.6)	0.00 (5.7)
Gross wage control	0.30 (7.5)	0.13 (4.45)	0.30 (7.96)	0.12 (4.05)	0.30 (7.53)	0.18 (5.74)	0.28 (7.18)	0.12 (4.08)
Occupations:								
Occupation 1	0.66 (4.3)	0.43 (0.79)	0.65 (3.88)	0.22 (0.42)	1.23 (9.07)	1.98 (3.62)	1.15 (8.33)	1.51 (2.88)
Occupation 2	0.43 (2.3)	0.26 (0.46)	0.44 (2.34)	0.13 (0.4)	0.58 (3.48)	1.72 (3.9)	0.53 (3.20)	1.30 (4.4)
Occupation 3	0.49 (1.7)	-0.07 (0.60)	0.47 (1.65)	0.17 (1.91)	0.41 (1.46)	-0.06 (0.05)	0.41 (1.48)	0.07 (0.58)
Occupation 5	0.46 (2.7)	0.06 (0.12)	0.47 (2.16)	0.04 (0.1)	0.29 (1.34)	0.99 (3.3)	0.32 (1.52)	0.72 (4.0)

<i>Other individual variables</i>								
Age	0.017	(1.77)	0.034	(3.61)	0.015	(1.50)	0.021	(2.30)
Age squared (/100)	-0.021	(1.75)	-0.011	(3.59)	-0.018	(1.48)	-0.026	(2.32)
<i>Education:</i>								
Primary school	-0.018	(1.44)	0.008	(1.02)	-0.011	(1.53)	0.002	(0.72)
Secondary school	-0.027	(0.26)	0.024	(2.88)	-0.006	(0.65)	0.013	(1.98)
Vocational training	-0.005	(0.83)	0.011	(1.57)	-0.009	(0.88)	0.007	(0.97)
University degree (intermediate)	0.014	(1.27)	0.032	(4.06)	0.011	(1.07)	0.028	(3.45)
University degree (higher)	0.033	(2.55)	0.057	(6.23)	0.022	(1.87)	0.047	(5.21)
<i>Industrial sector:</i>								
Agriculture	0.000	(0.29)	0.000	(0.37)	0.004	(0.27)	0.009	(0.05)
Basic industry	0.052	(5.00)	0.034	(1.94)	0.050	(4.84)	0.024	(1.44)

Heavy industry	0. (3. 46 94)	0. (2. 36 04 4)	05 (4. 01 26)	0. (1. 31 86)
Light industry	0. (1. 19 86)	0. (1. 17 01 8)	0. (2. 20 00)	0. (0. 10 63 7)
Buildin g	0. (3. 46 38)	0. (1. 31 76 6)	0. (3. 47 49)	0. (1. 24 43 6)
Retailin g and Hotels	0. (3. 34 82)	0. (0. 09 52 1)	0. (3. 35 84)	0. (0. 03 18 0)
Transport	0. (3. 38 12)	0. (2. 38 11 0)	0. (2. 35 82)	0. (1. 32 86 1)
Econom ics services	0. (3. 32 01)	0. (2. 42 40 5)	0. (3. 33 14)	0. (1. 33 97 5)
Public Admini stration	0. (5. 53 98)	0. (1. 32 86 4)	0. (6. 56 23)	0. (1. 28 70 3)
Reprod uction	0. (4. 39 11)	0. (1. 21 16 0)	0. (4. 41 29)	0. (1. 26 50 0)
Social services	0. (3. 40 34)	0. (0. 01 07 5)	0. (3. 39 31)	- (0. 0 10 8)
<i>Region:</i>				
Andaluc ia	0. (2. 30 47)	- (3. 0 16 33)	0. (2. 26 09)	- (1. 0 73 18)
Aragón	0. (1. 18 41)	- (2. 0 53 29)	0. (0. 12 93)	- (1. 0 07 12)
Asturias	0. (0. 08 57)	- (1. 0 99 24)	- (0. 0 12)	- (0. 0 67 08)
Baleare s	0. (1. 26 84)	- (3. 0 19 44)	0. (1. 24 75)	- (1. 0 33 18)
		6	8	8

	- (0.	- (4.	- (0.	- (3.
Canarias	0. 73)	0. 55	0. 63)	0. 12
	10	58)	09	39)
	8	1	5	9
	0. (1.	- (2.	0. (1.	- (1.
Cantabria	27 54)	0. 78	22 27)	0. 33
	0	38)	7	18)
		3		3
	0. (0.	- (2.	0. (0.	- (1.
Castilla-La Mancha	05 34)	0. 72	00 03)	0. 23
	3	35)	5	16)
		9		1
	0. (0.	- (2.	0. (0.	- (1.
Castilla-León	03 26)	0. 91	00 04)	0. 85
	8	34)	6	22)
		4		2
	0. (2.	- (2.	0. (1.	0. (1.
Cataluña	26 24)	0. 61	19 57)	12 16
	3	27)	2	2)
		5		
	0. (1.	- (2.	0. (0.	- (0.
Comunidad Valenciana	16 36)	0. 73	10 88)	0. 99
	2	28)	7	10)
		9		7
	0. (0.	- (5.	- (0.	- (3.
Extremadura	01 09)	0. 01	0. 08)	0. 48
	4	63)	03	44)
		6	5	4
	- (0.	- (3.	- (0.	- (2.
Galicia	0. 07)	0. 56	0. 65)	0. 30
	00	39)	08	25)
	9	0	4	3
	0. (1.	- (1.	0. (1.	- (0.
Madrid	22 96)	0. 92	18 54)	0. 41
	8	20)	1	04)
		1		3
	0. (2.	- (2.	0. (2.	- (0.
Murcia	36 46)	0. 22	34 31)	0. 47
	3	34)	2	07)
		2		3
	0. (1.	- (2.	0. (0.	- (1.
Navarra	22 12)	0. 17	18 96)	0. 18
	0	31)	7	16)
		6		9
País	0. (2.	- (1.	0. (2.	- (0.

Vasco	31 54)	0. 92	28 20)	0. 93
	9	21)	3	10)
		3		3

**Job
characteristics:**

Autonomy:

Own work		0. (1. 04	0. (2. 18)	05 05		0. (0. 02	0. (2. 62)	08 70)
		3	8)			4	0	

Working pace

		0. (0. 01	0. (2. 31)	06 39		0. (0. 03	0. (2. 98)	08 83)
		1	9)			4	3	

Working hours

		0. (1. 05	0. (1. 10)	04 17		0. (1. 06	0. (0. 19)	03 99)
		4	4)			2	9	

Degree of control:

Quality control		0. (1. 02	- (0. 24)	0. 33		0. (2. 07	0. (0. 98)	01 73)
		9	00)	6		0	4	

Quantity control

		- (2. 0. 04	- (0. 08)	0. 58		- (1. 0. 03	- (1. 84)	0. 53)
		5	01)	1		8	0	

Directive power over employees

Director		0. (2. 24	0. (4. 31)	25 36		0. (1. 15	0. (4. 35)	27 62)
		1	1)			0	4	

Supervisor

		0. (2. 10	0. (3. 13)	13 83		0. (2. 12	0. (4. 34)	18 79)
		1	9)			0	0	

Power to make decisions:

Raising a		- (1. 1. 22	0. (1. 64)	25 84		- (1. 1. 56	0. (2. 94)	30 04)
			3)				0	

worker's salary	1			6				
	0. (0. 0. (0. 13 71) 06 63			0. (0. 0. (0. 05 30) 02 28)				
Penalising a worker	4	2		9	9			
	0. (0. 0. (0. 02 20) 06 95			0. (0. 0. (0. 07 55) 04 67)				
Changing methods	6	7		6	7			
<i>Educational mismatch:</i>								
Educational mismatch more than sufficient	0. (1. 10 69) 6	- (0. 0. 01 00) 1		0. (3. 20 37) 8	- (0. 0. 85) 03 8			
Educational mismatch sufficient	0. (2. 11 01) 5	- (0. 0. 34 01) 3		0. (2. 17 95) 2	- (1. 0. 29) 05 2			
Correct ion terms:								
<input type="checkbox"/>	- (1. 0. 09) 03 2	- (0. 0. 27 00) 7	- (1. 0. 09) 03 3	- (0. 0. 12 00) 3	- (1. 0. 40) 03 3	- (4. 0. 65 09) 8	- (1. 0. 18) 02 9	- (3. 0. 15) 06 3
$(\phi / F)_1$	- (2. 0. 33) 17 9	- (0. 0. 82 05) 8	- (2. 0. 24) 17 5	- (0. 0. 45 03) 1	- (5. 0. 94) 38 4	- (5. 0. 93 33) 8	- (5. 0. 56) 35 6	- (5. 0. 33) 28 8
$(\phi / F)_2$	- (0. 0. 12) 01 3	- (0. 0. 32 03) 4	- (0. 0. 34) 03 7	- (0. 0. 10 01) 0	0. (0. 08 77) 1	- (2. 0. 47 25) 3	0. (0. 07 74) 8	- (1. 0. 87) 18 2
$(\phi / F)_3$	- (1. 0. 60) 24	0. (1. 43 39 3)	- (1. 0. 61) 24	0. (1. 53 77 1)	- (0. 0. 52) 08	0. (2. 65 00 9)	- (0. 0. 72) 11	0. (2. 74 40) 9

	2		2		0		0		
$(\phi / F)_4$	0. (2. 21 21)	0. (0. 08 31 9)	0. (2. 20 15) 7	0. (0. 04 17 6)	0. (3. 30 17) 2	0. (2. 61 05 3)	0. (2. 26 80) 5	0. (1. 47 67) 5	
$(\phi / F)_5$	- (1. 0. 01) 13 5	0. (0. 02 40 7)	- (1. 0. 02) 14 0	0. (0. 04 76 9)	0. (0. 08 65) 6	0. (4. 24 46 5)	- (0. 0. 21) 02 8	0. (4. 21 17) 7	
Observations	651	968	651	968	651	968	651	968	
Adjusted R-squared	0.5247	0.4830	0.5387	0.5269	0.4142	0.3681	0.4402	0.4405	

Omitted dummies are: illiterate, domestic services, La Rioja and occupation 4.

Table 3: Wage gap decomposition (%)

	Discrimination	Differences in characteristics
<i>Model 1</i>	14.15	-0.61
<i>Model 2</i>	11.47	2.07
<i>Model 3</i>	14.81	-1.27
<i>Model 4</i>	10.95	2.59

Table A.1: Descriptive statistics for wage earners

	3. Females		4. Males	
<i>Variable</i>	<i>Mean</i>	<i>S.D.*</i>	<i>Mean</i>	<i>S.D.*</i>
Hourly wage (log)	6.266	0.517	6.401	0.505
Hourly wage (pesetas)	609.4	409.5	696.6	501.6
Indefinite contract	0.665	0.472	0.781	0.414
Age	33.48	10.25	38.50	11.63
Seniority	7.67	7.30	11.01	10.18
<i>Education</i>				
Illiterate	0.042	0.201	0.075	0.263
Primary school	0.496	0.500	0.515	0.500
Secondary school	0.109	0.312	0.109	0.312
Vocational training	0.088	0.284	0.097	0.296
University degree (interm.)	0.173	0.378	0.076	0.265
University degree (higher)	0.079	0.271	0.077	0.266
<i>Industrial sector</i>				
Agriculture	0.003	0.055	0.034	0.182
Basic industry	0.066	0.249	0.106	0.308
Heavy industry	0.039	0.194	0.115	0.319
Light industry	0.121	0.326	0.146	0.354
Building	0.021	0.144	0.103	0.305
Retailing and Hotels	0.192	0.394	0.117	0.322
Transport	0.027	0.162	0.069	0.254
Economics services	0.060	0.238	0.096	0.294
Public Administration	0.194	0.396	0.115	0.319
Reproduction	0.204	0.403	0.079	0.270
Social services	0.031	0.174	0.014	0.117
Domestic services	0.026	0.160	0.004	0.064
<i>Region</i>				
Andalucía	0.092	0.289	0.130	0.337
Aragón	0.045	0.207	0.043	0.204
Asturias	0.028	0.165	0.028	0.165
Baleares	0.028	0.165	0.016	0.126
Canarias	0.026	0.159	0.025	0.158
Cantabria	0.013	0.111	0.017	0.129
Castilla-La Mancha	0.022	0.147	0.020	0.139
Castilla-León	0.030	0.171	0.037	0.188
Cataluña	0.165	0.372	0.148	0.355
Comunidad Valenciana	0.110	0.313	0.133	0.340
Extremadura	0.024	0.153	0.025	0.155
Galicia	0.057	0.232	0.079	0.269
Madrid	0.213	0.410	0.175	0.380
Murcia	0.025	0.156	0.011	0.116

Navarra	0.008	0.091	0.013	0.114
País Vasco	0.093	0.291	0.085	0.280
La Rioja	0.012	0.110	0.009	0.196
<i>Occupation</i>				
Occupation1	0.312	0.464	0.185	0.389
Occupation2	0.273	0.446	0.182	0.386
Occupation3	0.102	0.302	0.056	0.230
Occupation4	0.173	0.378	0.072	0.259
Occupation5	0.141	0.348	0.505	0.500
<i>Job characteristics</i>				
<i>Autonomy</i>				
Own work	0.447	0.498	0.467	0.499
Working pace	0.352	0.479	0.341	0.474
Working hours	0.123	0.329	0.176	0.381
<i>Degree of control</i>				
Quality control	2.071	0.947	2.261	0.929
Quantity control	2.108	0.980	2.121	0.938
<i>Directive power over employees</i>				
Director	0.022	0.148	0.075	0.264
Supervisor	0.115	0.319	0.157	0.364
<i>Power to make decisions</i>				
Raising a worker's salary	0.001	0.019	0.012	0.120
Penalising a worker	0.007	0.083	0.024	0.154
Changing methods	0.013	0.114	0.035	0.183
<i>Education mismatch</i>				
More than sufficient	0.332	0.471	0.274	0.446
Sufficient	0.586	0.493	0.610	0.488
Gross wage control	0.188	0.391	0.220	0.415
Observations	651		968	

***Unadjusted standard errors in the case of binary variables. These have been computed using the expression $[p(1-p)]^{1/2}$ and do not account for sample size.**

Table A.2: Probit estimates of the participation equation

Variables	Females		Males	
	Coef.	5. z- sta t.	Coef.	6. z- sta t.
Age	0.026	(1.08)	0.132	(5.14)
Age squared (/100)	-0.034	(1.14)	-0.166	(5.43)
Married	-0.228	(1.77)	0.004	(0.02)
<i>Education of partner:</i>				
Primary education	-0.108	(0.89)	-0.019	(0.14)
Secondary education	0.354	(1.92)	-0.050	(0.24)
Vocational training	0.068	(0.34)	0.029	(0.09)
University diploma	0.565	(3.22)	0.524	(2.26)
University degree	0.428	(2.63)	0.524	(2.26)
<i>Education:</i>				
Primary school	0.336	(2.49)	0.229	(1.55)
Secondary school	0.778	(4.52)	0.327	(1.79)
Vocational training	0.530	(3.03)	0.462	(2.50)
University degree (intermediate)	0.899	(5.94)	0.173	(0.99)
University degree (higher)	0.643	(3.88)	0.274	(1.77)
<i>Household variables:</i>				
Head of household	0.503	(4.55)	0.456	(2.96)
Number of children	-0.092	(2.46)	0.056	(1.27)
Number of dependent adults	0.110	(2.92)	0.176	(4.74)
Number of income earners	-0.289	(5.16)	-0.596	(11.80)
<i>Region:</i>				
Andalucía	-1.110	(2.21)	-0.662	(1.07)
Aragón	-0.865	(1.65)	-0.041	(0.06)
Asturias	-0.909	(1.79)	-0.175	(0.26)
Baleares	-0.248	(0.43)	-0.173	(0.25)
Canarias	-1.167	(2.18)	-0.775	(1.20)
Cantabria	-0.515	(0.80)	-0.024	(0.03)
Castilla-La Mancha	-1.205	(2.29)	-0.274	(0.42)
Castilla-León	-1.365	(2.66)	-0.309	(0.49)
Cataluña	-0.605	(1.21)	0.058	(0.09)
Comunidad Valenciana	-0.773	(1.53)	-0.354	(0.57)
Extremadura	-1.121	(2.06)	-0.501	(0.77)
Galicia	-1.020	(1.98)	-0.240	(0.38)
Madrid	-0.751	(1.51)	-0.139	(0.22)
Murcia	-0.715	(1.33)	-0.869	(1.32)
Navarra	-0.872	(1.48)	0.0222	(0.03)

País Vasco	-0.922	(1.80)	-0.533	(0.84)
Constant	0.278	(0.42)	-1.510	(1.91)
Observations	1701		1986	
Log L	-994.765		-708.037	

Omitted dummies: Not married, illiterate, illiterate partner, not being head of the household and La Rioja.

Table A.3: Multinomial logit for occupational choice (females)

Absolute value of z-statistics in parenthesis

Variables	$Ln \left[\frac{Pr.(Oc. 1)}{Pr.(Oc. 5)} \right]$	$Ln \left[\frac{Pr.(Oc. 2)}{Pr.(Oc. 5)} \right]$	$Ln \left[\frac{Pr.(Oc. 3)}{Pr.(Oc. 5)} \right]$	$Ln \left[\frac{Pr.(Oc. 4)}{Pr.(Oc. 5)} \right]$
Age	0.078 (3.72)	-0.010 (0.36)	-0.020 (0.76)	0.011 (0.54)
Years of education	0.667 (6.72)	0.416 (4.64)	0.146 (1.38)	-0.028 (0.33)
Father's years of education	0.123 (2.13)	0.107 (1.95)	0.076 (1.18)	-0.032 (0.52)
Married	0.276 (0.54)	0.337 (0.71)	-0.412 (0.75)	-0.522 (1.19)
Constant	-9.929 (7.88)	-3.892 (3.65)	-1.160 (1.01)	0.423 (0.39)
Observations	651			
Log L	-787.688			

Table A.4: Multinomial logit for occupational choice (males)

Absolute value of z-statistics in parenthesis

Variables	$Ln \left[\frac{Pr.(Oc.1)}{Pr.(Oc..5)} \right]$	$Ln \left[\frac{Pr.(Oc.2)}{Pr.(Oc..5)} \right]$	$Ln \left[\frac{Pr.(Oc.3)}{Pr.(Oc..5)} \right]$	$Ln \left[\frac{Pr.(Oc.4)}{Pr.(Oc..5)} \right]$
Age	0.059 (2.99)	0.033 (2.44)	0.001 (0.03)	-0.016 (0.86)
Years of education	0.593 (11.38)	0.338 (8.04)	0.163 (3.17)	0.050 (1.14)
Father's years of education	0.145 (3.77)	0.096 (3.25)	0.084 (2.30)	-0.002 (0.04)
Married	0.172 (0.44)	0.332 (1.10)	0.032 (0.08)	0.396 (0.82)
Constant	-10.743 (10.61)	-6.166 (8.61)	-4.014 (5.23)	-2.002 (2.73)
Observations	968			
Log L	-1006.882			

